

International Conference on Fuzzy Sets and
Soft Computing in Economics and Finance
FSSCEF 2004

Proceedings

Volume I

Saint-Petersburg, Russia
June 17-20, 2004

ISBN 968-489-028-1
968-489-029-X (Printed version)
968-489-030-3 (Electronic version)

Copyright © 2004

Instituto Mexicano del Petróleo

Eje Central Lázaro Cárdenas No. 152
Col. San Bartolo Atepehuacan
C.P. 07730 México D.F.

Russian Fuzzy Systems Association

Vavilova 40, GSP-1, Moscow 119991
Russia

<http://sedok.narod.ru/fsef.html>

Editors :

Ildar Batyrshin
Janusz Kacprzyk
Leonid Sheremetov

Design and Format:

Edgar J. Larios

This work is subject to copyright. All rights are reserved. Reproduction of this publication in any form by any electronic or mechanical means (including photocopying, recording or information storage and retrieval) is permitted only under the provisions of the Mexican Federal Copyright Law and the prior permission in writing of Mexican Petroleum Institute.

Edited and Formatted in Mexico
Printed in Russia
(500)

Greetings to Participants in the Conference on Fuzzy Sets and Soft Computing in Economics and Finance

I deeply regret my inability, due to a scheduling conflict, to participate in FSSCEF 2004. I have no doubt that FSSCEF 2004, held in the handsome city of St. Petersburg, will be an important event, providing the participants with opportunity to exchange views and ideas relating to application of fuzzy logic and soft computing in the realms of economics and finance.

In my view, it is inevitable that applications of fuzzy logic and soft computing to economics and finance will grow in visibility and importance in the years ahead. The reason is that traditional, bivalent-logic-based approaches, are not a good fit to reality—the reality of pervasive imprecision, uncertainty and partiality of truth. The centrepiece of bivalent logic is the principle of the excluded middle: truth is bivalent, meaning that every proposition is either true or false, with no shades of truth allowed. By contrast, in fuzzy logic everything is, or is allowed to be, a matter of degree. It is this characteristic of fuzzy logic that makes it a much better fit to reality than bivalent logic. In particular, it is this characteristic of fuzzy logic that makes it possible for fuzzy logic to deal with perception-based information. Such information plays an essential role in economics, finance and, more generally in all domains in which human perceptions and emotions are in evidence.

With warmest regards to all,

Lotfi A. Zadeh

Professor in the Graduate School
Director, Berkeley Initiative in Soft Computing (BISC)

**Berkeley, CA
May 2004**

Message From the Program Chairs

Following the traditions on organizing international forums in St. Petersburg, the cradle of the Russian Academy of Sciences, the First International Conference on Fuzzy Sets and Soft Computing in Economics and Finance, FSSCEF 2004, was held this year in Northern Palmira. The conference is organized as a platform for exchange of ideas, experiences, and opinions among the academicians, professional engineers and financial community's practitioners on the applications of soft computing methods and techniques to economics and finance. Conference papers were carefully selected in accordance with best quality international standards. All papers were reviewed by international Program Committee. Finally, the International Program Committee accepted 64 papers from 24 countries. Also Bernard De Baets, Hans De Meyer, Alexander Yazenin, Arkady Borisov, Kaouru Hirota and Toshihiro Kaino accepted the invitation as keynote speakers to present their work.

Soft computing techniques have been applied to a number of systems in economics and finance showing in many cases better performance than competing approaches. At this conference, promising areas such as fuzzy data mining, fuzzy game theory, multi-agent systems, fuzzy and neural modeling have been studied for macro-economic analysis, investment and risk management, time series analysis, portfolio optimization and other applications.

Conference Program is composed of 6 sections reflecting the main problem areas and application domains studied in the conference papers: "Fuzzy Data Mining in Economics and Finance", "Fuzzy Games, Decisions and Expert Systems", "Fuzzy Mathematical Structures and Graph Theory", "Multi-Agent Systems and Soft Computing Applications in Economical and Financial Systems", "Soft Computing Methods in Investment and Risk Analysis and in Portfolio Optimization", and "Fuzzy Economic and Information Systems".

Many persons contributed numerous hours to organize the Conference. First of all, we would like to thank all the authors and participants for contributing to the Conference and stimulating the technical discussions. Special thanks to the Program and Organizing Committees members and all the institutions supporting this event. We hope that all participants enjoyed the Conference and their stay in St. Petersburg.

Alexey Averkin, Ildar Batyrshin, Janusz Kacprzyk
Program Committee Co-Chairs

St. Petersburg
June 2004

FSSCEF 2004 Committees

Honorary Chairman

Lotfi A. Zadeh
University of California
Berkeley, USA

Co-Chairs of Program Committee

Alexey Averkin
Dorodnicyn Computing Center, Russian Academy of Sciences
Russia

Ildar Batyrshin
Mexican Petroleum Institute, Mexico
Institute of Problems of Informatics, Academy of Sciences of Tatarstan, Russia

Janusz Kacprzyk
Polish Academy of Sciences
Poland

Co-Chairs of Organizing Committee

Alexey Nedosekin
Siemens Business Services Russia
Russia

Lev Utkin
St.Petersburg State Forest Technical
Academy
Russia

Program Committee

R. Aliev, Azerbaijan
A. Altunin, Russia
L. Bershtein, Russia
J. Bezdek, USA
A. Borisov, Latvia
J. J. Buckley, USA
Ch. Carlsson, Finland
J. L. Castro, Spain
Sh. Chabdarov, Russia
B. De Baets, Belgium
S. Diakonov, Russia
D. Dubois, France
V. Emeljanov, Russia
A. Eremeev, Russia
F. Esteva, Spain
B. Fioleau, France
J. Fodor, Hungary
T. Fukuda, Japan
T. Gavrilova, Russia
J. Gil-Aluja, Spain
Ju. N. Zhuravlev, Russia
K. Hirota, Japan
R. M. Kachalov, Russia
O. Kaynak, Turkey
E. Kerre, Belgium
G. B. Klejner, Russia
E. P. Klement, Austria
R. Klempous, Poland
G. J. Klir, USA
L. T. Koczy, Hungary
V. M. Kurejchik, Russia

M. Mares, Czech Republic
I. Nasyrov, Russia
A. Nedosekin, Russia
V. Novak, Czech Republic
G. Osipov, Russia
P. Osmera, Czech Republic
I. Perfilieva, Czech Republic
D. Pospelov, Russia
H. Prade, France
S. V. Prokopchina, Russia
A. Ryjov, Russia
I. J. Rudas, Hungary
T. Rudas, Hungary
D. Rutkowska, Poland
L. Rutkowski, Poland
R. Setiono, Singapore
P. Sevastjanov, Poland
L. Sheremetov, Mexico
P. Sincak, Slovakia
R. Slowinski, Poland
V. Stefanjuk, Russia
R. Suarez, Mexico
M. Sugeno, Japan
V. Tarasov, Russia
I. B. Turksen, Canada
L. Utkin, Russia
V. Vagin, Russia
M. Wagenknecht, Germany
R. Yager, USA
A. Yazenin, Russia
C. Zopounidis, Greece

Invited Speakers

Arkady Borisov

Institute of Information Technology
LATVIA

Kaoru Hirota

Tokyo Institute of Technology
JAPAN

Bernard De Baets

Ghent University
BELGIUM

Toshihiro Kaino

Aoyama-Gakuin University
JAPAN

Hans De Meyer

Ghent University
BELGIUM

Alexander Yazenin

Tver State University
RUSSIA

FSSCEF Sponsorship

Siemens Business Services Russia
Mexican Petroleum Institute
Russian Fuzzy Systems Association
Institute of Problems of Informatics, Academy of Sciences Republic of Tatarstan
Kazan State Technological University
University of Economics and Finance
Economic Faculty of Moscow State University
Dorodnicyn Computing Centre of the Russian Academy of Science, Russia
International Fuzzy Systems Association IFSA
International Association for Fuzzy-Set Management and Economy SIGEF
European Society for Fuzzy Logic and Technology EUSFLAT
Journal of Audit and Financial Analysis
Journal of News of Artificial Intelligence
Journal of Mathware and Soft Computing
Financial company "Inkastrast"

Table of Contents

VOLUME I

INVITED LECTURES

A New Approach to Stochastic Dominance	3
<i>De Baets Bernard and De Meyer Hans</i>	
Optimization with Fuzzy Random Data and its Application in Financial Analysis <i>Yazenin A.V.</i>	16
Machine Learning in Fuzzy Environment	33
<i>Borisov Arkady</i>	
Non-Stochastic-Model Based Finance Engineering	35
<i>Hirota Kaouru and Kaino Toshihiro</i>	

FUZZY DATA MINING IN ECONOMICS AND FINANCE

Plenary Report

Mining Fuzzy Association Rules and Networks in Time Series Databases	39
<i>Batyrshin I., Herrera-Avelar R., Sheremetov L. and Suarez R.</i>	

Perceptual Time Series Data Mining

A Clear View on Quality Measures for Fuzzy Association Rules	54
<i>De Cock Martine, Cornelis Chris and Kerre Etienne</i>	
Moving Approximations in Time Series Data Mining	62
<i>Batyrshin I., Herrera-Avelar R., Sheremetov L. and Suarez R.</i>	
On Qualitative Description of Time Series Based on Moving Approximations	73
<i>Batyrshin I., Herrera-Avelar R., Sheremetov L. and Suarez R.</i>	
Generating Fuzzy Rules for Financial Time Series by Neural Networks with Supervised Competitive Learning Techniques	81
<i>Marček Dušan</i>	
Pattern Recognition through Perceptually Important Points in Financial Time Series	89
<i>Zaib Gul, Ahmed Uzair and Ali Arshad</i>	

Fuzzy Classification and Pattern Recognition

Soft Clustering for Funds Management Style Analysis: Out-of-Sample Predictability	97
<i>Lajbcygier Paul and Yahya Asjad</i>	
Tuning the Fuzzy Classification Models with Various Learning Criteria: the Case of Credit Data Classification	103
<i>Shtovba Serhiy, Pankevich Olga and Dounias Georgios</i>	
prInvestor: Pattern Recognition Based Financial Time Series Investment System ..	111
<i>Ruta Dymitr</i>	
On General Scheme of Invariant Clustering Procedures Based on Fuzzy Similarity Relation	122
<i>Batyrshin I.Z., Rudas T. and Klimova A.</i>	
Evolutionary Procedures of Visualization of Multidimensional Data	130
<i>Angelica Klimova</i>	

FUZZY GAMES, DECISIONS AND EXPERT SYSTEMS

Plenary Report

Vague Utilities in Cooperative Market	143
<i>Mareš Milan</i>	

Fuzzy Games and Decision Making

The Shapley Value for Games on Lattices (L-Fuzzy Games)	154
<i>Grabisch Michel</i>	
Optimal Strategies of Some Symmetric Matrix Games	162
<i>De Schuymer Bart, De Meyer Hans and De Baets Bernard</i>	
On Strict Monotonic t -norms and t -conorms on Ordinal Scales	170
<i>Batyrshin I.Z. and Batyrshin I.I.</i>	
Belief Functions and the Imprecise Dirichlet Model	178
<i>Utkin Lev V.</i>	
Fuzzy Decision Making Using the Imprecise Dirichlet Model	186
<i>Utkin Lev V. and Augustin Thomas</i>	

Fuzzy Expert Systems

A Fuzzy Expert System for Predicting the Effect of Socio-Economic Status on Noise-Induced Annoyance	194
<i>Zaheeruddin and Jain V. K.</i>	

The Knowledge Representation Model in Fuzzy Expert Systems Provided by Computing with Words and PRUF Language	202
<i>Glova V.I., Anikin I.V., Katasev A.S. and Pheoktistov O.N</i>	

Decision Trees and Optimization

Construction of Incremental Fuzzy Decision Tree	210
<i>Borisov Arkady and Bikesheva Gulnara</i>	
Fuzzy Binary Tree Model for European-Style Vanilla Options	222
<i>Muzzioli Silvia and Reynaerts Hugnette</i>	
Tree-Structured Smooth Transition Regression	230
<i>Correa da Rosa Joel, Veiga Alvaro, Medeiros Marcelo</i>	
Discrete Dynamic Programming With Outcomes In Fuzzy Ordered Structures	238
<i>Trzaskalik Tadeusz and Sitarz Sebastian</i>	

FUZZY MATHEMATICAL STRUCTURES AND GRAPH THEORY

Plenary Report

Fractal Methods of Graphs Partitioning	249
<i>Kureichik V.V. and Kureichik V.M.</i>	

Fuzzy Mathematical Structures

A Method for Defining and Fuzzifying Mathematical Structures	257
<i>Ionin V.K. and Plesniewicz G.S.</i>	
On Symmetric MV-Polynomials	265
<i>Di Nola Antonio, Lettieri Ada and Belluce Peter</i>	
Lattice Products, Bilattices and Some Extensions of Negations, Triangular Norms and Triangular Conorms	272
<i>Tarassov Valery B.</i>	

Soft Computing Methods in Graphs Theory and Applications

Definition of Optimum Allocation of the Service Centers	283
<i>Bershtein Leonid, Bozhenyuk Alexander and Rozenberg Igor</i>	
Evolutionary Algorithm of Minimization of Intersections and Flat Piling of the Graph Mathematic Models.....	291
<i>Gladkov L.A. and Kureichik V.M.</i>	
Use of a Fuzzy Strategy to Create the Shortest Hamiltonian Cycle	299
<i>Piech Henryk and Ptak Aleksandra</i>	

VOLUME II

MULTI-AGENT SYSTEMS AND SOFT COMPUTING APPLICATIONS IN ECONOMICAL AND FINANCIAL SYSTEMS

Plenary Report

- Combining Multi-agent Approach with Intelligent Simulation in Resources Flow
Management 311
Emelyanov Viktor V.

Multi-Agent and Soft Computing Models of Economical Systems

- Agent-based Collective Intelligence for Inter-bank Payment Systems 321
Rocha-Mier Luis, Villareal Francisco and Sheremetov Leonid
- Macroeconomic Fuzzy Model of Azerbaijan 332
Imanov Q.C., Mamedov F.C and Akbarov R.M.
- Application of Fuzzy Cognitive Maps to the Political-Economic Problem of
Cyprus 340
Neocleous Costas and Schizas Christos
- Fuzzy Linear Regression Application to the Estimation of Air Transport Demand ..
Charfeddine Souhir, Mora-Camino Félix and De Coligny Marc 350
- Methodology of the Estimation of Quality of Objects with Complex Structure
Under Conditions of Non-Stochastic Uncertainty 360
*Zhelezko Boris A., Siniavskaya Olga A., Ahrameiko Alexey A. And
Berbasova Natalya Y.*

Neural Networks Applications in Economics and Finance

- Decision Making in Stock Market with NN and Fuzzy Logic 368
Aliev R.A. and Mammadli S. F.
- The Application of Non-linear Model and Artificial Neural Network to Exchange
Rate Forecasting 378
Barbouch Rym and Abaoub Ezzeddine
- Financial Implications of Artificial Neural Networks in Automobile Insurance
Underwriting 386
Kitchens Fred L.
- A Neural-Fuzzy Approach to Economics Data Classification 394
Salakhutdinov R.Z., Ismagilov I. and Rubtsov A.V.
-

SOFT COMPUTING METHODS IN INVESTMENT AND RISK ANALYSIS AND IN PORTFOLIO OPTIMIZATION

Plenary Report

- Portfolio Optimization System (Siemens Business Services Russia) 403
Alpatsky V.V. and Nedosekin A.O.

Soft Computing in Investment and Risk Analysis

- Optimal Investment Using Technical Analysis and Fuzzy Logic 415
Sevastianov P. and Rozenberg P.
- Investment Risk Estimation for Arbitrary Fuzzy Factors of Investments Project 423
Nedosekin A. and A. Kokosh
- Robust Selection of Investment Projects 438
Kuchta Dorota
- Analyzing the Solvency Status of Brazilian Enterprises with Fuzzy Models 446
Pinto Dias Alves Antonio
- Opportunities Within the New Basel Capital Accord for Assessing Banking Risk by Means of Soft Computing 457
Canfora Gerardo, D'Alessandro Vincenzo and Troiano Luigi
- Risk Analysis in Granular-Information-Based Decision Aid Models 466
Valishevsky Alexander and Borisov Arkady

Fuzzy Portfolio Optimization

- A New Approach to Optimizing Portfolio Funding in an Fuzzy Environment 474
Nedosekin Alexey and Korchunov Valentin
- Comparative Study of Aggregation Methods in Bicriterial Fuzzy Portfolio Selection 484
Sewastianow P. and Jończyk M.
- On One Method of Portfolio Optimization With Fuzzy Random Data 493
Grishina E. N.

FUZZY ECONOMIC AND INFORMATION SYSTEMS

Plenary Report

- Towards Human-Consistent Data-Driven Decision Support Systems Via Fuzzy Linguistic Data Summaries 501
Kacprzyk Janusz and Zadrozny Slawomir

Fuzzy Information Systems

Information Monitoring Systems as a Tool for Strategic Analysis and Simulation in Business	511
<i>Rylov Alexander</i>	
How to Select a Corporate Information System Using Fuzzy Sets	521
<i>Korolkov Mikhail, Nedosekin Alexey and Segeda Anna</i>	
Application of a Fuzzy Relational Server to Concurrent Engineering	530
<i>Yarushkina Nadezhda</i>	
Toward Problem of Information Retrieval from Internet	538
<i>Gilmudinov R.</i>	

Fuzzy Models of Economic Systems

New Method for Interval Extension of Leontief's Input-Output Model Using Parallel Programming	549
<i>Dymova L., Gonera M., Sevastianov P. and Wyrzykowski R.</i>	
The Solution of Transport Problem in Fuzzy Statement on the Basis of Platform Anylogic	557
<i>Karpov Yury, Lyubimov Boris and Nedosekin Alexey</i>	
On the Application of Fuzzy Sets Theory to Russian Banking System Fragility Monitoring	565
<i>Sergei V. Ivliev and Velisava T. Sevrouk</i>	
A Fuzzy Model of Productivity Control	570
<i>Piech Henryk and Leks Dariusz</i>	
Regression Prediction Models with Fuzzy Structure of Adaptive Mechanism	578
<i>Davnis V.V. and Tinyakova V.I.</i>	

INVITED LECTURES

A New Approach to Stochastic Dominance

De Baets Bernard and De Meyer Hans

**Optimization with Fuzzy Random Data and its Application in Financial
Analysis**

Yazenin A.V.

Machine Learning in Fuzzy Environment

Borisov Arkady

Non-Stochastic-Model Based Finance Engineering

Hirota Kaouru and Kaino Toshihiro

A New Approach to Stochastic Dominance

Bernard De Baets¹ and Hans De Meyer²

¹ Department of Applied Mathematics, Biometrics and Process Control,
Ghent University, Coupure links 653, B-9000 Gent, Belgium

² Department of Applied Mathematics and Computer Science,
Ghent University, Krijgslaan 281 (S9), B-9000 Gent, Belgium

Summary. We establish a pairwise comparison method for random variables. This comparison results in a probabilistic relation on the given set of random variables. The transitivity of this probabilistic relation is investigated in the case of independent random variables, as well as when these random variables are pairwise coupled by means of a copula, more in particular the minimum operator or the Łukasiewicz t-norm. A deeper understanding of this transitivity, which can be captured only in the framework of cycle-transitivity, allows to identify appropriate strict or weak cutting levels, depending upon the copula involved, turning the probabilistic relation into a strict order relation. Using 1/2 as a fixed weak cutting level does not guarantee an acyclic relation, but is always one-way compatible with the classical concept of stochastic dominance. The proposed method can therefore also be seen as a way of generating graded as well as non-graded variants of that popular concept.

1 Introduction

We denote the joint cumulative distribution function (c.d.f.) of a random vector (X_1, X_2, \dots, X_m) as F_{X_1, X_2, \dots, X_m} . This joint c.d.f. characterizes the random vector almost completely. Nevertheless, it is known from probability theory and statistics that practical considerations often lead one to capture the properties of the random vector and its joint c.d.f. as much as possible by means of a restricted number of (numerical) characteristics. The expected value, variance and other (central) moments of the components X_i belong to the family of characteristics that can be computed from the marginal c.d.f. F_{X_i} solely. A second family consists of characteristics that measure dependence or association between the components of the random vector. Well-known members of this family are the linear correlation coefficient, also known as Pearson's product-moment correlation coefficient, Kendall's τ and Spearman's ρ . In general, their computation only requires the knowledge of the bivariate c.d.f. F_{X_i, X_j} . The function C that joins the one-dimensional marginal c.d.f. F_{X_i} and F_{X_j} into the bivariate marginal c.d.f. F_{X_i, X_j} is known as a copula [10]:

$F_{X_i, X_j} = C(F_{X_i}, F_{X_j})$. Although in general not required, we shall assume that the copula C is the same for all (i, j) .

Our goal in this contribution is to establish a new method for comparing the components of a random vector in a pairwise manner. More in particular, with any given random vector we will associate a so-called probabilistic relation. Our main concern is to study the type of transitivity exhibited by this probabilistic relation and to analyze to what extent it depends upon the copula that pairwise couples the components of the random vector. To that end, we need a framework that allows to describe a sufficiently broad range of types of transitivity. The one that will prove to be the best suited is the framework of cycle-transitivity, which has been laid bare by the present authors [2].

This paper is organized as follows. In Section 2, we propose a new method for generating a probabilistic relation from a given random vector and indicate in what sense this relation generalizes the concept of stochastic dominance [9]. One of our aims is to characterize the type of transitivity exhibited by this relation. To that end, we give a brief introduction to the framework of cycle-transitivity in Section 3. In Section 4, we consider a random vector with pairwise independent components and analyze the transitivity of the generated probabilistic relation, while in Section 5 we are concerned with random vectors that have dependent components. In the latter section, we first briefly review the important concept of a copula. Then we study two extreme types of couplings between the components of a random vector, namely by means of one of the copulas in between which all other copulas are situated, i.e. the minimum operator and the Łukasiewicz t-norm [7, 10]. Finally, in Section 6, we explain how the results presented lead to a whole range of methods for comparing probability distributions and identify proper ways of defining a strict order on them, thus offering valuable alternatives to the usual notion of stochastic dominance.

2 A method for comparing random variables

An immediate way of comparing two real random variables X and Y is to consider the probability that the first one takes a value greater than the second one. Proceeding in this way, a random vector (X_1, X_2, \dots, X_m) generates a probabilistic relation (also called reciprocal relation or ipsodual relation), as follows.

Definition 1. *Given a random vector (X_1, X_2, \dots, X_m) , the binary relation Q defined by:*

$$Q(X_i, X_j) = \text{Prob}\{X_i > X_j\} + \frac{1}{2} \text{Prob}\{X_i = X_j\} \quad (1)$$

is a probabilistic relation, i.e. for all (i, j) it holds that:

$$Q(X_i, X_j) + Q(X_j, X_i) = 1.$$

Note that $Q(X, Y)$ is not the probability that X takes a greater value than Y , since in order to make Q a probabilistic relation, we also take half of the probability of a tie into account. It is clear from this definition that the relation Q can be computed immediately from the bivariate joint cumulative distributions F_{X_i, X_j} as:

$$Q(X_i, X_j) = \int_{x>y} dF_{X_i, X_j}(x, y) + \frac{1}{2} \int_{x=y} dF_{X_i, X_j}(x, y). \quad (2)$$

If we want to further simplify (2), it is appropriate to distinguish between the following two cases. If the random vector is a discrete random vector, then

$$Q(X_i, X_j) = \sum_{k>l} p_{X_i, X_j}(k, l) + \frac{1}{2} \sum_k p_{X_i, X_j}(k, k), \quad (3)$$

with p_{X_i, X_j} the joint probability mass function of (X_i, X_j) , and if it is a continuous random vector, then

$$Q(X_i, X_j) = \int_{-\infty}^{+\infty} dx \int_{-\infty}^x f_{X_i, X_j}(x, y) dy, \quad (4)$$

with f_{X_i, X_j} the joint probability density function of (X_i, X_j) . Note that in the transition from the discrete to the continuous case, the second contribution to $Q(X_i, X_j)$ in (2) has disappeared in (4), since in the latter case it holds that $\text{Prob}\{X_i = X_j\} = 0$.

The probabilistic relation Q generated by a random vector yields a recipe for comparison that takes into account the bivariate joint probability distribution function, hence to some extent the pairwise dependence of the components. The information contained in the probabilistic relation is therefore much richer than if for the pairwise comparison of X_i and X_j we would have used, for instance, only their expected values $\mathbf{E}[X_i]$ and $\mathbf{E}[X_j]$.

For two random variables X and Y , one says that X is weakly statistically preferred to Y , denoted as $X \supseteq Y$, if $Q(X, Y) \geq 1/2$; if $Q(X, Y) > 1/2$, then one says that X is statistically preferred to Y , denoted $X \triangleright Y$. Of course, we would like to know whether the relations \supseteq or \triangleright are transitive. To that aim, let us consider the following example of a discrete random vector (X, Y, Z) with three pairwise independent components, uniformly distributed over the integer sets

$$D_X = \{1, 3, 4, 15, 16, 17\}, D_Y = \{2, 10, 11, 12, 13, 14\}, D_Z = \{5, 6, 7, 8, 9, 18\}.$$

We can apply (3) with all joint probability masses equal to $1/36$. More precisely, we obtain $Q(X, Y) = 20/36$, $Q(Y, Z) = 25/36$ and $Q(X, Z) = 15/36$, from which it follows that $X \triangleright Y$, $Y \triangleright Z$ and $Z \triangleright X$, and it turns out that in this case the relation \triangleright (and hence also \supseteq) forms a cycle, and hence is not transitive.

An alternative concept for comparing two random variables, or equivalently, two probability distributions, is that of stochastic dominance [9], which is particularly popular in financial mathematics.

Definition 2. *A random variable X with c.d.f. F_X weakly stochastically dominates in first degree a random variable Y with c.d.f. F_Y , denoted as $X \succsim_1 Y$, if it holds that $F_X \leq F_Y$. If, moreover, it holds that $F_X(t) < F_Y(t)$, for some t , then it is said that X stochastically dominates in first degree Y , denoted as $X \succ_1 Y$.*

Note that, as for any comparison method that relies only upon characteristics of the marginal distributions, the stochastic dominance relation \succsim_1 does not take into account any effects of the possible pairwise dependence of the random variables. Moreover, the condition for first-degree stochastic dominance is rather severe, as it requires that the graph of the c.d.f. F_X lies beneath the graph of the c.d.f. F_Y . The need to relax this condition has led to other types of stochastic dominance, such as second-degree and third-degree stochastic dominance. We will not go into more details here, since we just want to emphasize the following relationship between weak first-degree stochastic dominance and the relation \supseteq .

Proposition 1. *For any two random variables X and Y it holds that weak stochastic dominance implies weak statistical preference, i.e. $X \succsim_1 Y$ implies $X \supseteq Y$.*

The relation \supseteq therefore generalizes weak first-degree stochastic dominance \succsim_1 . Note that the same implication is not true in general for the strict versions \triangleright and \succ_1 . Since the probabilistic relation Q is a graded alternative to the crisp relation \supseteq , we can interpret it as a graded generalization of weak first-degree stochastic dominance. Unfortunately, as shown in the example above, the relation \supseteq is not necessarily transitive, while this is obviously the case for the relation \succsim_1 . Further on in this paper, we will show how this shortcoming can be resolved.

3 Cycle-transitivity

Cycle-transitivity was proposed recently by the present authors as a general framework for studying the transitivity of probabilistic relations [2]. The key feature is the cyclic evaluation of transitivity: triangles (i.e. any three points) are visited in a cyclic manner. An upper bound function acting upon the ordered weights encountered provides an upper bound for the ‘sum minus 1’ of these weights. Cycle-transitivity incorporates various types of fuzzy transitivity and stochastic transitivity.

For a probabilistic relation Q on A , we define for all a, b, c the following quantities:

$$\begin{aligned}\alpha_{abc} &= \min(Q(a, b), Q(b, c), Q(c, a)), \\ \beta_{abc} &= \text{median}(Q(a, b), Q(b, c), Q(c, a)), \\ \gamma_{abc} &= \max(Q(a, b), Q(b, c), Q(c, a)).\end{aligned}$$

Let us also denote $\Delta = \{(x, y, z) \in [0, 1]^3 \mid x \leq y \leq z\}$.

Definition 3. A function $U : \Delta \rightarrow \mathbb{R}$ is called an upper bound function if it satisfies:

- (i) $U(0, 0, 1) \geq 0$ and $U(0, 1, 1) \geq 1$;
- (ii) for any $(\alpha, \beta, \gamma) \in \Delta$:

$$U(\alpha, \beta, \gamma) + U(1 - \gamma, 1 - \beta, 1 - \alpha) \geq 1. \quad (5)$$

The function $L : \Delta \rightarrow \mathbb{R}$ defined by

$$L(\alpha, \beta, \gamma) = 1 - U(1 - \gamma, 1 - \beta, 1 - \alpha) \quad (6)$$

is called the dual lower bound function of a given upper bound function U . Inequality (5) then simply expresses that $L \leq U$.

Definition 4. A probabilistic relation Q on A is called cycle-transitive w.r.t. an upper bound function U if for any $(a, b, c) \in A^3$ it holds that

$$L(\alpha_{abc}, \beta_{abc}, \gamma_{abc}) \leq \alpha_{abc} + \beta_{abc} + \gamma_{abc} - 1 \leq U(\alpha_{abc}, \beta_{abc}, \gamma_{abc}), \quad (7)$$

where L is the dual lower bound function of U .

Due to the built-in duality, it holds that if (7) is true for some (a, b, c) , then this is also the case for any permutation of (a, b, c) . In practice, it is therefore sufficient to check (7) for a single permutation of any $(a, b, c) \in A^3$. Alternatively, due to the same duality, it is also sufficient to verify the right-hand inequality (or equivalently, the left-hand inequality) for two permutations of any $(a, b, c) \in A^3$ (not being cyclic permutations of one another), e.g. (a, b, c) and (c, b, a) .

Proposition 2. A probabilistic relation Q on A is cycle-transitive w.r.t. an upper bound function U if for any $(a, b, c) \in A^3$ it holds that

$$\alpha_{abc} + \beta_{abc} + \gamma_{abc} - 1 \leq U(\alpha_{abc}, \beta_{abc}, \gamma_{abc}). \quad (8)$$

4 The case of independent random variables

In this section, we consider the case of a random vector with pairwise independent components X_i . In fact, as is well known in probability theory, this does not mean that the random variables X_i are necessarily mutually independent. However, since our comparison method only involves the bivariate c.d.f., the distinction between pairwise and mutually independent random variables is superfluous in the present discussion. An important consequence of the assumed pairwise independence is that the bivariate distribution functions become factorizable into the univariate marginal distributions, in particular $F_{X_i, X_j} = F_{X_i} F_{X_j}$, for a discrete random vector $p_{X_i, X_j} = p_{X_i} p_{X_j}$, and for a continuous random vector $f_{X_i, X_j} = f_{X_i} f_{X_j}$.

The first case in which we have been able to determine the type of transitivity of the probabilistic relation is that of a discrete random vector with pairwise independent components that are uniformly distributed on arbitrary integer (multi)sets. In this case the components X_i of the random vector can be regarded as hypothetical dice (with as many faces as elements in the corresponding multiset), whereas $Q(X_i, X_j)$ can then be seen as the probability that dice X_i wins from dice X_j . Such a discrete uniformly distributed random vector, together with the generated probabilistic relation, has therefore been called a standard discrete dice model [4]. The type of transitivity of the probabilistic relation Q can be very neatly described in the framework of cycle-transitivity [2].

Proposition 3. *The probabilistic relation generated by a random vector with pairwise independent components that are uniformly distributed on finite integer multisets is cycle-transitive w.r.t. the upper bound function U_D defined by:*

$$U_D(\alpha, \beta, \gamma) = \beta + \gamma - \beta\gamma.$$

This type of transitivity is called dice-transitivity.

In [5], the transitivity was investigated in the general case of discrete or continuous random vectors with arbitrary independent components. Somewhat surprisingly, the above case turned out, as far as transitivity of the probabilistic relation is concerned, to be generic for the most general situation.

Proposition 4. *A random vector with arbitrary pairwise independent components generates a probabilistic relation that is dice-transitive.*

5 The case of dependent random variables

5.1 Joint distribution functions and copulas

In this section, we focus on dependent random variables. We consider the general case of a random vector (X_1, X_2, \dots, X_m) with joint c.d.f. F_{X_1, X_2, \dots, X_m} ,

to which we associate a probabilistic relation Q , as defined in (1) or, equivalently, in (2).

Sklar's theorem [10, 11] tells us that if a joint c.d.f. F_{X_i, X_j} has marginal c.d.f. F_{X_i} and F_{X_j} , then there exists a copula C such that for all x, y :

$$F_{X_i, X_j}(x, y) = C(F_{X_i}(x), F_{X_j}(y)). \quad (9)$$

Let us recall that a copula is a binary operation $C : [0, 1]^2 \rightarrow [0, 1]$, that has neutral element 1 and absorbing element 0 and that satisfies the property of moderate growth [10]: for any $(x_1, x_2, y_1, y_2) \in [0, 1]^4$ it holds that

$$(x_1 \leq x_2 \wedge y_1 \leq y_2) \Rightarrow C(x_1, y_1) + C(x_2, y_2) \geq C(x_1, y_2) + C(x_2, y_1).$$

A copula C is called stable if for all $(x, y) \in [0, 1]^2$ it holds that [8]:

$$C(x, y) + 1 - C(1 - x, 1 - y) = x + y.$$

If the random variables X_i and X_j are continuous, then the copula C in (9) is unique; otherwise, C is uniquely determined on $\text{Ran}(F_{X_i}) \times \text{Ran}(F_{X_j})$. Conversely, if C is a copula and F_{X_i} and F_{X_j} are c.d.f., then the function defined by (9) is a joint c.d.f. with marginal c.d.f. F_{X_i} and F_{X_j} . For independent random variables, the copula C is the product copula $T_{\mathbf{P}}$ ($T_{\mathbf{P}}(x, y) = xy$).

In this section we will consider the two extreme copulas in between which all other copulas are situated, i.e. the Łukasiewicz copula $T_{\mathbf{L}}$ ($T_{\mathbf{L}}(x, y) = \max(x + y - 1, 0)$, also called the Fréchet-Hoeffding lower bound) and the minimum operator $T_{\mathbf{M}}$ ($T_{\mathbf{M}}(x, y) = \min(x, y)$, also called the Fréchet-Hoeffding upper bound).

In the case of pairwise independent random variables, the computation of Q was done by means of formula (3) (discrete case) or formula (4) (continuous case), taking into account that the joint probability mass function (joint probability density function) is simply the product of the marginal probability mass functions (marginal probability density functions). The question arises whether for the two extreme cases of dependent random variables, namely coupled by $T_{\mathbf{M}}$ or $T_{\mathbf{L}}$, there exist formulae that simplify the computation of Q .

5.2 The extreme copulas

We first consider the pairwise coupling of components by means of the copula $T_{\mathbf{M}}$:

$$F_{X_i, X_j}(x, y) = \min(F_{X_i}(x), F_{X_j}(y)). \quad (10)$$

We have demonstrated that [3]:

Proposition 5. *Let (X_1, X_2, \dots, X_m) be a continuous random vector with bivariate joint c.d.f. given by (10). Then the probabilistic relation Q^M , defined by $Q^M(X_i, X_j) = \text{Prob}\{X_i > X_j\}$, can be computed as:*

$$Q^M(X_i, X_j) = \int_{x:F_{X_i}(x) < F_{X_j}(x)} f_{X_i}(x) dx + \frac{1}{2} \int_{x:F_{X_i}(x) = F_{X_j}(x)} f_{X_i}(x) dx. \quad (11)$$

Note that if X_i and X_j are identically distributed, i.e. $F_{X_i} = F_{X_j}$, then $Q^M(X_i, X_j) = 1/2$, as expected. In Figure 1 we give a graphical interpretation of formula (11). The two curves correspond to the marginal c.d.f. F_X and F_Y .

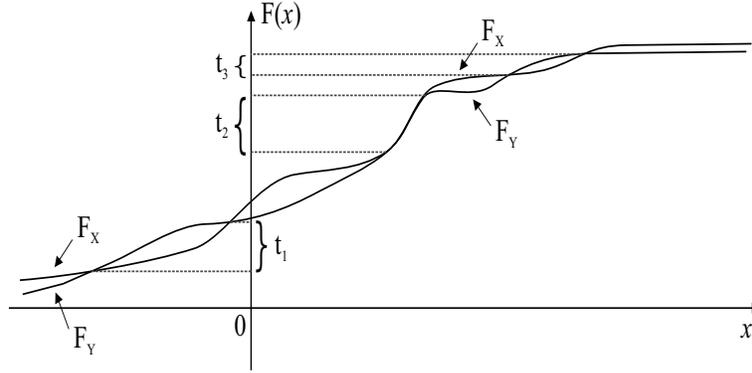


Fig. 1. Comparison of two continuous random variables coupled by T_M .

According to (11) we have to distinguish three domains: the domain where F_X lies beneath F_Y , the domain where F_X lies above F_Y , and the domain where F_X and F_Y coincide. The value of $Q^M(X, Y)$ is computed as the sum of the increment of F_X over the first domain and half of the increment of F_X (or F_Y) over the third domain. With the notations shown on the figure, we obtain for the example of Figure 1:

$$Q^M(X, Y) = t_1 + t_3 + \frac{1}{2} t_2.$$

The case of independent random variables has been treated before, and the computation of Q^P according to (4) can be concisely written as

$$Q^P(X_i, X_j) = \mathbf{E}_{X_i}[F_{X_j}]. \quad (12)$$

Next, we consider a continuous random vector (X_1, X_2, \dots, X_m) with arbitrary c.d.f. pairwise coupled by $T_{\mathbf{L}}$:

$$F_{X_i, X_j}(x, y) = \max(F_{X_i}(x) + F_{X_j}(y) - 1, 0). \quad (13)$$

In [3], we have also shown that:

Proposition 6. *Let (X_1, X_2, \dots, X_m) be a continuous random vector with bivariate joint c.d.f. given by (13). Then the probabilistic relation $Q^{\mathbf{L}}$, defined by $Q^{\mathbf{L}}(X_i, X_j) = \text{Prob}\{X_i > X_j\}$, can be computed as:*

$$Q^{\mathbf{L}}(X_i, X_j) = \int_{x: F_{X_i}(x) + F_{X_j}(x) \geq 1} f_{X_i}(x) dx, \quad (14)$$

or, equivalently:

$$Q^{\mathbf{L}}(X_i, X_j) = F_{X_j}(u) \text{ with } u \text{ such that } F_{X_i}(u) + F_{X_j}(u) = 1. \quad (15)$$

One can again easily verify that $Q^{\mathbf{L}}(X, Y) = 1/2$ when $F_X = F_Y$. Note that u in (15) may not be unique, in which case any u fulfilling the right equality may be considered. Then $Q^{\mathbf{L}}(X, Y)$ is simply the height of F_{X_j} in u . This is illustrated in Figure 2, where $Q^{\mathbf{L}}(X, Y) = F_Y(u) = t_1$, since $t_1 + t_2 = 1$.

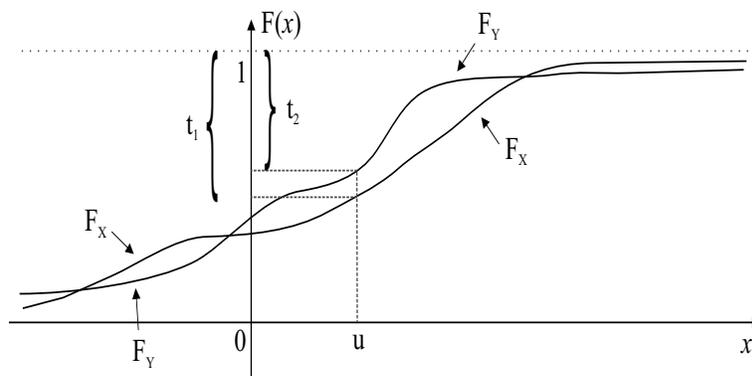


Fig. 2. Comparison of two continuous random variables coupled by $T_{\mathbf{L}}$.

5.3 Transitivity

In Section 3, we have indicated that the probabilistic relation of a discrete or continuous dice model is dice-transitive and we have demonstrated that the framework of cycle-transitivity is very well suited to express this type of transitivity in a concise manner. Now, we want to investigate the transitivity

of the probabilistic relation generated by a random vector, with components pairwise coupled by a same commutative copula C . We have been able to show that [3]:

Theorem 1. *The probabilistic relation Q generated by a random vector pairwise coupled by a commutative copula C is cycle-transitive w.r.t. to the upper bound function U^C , defined by:*

$$U^C(\alpha, \beta, \gamma) = \min(\beta + C(1 - \beta, \gamma), \gamma + C(\beta, 1 - \gamma)). \quad (16)$$

If C is stable, then

$$U^C(\alpha, \beta, \gamma) = \beta + C(1 - \beta, \gamma) = \gamma + C(\beta, 1 - \gamma). \quad (17)$$

Note that without the framework of cycle-transitivity, it would be extremely difficult to describe this type of transitivity in a compact manner. This theorem implies that the relation Q can be seen as a graded alternative to the notion of stochastic dominance.

The above theorem applies in particular to the Frank t-norm family. For $C = T_\lambda^{\mathbf{F}}$, it then holds that the probabilistic relation Q is cycle-transitive w.r.t. the upper bound function $U_\lambda^{\mathbf{F}}$ given by:

$$U_\lambda^{\mathbf{F}}(\alpha, \beta, \gamma) = \beta + T_\lambda^{\mathbf{F}}(1 - \beta, \gamma) = \beta + \gamma - T_{1/\lambda}^{\mathbf{F}}(\beta, \gamma) = S_{1/\lambda}^{\mathbf{F}}(\beta, \gamma). \quad (18)$$

It is well known that if $\lambda < \lambda'$, then $S_\lambda^{\mathbf{F}} < S_{\lambda'}^{\mathbf{F}}$, which implies that $U_\lambda^{\mathbf{F}} > U_{\lambda'}^{\mathbf{F}}$. Therefore, the lower the value of λ when the random variables are coupled by $T_\lambda^{\mathbf{F}}$, the weaker the type of transitivity exhibited by the probabilistic relation generated by these random variables. In particular, the strongest type of transitivity is observed when coupling by $T_{\mathbf{L}}$, the weakest when coupling by $T_{\mathbf{M}}$.

Let us discuss the three main copulas:

- (i) For $C = T_{\mathbf{L}} = T_\infty^{\mathbf{F}}$, we obtain from (18) that Q is cycle-transitive w.r.t. the upper bound function $U_\infty^{\mathbf{F}}$ given by:

$$U_\infty^{\mathbf{F}}(\alpha, \beta, \gamma) = \max(\beta, \gamma) = \gamma.$$

This upper bound function has not yet been encountered.

- (ii) For $C = T_{\mathbf{P}} = T_1^{\mathbf{F}}$, we retrieve the well-known case of independent variables, with

$$U_1^{\mathbf{F}}(\alpha, \beta, \gamma) = \beta + \gamma - \beta\gamma = U_D(\alpha, \beta, \gamma).$$

- (iii) For $C = T_{\mathbf{M}} = T_0^{\mathbf{F}}$, we obtain from (18) that Q is cycle-transitive w.r.t. the upper bound function $U_0^{\mathbf{F}}$ given by:

$$U_0^{\mathbf{F}}(\alpha, \beta, \gamma) = \min(\beta + \gamma, 1).$$

Although not immediately apparent, one can show that cycle-transitivity w.r.t. this upper bound function is equivalent to $T_{\mathbf{L}}$ -transitivity.

6 Alternative notions of stochastic dominance

The results from the foregoing sections can also be exploited to come up with non-graded alternatives to the concept of stochastic dominance. Indeed, consider m random variables X_1, X_2, \dots, X_m with associated marginal c.d.f. $F_{X_1}, F_{X_2}, \dots, F_{X_m}$, then we can pairwise couple them (in a virtual manner) by means of a copula C and come up with a probabilistic relation Q^C on the set of random variables which is cycle-transitive w.r.t. the upper bound function U^C . The latter knowledge allows us to identify more appropriate cutting levels resulting in a strict order relation.

Theorem 2. *Let X_1, X_2, \dots, X_m be m random variables. For the copula $C = T_{\mathbf{L}}$, it holds that the binary relation $>_{\mathbf{L}}$ defined by*

$$X_i >_{\mathbf{L}} X_j \Leftrightarrow Q^{\mathbf{L}}(X_i, X_j) > \frac{1}{2}$$

is a strict order relation.

For the probabilistic relations $Q^{\mathbf{P}}$ and $Q^{\mathbf{M}}$ things are more complicated.

Theorem 3. *Let X_1, X_2, \dots, X_m be m random variables and consider the copula $T_{\mathbf{P}}$. Let $k \in \mathbb{N}$, $k \geq 2$.*

(i) The binary relation $>_{\mathbf{P}}^k$ defined by

$$X_i >_{\mathbf{P}}^k X_j \Leftrightarrow Q^{\mathbf{P}}(X_i, X_j) > 1 - \frac{1}{4 \cos^2(\pi/(k+2))}$$

is an asymmetric relation without cycles of length k .

(ii) The binary relation $>_{\mathbf{P}}^{\infty}$ defined by

$$X_i >_{\mathbf{P}}^{\infty} X_j \Leftrightarrow Q^{\mathbf{P}}(X_i, X_j) \geq \frac{3}{4}$$

is an asymmetric acyclic relation.

(iii) The transitive closure $>_{\mathbf{P}}$ of $>_{\mathbf{P}}^{\infty}$ is a strict order relation.

Note that the above theorem resolves the dice problem in Section 2. Indeed, for the given example it only holds that $Y >_{\mathbf{P}}^3 Z$, since $\frac{22}{36} < \frac{\sqrt{5}-1}{2} < \frac{23}{36}$, and there is no longer a cycle. The appropriate cutting level in this case is nothing else but the *golden section* $(\sqrt{5}-1)/2$.

As can be expected from the above results, it is not easy to identify the appropriate cutting level for a given copula C leading to an acyclic relation. The above theorem expresses that for the product copula $T_{\mathbf{P}}$ there exists a sequence of cutting levels converging to $3/4$ and guaranteeing that the corresponding relation $>_{\mathbf{P}}^k$ contains no cycles of length k . Although $>_{\mathbf{P}}^{\infty}$ is not transitive in general, its transitive closure yields a strict order relation.

The same can be done for the copula $T_{\mathbf{M}}$, but the results are less exciting.

Theorem 4. Let X_1, X_2, \dots, X_m be m random variables and consider the copula $T_{\mathbf{M}}$. Let $k \in \mathbb{N}$, $k \geq 2$.

(i) The binary relation $>_{\mathbf{M}}^k$ defined by

$$X_i >_{\mathbf{M}}^k X_j \Leftrightarrow Q^{\mathbf{M}}(X_i, X_j) > \frac{k-1}{k}$$

is an asymmetric relation without cycles of length k .

(ii) The binary relation $>_{\mathbf{M}}$ defined by

$$X_i >_{\mathbf{M}} X_j \Leftrightarrow Q^{\mathbf{M}}(X_i, X_j) = 1$$

is a strict order relation.

The above theorem shows that also for $T_{\mathbf{M}}$ there exists a sequence of cutting levels. Unfortunately, here it converges to 1. It is easily seen that $>_{\mathbf{M}}$ is even more demanding than \succ_1 . Finally, note that none of the relations $>_{\mathbf{L}}$, $>_{\mathbf{P}}$ and $>_{\mathbf{M}}$ generalizes the relation \succ_1 .

7 Conclusion

We have developed a general framework for the pairwise comparison of the components of a random vector, expressed in terms of a probabilistic relation. The framework of cycle-transitivity has proven extremely suitable for characterizing the transitivity of this probabilistic relation. This transitivity has been studied for probabilistic relations generated by pairwise independent random variables as well as in the case of dependent random variables, although most of the discussion was focused on coupling by $T_{\mathbf{L}}$ or $T_{\mathbf{M}}$. This study has led to graded as well as non-graded alternatives to the classical concept of stochastic dominance.

Acknowledgments

H. De Meyer is a Research Director of the Fund for Scientific Research - Flanders. This work is supported in part by the Bilateral Scientific and Technological Cooperation Flanders–Hungary BIL00/51 (B-08/2000). Special thanks also goes to EU COST Action 274 named TARSKI: “*Theory and Applications of Relational Structures as Knowledge Instruments*”.

References

1. B. De Baets and H. De Meyer, *Cycle-transitivity versus FG-transitivity*, Fuzzy Sets and Systems, submitted.
2. B. De Baets, H. De Meyer, B. De Schuymer and S. Jenei, *Cyclic evaluation of transitivity of reciprocal relations*, Social Choice and Welfare, to appear.
3. H. De Meyer, B. De Baets and B. De Schuymer, *Extreme copulas and the comparison of ordered lists*, submitted.
4. B. De Schuymer, H. De Meyer, B. De Baets and S. Jenei, *On the cycle-transitivity of the dice model*, Theory and Decision **54** (2003), 264–285.
5. B. De Schuymer, H. De Meyer and B. De Baets, *Cycle-transitive comparison of independent random variables*, submitted.
6. J. García-Lapresta and B. Llamazares, *Aggregation of fuzzy preferences: some rules of the mean*, Social Choice and Welfare **17** (2000), 673–690.
7. E. Klement, R. Mesiar and E. Pap, *Triangular Norms*, Trends in Logic, Studia Logica Library, Vol. 8, Kluwer Academic Publishers, Dordrecht, 2000.
8. E. Klement, R. Mesiar and E. Pap, *Invariant copulas*, Kybernetika **38** (2002), 275–285.
9. H. Levy, *Stochastic Dominance*, Kluwer Academic Publishers, MA, 1998.
10. R. Nelsen, *An Introduction to Copulas*, Lecture Notes in Statistics, Vol. 139, Springer-Verlag, New York, 1998.
11. A. Sklar, *Fonctions de répartition à n dimensions et leurs marges*, Publ. Inst. Statist. Univ. Paris **8** (1959), 229–231.

Optimization with fuzzy random data and its application in financial analysis¹

A.V. Yazenin

Computer Science Department, Tver State University, Ul. Zhelyabova 33,
170000 Tver, Russia. Alexander.Yazenin@tversu.ru.

Abstract.

In the present paper an approaches to the definition of numerical characteristics of fuzzy random variables are analyzed and proposed. Appropriate methods of its calculation, in particular within the framework of shift-scaled representation are obtained. Principles of decision making in fuzzy random environment are formulated. Possibilistic-probabilistic models of portfolio analysis problems and general methods for its solving are developed.

Keywords:

Fuzzy random variable, variance and covariance, possibilistic-probabilistic optimization, financial analysis, portfolio selection

Introduction

Fuzzy random variable is a mathematical model of a probabilistic experiment with fuzzy outcome. A number of works are devoted to its definition and investigation of its properties – see, for example [1-4] and other. Besides of proper definition of fuzzy random variable, definitions of its expected value are introduced, its properties are investigated and methods

¹ This work was carried out with financial support of RFBR (project No. 02-01-011137)

of calculation are proposed for different particular cases. However, series of important questions such as a manner of representation of fuzzy random variable, definition of variance and covariance, development of calculus of fuzzy random variables are still remains open. This circumstance evidently restrains an application of fuzzy random variables for the modeling of such combined type of uncertainty in decision making.

Here we present results that reflect latest achievements in formulated above directions of researching. Representation of a fuzzy random variable allowing to explicate random and fuzzy factors and build an appropriate calculus is considered. Approaches to definition of the moments of the second order are analyzed. Developed mathematical apparatus and fuzzy random variable calculus are oriented for its application in problems of optimization and decision making, in particular in portfolio analysis problems.

The content of the paper is as follows.

In the first part the definition of the possibilistic variable is given. In common it inherits S. Nahmias approach [1, 2]. The possible value distributions of fuzzy random variables are introduced along with several conceptions necessary for following formation. In possibilistic-probabilistic context the definition of a fuzzy random variable and its interpretation are given.

In the second part method of representation of a fuzzy random variable on the ground of shift-scaled family of possibilistic distributions is considered. For such fuzzy random variable representation the calculation methods for expected value, variance and correlation coefficients with one of the approaches to its definition are obtained. As an illustration, it is shown how the calculations can be carried out in general with shift-scaled representation in the class of symmetrical triangular possibilistic distributions.

In the third part the principles and criteria of decision making in fuzzy random environment are formulated.

The fourth section is devoted to developing portfolio analysis problem models with fuzzy random data and to methods of their solution.

In conclusion the represented results and course of further researches are considered.

1. Fuzzy random variables and their distributions in possibility-probability context

Following [1,2] we introduce necessary definitions and notations.

Let Γ be a set of elements denoted as $\gamma \in \Gamma$, $P(\Gamma)$ is a power set of Γ , E^n denotes the n – dimensional Euclidean space.

Definition 1. A possibility measure is a set function $\pi : P(\Gamma) \rightarrow E^1$ with properties:

$$1. \pi\{\emptyset\} = 0, \pi\{\Gamma\} = 1; \quad 2. \pi\left\{\bigcup_{i \in I} A_i\right\} = \sup_{i \in I} \pi\{A_i\},$$

for any index set I and $A_i \in P(\Gamma)$.

Triplet $(\Gamma, P(\Gamma), \pi)$ is a possibilistic space.

Definition 2. A possibilistic (fuzzy) variable is a mapping $Z : \Gamma \rightarrow E^1$. Distribution of possibilistic values of variable Z is function $\mu_Z : E^1 \rightarrow E^1$, defined as

$$\mu_Z(z) = \pi\{\gamma \in \Gamma : Z(\gamma) = z\}, \quad \forall z \in E^1.$$

$\mu_Z(z)$ is a possibility that variable Z may accept value z .

From the last definition and possibility measure properties it follows

$$a) 0 \leq \mu_Z(z) \leq 1, \quad \forall z \in E^1; \quad b) \sup_{z \in E^1} \mu_Z(z) = 1.$$

Definition 3. The support of a possibilistic variable Z is given by $\text{supp}(Z) = \{z \in E^1 / \mu_Z(z) > 0\}$.

Definition 4. r -level set of a fuzzy variable Z is given by $Z_r = \{z \in E^1 / \mu_Z(z) \geq r\}, r \in (0, 1]$.

A necessity measure ν is a dual concept notion to a possibility measure and defined as

$$\nu(A) = 1 - \pi(A^c), \quad \text{where "c" means the complement of a set } A \in P(\Gamma).$$

Taking into consideration results [2], [5], give the definition of a fuzzy random variable and its interpretation.

Let (Ω, B, P) be a probability space.

Definition 5. Fuzzy random variable X is a real function $X(\cdot, \cdot) : \Omega \times \Gamma \rightarrow E^1$, such that for any fixed $\gamma \in \Gamma$ $X_\gamma = X(\omega, \gamma)$ is a random variable on (Ω, B, P) .

From the forecited definition follow two interpretations.

For a fixed $\omega \in \Omega$ we get a fuzzy variable $X_\omega = X(\omega, \gamma)$. The values of a random variable are fuzzy variables with probability distributions $\mu_X(x, \omega)$.

For a fixed γ X_γ can be considered as a random variable with possibility defined by a possibility measure.

Everything becomes clear when distribution $\mu_X(x, \omega)$ is defined as in case of fuzzy variable:

$$\mu_X(x, \omega) = \pi\{\gamma \in \Gamma : X(\omega, \gamma) = x\} \quad \forall x \in E^1.$$

For each ω corresponds possibilistic distribution that is a random choice of an expert who gives an indefinite subjective estimation with certain amount.

X is a random variable for a fixed γ , but we are not sure in its distribution value.

In the context of decision making expected value plays crucial role in explanation of random information. The expected value $E\{X(\omega, \gamma)\}$ of a random variable $X(\omega, \gamma)$ can be defined in different ways.

We define distribution of a random variable expected value according to [2] through averaged random variable:

$$\mu_{EX}(x) = \pi\{\gamma \in \Gamma : E\{X(\omega, \gamma)\} = x\} \quad \forall x \in E^1.$$

It's easy to show an expected value of a fuzzy random variable defined this way has the basic properties of random variable expected value.

2. Fuzzy random variables presentation and calculation of their characteristics

Let's consider fuzzy random variable $X(\omega, \gamma)$. Presentation [6] is interesting for applications:

$$X(\omega, \gamma) = a(\omega) + \sigma(\omega)X_0(\gamma), \quad (2.1)$$

where $a(\omega), \sigma(\omega)$ are random variables defined on probability space $(\Omega, \mathcal{B}, \mathbf{P})$, have finite moments of the second order, and $X_0(\gamma)$ is a fuzzy (possibilistic) variable defined on possibilistic space $(\Gamma, \mathbf{P}(\Gamma), \pi)$.

To simplify demonstration of basic ideas suppose $X_0 \in Tr(0,1)$, that is X_0 has triangular distribution function

$$\mu_{X_0}(t) = \begin{cases} 1 - |t|, & |t| \leq 1, \\ 0, & |t| > 1. \end{cases} \quad (2.2)$$

From (2.2) follows a fuzzy random variable X_0 has modal value and fuzziness coefficient equal, respectively, to 1 and 0.

Presentation (2.1) is shift-scale presentation of a fuzzy random variable. $a(\omega), \sigma(\omega)$ are shift and scale parameters which are modal value and fuzziness coefficient of a fuzzy random variable

$$X_\omega = X(\omega, \gamma) \in Tr(a(\omega), \sigma(\omega)).$$

Let $E(a) = a_0, E(\sigma) = \sigma_0$. Then according to [1,2] $E(X) = a_0 + \sigma_0 X_0$ and

$$\mu_{EX}(t) = \mu_{X_0}((t - a_0) / \sigma_0) \quad \forall t \in E^1.$$

Solving applied problems we are interested not only in expected value but also in variance and covariance of fuzzy random variables. There exist at least two approaches to their definition. The characteristics are fuzzy within the **first** approach [6] and nonfuzzy within the **second** one [7].

Consider the **first** approach. Variance and covariance are defined by probability theory formulae. Using presentation (2.1) we obtain formula for variance $D(X)$ of a fuzzy random variable $X(\omega, \gamma)$ as function of fuzzy variable X_0 :

$$\begin{aligned} D(X) &= D(a + \sigma X_0) = E(a + \sigma X_0 - a_0 - \sigma_0 X_0)^2 = \\ &= E(a - a_0 + (\sigma - \sigma_0) X_0)^2 = D(a) + 2 \text{cov}(a, \sigma) X_0 + D(\sigma) X_0^2 = \\ &= D(\sigma) \left[X_0 + \frac{\text{cov}(a, \sigma)}{D(\sigma)} \right]^2 + \frac{D(a) D(\sigma) - \text{cov}^2(a, \sigma)}{D(\sigma)}. \end{aligned} \quad (2.3)$$

Let

$$C_1^2 = D(\sigma); C_2 = \frac{\text{cov}(a, \sigma)}{D(\sigma)}, C_3 = \frac{D(a) D(\sigma) - \text{cov}^2(a, \sigma)}{D(\sigma)}.$$

Then formula (2.3) is the following $D(X) = C_1^2 [X_0 + C_2]^2 + C_3$. In force of Cauchy-Bunyakovski inequality $C_3 \geq 0$. Complexity of variance definition depends on fuzzy variable distribution.

The case when $C_2 = C_3 = 0$ will be considered for illustration. Then $D(X) = C_1^2 X_0^2$. The following result is valid for triangular symmetric possibilistic distribution.

Theorem 1 [6]. Let $X_0 \in Tr(0,1)$. Then

$$\mu_{D(X)}(t) = \begin{cases} 1 - \sqrt{t}/C_1, & \text{if } 0 < t < C_1^2 \\ 0, & \text{if } t \notin (0, C_1^2). \end{cases}$$

To describe collective behavior of $E(X)$ and $D(X)$ it's convenient to introduce parametric description. Let parameter $t \in \text{supp}(X_0)$. Then pair $(E(X), D(X))$ accepts value $(a_0 + \sigma_0 t, C_1^2(t + C_2)^2 + C_3)$ with possibility $\mu_{X_0}(t)$.

Based on the results obtained describe the collective behavior of fuzzy random variables X_1, X_2, \dots, X_n . We come to the following model:

$X_k(\omega, \gamma) = a_k(\omega) + \sigma_k(\omega)X_k^0(\gamma)$, where (X_1^0, \dots, X_n^0) is a fuzzy vector.

Introduce the following notations:

$$\begin{aligned} a_k^0 &= E(a_k), \quad \sigma_k^0 = E(\sigma_k), \\ C_k^2 &= D(\sigma_k), \quad C_{ij} = \text{cov}(\sigma_i, \sigma_j), \quad f_{ij} = -\frac{\text{cov}(\sigma_i, a_j)}{\text{cov}(\sigma_i, \sigma_j)}, \\ d_{ij} &= \text{cov}(a_i, a_j) - \frac{\text{cov}(\sigma_i, a_j) \cdot \text{cov}(\sigma_j, a_i)}{\text{cov}(\sigma_i, \sigma_j)}. \end{aligned}$$

Applying ordinary random variables numeric characteristics calculation rules we obtain formulae that represent characteristics of fuzzy random variables X_1, \dots, X_n :

$$m_k = E(X_k) = a_k^0 + \sigma_k^0 \cdot X_k^0; \quad (2.4)$$

$$D_k^2 = D(X_k) = C_k^2 [X_k^0 + f_{kk}]^2 + d_{kk}; \quad (2.5)$$

$$\Sigma_{ij} = \text{cov}(X_i, X_j) = C_{ij} (X_i^0 - f_{ij})(X_j^0 - f_{ji}) + d_{ij} \quad (2.6)$$

Let $m = (m_1, \dots, m_n)$ be a mean vector, $\Sigma = (\Sigma_{ij})$ be a covariance matrix of fuzzy vector $(X_1(\omega, \gamma), \dots, X_n(\omega, \gamma))$. Calculate collective possibilistic distribution of m and Σ . Let $t = (t_1, \dots, t_n)$ be a point from a set of possible values of a fuzzy vector $X^0 = (X_1^0, \dots, X_n^0)$. According to forecited results pair (m, Σ) accepts value $(m(t), \Sigma(t))$ with possibility $\mu_{X^0}(t)$. Elements $m_k(t)$ and $\Sigma_{ij}(t)$ can be calculated by formulae (2.4) - (2.6).

Actually, as $X_i^0 = t_i$, $X_j^0 = t_j$, then

$$m_k(t) = a_k^0 + \sigma_k^0 \cdot t_k; \quad \Sigma_{ij}(t) = C_{ij} (t_i - f_{ij})(t_j - f_{ji}) + d_{ij}.$$

If elements of the vector X^0 are min-related [5] then

$$\mu_{X^0}(t) = \min_{1 \leq i \leq n} \{\mu_{X_i^0}(t_i)\}.$$

The **second** approach. Omitting all the technical details connected with fuzzy random variable values definition in space L_∞ [7] in accepted notation, the corresponding formulae are:

$$\text{cov}(X, Y) = \frac{1}{2} \int_0^1 (\text{cov}(X_\omega^-(r), Y_\omega^-(r)) + \text{cov}(X_\omega^+(r), Y_\omega^+(r))) dr,$$

where $X_\omega^-(r), Y_\omega^-(r), X_\omega^+(r), Y_\omega^+(r)$ are r -level set endpoints of fuzzy variables X_ω, Y_ω respectively. It's obvious variance $D(X) = \text{cov}(X, X)$ and moments of the second order are without fuzziness. It's important that the definition methods of the second order moments in the **first** and **second** approaches are different in principle. In the **first** approach we identify possibility distribution and in the **second** one we make numeric calculations.

3. Possibility-probability optimization models and decision making

Within fuzzy random data functions that form goals and restrictions of decision making problem make sense of mapping $R_i(\cdot, \cdot, \cdot): W \times \Omega \times \Gamma \rightarrow E^1, i = \overline{0, m}$, where W is a set of acceptable solutions, $W \subset E^n$. Thus, a set of acceptable outcomes can be obtained by combination of solution set elements with elements of random and fuzzy parameter sets. That's why any concrete solution can't be directly connected either with goal achievement degree no with restriction system execution degree.

Existence of two different types of uncertainty in efficiency function complexifies reasoning and formalization of solution selection optimality principles. However decision making procedure based on **expected possibility** principle is quite natural. Its content is elimination of two types of uncertainty that is a realization of two types of decision making principles [5], [8,9]:

- **averaging of fuzzy random data that allows to get to decision making problem with fuzzy data;**
- **choice of optimal solution with more possible values of fuzzy parameters or with possibility not lower than preset level.**

Adequate mean of suggested optimality principle formalization is a mathematical apparatus of fuzzy random variables.

Let τ be a possibility or necessity measure that is $\tau \in \{\pi, \nu\}$. Taking into consideration stated decision making principles we came to the following optimization problem settings within fuzzy random factors.

Problem of maximizing goal achievement measure with liner possibility (necessity) restrictions

$$\begin{cases} \tau\{ER_0(w, \omega, \gamma) \mathfrak{R}_0 \ 0\} \rightarrow \max, \\ \tau\{ER_i(w, \omega, \gamma) \mathfrak{R}_i \ 0\} \geq a_i, \ i = \overline{1, m}, \\ w \in W. \end{cases}$$

Problem of level optimization with liner possibility (necessity) restrictions

$$k \rightarrow \max,$$

$$\begin{cases} \tau\{ER_0(w, \omega, \gamma) \mathfrak{R}_0 k\} \geq a_0, \\ \tau\{ER_i(w, \omega, \gamma) \mathfrak{R}_i 0\} \geq a_i, i = \overline{1, m}, \\ w \in W. \end{cases}$$

In the stated problems $\mathfrak{R}_0, \mathfrak{R}_i$ are binary relations, $\mathfrak{R}_0, \mathfrak{R}_i \in \{\leq, \geq, =\}$, $\alpha_i \in (0, 1]$, k is an additional (level) variable.

Possibility-probability optimization models introduced define an approach to portfolio analysis model construction with combination of fuzzy and random uncertainties.

4. Models and methods of portfolio analysis in fuzzy random environment

Portfolio selection problem [10] is a central problem of financial and investment analysis. It's still interesting for researchers. As some researchers equitably denote the main drawback of Markowitz approach to portfolio selection problem is an absence of statistic data which are used for model parameters estimation. Expert estimations are used in such situations. Possibility and fuzzy sets theory gave further impetus to problem developing [11,12]. In [13,14] analysis of portfolio analysis problems is conducted when appropriate probability characteristics of financial market model according to Markowitz [10] are replaced by fuzzy expert estimations.

However financial market is instable and changeable so investment decision making leans on both expert estimations which are tolerant and fuzzy and statistic information. In some instances profitabilities and prices of separate financial assets are characterized by tolerant time series. In this case a fuzzy random variable is an adequate model of profitability.

4.1. Expected value and risk of portfolio with fuzzy random data

Let $R_i(\cdot, \cdot): \Omega \times \Gamma \rightarrow E^1$ be a fuzzy random variable that represents profitability of an i -asset and Γ are elements of probability space $(\Omega, \mathcal{B}, \mathcal{P})$ and possibilistic space $(\Gamma, \mathcal{P}(\Gamma), \pi)$ respectively. Then expected value of portfolio is a fuzzy random variable

$$R_p(w, \omega, \gamma) = \sum_{i=1}^n w R_i(\omega, \gamma).$$

Here $w = (w_1, \dots, w_n)$ is the vector representing the portfolio:

$$w \geq 0, \quad \sum_{i=1}^n w_i = 1.$$

Expected profit and risk of the portfolio under fixed w are presented by following fuzzy variables:

$$\widehat{R}_p(w, \gamma) = ER_p(w, \omega, \gamma), \quad \widehat{V}_p(w, \gamma) = E(R_p(w, \omega, \gamma) - \widehat{R}_p(w, \gamma))^2.$$

Hereinafter we assume that considering fuzzy random variables can be presented in following form:

$$R_i(\omega, \gamma) = a_i(\omega) + \sigma_i(\omega)X_i(\gamma),$$

where $a_i(\omega), \sigma_i(\omega)$ are random variables, defined on probabilistic space $(\Omega, \mathcal{B}, \mathcal{P})$ with mathematical expectations a_i^0, σ_i^0 ; $X_i(\gamma)$ is a fuzzy variable, defined on possibilistic space $(\Gamma, \mathcal{P}(\Gamma), \pi)$.

After calculation of the mathematical expectation characteristics of portfolio profit (expected profit and risk) takes form

$$\widehat{R}_p(w, \gamma) = \sum_{i=1}^n (a_i^0 + \sigma_i^0 X_i(\gamma)) w_i,$$

$$\widehat{V}_p(w, \gamma) = \sum_{i=1}^n D(R_i) w_i^2 + \sum_{k=1}^n \sum_{\substack{j=1 \\ k \neq j}}^n w_k w_j \text{cov}(R_k, R_j),$$

where $D(R_i)$ is the dispersion of fuzzy random variable $R_i(\omega, \gamma)$, $\text{cov}(R_k, R_j)$ is a covariance coefficient of fuzzy random variables $R_k(\omega, \gamma)$, $R_j(\omega, \gamma)$.

In accordance with results obtained earlier these characteristics are functions of fuzzy variables therefore are fuzzy variables. Let below

$$\hat{R}_i(\gamma) = a_i^0 + \sigma_i^0 X_i(\gamma);$$

$$d_i(\gamma) = D(R_i) = C_i^2 [X_i - f_{ii}]^2 + d_{ii};$$

$$\Theta_{kj}(\gamma) = \text{cov}(R_k, R_j) = C_{kj} (X_k - f_{kj})(X_j - f_{jk}) + d_{kj}.$$

4.2. Basic models of portfolio analysis in probabilistic-possibilistic context

Basing on results presented in section 3 and classical results [10] we propose the generalized models of portfolio analysis oriented on processing of fuzzy random data. Consider the following as basic ones.

Maximum effectiveness portfolio:

$$\begin{aligned} k &\rightarrow \max, \\ \tau\{\tilde{R}_p(w, \gamma) \mathfrak{R}_0 k\} &\geq \pi_0, \\ \left\{ \begin{array}{l} \tau\{\tilde{V}_p(w, \gamma) \mathfrak{R}_1 r_p(\gamma)\} \geq \pi_1, \\ \sum_{i=1}^n w_i = 1, \quad w_1, \dots, w_n \geq 0. \end{array} \right. \end{aligned}$$

Minimal risk portfolio:

$$\begin{aligned} k &\rightarrow \min, \\ \tau\{\tilde{V}_p(w, \gamma) \mathfrak{R}_0 k\} &\geq \pi_0, \end{aligned}$$

$$\begin{cases} \tau \{ \tilde{R}_p(w, \gamma) \mathfrak{R}_1 m_p(\gamma) \} \geq \pi_1, \\ \sum_{j=1}^n w_j = 1, \quad w_1, \dots, w_n \geq 0. \end{cases}$$

Maximization of possibility (necessary) achievement of portfolio acceptable profitability level:

$$\begin{cases} \tau \{ \hat{R}_p(w, \gamma) \mathfrak{R}_0 m_p(\gamma) \} \rightarrow \max, \\ \tau \{ \hat{V}_p(w, \gamma) \mathfrak{R}_1 r_p(\gamma) \} \geq \pi_0, \\ \sum_{i=1}^m w_i = 1, \quad w_1, \dots, w_n \geq 0, \end{cases}$$

In presented models $\tau \in \{\pi, \nu\}$, m_p is a fuzzy profitability level acceptable for investor, $r_p(\gamma)$ is a level of possible risk, $\pi_0, \pi_1 \in (0, 1]$ are given levels of possibility (necessity).

4.3. Solving methods

In papers [16-18] solving methods for portfolio analysis problems are developed in correspondence with models presented in section 4.2 with possibility measure. These models are the following:

Maximum effectiveness portfolio:

$$k \rightarrow \max, \quad (4.1)$$

$$\pi \{ \tilde{R}_p(w, \gamma) = k \} \geq \pi_0, \quad (4.2)$$

$$\begin{cases} \pi \{ \tilde{V}_p(w, \gamma) \leq r_p(\gamma) \} \geq \pi_1, \\ \sum_{i=1}^n w_i = 1, \quad w_1, \dots, w_n \geq 0. \end{cases} \quad (4.3)$$

Minimal risk portfolio:

$$k \rightarrow \min, \quad (4.4)$$

$$\pi \{ \tilde{V}_p(w, \gamma) = k \} \geq \pi_0, \quad (4.5)$$

$$\begin{cases} \pi \{ \tilde{R}_p(w, \gamma) \geq m_p(\gamma) \} \geq \pi_1, \\ \sum_{j=1}^n w_j = 1, \quad w_1, \dots, w_n \geq 0. \end{cases} \quad (4.6)$$

Maximization of possibility achievement of portfolio acceptable profitability level:

$$\pi \{ \hat{R}_p(w, \gamma) = m_p(\gamma) \} \rightarrow \max, \quad (4.7)$$

$$\begin{cases} \pi \{ \hat{V}_p(w, \gamma) \geq r_p(\gamma) \} \geq \pi_0, \\ \sum_{i=1}^m w_i = 1, \quad w_1, \dots, w_n \geq 0. \end{cases} \quad (4.8)$$

The essence of solution methods developed in [16-18] for problems of portfolio analysis consists in construction of its equivalent deterministic analogues. Such methods can be classified as indirect ones. Further we formulate the most important results of these works.

Theorem 2. Let in problem (4.1)-(4.3) random variable are unrelated and characterized by covariance matrix $\{\Theta_{kl}(\gamma)\}_{k,l=1}^n$, fuzzy variables $X_i(\gamma) \in Tr(m_i, 1)$, $i = 1, \dots, n$; $r_p(\gamma) \in Tr(\bar{m}, \bar{d})$ are min-related. Then problem (4.1)-(4.3) is equivalent to

$$\sum_{i=1}^n X_i^+(\pi_0) w_i \rightarrow \max,$$

$$\begin{cases} \sum_{i=1}^n d_i^-(\pi_1) w_i + \sum_{\substack{k,l=1 \\ k \neq l}}^n \Theta_{kl}^-(\pi_1) w_k w_l \leq r_p^+(\pi_1), \\ \sum_{i=1}^n w_i = 1, w_1, \dots, w_n \geq 0, \end{cases}$$

where $d_i^-(\pi_1), \Theta_{kl}^-(\pi_1)$ are the left endpoints of level sets of appropriate fuzzy variables,

$$r_p^+(\pi_1) = \bar{m} + \bar{d}(1 - \pi_1), X_i^+(\pi_0) = a_i^0 + \sigma_i^0(1 - \pi_0).$$

Theorem 3. Let in problem (4.4)-(4.6) probabilistic variables are unrelated and characterized by covariance matrix $\{\Theta_{kl}(\gamma)\}_{k,l=1}^n$, fuzzy variables $X_i(\gamma) \in Tr(m_i, 1), i = 1, \dots, n; m_p(\gamma) \in Tr(\bar{m}, \bar{d})$ are min-related. Then problem (4.4)-(4.6) is equivalent to

$$\sum_{i=1}^n d_i^-(\pi_0) w_i + \sum_{\substack{k,j \\ k \neq j}}^n \Theta_{kl}^-(\pi_0) w_k w_j \rightarrow \min,$$

$$\begin{cases} \sum_{i=1}^n X_i^+(\pi_0) w_i \geq m_p^-(\pi_1), \\ \sum_{i=1}^n w_i = 1, w_1, \dots, w_n \geq 0, \end{cases}$$

where $d_i^-(\pi_1), \Theta_{kl}^-(\pi_1)$ are the left endpoints of level sets of appropriate fuzzy variables,

$$m_p^-(\pi_1) = \bar{m} + \bar{d}(\pi_1 - 1), X_i^+(\pi_0) = a_i^0 + \sigma_i^0(1 - \pi_0).$$

Theorem 4. Let in problem (4.7), (4.8) fuzzy variables $m_p(\gamma), r_p(\gamma), X_i(\gamma)$ are convex, min-related and characterized by upper semi-continuous distributions with finite supports. Then problem (4.7), (4.8) is equivalent to

$$x_0 \rightarrow \max,$$

$$\left\{ \begin{array}{l} x_0 \leq \mu_{\hat{R}_i}(v_i), \quad i = 1, \dots, n; \\ x_0 \leq \mu_{m_p}(t), \\ \sum_{i=1}^n w_i v_i = t, \\ \sum_{i=1}^n d_i^-(\pi_0) w_i + \sum_{\substack{k,l=1 \\ k \neq l}}^n \Theta_{kl}^-(\pi_0) w_k w_l \leq r_p^+(\pi_0), \\ \sum_{i=1}^n d_i^+(\pi_0) w_i + \sum_{\substack{k,l=1 \\ k \neq l}}^n \Theta_{kl}^+(\pi_0) w_k w_l \geq r_p^-(\pi_0), \\ \sum_{i=1}^n w_i = 1, \quad w_1, \dots, w_n \geq 0, \end{array} \right.$$

where $\mu_{\hat{R}_i}$, μ_{m_p} are distribution functions of appropriate fuzzy variables; $d_i^-(\pi_0)$, $d_i^+(\pi_0)$, $\Theta_{kl}^-(\pi_0)$, $\Theta_{kl}^+(\pi_0)$, $r_p^-(\pi_0)$, $r_p^+(\pi_0)$ are endpoints of π_0 -level sets of fuzzy variables $d_i(\gamma)$, $\Theta_{ij}(\gamma)$ and r_p .

To prove the formulated above theorems mathematical apparatus developed in [19,20] can be used.

Conclusion

In the present paper the approach to analysis of portfolio selection problems based on possibilistic-probabilistic optimization is described. Principles of decision making in fuzzy random environment are formulated. It gives ground for developed generalized models of portfolio analysis. Indirect methods of portfolio optimization with that models in fuzzy random environment are represented. It is based on construction of the equivalent determined analogues. Its realization can be carried out in frames of quadratic and, in some cases, of separable programming. Representation of fuzzy random data is implemented on the basis of shift-scaled family of possibilistic variables distributions. It allows the explication of the probability on the level of shift and scale parameters. This model of fuzzy random variable is convenient for applications. The calculus of fuzzy random

variables is presented with definition of its moments of the second order in fuzzy form. However, in the frames of proposed schema of possibilistic-probabilistic optimization the models of portfolio analysis and optimization methods can be developed in case the second order moments of fuzzy random variables are defined in certain form. As a direction of a further research a comparative investigation of these two designated approaches and determination of the bounds of its adequate application can be considered.

References

1. S.Nahmias. Fuzzy variables, Fuzzy sets and systems 1(1978) 97-110.
2. S.Nahmias. Fuzzy variables in random environment, In: Gupta M.M. et. al. (eds.). Advances in fuzzy sets Theory. NHCP. 1979.
3. H.Kwakernaak. Fuzzy random variables – 1.Definitions and theorems, Inf. Sci.. 15(1978) 1-29.
4. M.D.Puri, D.Ralesky. Fuzzy random variables, J. Math. Anal. Appl. 114(1986) 409-422.
5. A.V.Yazenin, M.Wagenknecht, Possibilistic optimization. Brandenburgische Technische Universitat. Cottbus. Germany. 1996.
6. M.Yu.Khokhlov, A.V.Yazenin. The calculation of numerical characteristics of fuzzy random data, Vestnik TvGU, No.2. Series «Applied mathematics and cybernetics», 2003, pp.39-43.
7. Y.Feng, L.Hu, H.Shu. The variance and covariance of fuzzy random variables, Fuzzy sets and systems 120(2001) 487-497.
8. A.V.Yazenin. Linear programming with fuzzy random data, Izv. AN SSSR. Tekhn. kibernetika 3(1991) 52-58.
9. A.V.Yazenin. On the method of solving the linear programming problem with fuzzy random data, Izv. RAN, Teoriya i sistemy upravleniya 5(1997) 91-95.
- 10.H.Markowitz. Portfolio selection: efficient diversification of investments. Wiley. New York, 1959.
- 11.L.A.Zadeh. Fuzzy sets as a basis for a theory of possibility, Fuzzy sets and systems 1(1978) 3-28.
- 12.Dubois D., Prade H. Possibility theory: an approach to computerized processing of uncertainty, Plenum press, 1988.
- 13.M. Inuiguchi, J.Ramik. Possibilistic linear programming: a brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem, Fuzzy sets and systems 111(2000) 3-28.

- 14.M. Inuiguchi, T.Tanino. Portfolio selection under independent possibilistic information, Fuzzy sets and systems 115(2001) 83-92.
- 15.I.A.Yazenin. Minimal risk and efficiency portfolios for fuzzy random data, XXI Seminar on stability problems of stochastic models. Abstracts, Eger, Hungary, 2001, P. 182.
- 16.I.A.Yazenin, Minimal risk portfolio and maximum effectiveness portfolio in fuzzy random environment, Slozhnye sistemy: modelirovanie I optimizatsiya, Tver, TvGU, 2001, pp.59 – 63.
- 17.I.A.Yazenin, On the methods of investment portfolio optimization in fuzzy random environment, Slozhnye sistemy: modelirovanie I optimizatsiya, Tver, TvGU, 2002, pp. 130 – 135.
- 18.I.A.Yazenin, On the model of investment portfolio optimization, Vestnik TvGU, No.2. Series «Applied mathematics and cybernetics», 2003, pp. 102 – 105.
- 19.A.V.Yazenin, On the problem of maximization of attainment of fuzzy goal possibility, Izv. RAN, Teoriya i sistemy upravleniya 4(1999) pp.120 – 123.
- 20.A.V.Yazenin, On the problem of possibilistic optimization, Fuzzy sets and systems 81(1996) 133-140.

Machine Learning in Fuzzy Environment

Prof., dr.habil.sci.comp. Arkady Borisov
Institute of Information Technology
Riga Technical University
1 Kalku Street, Riga LV-1658, Latvia
E-mail:aborisov@egle.cs.rtu.lv

Conditional rule mining (generation) exemplifies one of successful applications of machine learning. The use of conditional rules is caused by simplicity of input-output relationship description of the task (object) under study. At the same time, the presence of a set of subject area rules is sufficient for solving applied tasks in different areas. First of all, these are tasks of classification (diagnostics), quality control, motion planning, etc.

The main technique of deriving rules of that kind is inductive inference. During the past years a number of effective inductive algorithms were developed. These are, first of all, decision tree generation based rule acquisition techniques suggested in [1,2] as well as the algorithms described in [3,4] that enable one to derive conditional rules escaping the stage of decision tree construction.

One of principal problems in decision tree construction is their size and depth. Major efforts in that area are directed towards obtaining algorithms for constructing trees of a small size that at the same ensure high classification quality. The algorithms differ in computation method of the most informative attribute of the initial data table. Basic methods are entropy measure, information gain, chi-square criterion, GINI index of diversity, gain-ratio, etc.

However, traditional inductive learning methods have proved to be useless for those applied areas that contain vagueness and ambiguity. To cope with that drawback, fuzzy inductive learning algorithms [5] have been developed. They differed in methods of attribute informativeness determination of the initial instance table. One of the first methods suggested were entropy analogue, minimum ambiguity, etc. Besides that, fuzzy rule generation methods were suggested that were directly based on the table of initial fuzzy data [6].

The aforementioned methods have prospects not only in static tasks. Their main task is to serve as self-learning blocks in dynamic conditions.

References

1. Quinlan J.R. (1993). *C4.5: Program for Machine Learning*, Morgan Kaufman, San Mateo, CA.
2. Breiman L., Friedman J.H., Olshen R.A., Stone C.J. (1984). *Classification and Regression Tree*, Wadsworth International Group, Belmont, CA.
3. Michalski R.S., Chilausky R.L. (1980). Knowledge acquisition by encoding expert rules versus computer induction from examples: a case study involving soybean pathology. *International Journal of Man-Machine Studies*, N 12, P. 63-87.
4. Clark P., Niblett T. (1989). The CN2 induction algorithm. *Machine Learning*, 3(4), Kluwer, P.261-283.
5. Yuan Y., Shaw M.J. (1995). Induction of fuzzy decision trees. *Fuzzy Sets and Systems*, Vol.69, N2, P. 125-139.
6. Ching-Hung Wang, Jau-Fu Liu, Tzung-Pei Hong, Shian-Shyong Tseng. (1999). A fuzzy inductive learning strategy for modular rules. *Fuzzy Sets and Systems 103*, P. 91-105.

Non-Stochastic-Model Based Finance Engineering

Kaoru Hirota (Tokyo Institute of Technology)

Toshihiro Kaino (Aoyama-Gakuin University)

e-mail: hirota@hrt.dis.titech.ac.jp

Most of the models in the field of finance engineering are proposed based on the stochastic theory, e.g., the well known option pricing model proposed by F. Black and M. Scholes in 1973 is premised on following log normal distribution by the underlying price. Many researchers have also pointed out that this assumption is not always valid for real world financial problems. Although various kinds of improvements have been done, there still exists an application limit with respect to the statistical distribution and the additivity of probability measure, e.g., in evaluation of venture, small, and medium companies, underlying assets are a company and an enterprise, the distribution of the value of underlying assets is not a probability distribution.

A new corporate evaluation model that is able to deal with ambiguous and discrete data better is proposed based on Choquet Integral to overcome the gap mentioned above. First the differentiation of the Choquet integral of a nonnegative measurable function with respect to a fuzzy measure on a fuzzy measure space is proposed and it is applied to the capital investment decision-making problem. Then the differentiation of the Choquet integral of a nonnegative measurable function is extended to differentiation of the Choquet integral of a measurable function, and its properties are shown. The Choquet integral is applied to the long-term debt ratings model, where the input is qualitative and quantitative data of the corporations, and the output is the Moody's long-term debt ratings. The fuzzy measure, that is given as the importance of an each qualitative and quantitative data, is

derived from a neural net method. Moreover, differentiation of the Choquet integral is applied to the long-term debt ratings, where this differentiation indicates how much evaluation of each specification influences to the rating of the corporation.

**FUZZY DATA MINING IN
ECONOMICS AND FINANCE**

Plenary Report

Mining Fuzzy Association Rules and Networks in Time Series Databases
Batyrshin I., Herrera-Avelar R., Sheremetov L., Suarez R.

Perceptual Time Series Data Mining

A clear view on quality measures for fuzzy association rules

De Cock Martine, Cornelis Chris, Kerre Etienne

Moving Approximations in Time Series Data Mining

Batyrshin I., Herrera-Avelar R., Sheremetov L., Suarez R.

On Qualitative Description of Time Series Based on Moving Approximations

Batyrshin I., Herrera-Avelar R., Sheremetov L., Suarez R.

Generating Fuzzy Rules for Financial Time Series by Neural Networks with Supervised Competitive Learning Techniques

Marček Dušan

Pattern Recognition through Perceptually Important Points in Financial Time Series

Zaib Gul, Ahmed Uzair and Ali Arshad

Fuzzy Classification and Pattern Recognition

Soft Clustering for Funds Management Style Analysis: Out-of-sample Predictability

Lajbcygier Paul and Yahya Asjad

Tuning the Fuzzy Classification Models with Various Learning Criteria: the Case of Credit Data Classification

Shtovba Serhiy, Pankevich Olga, and Dounias Georgios

prInvestor: Pattern Recognition based Financial Time Series Investment System

Ruta Dymitr

On general Scheme of Invariant Clustering Procedures Based on Fuzzy Similarity Relation

Batyrshin I.Z., Rudas T. and Klimova A.

Evolutionary Procedures of Visualization of Multidimensional Data

Angelica Klimova

Mining Fuzzy Association Rules and Networks in Time Series Databases

Batyrshin^{1,2} I., Herrera-Avelar¹ R., Sheremetov^{1,3} L., Suarez¹ R.

¹Mexican Petroleum Institute, Mexico

²Institute of Problems of Informatics, Academy of Sciences of Tatarstan, Russia

³St. Petersburg Institute for Informatics and Automation Russian Academy of Sciences (SPIIRAS), Russia

batyr@imp.mx, batyr1@hotmail.ru

Abstract. A new approach to time series data mining (TSDM) is proposed in this paper. Time series database (TSDB) is considered as a description of dynamics of some system and the goal of TSDM is to find relationships between dynamics of the elements of this system. As such a system it may be considered economic, financial, industrial or natural (Earth Science) systems with the elements described by the sets of time series. The intersystem relationships are formulated as association rules in linguistic form and may be used in perception based reasoning. The general approach to association rules extraction from (crisp) TSDB is based on a consideration of partial (conditional) time series defined in general by perceptual (fuzzy) conditions fulfilled for some time series. Several types of perceptual association rules based on this approach to TSDM are considered in this paper. A new type of association rules based on correlations between partial time series is proposed. These association rules define association network on the elements of the system which may be considered together with spatial, causal, physical etc relations existing in the system. The proposed methods are demonstrated on examples of TSDM in economics and petroleum industry. It is argued that TSDB give regular resources for mining perception based information and for construction, examination and tuning of fuzzy models of knowledge extraction and processing in perception based reasoning.

1. Introduction

The success of fuzzy logic applications in control, technical systems modeling and pattern recognition is based on synergism of linguistic descriptions and numerical data available for these application areas. Fuzzy logic serves here as a bridge between linguistic and numerical information. One of the reason of the success of fuzzy logic applications in these areas is the existence of regular resources of numerical data obtained from traditional mathematical models, experiments or measurements which can used as a basis for construction, examination and tuning of fuzzy models. In his recent works and lectures Lotfi Zadeh called attention to decision making applications of fuzzy logic in economics, finance, Earth sciences etc. with the central role of human perceptions [13-15]. Perception based propositions like *“The price of gas is low and declining”*, *“It is very unlikely that there will be a significant increase in the price of oil in the near future”* etc. are used by peoples in many decision making procedures. Usually perceptions use fuzzy granulation of information obtained from observations, measurements, life experience, mathematical analysis, visual perceptions about curves etc [13, 6]. The formation of perceptions is a process of knowledge extraction from different resources. In this paper we propose the formal methods of extracting such knowledge from time series databases (TSDB). The approach is based on methods developed in data mining [2, 7, 8, 11, 12, 16] for generation of association rules and new methods for generation of crisp and fuzzy association rules from (crisp) TSDB are considered in this paper.

The paper proposes new approach to application of fuzzy logic in economics, finance, medicine, biology, natural and Earth sciences based on analysis of time series data bases existing in these areas. The terabytes of information were collected here during many years and TSDB can give regular and unified resources for fuzzy logic applications in these areas. Linguistic descriptions of patterns of time series like *“High value”*, *“Quickly increasing”*, *“End of the day”*, *“Small time period”* etc have fuzzy nature and may be described by fuzzy sets defined on the time domain and on the sets of time series values. TSDB can be used also as regular recourses for construction, examination and tuning of perceptual models of knowledge extraction and reasoning.

The concept of association rule is one of the basic concepts of data mining (DM) [2, 11]. One of the well-studied problems in DM is the search for association rules in market basket data. The database in this problem contain records of items bought by customers and the goal of DM is to discover buying patterns of two or more items (e.g. “beer” and “chips”) often bought together.

Such patterns are usually represented in the form of association rules like “If beer (is bought) then chips (is bought), (W)” where W is a measure of association between these items usually given by support and confidence measures evaluating how often these items bought together.

In spite of databases (DB) with market basket structure which contain transactions or records with subsets of items the time series data bases (TSDB) contain time series (TS) consisting of sequences of values of some parameter given for increasing time moments. The methods of association rules generation for TSDB usually search the patterns in time series which often followed by other patterns [7, 8].

Two types of TSDB may be differentiated. One type of TSDB contains results of measurement of some parameter or attribute values during given time interval for objects sampled from some population. For example it may be TS of cardiograms measured for the sample of patients, or the TS of annual profit of companies sampled from some sector of economics. Usually in TSDM such type of TSDB is considered and the goal of TSDM consists in extraction of association rules with sequences of patterns often presented in TS, classification of TS (and hence objects) on classes with similar shapes of TS, description of these classes in terms of patterns etc [7, 8, 16]. The methods of solution of such type of problems based on a new technique of moving (sliding) approximations will be considered also in [5]. The traditional problem of TS modeling and forecasting its value also considered in TSDM.

The second type of TSDB which will be considered in this paper contains results of measurements of different parameters measured at the same time moments. We propose a new approach to time series data mining based on consideration of such TSDB as information describing some system of elements evolving in time. As such a system it may be considered the set of countries, regions, industries, companies, petroleum wells etc described by the sets of time series of economical or financial indicators, physical parameters or attributes. Time series data mining (TSDM) is considered in such context as an extraction of information about relationships between elements of this system existing during the evolution of the system. These relationships may be extracted in the form of association rules tying together perceptual patterns of TS for different attributes of the system elements. Such perceptual patterns may be given by crisp or fuzzy granulation of time moments and time series values. Perceptual association rules will define an association network on the set of subsystems which may be considered as information additional to spatial, causal, physical and other relationships existing in the system.

If for the first type of TSDB the association rules analyze mainly associations between patterns which often registered together in one time series in neighboring time intervals, for the second type of TSDB the association rules analyze associations between patterns in different time series which happen at the same time intervals. For a mixture of two considered types of TSDB the combination of two types of associations may be studied.

Our approach is based on the analysis of associations between partial or conditional TS given as sets of parts of TS where some condition is fulfilled. The partial TS analysis gives possibility to find associations which are hidden or not clear in the analysis of the entire time series. We demonstrate this approach on the examples of system of economic indicators and system of petroleum wells given by TS of oil and gas production.

The paper is organized as follows. In Section 2 and 3 we consider partial and conditional time series and methods of their definition by simple and composite perceptual patterns (conditions) fulfilled for elements of these time series. In Section 4 we consider the methods of construction of perceptual association rules based on support and confidence measures and demonstrate them on the example of the system of economic indicators. In Section 5 we consider association rules and networks based on partial correlation analysis. In conclusion we discuss the main results and the future directions of the work.

2. Partial time series

Suppose $t = (t_1, \dots, t_n)$, ($n > 1$) is a sequence of time moments such that $t_i < t_{i+1}$ for all $i = 1, \dots, n-1$. A time series (TS) y is a sequence of real numbers $y = (y_1, \dots, y_n)$ corresponding to the sequence of time moments t . Denote $I = (1, \dots, n)$ the sequence of indexes of TS. Time series are usually obtained as a result of measurement of some parameter (economical, physical, etc) followed at equal intervals of time. In this case for simplicity it is supposed that $t = I = (1, \dots, n)$.

A window W_i of a length $k - 1$ is a sequence of indexes $W_i = (i, i+1, \dots, i+k-1)$, if $k > 1$, and $W_i = (i)$ if $k = 1$. Denote $y_{W_i} = (y_i, y_{i+1}, \dots, y_{i+k-1})$ the corresponding values of TS y in the window W_i . Suppose $J = (W_{i_1}, \dots, W_{i_m})$ is a sequence of windows such that $i_1 < \dots < i_m$. A sequence $y_J = (y_{W_{i_1}}, \dots, y_{W_{i_m}})$ will be called a partial time series (PTS) of y defined by the sequence of windows J .

Two special cases of partial time series will be considered. If the lengths of all windows in J are equal to 1 then $J = (i_1, \dots, i_m)$, $i_1 < \dots < i_m$, and the partial

time series $y_J = (y_{i_1}, \dots, y_{i_m})$ is a subsequence of TS y . Such partial TS were discussed in [4]. Suppose $J = (W_1, W_2, \dots, W_{n-k+1})$ is a sequence of all windows of the length k , ($1 \leq k \leq n$). When $1 < k < n$ the set J contains a set of “moving” or “sliding” windows. Such moving windows are used in statistics in moving average procedure for smoothing time series when the value in the middle of the window replaced by the mean of values from this window [9]. Moving windows will be considered in data mining procedures based on moving approximations [5]. When $k = 1$ then J is simply a set of all indexes $J = I$ and when $k = n$ then J contains only one window $W_1 = I$.

Two types of fuzzy partial TS will be considered. Fuzzy set of windows is the sequence of windows W_i accompanied by membership values μ_i . Set of fuzzy windows J contains fuzzy subsets of I .

A set of elements $S = \{s_1, \dots, s_m\}$ described by some set of time series will be called a system and his elements will be called subsystems. We will suppose that TSDB contains time series $P^i_s = (P^i_s(t_1), \dots, P^i_s(t_n))$, obtained as results of measurement of values of parameters (properties, attributes) $P = \{P^1, \dots, P^v\}$ for elements $s \in S$ in time moments $t = (t_1, \dots, t_n)$. For P^i_s we use also notation $P^i(s)$. For example, S is a system of countries described by time series of economical indicators. *Consumer_Price_Index(USA)* and *Consumer_Price_Index(Mexico)* will denote time series of consumer price index values for Mexico and USA. Another example gives a petroleum reservoir with a number of producing wells characterized by time series of oil and gas production during given time period [4], denoted as *Oil_Production(Well_N)* and *Gas_Production(Well_M)*. The set of all time series describing the subsystems of the system will be denoted as $T = \{x, y, z, \dots\}$. If each element s_k described by only one time series then the set of all TS may be considered as a system. For the set of time series T the sequence of windows J will define the set of partial time series $T_J = \{x_J, y_J, z_J, \dots\}$. Note that the sets of indexes I and windows J play the role of unified domains for joint analysis of different time series in TSDB describing the given system S .

3. Conditional time series

Suppose a TS $y = (y_1, \dots, y_n)$ denotes $P^i_s = (P^i_s(t_1), \dots, P^i_s(t_n))$. We will say that partial time series y_J is defined by condition $P^i_s = A$ if it consists of all windows where y satisfies this condition. Such partial time series will be called a conditional time series (CTS). We will consider two types of conditions which will define different types of CTS.

A first type of conditions is fulfilled on the elements of time series and will be called point-wise (pw-) conditions. For example the condition “*High price of gas in USA*” or $Price_of_Gas(USA) = High$ will define crisp partial TS if $High$ is a crisp interval $[a,b]$ of gas prices. Then $J = \{j | P_s^i(t_j) \in [a,b], j \in \{1, \dots, n\}\}$, where $P_s^i = Price_of_Gas(USA)$, is the set of time moments when this condition is fulfilled, and corresponding partial TS y_j consists of all elements $P_s^i(t_j) \in [a,b]$. If $High$ is defined as a fuzzy interval then J will be a fuzzy subset of $I: J = \{(\mu_j, j) | \mu_j = \mu_{High}(P_s^i(t_j)), j \in \{1, \dots, n\}\}$ and $y_j = \{(\mu_j, y_j) | \mu_j = \mu_{High}(P_s^i(t_j)), (j = 1, \dots, n)\}$ will be a fuzzy partial TS. The conditions defined on time domain will be not related with the elements of system and parameters values but only with time values. For example a condition “*End of the day*” which may be written as $Time = End_of_the_day$ will define a crisp $J = \{j | t_j \in [a,b], j \in \{1, \dots, n\}\}$ or fuzzy $J = \{(\mu_j, j) | \mu_j = \mu_{End\ of\ the\ day}(t_j), j \in \{1, \dots, n\}\}$ sets of indexes if this condition defined as a crisp time interval $[a,b]$ or as a fuzzy set.

The second type of conditions will be called window-wise (ww-) conditions. Windows-wise conditions define a set of windows J where condition is fulfilled. PW-condition may be considered as a special case of ww-condition when the length of window equals to 1. For example ww-condition may be given by “*Slowly Increasing price of gas in USA*”, describing the property of a sequence of the values of TS in some window $W_j = (P_s^i(t_j), P_s^i(t_{j+1}), \dots, P_s^i(t_{j+k}))$. This condition may be written as $Price_of_Gas(USA) = Slowly_Increasing$. Explication of such conditions also may be crisp or fuzzy. Suppose the following Increasing-Decreasing scale (ID-scale) is used [3]: $IDS = \{QDE: Quickly_decreasing, DEC: Decreasing, SDE: Slowly_decreasing, CON: Constant, SIN: Slowly_increasing, INC: Increasing, QIN: Quickly_increasing\}$. The crisp granulation of the grades of this scale may be done as follows.

Suppose the data in all windows $W_j = (P_s^i(t_j), P_s^i(t_{j+1}), \dots, P_s^i(t_{j+k}))$ from considered set of windows are approximated by the linear functions $y_j = p_j t + q_j$. Denote M the maximal absolute slope value from $\{|p_j|\}$. The interval $D_p = [-dM, dM]$, where d is a positive constant, e.g. $d = 1.2$, will be considered as a domain of slope values. The granulation of ID-scale may be obtained as a result of suitable partition of D_p on intervals. For example, if $D_p = [-10, 10]$, then the crisp granulation of ID-scale may the following: $QDE = [-10, -7]$; $DEC = [-7, -4]$; $SDE = [-4, -1]$; $CON = [-1, 1]$; $SIN = [1, 4]$; $INC = [4, 7]$; $QIN = [7, 10]$. As a result, the numerical slope value p_j of linear function approximating TS data in window W_j can be retranslated into corresponding linguistic grade of ID-scale.

Windows-wise property may be given also by conditions defined on a time domain D_t like “*During the year*”. Depending on the meaning of this condition D_t will be partitioned on the time intervals corresponding to calendar years or the set of sliding windows with the length of 365 days will be defined.

A fuzzy ww-conditions will define a fuzzy set of windows or a set of fuzzy windows. For example if the granulation of ID-scale considered above is given by fuzzy intervals [3] defined on D_p , then the numerical slope value p_j of linear function $y_j=p_jt+q_j$ approximating TS data in window W_j will belong to fuzzy grade *SIN: Slowly_increasing* with some membership value \mathbf{m} and the fuzzy condition $Price_of_Gas(USA) = Slowly_Increasing$ will define the fuzzy set of windows $\{(\mathbf{m}, W_j)\}$ where \mathbf{m} may be considered as a degree of fulfillment of this condition on the window W_j . A fuzzy condition “*During a small time period*” will define a sliding set of fuzzy windows defined on time domain. Example of association rules for ww-conditions will be considered further.

The third type of conditions will be called external conditions. It is related with the properties of subsystems described by TS. If a subsystem satisfies to this property with some degree a , then we will suppose that all elements of time series related with this subsystem satisfy to this property with the same degree $\mathbf{m}(i) = a$, ($i=1, \dots, n$). For example if TS describes the sale volumes of some article during one year then the properties like “*Expensive article*” or “*High quality article*” will be external for this TS.

The condition defining the set of windows J and corresponding conditional time series may be simple like conditions considered above or composite one obtained as a result of logical combination of simple conditions. For example the conditions $P =$ “*The high level of oil production in well number 1 in winter months*” and $Q =$ “*Quick increase of price on oil or low level of currency reserve*” may be written as $P = (P_1 \text{ AND } P_2)$ and $Q = (Q_1 \text{ OR } Q_2)$ where P_1 : “*Level_of_oil_production(Well_1) = High*”, P_2 : “*Time= Winter_month*”, Q_1 : “*Oil_Price= Quick_increase*”, $Q_2 =$ “*Level_of_currency_reserve= Low*”. Linguistic connectives *AND* and *OR* will define corresponding (fuzzy) set theoretic operations on the set of (fuzzy) windows .

4. Association rules with support and confidence measures

The basic framework of the association rule mining in DB is formalized as follows [2,11,12]. Let $I = \{i_1, i_2, \dots, i_m\}$ be a set of distinct literals called items

and $D = \{T_1, T_2, \dots, T_n\}$ be a database of transactions or baskets of items such that $T_i \subseteq I$. In general, association rules have the form $A \Rightarrow B$ or

$$R: \text{If } A \text{ then } B, (W), \quad (1)$$

where $A \subset I, B \subset I$ and $A \cap B = \emptyset$. For market basket data, for example, A and B denote some groups of items bought together. For evaluating significance, informative usefulness or strength of the association between A and B some measures W like support and confidence are used. Support and confidence measures are defined as follows:

$$\text{supp}(R) = \frac{|A \cap B|}{|D|}, \text{conf}(R) = \frac{|A \cap B|}{|A|}, \quad (2)$$

where $|A \cap B|$ is the number of transactions containing items A and B together and $|D|$ is a total number of transactions in database D . In some approaches instead of intersection $A \cap B$ of sets A and B it is used union $A \cup B$. Association rules usually do not causal but associative, i.e. B takes place together which A but may be not caused by A . Additionally to these measures some measures for selecting of informative rules are used [11,12]. J -measure is defined as follows:

$$J(B; A) = P(A) \left[P(B/A) \log \frac{P(B/A)}{P(B)} + P(\neg B/A) \log \frac{P(\neg B/A)}{P(\neg B)} \right], \quad (3)$$

where $P(A)$ is a probability of the set of items A in D , etc.

The extension of association rules and association measures on time series databases may be done as follows. Suppose A and B are crisp point-wise conditions (simple or composite) and J_A and J_B are the corresponding sets of indexes. Support and confidence measure will be defined as follows:

$$\text{supp}(R) = \frac{|J_A \cap J_B|}{N}, \text{conf}(R) = \frac{|J_A \cap J_B|}{|J_A|}, \quad (4)$$

where $N = n$ is a number of time moments. If A and B define fuzzy sets of indexes J_A and J_B then fuzzy cardinality is used in (4):

$$\text{supp}(R) = \frac{1}{N} \sum_{i=1}^N (\mathbf{m}_A(i) \wedge \mathbf{m}_B(i)), \text{conf}(R) = \frac{\sum_{i=1}^N (\mathbf{m}_A(i) \wedge \mathbf{m}_B(i))}{\sum_{i=1}^N \mathbf{m}_A(i)}. \quad (5)$$

where \wedge in (5) denotes some t -norm, e.g. \min . The following rule gives an example of fuzzy association rule with point-wise conditions: R : *If Oil_price= High then Profit_in_cosmetics_industry= Low*, (W). The terms *High* and *Low* may be defined as crisp or fuzzy intervals of possible values of oil price and profit values. Then the measures (4) or (5) correspondingly should be used.

For windows-wise crisp conditions the support and confidence measures are defined by (4) where N denotes the number of windows considered in model. For conditions defining a fuzzy set of windows the (5) will transform to

$$supp(R) = \frac{1}{N} \sum_{i=1}^N (\mathbf{m}_A(W_i) \wedge \mathbf{m}_B(W_i)), \quad conf(R) = \frac{\sum_{i=1}^N (\mathbf{m}_A(W_i) \wedge \mathbf{m}_B(W_i))}{\sum_{i=1}^N \mathbf{m}_A(W_i)}. \quad (6)$$

For conditions defining set of fuzzy windows we obtain:

$$supp(R) = \frac{1}{Nn} \sum_{i=1}^N \sum_{j=1}^n (\mathbf{m}_{A_i}(j) \wedge \mathbf{m}_{B_i}(j)), \quad (7)$$

$$conf(R) = \frac{\sum_{i=1}^N \sum_{j=1}^n (\mathbf{m}_{A_i}(j) \wedge \mathbf{m}_{B_i}(j))}{\sum_{i=1}^N \sum_{j=1}^n \mathbf{m}_{A_i}(j)}.$$

where $\mu_{A_i}(j)$ denotes membership value of time moment j in fuzzy window W_i corresponding to condition A .

Generation of association rules with crisp ww-conditions is illustrated by the example of TSDB containing the following 13 indicators of Mexican economics dynamics during 13 years: 1: *Unemployment Rate*; 2: *Real average*; 3: *Consumer Price Index*; 4: *Yearly Percent change*; 5: *Retail sales*; 6: *Industrial Production*; 7: *Exports (Monthly % Change)*; 8: *Imports (Monthly % Change)*; 9: *Merchandise Trade Balance (millions of US\$)*; 10: *Interbank Cetes*; 11: *MSE Market Index*; 12: *Mexican pesos per US \$*; 13: *Mexican pesos per Canadian \$*. The corresponding time series are presented in Fig. 1.

Suppose we want to receive the reply on the query: “*How the indicators are changed if (when) “Yearly Percent change” is decreasing?*” To receive the reply on this question we searched for association rules in the form:

If *Yearly_Percent_change* = *Decreasing* then $K = \text{Decreasing}$, (8)

If *Yearly_Percent_change* = *Increasing* then $K = \text{Increasing}$, (9)

where K denotes indicators different from “*Yearly Percent change*”.

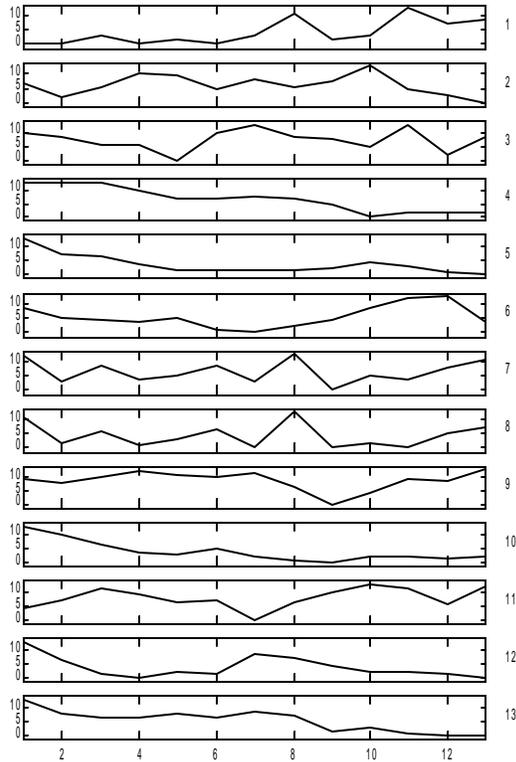


Fig. 1. Time series of 13 indicators of Mexican economics

For finding such association rules we considered all set of windows of size 2 such that the total number of windows equals to $N = 12$. For each window $W_i=(i,i+1)$ the value of considered indicator K is determined by the values y_i of indicator as follows: $K = \text{Increasing}$, if $y_{i+1} > y_i$, and $K = \text{Decreasing}$, if $y_{i+1} < y_i$. The similar approaches to evaluating ID-value considered also in [1,3]. The set J_A in (4) is a set of windows W_i where the condition *Yearly_Percent_change* = *Decreasing* is fulfilled. The set J_B in (4) is a set of windows W_i where the

condition $K = Decreasing$ is fulfilled. The calculation of support and confidence measures was done for rules (8), (9) for all indicators K different from “Yearly Percent change” and the rules with confidence and support values greater than 0.5 were selected as more valuable. As result three rules of type (8) for indicators $K= 3, 10, 12$, and one rule of type (9) for indicator $K = 11$ were selected. The confidence and support values were equal ($c= 0.75, s= 0.5$) for these rules with $K= 3, K= 10$ and $K= 11$, and ($c= 0.875, s= 0.583$) for the rule with $K= 12$. But a calculation of the measure of informative value (3) for these rules give the maximal value for rules with indicators $K= 3$ and $K= 11$. Finally the following two most informative rules were obtained:

If Yearly_Percent_change = Decreasing then Consumer_Price_Index = Decreasing,

If Yearly_Percent_change = Decreasing then MSE_Market_Index = Increasing.

It should be noted that the rule (8) with indicator $K= 12$: “Mexican pesos per US” received the maximal values of confidence and support and almost the minimal value of J -measure in comparison with other rules. It may be explained by the fact that this indicator is decreasing in almost all windows.

The use of Increasing-Decreasing scale considered in Section 3 gives possibility to receive more fine crisp or fuzzy rules like

If Yearly_Percent_change = Quickly_Decreasing then K = Slowly_Increasing.

Generally more complex patterns of TS like convex-concave patterns [6] may be considered.

5. Association rules based on correlation measure

In [4] it was proposed the method of generation of rules like

$$R: \text{If } A(s_i) \text{ then } P(s_j) \text{ associated with } Q(s_k), (W), \quad (10)$$

where $A(s_i)$ is a condition fulfilled for subsystem s_i and $P(s_j), Q(s_k)$ are some parameters evaluated for subsystems s_j, s_k . The following is an example of such association rule:

R: If Oil_production(Well_T101)= High then Oil_production(Well_T25) highly associated with Gas_production(Well_T03),(W),

where *High* represented generally by a crisp or fuzzy interval of oil production values. Then the condition of rule will define a crisp or fuzzy set of indexes J .

Suppose the condition $A(s_i)$ in (10) defines the crisp set of indexes J , and x, y are time series corresponding to $P(s_j)$ and $Q(s_k)$. As a measure of association between $P(s_j)$ and $Q(s_k)$ a correlation [9,16] of PTS x_j and y_j was considered:

$$r(x_J, y_J) = \frac{1}{|J|-1} \sum_{i \in J} \left(\frac{x_i - \bar{x}_J}{\mathbf{s}_{x_J}} \right) \left(\frac{y_i - \bar{y}_J}{\mathbf{s}_{y_J}} \right) \quad (11)$$

where $\bar{z}_J = \frac{1}{|J|} \sum_{i \in J} z_i$, $\mathbf{s}_{z_J} = \sqrt{\frac{1}{|J|-1} \sum_{i \in J} (z_i - \bar{z}_J)^2}$, for $z = x$ or $z = y$. The

association between x_j and y_j is considered as high if t -test [9] shows high level of significance of this association, where t calculated as follows:

$$t = r \sqrt{(|J|-2)/(1-r^2)}. \quad (12)$$

Here r is a correlation coefficient and $|J|-2$ is called a degree of freedom.

The set of association rules which may be generated for the condition $A(s_i)$ defines some weighted relation on the set of time series and subsystems and considered as an association network defined on the set of subsystems. This association network may be combined and compared with other relations existing between subsystems, such as spatial relations, causal relations etc and will give additional information about the structure of the considered system.

This approach was applied to association network construction for the system of 4 wells {T03, T25, T43, T101} from a Mexican reservoir based on the analysis of partial time series of oil and gas production in these wells registered during several years. For example, the following two association rules with high significance value were generated for condition $Oil_production(Well_T43) = High$:

R: If $Oil_production(Well_T43) = High$ then $Oil_production(Well_T25)$ highly associated with $Oil_production(Well_T03)$,

R: If $Oil_production(Well_T43) = High$ then $Oil_production(Well_T25)$ highly associated with $Oil_production(Well_T101)$.

The analysis of association networks obtained for different conditions showed high association between T3 well oil production and T25 well oil and gas production. After comparison of association networks with spatial location of wells in reservoir this fact obtained a natural explanation since T3 well is located above T25 well and not far from it. But association between these wells

was not evident before this analysis because, actually, T3 well belongs to another rock formation than the wells T25, T43 and T101. This analysis showed also that the pairs of wells with similar distances between them have different pair wise associations. This fact gives possibility to raise hypothesis about heterogeneity of rock properties in reservoir.

The proposed approach may be extended on the case of fuzzy conditions. Suppose condition $A(s_i)$ in (10) defines fuzzy set of indexes J . Define the sequence of levels $L = \{a_1 < a_2 < \dots < a_c\}$, $a_k \in [0,1]$ and corresponding sequence of a -cuts of fuzzy set J : $J_1 \supseteq J_2 \supseteq \dots \supseteq J_c$. For each level set J_k we calculate correlation r_k by (11) and corresponding value of t by (12). Denote $R = \{r_k\}$. Applying extension principle of Zadeh we determine a fuzzy subset of correlation values on the set R as follows: $\mu(r) = \max_{r_k=r} \{a_k\}$. Degree of freedom of correlation value r from R is obtained by extension principle of Zadeh as follows: $N_r = \max_{r_k=r} \{|J_k|\} - 2$ which will define by t -test a significance of corresponding association. The extension principle may be applied also to the levels of significance of t -test. Fuzzy correlations are considered also in [10].

Conclusions

In this paper, a novel data mining approach for generation of fuzzy and crisp association rules from (crisp) TSDB based on partial time series analysis is proposed. Partial time series given by crisp or fuzzy conditions are determined by the sets of crisp or fuzzy windows where this condition is fulfilled. The methods of calculation of confidence and support measures for new type of association rules are proposed in the paper. The efficiency of this measure was illustrated by the example of analysis of TSDB describing the dynamic of the system of economical parameters of Mexico.

A new type of association rules based on partial correlation measure considered also in the paper. The approach is illustrated using an example of the time series of petroleum wells oil and gas production from a Mexican reservoir. The initial correlations (obtained in a classical way) between time series did not show high associations between production volumes of the analyzed wells. However, the analysis of the partial time series gave the possibility to find such associations. These associations were described by rules and presented as association networks in 3D representation of wells in reservoir. Later on, these association networks gave the possibility to compare

the obtained associations with the geophysical information and other reservoir properties. Such analysis offers the possibility of better understanding of the existing relationships in analyzed system of wells which are hidden or not clear enough before this analysis. The possible generalization of this approach on fuzzy partial correlation analysis is discussed.

Industrial and economic systems are often characterized by the sets of time series describing the change in time of economical, financial or technological parameters. The analysis of associations between these TS, e.g. in macro economical analysis, can give the possibility to understand the relationships between subsystems of the system and their subsystems, which may be represented in linguistic or graphic form and used in perceptual decision making system or in informal analysis. Being a bridge between linguistic and numerical information, fuzzy logic gives us the tools for description and extraction of knowledge in crisp TSDB and for the further use of this knowledge in perception based reasoning. Generally the algorithms of generation of association rules in large DB require a lot of operations and one of the main problems in DM is development of algorithms with minimal complexity. This problem exists also in fuzzy TSDB but the main advantage of the perception based queries to TSDB is that they can essentially decrease the complexity of the algorithms decreasing the search space. TSDB form an efficient environment for construction, examination and tuning of fuzzy perceptual models.

Acknowledgements

The support for this research work has been provided by the IMP, projects D.00006 “Distributed Intelligent Computing”.

References

- [1] Alvarez-Ramirez J, Soriano A, Cisneros M, Suarez R (2003) Symmetry/ anti-symmetry phase transitions in crude oil markets. *Physica A: Statistical Mechanics and its Applications*, vol. 322(C), pp. 583-596
- [2] Agrawal R, Imelinski T, Swani A (1993) Mining association rules between sets of items in large database. In *Proc. ACM SIGMOD Int. Conf. Management of Data*, pp. 207-216
- [3] Batyrshin I (2002) “On granular derivatives and the solution of a granular initial value problem,” *Int. J. Appl. Math. Comp. Sci.* (Special Issue on Computing with

- Words and Perceptions, D. Rutkowska, J. Kacprzyk and L.A. Zadeh, Eds.) vol. 12, pp. 403 – 410. (<http://matwbn.icm.edu.pl/ksiazki/amc/amc12/amc1236.pdf>)
- [4] Batyrshin I, Cosultchi A, Herrera-Avelar R (2004) Association Network of Petroleum Wells Based on Partial Time Series Analysis. Proc. Workshop on Intelligent Computing. (Ed. by L. Sheremetov and M. Alvarado), Mexico City, Mexico, pp. 17 – 26
 - [5] Batyrshin I, Herrera-Avelar R, Sheremetov L, Suarez R (2004) Moving Approximations in Time Series Data Mining. FSSCEF 2004, (in this volume)
 - [6] Batyrshin I, Panova A (2003) “On convex-concave perception based functions,” in Proceedings of EUSFLAT 2003, Zittau, Germany, pp. 485– 490
 - [7] Das G, Lin KI, Mannila H, Renganathan G, Smyth P (1998) Rule Discovery from Time Series. KDD, pp. 16-22
 - [8] Last, M, Klein Y, Kandel A (2001) Knowledge Discovery in Time Series Databases. IEEE Transactions on Systems, Man, and Cybernetics, vol. 31B, pp. 160-169
 - [9] Lindgren BW (1998) Statistical Theory (Fourth Edition). Chapman-Hall
 - [10] Pedrycz W, Smith MH (1999) Granular correlation analysis in Data Mining. Proc. IEEE Int. Fuzzy Systems Conf. Korea, pp. III-1235 – III-1240
 - [11] Piatetsky-Shapiro G (1991) Discovery, analysis and presentation of strong rules. In Piatetsky-Shapiro G., and Frawley W.J. eds. Knowledge Discovery in Databases. Menlo Park, CA: AAAI Press. pp. 229-248
 - [12] Smyth P, Goodman RM (1992) An information theoretic approach to rule induction from databases. IEEE Trans. Knowledge Data Engineering 4, vol. 4, pp. 301-316
 - [13] Zadeh LA (1997) Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, Fuzzy Sets and Systems, vol. 90, pp. 111-127
 - [14] Zadeh LA (1999) From computing with numbers to computing with words - from manipulation of measurements to manipulation of perceptions, IEEE Trans. on Circuits and Systems - 1: Fundamental Theory and Applications, vol. 45, pp. 105-119
 - [15] Zadeh LA (2001) A New Direction in AI: Toward a Computational Theory of Perceptions, AI Magazine, Spring 2001 issue.
 - [16] Zhang P, Huang Y, Shekhar S, Kumar V(2003) Correlation analysis of spatial time series datasets: a filter-and-refine approach. In: Proc. PAKDD '03, Lecture Notes in AI. Vol. 2637, Springer-Verlag, Seoul, Korea, pp. 532-544

A Clear View on Quality Measures for Fuzzy Association Rules

Martine De Cock, Chris Cornelis, and Etienne E. Kerre

Fuzziness and Uncertainty Modelling Research Unit
Department of Applied Mathematics and Computer Science
Ghent University, Krijgslaan 281 (S9), 9000 Gent, Belgium
tel: +32 9 2644742, fax: +32 9 2644995, <http://fuzzy.UGent.be>
{Martine.DeCock, Chris.Cornelis, Etienne.Kerre}@UGent.be

1 Introduction

In today's information-driven economy, companies may benefit a lot from suitable knowledge management. Although knowledge management is not just a technology-based concept but rather a business practice in general, the possible and even indispensable support of IT-tools in this context is obvious. Because of the large data repositories many firms maintain nowadays, an important role is played by data mining techniques that dig up useful knowledge from these large data volumes. Among them, association rules [1] provide a convenient and effective way to identify and represent certain dependencies between attributes in a database. Originally, association rules emerged in the domain of shops and customers; the basic idea is to identify frequent itemsets in market baskets, i.e. groups of products frequently bought together. Storekeepers may use this information to decide on how to place merchandise on shelves to maximize a cross-selling effect, how to advertise, what to put on sale (for instance lowering the price of product A to attract customers, meanwhile increasing the price of product B that is frequently bought together with A), ... Evidently, the application of association rules can shed light on a wide range of decision making and marketing problems going beyond the scope of straightforward storekeeping.

Association rule mining is traditionally performed on a data table with binary attributes. Conceptually, a record x in the data table represents a customer transaction, whereas the attributes represent items that may be either purchased in that transaction, or not. Therefore, for each attribute A , $A(x)$ is either 1 or 0 indicating whether or not item A was bought in transaction x . An association rule is an expression of the form $A \Rightarrow B$ in which A and B are attributes, such as *cheese* \Rightarrow *bread*. The meaning is that when A is bought in a transaction, B is likely to be bought as well.

In most real life applications, databases contain many other attribute values besides 0 and 1. Very common for instance are quantitative attributes such as *age* or *income*, taking values from a partially ordered, numerical scale, often a subset of the real numbers. One way of dealing with a quantitative attribute like *cost* is to replace it by a few other attributes that form a crisp partition of the range of the original one, such as *low* = $[0, 100[$, *medium* = $[100, 300[$ and *high* = $[300, +\infty[$. Now we can consider these new attributes as binary ones that have value 1 if the *cost* attribute equals a value within their range, and 0 otherwise. In this way, the problem is reduced to the mining procedure described above [9]. From an intuitive viewpoint, it makes more sense however to draw values from the interval $[0, 1]$ (instead of just $\{0, 1\}$), to allow records to exhibit a given attribute to a certain extent only. In this way binary attributes are replaced by fuzzy ones. The corresponding mining process yields fuzzy association rules (see e.g. [2, 4, 5, 6, 7]).

Association rules can be rated by a number of quality measures, among which *support* and *confidence* stand out as the two essential ones. Support measures the statistical significance of a candidate rule $A \Rightarrow B$, whereas confidence assesses its strength. The basic problem of mining association rules is then to generate all association rules $A \Rightarrow B$ that have support and confidence greater than user-specified thresholds. These measures can be generalized for fuzzy association rules in several ways.

The goal of this paper is not to introduce yet another series of quality measures, but to shine a bright light on what has been proposed so far. Section 2 deals with the first pillar of our argument: the identification of transactions in a database as positive or negative examples of an association between attributes. Along the way we recall the basic concepts of support and confidence, initially in the framework of crisp association rules. Soon however we move on to the mining of fuzzy association rules as it is specifically in this setting that new and seemingly aberrant quality measures have been proposed recently, such as non-symmetrical measures of support. The second important pillar in this paper is that support and confidence measures should actually be thought of as compatibility and inclusion measures respectively (Section 3). Leaning on both pillars, in Section 4 we take the mystery out of some recently proposed quality measures for fuzzy association rules by providing clear insight into their true semantics.

2 Positive and Negative Examples

2.1 Crisp Association Rules

Let X be a non-empty data table containing records described by their values for binary attributes A belonging to a set \mathcal{A} . For an attribute A and a record $x \in X$, $A(x) = 1$ means item A was purchased in transaction x , while $A(x) = 0$ means A was not bought. In this way, A can also be thought of as the set

of transactions containing the item, i.e. $x \in A$ iff $A(x) = 1$, and $x \notin A$ iff $A(x) = 0$. Likewise coA is the set of transactions not containing the item, i.e. $x \in coA$ iff $A(x) = 0$, and $x \notin coA$ iff $A(x) = 1$. Let $A, B \in \mathcal{A}$. The support of an association rule $A \Rightarrow B$ is usually defined as

$$\text{supp}(A \Rightarrow B) = |A \cap B|/|X| \quad (1)$$

i.e. the number of elements belonging to both A and B , scaled to a value between 0 and 1. The idea behind the definition of support is to measure the statistical significance by counting *positive examples*, i.e. transactions that explicitly support the hypothesis expressed by the association rule. It is worth noting that the positive examples of $A \Rightarrow B$ are also those of the rule $B \Rightarrow A$, i.e. support is a symmetric measure. Hence, as can be expected, it only reveals part of the global picture. This is why we also need the confidence measure, to assess the strength of a rule. Traditionally, if a rule $A \Rightarrow B$ generates a support exceeding a user-specified threshold, it is meaningful to compute its confidence, i.e. the proportion of correct applications of the rule.

$$\text{conf}(A \Rightarrow B) = |A \cap B|/|A| \quad (2)$$

Note that $|A|$ will not be 0 if we assume that the confidence is computed only when the support exceeds a certain threshold (which should be greater than 0 to be meaningful).

Having identified the “supporters” of $A \Rightarrow B$ as positive examples, we can ask ourselves what a *negative example* of the same rule might look like. It is clear that a transaction violates the rule $A \Rightarrow B$ as soon as it contains A but not B . As opposed to positive examples, a negative example of $A \Rightarrow B$ is no negative example of $B \Rightarrow A$, and vice versa. Also, the complement of the set of positive examples does not necessarily equal that of negative examples, just like a “non-negative example” differs from a “positive example”. This is summarized in Table 1 (see also [4]). It is interesting that Dubois et al. [5]

Table 1. The nature of transaction x w.r.t. rules $A \Rightarrow B$ and $B \Rightarrow A$

x	$A \Rightarrow B$	$B \Rightarrow A$
positive example	$x \in A \wedge x \in B$	$x \in A \wedge x \in B$
non-positive example	$x \notin A \vee x \notin B$	$x \notin A \vee x \notin B$
negative example	$x \in A \wedge x \notin B$	$x \notin A \wedge x \in B$
non-negative example	$x \notin A \vee x \in B$	$x \in A \vee x \notin B$

also distinguish between positive and negative examples that are grouped into sets they call S_+ and S_- respectively. Furthermore, they introduce the class of irrelevant examples S_{\pm} as $S_{\pm} = \{x \in X \mid x \notin A\}$. One can easily verify that our classes of non-positive and non-negative examples are obtained as unions

of S_{\pm} with the set of positive and negative examples, respectively. Also, while S_{-} , S_{+} and S_{\pm} form a partition of X , this is clearly not the case for the four classes we defined. The most important reason we choose to consider them is that they all give rise to different measures:

Definition 1. The quality measures M_1, M_2, M_3 , and M_4 of the rule $A \Rightarrow B$ are respectively defined as

$$M_1(A \Rightarrow B) = |A \cap B|/|X| \qquad M_3(A \Rightarrow B) = |A \cap coB|/|X|$$

$$M_2(A \Rightarrow B) = |coA \cup coB|/|X| \qquad M_4(A \Rightarrow B) = |coA \cup B|/|X|$$

It can be easily verified that

$$M_2(A \Rightarrow B) = 1 - M_1(A \Rightarrow B) \quad \text{and} \quad M_3(A \Rightarrow B) = 1 - M_4(A \Rightarrow B) \quad (3)$$

Hence, only two measures are independent. We can for instance choose to work with M_1 and M_4 . The measure M_1 corresponds to the symmetrical support measure (supp) of Formula (1), while M_4 is a non-symmetrical measure taking into account all examples that do not violate the rule $A \Rightarrow B$.

2.2 Fuzzy Association Rules

Recall that a fuzzy set A in X is an $X \rightarrow [0, 1]$ mapping. Fuzzy-set-theoretical counterparts of complementation, intersection, and union are defined, as usual, by means of a negator, a t-norm, and a t-conorm. Recall that an increasing, associative and commutative $[0, 1]^2 \rightarrow [0, 1]$ mapping is called a t-norm \mathcal{T} if it satisfies $\mathcal{T}(x, 1) = x$ for all x in $[0, 1]$, and a t-conorm \mathcal{S} if it satisfies $\mathcal{S}(x, 0) = x$ for all x in $[0, 1]$. A negator \mathcal{N} is a decreasing $[0, 1] \rightarrow [0, 1]$ mapping satisfying $\mathcal{N}(0) = 1$ and $\mathcal{N}(1) = 0$. For A and B fuzzy sets in X we define $co_{\mathcal{N}}A(x) = \mathcal{N}(A(x))$, $A \cap_{\mathcal{T}} B(x) = \mathcal{T}(A(x), B(x))$, and $A \cup_{\mathcal{S}} B(x) = \mathcal{S}(A(x), B(x))$ for all x in X .

Let $A(x)$ be the degree to which an attribute A is bought in a transaction x (or in a broader context: the degree to which x satisfies the attribute). This way A can be thought of as a fuzzy set in the universe of transactions, and the measures discussed above have to be generalized accordingly. The cardinality of a fuzzy set in a finite universe X is defined as usual as the sum of the individual membership degrees. Replacing the set-theoretical operations in Definition 1 by their fuzzy-set-theoretical counterparts (defined by means of a negator \mathcal{N} , a t-norm \mathcal{T} , and a t-conorm \mathcal{S}), we obtain

Definition 2. The quality measures M_1, M_2, M_3 , and M_4 of the rule $A \Rightarrow B$ are respectively defined as

$$M_1(A \Rightarrow B) = \frac{1}{|X|} \sum_{x \in X} (A \cap_{\mathcal{T}} B)(x) \qquad M_3(A \Rightarrow B) = \frac{1}{|X|} \sum_{x \in X} (A \cap_{\mathcal{T}} co_{\mathcal{N}}B)(x)$$

$$M_2(A \Rightarrow B) = \frac{1}{|X|} \sum_{x \in X} (co_{\mathcal{N}}A \cup_{\mathcal{S}} co_{\mathcal{N}}B)(x) \quad M_4(A \Rightarrow B) = \frac{1}{|X|} \sum_{x \in X} (co_{\mathcal{N}}A \cup_{\mathcal{S}} B)(x)$$

The natural extension of Formula (3) holds when \mathcal{N} is the standard negator \mathcal{N}_s (defined by $\mathcal{N}_s(x) = 1 - x$ for all x in $[0, 1]$) and $(\mathcal{T}, \mathcal{S}, \mathcal{N}_s)$ is a de Morgan triplet, i.e. $\mathcal{T}(x, y) = \mathcal{N}_s(\mathcal{S}(\mathcal{N}_s(x), \mathcal{N}_s(y)))$ for all x and y in $[0, 1]$. Generalizing the confidence measure listed above to the fuzzy case, the following formula is obtained:

$$\text{conf}(A \Rightarrow B) = \frac{\sum_{x \in X} (A \cap_{\mathcal{T}} B)(x)}{\sum_{x \in X} A(x)} \quad (4)$$

3 Inclusion and Compatibility of Fuzzy Sets

Typically, to define fuzzy subsethood one takes a definition of classical set inclusion and tries to extend (“fuzzify”) it to apply to fuzzy sets. Below we quote three distinct, but essentially equivalent¹, definitions of the inclusion of A into B , where A and B are crisp subsets of X :

$$A \subseteq B \iff (\forall x \in X)(x \in A \Rightarrow x \in B), \quad (5)$$

$$\iff A = \emptyset \text{ or } \frac{|A \cap B|}{|A|} = 1, \quad (6)$$

$$\iff \frac{|coA \cup B|}{|X|} = 1 \quad (7)$$

While (5) is stated in strictly logical terms, the other two are based on counting the elements of a set, i.e. on cardinality, and have a probabilistic (i.e. frequentist) flavour. It is therefore not surprising that their respective generalizations to fuzzy set theory cease to be equivalent and give rise to cardinality-based and logical inclusion measures, respectively [3]. For instance, formula (5) can be generalized to fuzzy sets by replacing the two-valued implication by a $[0, 1]$ -valued implicator. Recall that an implicator \mathcal{I} is a $[0, 1]^2 \rightarrow [0, 1]$ mapping such that $\mathcal{I}(x, \cdot)$ is increasing and $\mathcal{I}(\cdot, x)$ is decreasing, and $\mathcal{I}(1, x) = x$ for all x in $[0, 1]$, and $\mathcal{I}(0, 0) = 1$. An inclusion measure satisfying desirable properties is then given by

$$Inc_1(A, B) = \inf_{x \in X} \mathcal{I}(A(x), B(x))$$

However this approach has certain disadvantages in applications. Indeed, if two fuzzy sets A and B are equal everywhere, except in the point x for which $A(x) = 1$ and $B(x) = 0$, then $Inc_1(A, B) = 0$. One can think of very concrete instances in which this indeed makes no sense. Imagine for instance that we are to evaluate to what extent the young people in a company are also rich. Testing subsethood of the fuzzy set of young workers into that of rich workers should then be based on the relative fraction (i.e. the *frequency*) of good earners

¹ Arguably, (5) is more general since it can also deal with infinite sets.

among the youngsters, and not on whether there exists or does not exist one poor, young employee. This observation has led researchers to consider extensions to definition (6) of crisp subsethood. If A and B are fuzzy sets, then one can define the subsethood of A into B as

$$Inc_2(A, B) = \frac{|A \cap_{\mathcal{T}} B|}{|A|}$$

if $A \neq \emptyset$, and 1 otherwise.

In formula (7) the presence of implication is also very clear. For propositions p and q in binary logic, $p \Rightarrow q$ has the same truth value as $\neg p \vee q$. The counterpart in fuzzy logic is the so-called S-implicator induced by \mathcal{S} and \mathcal{N} , defined by $\mathcal{I}_{\mathcal{S}, \mathcal{N}}(x, y) = \mathcal{S}(\mathcal{N}(x), y)$ for all x and y in $[0, 1]$. Generalizing formula (7) hence gives rise to a softened version of Inc_1 in which the supremum is replaced by taking the average over all elements of X :

$$Inc_3(A, B) = \frac{1}{|X|} \sum_{x \in X} \mathcal{I}_{\mathcal{S}, \mathcal{N}}(A(x), B(x))$$

Another well-studied class of implicators are the residual implicators $\mathcal{I}_{\mathcal{T}}$, induced by a t-norm \mathcal{T} in the following way: $\mathcal{I}_{\mathcal{T}}(x, y) = \sup\{\lambda \mid \lambda \in [0, 1] \text{ and } \mathcal{T}(x, \lambda) \leq y\}$ for all x and y in $[0, 1]$.

Another important kind of comparison measures for fuzzy sets, the so-called compatibility measures, assess their degree of overlap (see e.g. [10]). The so-called simple matching coefficient

$$Com_1(A, B) = \frac{|A \cap_{\mathcal{T}} B|}{|X|} = \frac{1}{|X|} \sum_{x \in X} \mathcal{T}(A(x), B(x))$$

is the average degree to which the fuzzy sets A and B together span the universe X . It is a softened version of

$$Com_2(A, B) = \sup_{x \in X} \mathcal{T}(A(x), B(x))$$

which is the height of the \mathcal{T} -intersection of fuzzy sets A and B . Compatibility measures are symmetrical but in general not reflexive.

4 A Clear View on the Semantics of the Measures

Throughout the literature on fuzzy association rules, the quality measures listed in Table 2 are prominent. The first and the third measure are generally accepted as measures of support and confidence respectively. They assess the significance and the strength of a fuzzy association rule. They coincide with a compatibility measure (Com_1) and an inclusion measure (Inc_2) from fuzzy set theory.

Table 2. Quality measures for fuzzy association rules

(1)	$M_1(A \Rightarrow B)$ or $\text{supp}(A \Rightarrow B)$	$\frac{1}{ X } \sum_{x \in X} \mathcal{T}(A(x), B(x))$	Com_1
(2)	$M_4(A \Rightarrow B)$	$\frac{1}{ X } \sum_{x \in X} \mathcal{I}_{S, \mathcal{N}}(A(x), B(x))$	Inc_3
(3)	$\text{conf}(A \Rightarrow B)$	$\frac{1}{ A } \sum_{x \in X} \mathcal{T}(A(x), B(x))$	Inc_2

The second measure $M_4(A \Rightarrow B)$ corresponds to the number of non-negative examples of the rule, and coincides with the inclusion measure Inc_3 for \mathcal{I} an S-implicator. In [4] we tackled the question whether we can substitute the S-implicator in M_4 by a residual implicator, and concluded that such a replacement is not desirable. This can be roughly explained as follows: an example can be called non-negative if it does not contradict the rule; so either if it is in favour of the rule, or if it does not say anything about the rule. The latter situation arises when $A(x)$ is small. In this case S-implicators tend to always identify x correctly as a non-negative example, while some residual implicators overlook it for low $B(x)$ values. In [6], Hüllermeier suggests the following implication-based measure of support for a fuzzy association rule $A \Rightarrow B$:

$$\text{supp}_1(A \Rightarrow B) = \sum_{x \in X} \mathcal{I}(A(x), B(x))$$

where \mathcal{I} is an implicator. Note that by dividing it by $|X|$ we obtain a formula similar to Inc_3 . The rationale behind it is that a transaction x with $A(x) = 0.6$ and $B(x) = 0.4$ only contributes to degree 0.4 to the commonly used support (which is our Formula (2) defined by means of $\mathcal{T} = \min$). This is considered to be low since, in the words of [7] “*x does hardly violate (and hence supports) the rule*”. We fully agree on the first claim (x is a non-negative example to a high degree) but not on the second one (being a non-negative example does not imply being a positive example). Indeed the fundamental difference between positive and non-negative examples does not seem to be respected in [7], which becomes evident when examining those transactions that do not really tell us something about the rule (i.e. that have a low membership degree in A). To deal with this problem of “*trivial support*”, Hüllermeier suggests to extend the measure of support to

$$\text{supp}_2(A \Rightarrow B) = \sum_{x \in X} \mathcal{T}(A(x), \mathcal{I}(A(x), B(x)))$$

Furthermore he is in favour of using residual implicators over S-implicators, which seems to be in conflict with our findings. In [4] we go into this in detail. However if \mathcal{I} is the residual implicator induced by a continuous t-norm \mathcal{T} then $\text{supp}_2(A \Rightarrow B) = \sum_{x \in X} \min(A(x), B(x))$ (see e.g. [8]) as is also noted in [7]. Therefore in this case the new measure of support introduced in [6] reduces

to the commonly used one, and hence does not offer anything new. For this reason we disagree with the claim of [5] that whereas the traditional support measure (i.e. supp or M_1) is in line with the conjunction-based approach to modelling fuzzy rules, the above-defined measure supp_2 follows the tradition of implication-based fuzzy rules. Within the literature on fuzzy association rules there exists another view on the use of Inc_3 as well. Chen et al. [2] call this measure “degree of implication” and use it to replace the traditional confidence measure. This should not come as a great surprise, since their reliance on Inc_3 yields just another way of expressing the subsethood of A into B . For this reason we also prefer to view the non-symmetrical measure M_4 as a confidence measure. Finally, since association rule mining is concerned with finding frequent patterns in databases, it seems more natural to use cardinality-based rather than logical compatibility and inclusion measures, which explains why Inc_1 and Com_2 are not met in literature on fuzzy association rules.

Acknowledgements The results presented in this paper were obtained within the bilateral China–Flanders cooperation project 174B0201. Martine De Cock and Chris Cornelis would like to thank the Fund for Scientific Research Flanders–FWO for funding the research reported in this paper.

References

1. Agrawal R, Imielinski T, Swami A (1993) Mining Association Rules between Sets of Items in Large Databases. Proceedings of ACM SIGMOD conference on Management of Data, 207–216
2. Chen G, Yan P, Kerre E E (2004) Computationally Efficient Mining for Fuzzy Implication-based Association Rules in Quantitative Databases. To appear in: International Journal of General Systems
3. Cornelis C, Kerre E E (2003) Inclusion Measures in Intuitionistic Fuzzy Set Theory. Lecture Notes in Computer Science 2711:345–356
4. De Cock M, Cornelis C, Kerre E E (2003) Fuzzy Association Rules: a Two-Sided Approach. Proceedings of FIP2003, 385–390
5. Dubois D, Hüllermeier E, Prade H (2003) A Note on Quality Measures for Fuzzy Association Rules. Lecture Notes in Computer Science 2715:346–353
6. Hüllermeier E (2001) Fuzzy Association Rules: Semantic Issues and Quality Measures. Lecture Notes in Computer Science 2206:380–391
7. Hüllermeier E (2002) Mining Implication-Based Fuzzy Association Rules in Databases. Proceedings of IPMU2002, Vol. I:101–108
8. Novák V, Perfilieva I, Močkoř J (1999) Mathematical Principles of Fuzzy Logic. Kluwer
9. Srikant R, Agrawal R (1996) Mining Quantitative Association Rules in Large Relational Tables. Proceedings of ACM SIGMOD Conference on Management of Data, v.25 n.2:1–12
10. Tsiporkova E, Zimmermann H-J (1998) Aggregation of Compatibility and Equality: a New Class of Similarity Measures for Fuzzy Sets. Proceedings of IPMU1998, 1769–1776

Moving Approximations in Time Series Data Mining

Batyrshin^{1,2} I., Herrera-Avelar¹ R., Sheremetov¹ L., Suarez¹ R.

¹Mexican Petroleum Institute, Mexico

²Institute of Problems of Informatics, Academy of Sciences of Tatarstan,

Russia

batyr@imp.mx

Abstract. In this paper we propose a new technique for time series (TS) analysis called moving approximations. It defines a new procedure of TS transformation called Moving Approximations Transform (MAPT). Based on this transformation of TS two new measures of association between TS are introduced: a measure of local trend associations and a measure of local trend distances. The properties of invariance of MAPT and association measures to linear transformations of TS are studied. An example of association analysis of the set of TS from economics is described. As result of such analysis the association network and classification of TS is obtained.

1. Introduction

Time series data mining (TSDM) is actively developed research area [6, 7]. Traditional approaches to time series (TS) analysis study the problems of time series modelling and forecasting its values [5]. TSDM consider such problems as association rules extraction from TS data bases (TSDB), search of patterns in TS, TS classification, piece-wise linear representation (PLR), linguistic summarization of TS etc. [4, 6-9, 12]. Such technique like Fourier and Wavelet transforms are very important tools in TS analysis and TSDM [4,8].

In this paper we propose new approach to TSDM which use analysis of local trends in TS. This approach is based on the concepts of partial time series analysis when the parts of two time series are compared and on consideration

of TSDB as a description of the dynamics of the system [2]. In such contexts TSDM consists in analysis of local associations between dynamics of TS describing elements of the system. Local dynamic of TS maybe described by terms like “*increasing*”, “*decreasing*”, “*constant*” which may be evaluated by the slope value of linear regression of TS on considered segment. The comparison of such slope values for two TS on the same segments gives possibility to analyze associations between the changes of these TS. In this paper we consider a sequence of moving windows like in moving average procedure used for TS analysis and extract slope values of linear approximations of TD values on these windows. Moving approximations transform (MAPT) replace the sequence of TS values by the sequence of slope values. Based on MAPT new association measures are introduced. This measures may be used for construction of association network on the set of TS and for classification of TS on the classes of highly associated TS. The distance measure between slope values for short TS and TS classification based on this measure is considered also in [9].

The paper is organized as follows. In Section 2 MAPT is introduced and studied. In Section 3 new measures of TS association based on MAPT are introduced and studied. An example of application of these measures to analysis associations between TS in economics is considered in Section 4. In Conclusion main results and future work is discussed.

2. Moving approximations transform

Suppose $t = (t_1, \dots, t_n)$, ($n > 1$) is a sequence of time moments such that $t_i < t_{i+1}$ for all $i = 1, \dots, n-1$. A time series (y, t) is a sequence of real numbers $y = (y_1, \dots, y_n)$ corresponding to the sequence of time moments t . Denote $I = (1, \dots, n)$ the sequence of indexes of time series (TS). Time series are usually obtained as a result of measurement of some parameter (economical, physical, etc) followed at equal intervals of time.

A window W_i of a length $k > 1$ is a sequence of indexes $W_i = (i, i+1, \dots, i+k-1)$. Denote $y_{W_i} = (y_i, y_{i+1}, \dots, y_{i+k-1})$ the corresponding values of TS y in the window W_i . Suppose $J = (W_{i_1}, \dots, W_{i_m})$ is a sequence of windows such that $i_1 < \dots < i_m$. A sequence $y_J = (y_{W_{i_1}}, \dots, y_{W_{i_m}})$ will be called a partial time series

of y defined by the sequence of windows J . A sequence of all windows of the length k , $J = (W_1, W_2, \dots, W_{n-k+1})$, ($1 < k \leq n$), will be called a moving (or sliding) windows. We will denote it also as J_k and the set of indexes of all windows in J_k as $I_k = (1, 2, \dots, n-k+1)$. Such moving windows are used in statistics in moving average procedure for smoothing time series when the value in the middle of the window replaced by the mean of values from this window [5]. The set of all time series in TSDB will be denoted as $T = \{x, y, z, \dots\}$. For the set of time series T the sequence of windows J will define the set of partial time series $T_J = \{x_J, y_J, z_J, \dots\}$. Here the sets of windows J plays the role of unified domain for joint analysis of different time series in TSDB describing the given system S .

Suppose $J = (W_1, W_2, \dots, W_{n-k+1})$ are moving windows $W_i = (i, i+1, \dots, i+k-1)$ of size k and $y_{W_i} = (y_i, y_{i+1}, \dots, y_{i+k-1})$, $i \in I_k = (1, 2, \dots, n-k+1)$ are corresponding partial time series. A linear functions $f_i = a_i t + b_i$ with parameters $\{a_i, b_i\}$ minimizing the criterion

$$Q(f_i, y_{W_i}) = \sqrt{\sum_{j=i}^{i+k-1} (f_i(t_j) - y_j)^2} = \sqrt{\sum_{j=i}^{i+k-1} (a_i t_j + b_i - y_j)^2}, \quad (1)$$

will be called a moving least squares approximation of y_{W_i} . The parameters a_i, b_i of moving least squares approximations can be calculated as follows:

$$a_i = \frac{\sum_{j=i}^{i+k-1} (t_j - \bar{t}_i)(y_j - \bar{y}_i)}{\sum_{j=i}^{i+k-1} (t_j - \bar{t}_i)^2}, \quad b_i = \bar{y}_i - a_i \bar{t}_i, \quad (2)$$

where $\bar{t}_i = \frac{1}{k} \sum_{j=i}^{i+k-1} t_j$, $\bar{y}_i = \frac{1}{k} \sum_{j=i}^{i+k-1} y_j$,

Definition 1. Suppose $a = (a_1, \dots, a_{n-k+1})$ is a sequence of slope values obtained as a result of moving approximations of time series (y, t) in moving windows of size k . A transformation $MAP_k(y, t) = a$ will be called a moving approximation transform (MAPT) of time series y .

The slope values $a = (a_1, \dots, a_{n-k+1})$ will be called local trends.

Suppose p, q, r, s are real values, $r \neq 0$ and y, z are time series given in the same time moments $t = (t_1, \dots, t_n)$. Denote $py+q = (py_1+q, \dots, py_n+q)$ and $y+z = (y_1+z_1, \dots, y_n+z_n)$.

Proposition 2. A MAPT satisfies for all real values p, q, r, s , ($r \neq 0$), and time series z the following properties:

- 1) $MAP_k(py+q, t) = pMAP_k(y, t)$;
- 2) $MAP_k(y, rt+s) = (1/r)MAP_k(y, t)$;
- 3) $MAP_k(y+z, t) = MAP_k(y, t) + MAP_k(z, t)$.

Corollary 3. MAPT is invariant to equal simultaneous linear transformations of time values and time series values, i.e.

$$MAP_k(ry+s, rt+s) = MAP_k(y, t).$$

Corollary 4. Suppose time moments $t = (t_1, \dots, t_n)$ are increasing with a constant step h such that $t_{i+1} - t_i = h$ for all $i = 1, \dots, n-1$. Then in MAPT the set of time moments $t = (t_1, \dots, t_n)$ can be replaced by the set of indexes $I = (1, \dots, n)$ as follows:

$$MAP_k(y, t) = (1/h)MAP_k(y, I).$$

Corollary 5. Suppose $mean(y^1, \dots, y^m) = (mean(y_1^1, \dots, y_1^m), \dots, mean(y_n^1, \dots, y_n^m))$ is a mean value of m time series calculated element-wise in time moments $t = (t_1, \dots, t_n)$, then $mean(MAP_k(y^1, t), \dots, MAP_k(y^m, t)) = MAP_k(mean(y^1, \dots, y^m), t)$.

From Proposition 2 it follows also: $MAP_k(-y, t) = -MAP_k(y, t)$.

Further we will replace time series by corresponding MAP transform. Considered above properties show how the operations on time series can be replaced by corresponding operations on their MAPT. Because the conditions of the Corollary 4 as usually are fulfilled for time series we will suppose that $t = I = (1, \dots, n)$ if time moments are not mentioned and use notation $MAP_k(y)$. Also, if the size of windows in moving approximations does not important or fixed we will write simply $MAP(y)$.

3. New measures of time series association based on local trend analysis

Our goal is to introduce the measure of association between time series which could evaluate the relationships between the changes in time of time series describing the system evolution during considered time period. We say that y is

“increasing”, “decreasing” or “constant” in the window W_i if the corresponding slope value of linear approximation $f_i = a_i t + b_i$ satisfies condition $a_i > 0$, $a_i < 0$ or $a_i = 0$, respectively. We will say that 2 time series have similar local trend behavior in some window if their slope values of approximating lines in this window have equal signs.

Traditional correlation measure is suitable for measuring synchronous deviations of two time series values around mean values. But it can not be used for analysis of local trend associations. Fig. 1 shows 2 time series which are synchronously increasing and decreasing in local parts of time series which means that they are highly mutually associated but the correlation value between these time series equals to -0.0948 and shows the absence of association.

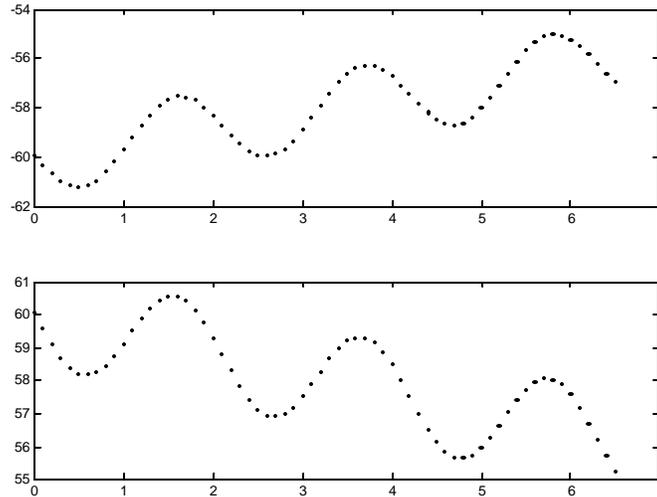


Fig. 1. Example of 2 time series with similar local trends and with correlation value $r = -0.0948$

For the analysis of local trend associations we introduce here a new measure which will be called a measure of local trend associations (*mlta*).

Definition 6. A measure of local trend associations between time series (y,t) and (x,t) is a cosine of angle between two vectors $MAP_k(y) = (a_{y1}, \dots, a_{ym})$ and $MAP_k(x) = (a_{x1}, \dots, a_{xm})$, where $m = n - k + 1$ and $MAP_k(y), MAP_k(x) \neq 0$:

$$mlta_k(y, x) = \cos(MAP_k(y), MAP_k(x)) = \frac{\sum_{i=1}^m a_{yi} \cdot a_{xi}}{\sqrt{\sum_{i=1}^m a_{yi}^2 \cdot \sum_{j=1}^m a_{xj}^2}}, \quad (3)$$

From the definition the properties of this measure are following:

$$\begin{aligned} mlta_k(y, x) &= mlta_k(x, y); & -1 \leq mlta_k(y, x) \leq 1; \\ mlta_k(y, y) &= 1; & mlta_k(y, -y) = -1; \\ mlta_k(y, -x) &= -mlta_k(y, x); & mlta_k(-y, -x) = mlta_k(y, x). \end{aligned}$$

Generally, for two given time series the measure of local trend associations defines a sequence of association values $(mlta_2, \dots, mlta_n)$ depending on the size of windows. The selection of the size of the moving windows will define the level of detailing of the analysis of associations between time series. In the presence of random fluctuations in time series values the size of windows k should not be very small. For time series presented in Fig. 1 the value of this measure equal to 0.923 for moving approximations with window size equal to 5. The high value of $mlta$ reflects the fact that these time series are highly associated due to synchronous increase and decrease in local parts of TS.

Suppose p, q, r, s are real constants and $p, r \neq 0$. A time series (z, u) obtained as a result of a linear transformation of time series (y, t) , such that $z_i = py_i + q$ and $u_i = rt_i + s$, $i = 1, \dots, n$ will be denoted as $(z, u) = L(y, t)$.

Proposition 7. Suppose (x, t) and (y, t) are two time series and L_1 and L_2 are two linear transformations of time series given by the sets of parameters (p_1, q_1, r_1, s_1) and (p_2, q_2, r_2, s_2) respectively, then

$$mlta_k(L_1(x, t), L_2(y, t)) = \text{sign}(p_1) \cdot \text{sign}(r_1) \cdot \text{sign}(p_2) \cdot \text{sign}(r_2) \cdot mlta_k((x, t), (y, t)).$$

Theorem 8. A measure $|mlta|$ is invariant to linear transformations of time series.

Theorem 8 shows very nice property of introduced association measure. It means particularly that time series may be normalized independently and the absolute measure of local trend associations $|mlta|$ will be not changed.

Definition 9. A measure of local trend distances between time series (y, t) and (x, t) is defined as an Euclidean distance between two vectors $MAP_k(y) = (a_{y1}, \dots, a_{ym})$ and $MAP_k(x) = (a_{x1}, \dots, a_{xm})$, $(m = n - k + 1)$:

$$mld_k(y, x) = d(MAP_k(y), MAP_k(x)) = \sqrt{\sum_{i=1}^m (a_{yi} - a_{xi})^2}, \quad (4)$$

From the definition it follows the metric properties of mld :

$$mld_k(y, x) = mld_k(x, y);$$

$$mld_k(y, x) = 0 \text{ and } mld_k(y, x) = 0 \text{ if and only if } y = x;$$

$$mld_k(y, x) \leq mld_k(y, z) + mld_k(z, x).$$

Proposition 10. Suppose p, q, r, s , ($r \neq 0$), are real values and (z, t) is a time series. A measure of local trend distances satisfies for all time series (y, t) and (x, t) the following properties:

$$mld_k(y+q, x+s) = mld_k(y, x);$$

$$mld_k(y+z, x+z) = mld_k(y, x);$$

$$mld_k(py, px) = p \cdot mld_k(y, x);$$

These introduced measures were tested on many artificial and real examples of time series and showed very good results. Examples of application of this measure to generation of association networks of real time series of economical data are considered in the following section.

4. Example of association analysis

The proposed approach to local trend analysis was applied to analysis associations between 10 time series of gross internal product in Mexico checked quarterly during the period 1980 – 2003 [1]. Each time series contains 96 data. The mld_3 measure with windows size equal to 3 was used for calculation pair-wise associations between time series. Fig. 2 shows association networks constructed sequentially for different levels of $|mld|$ values such that on each level only associations with the value greater than this level are shown.

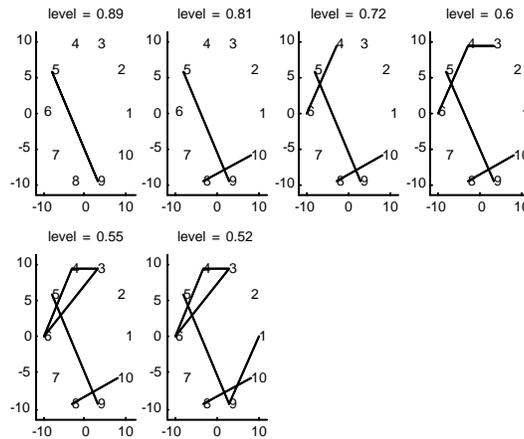


Fig. 2. Association networks with different levels of association

Below are the classes of time series obtained as sets of nodes of connected subgraphs shown in Fig. 2.

C1. {5: Electricity, Gas and Water; 9: Social and Personal Communal Services; 1: Farming, Forestry and Fishes};

C2. {8: Financial Services, Insurance, Real Estate Activities and Rent; 10: Liability to Banking Services Allocate};

C3. {3: Manufacturing Industry; 4: Construction; 6: Commerce, Restaurants and Hotels};

C4. {2: Mining};

C5. {7: Transport, Storage and Communications}.

On more low levels time series number 2 have more associations with the class C3 and time series 7 with the class C2.

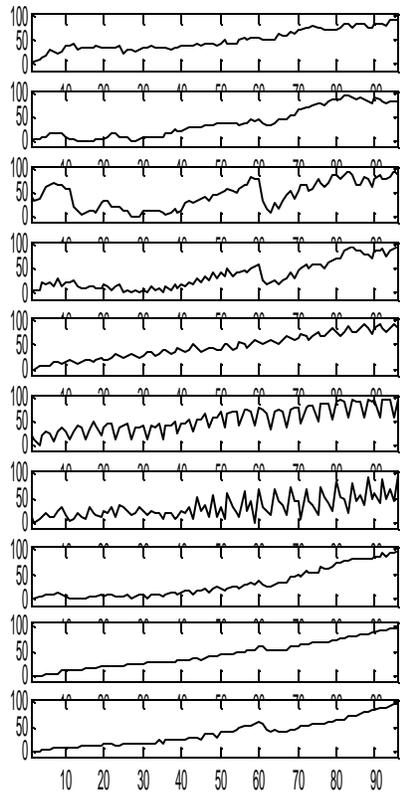


Fig. 3. Time series grouped in correspondence with the obtained classification. The number of time series is shown on its right side.

Evidently, the obtained classes have good interpretation and the time series from these classes shown in Fig. 3 have the similar shapes. The TS in Fig. 3 grouped in correspondence with obtained classifications.

Conclusion

The paper introduces a new technique for time series (TS) analysis based on moving approximation transform. In MAPT time series replaced by the

sequence of slopes of approximating lines. Such MART representation of TS gives possibility to introduce measures of local trend associations and distances which gives possibility to analyze and compare local dynamic of TS. The approach is demonstrated on the example of association analysis and classification of TS of economical data.

The developed technique of local trend analysis gives possibility to introduce perceptive fuzzy granulation [3,10,11] of the set of slope values and generate linguistic descriptions like “slowly increasing”, “quickly decreasing” etc. The rules and descriptions of such type may be used further for perception based reasoning about systems described by the set of time series. The methods of generation of such type of rules are considered in [3].

Acknowledgements

The support for this research work has been provided by the IMP, projects D.00006 “Distributed Intelligent Computing”.

References

- [1] Banco de Información Económica
<http://dgcnesyp.inegi.gob.mx/bdine/bancos.htm>
- [2] Batyrshin I, Herrera-Avelar R, Sheremetov L, Suarez R (2004) Mining Fuzzy Association Rules and Networks in Time Series Databases. International Conference on Fuzzy Sets and Soft Computing in Economics and Finance, FSSCEF 2004, St. Petersburg, Russia, 2004 (in this volume).
- [3] Batyrshin I, Herrera-Avelar R, Sheremetov L, Suarez R (2004) On Qualitative Description of Time Series Based on Moving Approximations. International Conference on Fuzzy Sets and Soft Computing in Economics and Finance, FSSCEF 2004, St. Petersburg, Russia, 2004 (in this volume).
- [4] Boyd S (1998) TREND: a system for generating intelligent descriptions of time-series data. In Proceedings of the IEEE International Conference on Intelligent Processing Systems (ICIPS-1998).
- [5] Brockwell PJ, Davis RA, Fienberg SE (1991) Time Series: Theory and Methods (Springer Series in Statistics) February, 1991.
- [6] Keogh E, Kasetty S (2002) On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration. In SIGKDD '02, July, 23-26 2002

- [7] Last M, Klein Y, Kandel A (2001) Knowledge Discovery in Time Series Databases. in IEEE Transactions on Systems, Man, and Cybernetics Vol. 31B.
- [8] Mörchen F (2003) Time series feature extraction for data mining using DWT and DFT. Data Bionics, Philipps-University Marburg, Hans-Meerwein-Strasse, 35032 Marburg, Germany, October 9, 2003.
- [9] Möller-Levet CS, Klawonn F, Cho KH, Wolkenhauer O (2003) Fuzzy Clustering of Short Time-Series and Unevenly Distributed Sampling Points. IDA 2003 pp. 330-340.
- [10] Zadeh LA (1997) Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets and Systems, vol. 90, pp. 111-127.
- [11] Zadeh LA (1999) From computing with numbers to computing with words - from manipulation of measurements to manipulation of perceptions," IEEE Trans. on Circuits and Systems - 1: Fundamental Theory and Applications, vol. 45, pp. 105-119.
- [12] Zhang P, Huang Y, Shekhar S, Kumar V (2003) Correlation analysis of spatial time series datasets: a filter-and-refine approach. In the Proc. Seventh Pacific-Asia Conf. Knowledge Discovery Data Mining, (PAKDD '03), Lecture Notes in Artificial Intelligence Vol. 2637, Springer-Verlag, Apr. 30-May 2, Seoul, Korea pp. 532-544

On Qualitative Description of Time Series Based on Moving Approximations

Batyrshin^{1,2} I., Herrera-Avelar¹ R., Sheremetov¹ L., Suarez¹ R.

¹Mexican Petroleum Institute, Mexico

²Institute of Problems of Informatics, Academy of Sciences of Tatarstan,
Russia

{batyr,herrerar,she,rsuarez}@imp.mx

Abstract. In this paper we propose a new method of piece-wise linear representation (PLR) of time series. Such representation is based on a moving approximations transform (MAPT) of time series in the sequence of local slope values. PLR is based on granulation of the set of slope values and on the sequential merging and splitting of segments of TS. Resulting PLR is retranslated in the form of rules containing linguistic terms like “quickly increasing”, “slowly decreasing” etc. The possible applications of PLR and rule based representation of TS in classification of TS and in perceptive reasoning are discussed.

1. Introduction

Time series (TS) are important sources for decision making in business and in science. The problem of qualitative analysis of TS arises in such areas like data mining, process monitoring, diagnosis and control, reasoning about processes, etc. The qualitative analysis of TS is related with coding, compressing, classification and search TS in databases, search of valuable patterns in data, linguistic description for knowledge extraction, inductive generalization, perceptive reasoning and decision making in knowledge based systems [1-3, 5-9]. The importance of development of such methods is increasing with the exponential growth of the size of TS databases in finance, economics, medicine, manufacturing, science, geophysics, technology etc. used in research and decision making procedures.

Piecewise linear representation (PLR) of time series gives possibility to reduce the complexity of TS data, describe them qualitatively as a sequence of straight lines, generate linguistic descriptions etc [2, 5 – 7, 9].

In this paper a new method of adaptive piece-wise linear representation of time series based on *MAP* transform [1] of TS is proposed. This method is based on scaling of slope values of linear segments which gives possibility to retranslate them in perceptive evaluations like “Quickly increasing”, “Slowly decreasing” etc. The method described in Section 2. His work is demonstrated in Section 3 on 2 examples of real time series from economics. In Conclusion the obtained results are discussed.

2. Adaptive PLR of Time Series Based on MAPT

Our goal is to transform PLR of TS in rules

$$R_k: \text{If } X \text{ is } T_k \text{ then } Y \text{ is } S_k \quad (1)$$

where T_k is a linguistic term like *LESS THAN 5*, *GRATER THAN 40*, *BETWEEN 10 AND 15* describing some (may be fuzzy) interval A_k [2]. The S_k is a linguistic term like *DECREASING*, *CONSTANT* and *QUICKLY INCREASING*, which describes the speed of function's y change on A_k . The rule set (1) will give qualitative description of given TS. Term S_k will code the slope p_k of the function $y_k = p_k x + q_k$ that approximates the data y_i on A_k .

We will suppose that on a set of possible slope values P a scale $L_p = \{P_1, \dots, P_M\}$, ($P_j \in P$) such that $P_j < P_{j+1}$ for all $j = 1, \dots, M-1$ is defined. This numerical scale defines a linguistic scale $L_s = \{s_1, \dots, s_{M+1}\}$ with linguistic terms such that for each numerical slope value p , a corresponding linguistic term from L_s may be pointed out as follows: $S(p) = s_1$, if $p \leq P_1$, $S(p) = s_j$, if $P_{j-1} < p \leq P_j$, for $j=2, \dots, M$ and $S(p) = s_{M+1}$, if $P_M < p$. It should be noted that the scale of numerical slope values L_p is usually context dependent because the slope values may vary for different TS. In these cases different numerical scales of slope values may be used. But linguistic scale L_s as usually context insensitive and depends only on the number of grades in the scale L_p . For example, for different sets of TS two different numerical scales may be obtained: $L_p = \{-5, -3, -1, 1, 3, 5\}$ and $L_p = \{-0.07, -0.04, -0.01, 0.01, 0.04, 0.07\}$ but in both cases for coding the numerical slope values, the same linguistic scale may be used:

$L_s = \{QDC: \text{Quickly Decreasing}, DEC: \text{Decreasing}, SDC: \text{Slowly Decreasing}, CON: \text{Constant}, SIN: \text{Slowly Increasing}, INC: \text{Increasing}, QIN: \text{Quickly Increasing}\}$.

The essential part of our PLR algorithm is the use of linguistic scales. The adaptive PLR of TS algorithm consists in the following steps: *MAP* transform of TS with small windows, linguistic coding of slopes obtained by *MAPT*, merging windows with similar slopes and finally separate intersected windows on non-intersected windows, which finally will give PLR of TS. Here we use some ideas for merging segments from [2]. But instead of initial partition on small intervals we use initial cover of points by windows obtained in *MAPT*. The idea of the proposed method is to transform the system of highly intersected windows in the set of non-intersected segments (windows) by merging and splitting segments obtained from initially constructed windows.

The general scheme of proposed procedure consists of the following six steps.

1. *Apply Moving Approximations Transform.* *MAPT* generates windows (segments) $G_i = (i, \dots, i+D-1)$, ($i = 1, \dots, n-D+1$) with D points. Each segment approximated by linear function $y_i = p_i x + q_i$ and TS replaced by the sequence of slopes (p_1, \dots, p_{n-D+1}) . Determine the codes $S(p_i)$ of slope values for all segments.

2. *Join similar segments.* Compare the slope codes of each two neighboring segments starting from the first segment. If some pair of segments has equal slope codes, join them in a new segment. Calculate a linear approximation and correspondent linguistic slope code for the new segment. Repeat this step while the neighboring segments with equal slope codes exist.

3. *Delete covered segments.* Denote Q the set of all segments and R an empty set: $R = \emptyset$. Find in Q a segment G_i with maximum number of points. Sequentially delete from Q segments G_{i+k} , ($k=1, \dots$), if all points of them covered by G_i and G_{i+k+1} , i.e. if $G_{i+k} \subseteq G_i \cup G_{i+k+1}$. Similarly delete all segments G_{i-k} , ($k=1, \dots$), if all points of them are covered by G_i and G_{i-k-1} . Delete segment G_i from Q and write it in R . Repeat this procedure till Q is empty. Rewrite all segments from R in Q . After each deleting of covered segment *join similar segments*.

4. *Delete segments with small uncovered sub-segments.* For each segment determine the set of points which do not belong to neighboring segments. This set is called an uncovered sub-segment (USS). Determine the size of all USS. While small USS with the size less than given threshold H exists, use one of the following transformations of segments. Denote the segment with small USS as G_s and his left and right neighboring segments as G_l and G_r respectively. Check the following two possibilities.

P1. Delete G_s and add its uncovered points to G_l and G_r such that the total error of linear approximations of all points of new two segments E_2 is minimum. Resulting two new segments are non-intersected.

P2. Suppose all points of three segments G_s , G_l and G_r may be divided in three new non-intersected segments G_s^* , G_l^* and G_r^* such that $G_s^* \subseteq G_s$, $G_l^* \subseteq G_l$ and $G_r^* \subseteq G_r$ and the size of each new segment is greater or equal than H . Construct such three non-intersected segments which minimize the total error E_3 of linear approximation of all points of G_s , G_l and G_r .

If $E_2 \leq E_3$ then build new two segments by P1, otherwise build new three segments by P2. After obtaining new segments *join similar segments*.

5. *Separate intersected segments*. If after application of all previous steps some two segments are still intersected then disjoint the points from this intersection between these segments such that the total error of linear approximations of new two non-intersected segments will be minimal. After obtaining new segments *join similar segments*.

6. *Re-translate the obtained PLR*. The obtained PLR of data on non-intersected segments G_1, \dots, G_m is fully represented by the set of parameters $\{(x_{k1}, x_{k2}, p_k, q_k)\}$, ($k=1, \dots, m$), where x_{k1} and x_{k2} are the first and the last points of G_k and p_k, q_k are parameters of linear function $y_k = p_k x + q_k$ approximating data from G_k . The rule (1) may be in the simplest case constructed as follows:

$$R_k: \text{If } X \text{ is BETWEEN } x_{k1} \text{ and } x_{k2} \text{ then } Y \text{ is } S(p_k). \quad (2)$$

Often, in qualitative description of data exact values of x_{k1} and x_{k2} are not meaningful and may be replaced by nearest values of some grid defined on X [2]. For example, if values of X denote weights in kilograms but qualitative descriptions operate by tons then exact values x_{k1} and x_{k2} will be rounded till some new numbers X_{k1} and X_{k2} and (2) will be transformed in the form:

$$R_k: \text{If } X \text{ is BETWEEN } X_{k1} \text{ and } X_{k2} \text{ then } Y \text{ is } S(p_k) \quad (3)$$

The meaning of intervals *BETWEEN* X_{k1} and X_{k2} in rule base (3) is fuzzy because they denote some uncertain set of crisp intervals “similar” to these intervals. Moreover, these fuzzy intervals may be further generalized and replaced by fuzzy terms like *SMALL WEIGHT, END OF THE DAY, HIGH* etc. Such hierarchical fuzzy granulation of values of parameters is typical for qualitative descriptions [2]. The qualitative generalization of obtained PLR of TS may be used for further aggregation of rules in more general and compact form and for inference of conclusions about dependencies. For example, the qualitative evaluation of width and localization of segments may be used in perception based decision-making procedure based on generalized rule base.

Contrary, when the information about segments does not important in qualitative description of TS then the set of rules (1) may be written in simplified form as the sequence of linguistic labels of slopes (S_1, \dots, S_m) from

consequent parts of rules or by the sequence of slope values (p_1, \dots, p_m) of. Such coding of TS may be considered as generalized *MAP* transform of TS and used in further TSDM.

3. Examples of PLR of TS

The proposed *MAR-PLR* procedure depends on a set of following parameters: the size of window, the size H of minimal uncovered segment and the scale L_p of slope values. The variation of these parameters gives possibility to analyze possible structures of TS. *MAR-PLR* showed good results on a large set of both artificial and real data downloaded from INTERNET. For example, on artificial data from [2], *MAR-PLR* shows the same results but works more quickly than evolutionary algorithm considered in [2]. Below we discuss the results of application of *MAR-PLR* on two TS downloaded from INTERNET.

First TS contains information about Industrial Production Index published by Board of Governors of the Federal Reserve System [4]. It includes monthly data of time period from 1940-01-01 to 2003-07-01. Data are presented on the upper part of Fig. 1.

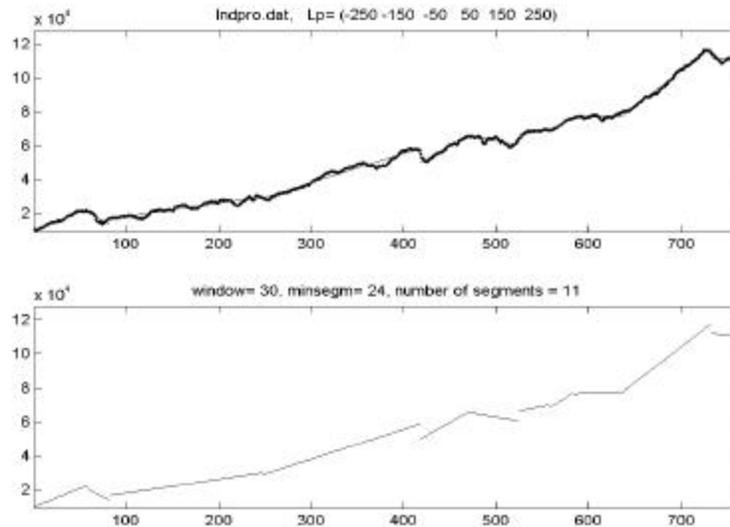


Fig. 1. Industrial Production Index.

We used the numerical slope scale: $\{-250, -150, -50, 50, 150, 250\}$ corresponding to linguistic scale L_s considered in Section 2. The size of windows equals 30 months and size of minimal segment equals to 24 months. The piece-wise linear representation of data obtained by *MAR-PLR* is presented on the upper and separately on the lower part of Fig. 1. The following sequence of slope values were obtained: (217.1, -281.0, 81.2, 174.1, 296.9, -94.2, 117.0, 346.1, 22.7, 416.8, -75.5). The retranslation of these slope values gives the following sequence of linguistic codes: (*INC*, *QDC*, *SIN*, *INC*, *QIN*, *SDC*, *SIN*, *QIN*, *CON*, *QIN*, *SDC*). The last two rules have the form:

R10: If X is BETWEEN year 1993 and year 2000 then Y is Quickly Increasing,

R11: If X is BETWEEN year 2000 and year 2003 then Y is Slowly Decreasing.

As a result of qualitative analysis it may be further concluded that “*During last 2,5 years the Index Slowly Decreased, whereas during 8 years before, it Quickly Increased*”. Qualitative description of this TS may be used also for finding association between this TS and other TS in TS database.

Fig. 2 shows results of PLR of data of Civilian Unemployment Rate monthly evaluated during time period 1948-01-01 to 2003-07-01 and published by the U.S. Department of Labor [10]. As it may be seen from Fig. 2, MAP can build good enough PLR of TS with sufficiently complex shapes.

Conclusions

The new method of qualitative description of time series is developed. It builds piece-wise linear representation of time series and transforms it in the rule set which may be used in perceptive decision making procedures [2, 11, 12]. Other positive features of the proposed procedure are the adaptive selection of the number of segments used for PLR and the use of numerical and linguistic scales of slope values for qualitative granulation of the speed of the function change. The first feature is overcome some known methods of PLR of TS which require a priory definition of the number of segments. The use of the scales of slope values gives possibility to change the qualitative description of data depending on the context of the problem and translate them in context insensitive perceptive linguistic descriptions. The proposed method based on

moving approximations transform [1] is faster than the similar method based on evolutionary generated fuzzy partitions [2].

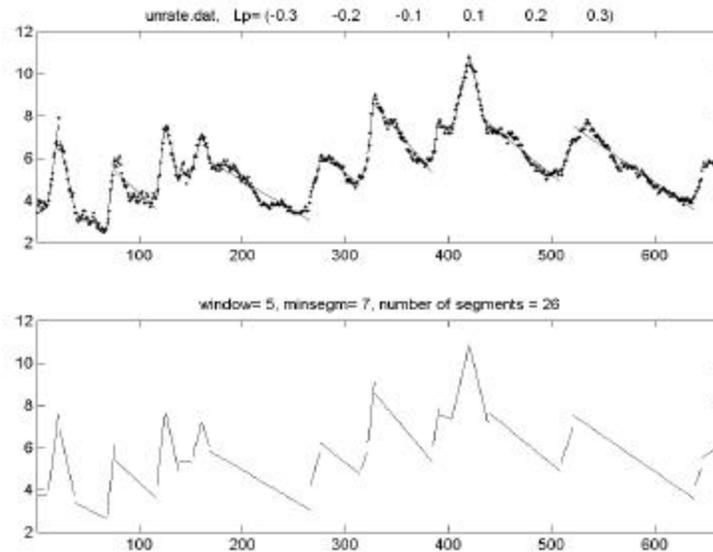


Fig. 2. Civilian Unemployment Rate.

Acknowledgements

The support for this research work has been provided by the IMP, projects D.00006 “Distributed Intelligent Computing”.

References

- [1] Batyrshin I, Herrera-Avelar R, Sheremetov L, Suarez R (2004) Moving Approximations in Time Series Data Mining. International Conference on Fuzzy Sets and Soft Computing in Economics and Finance, FSSCEF 2004, St. Petersburg, Russia, 2004 (in this volume).

- [2] Batyrshin, I., Wagenknecht, M. (2002) Towards a linguistic description of dependencies in data. *Int. J. Appl. Math. Computer Science* 12 pp. 391–401. (matwbn.icm.edu.pl/ksiazki/amc/amc12/amc1235.pdf)
- [3] Boyd S. (1998) TREND: a system for generating intelligent descriptions of time-series data. In *Proceedings of the IEEE International Conference on Intelligent Processing Systems (ICIPS-1998)*.
- [4] Federal Reserve Board, <http://www.federalreserve.gov/rnd.htm>
- [5] Keogh E, Kasetty S (2002) On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration. In *SIGKDD '02, July, 23-26, 2002*
- [6] Keogh E, Chu S, Hart D, Pazzani M (2001) An Online Algorithm for Segmenting Time Series. In *Proceedings of IEEE International Conference on Data Mining*. pp 289-296.
- [7] Keogh E (1997) A Fast and Robust Method for Pattern Matching in Time Series Databases. In *Proceedings of 9th Int. Conf. on Tools with AI (TAI 97)*.
- [8] Last M, Klein Y, Kandel A (2001) Knowledge Discovery in Time Series Databases. in *IEEE Transactions on Systems, Man, and Cybernetics* Vol. 31B.
- [9] Sripada SG, Reiter E, Hunter J, Yu J (2002) Segmenting Time Series for Weather Forecasting. In *Macintosh, A., Ellis, R. and Coenen, F. (ed) Proceedings of ES2002*, pp. 193-206
- [10] U.S. Department of Labor. Bureau of Labor Statistics, <http://www.bls.gov/ces>
- [11] Zadeh LA (1997) Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets and Systems*, vol. 90, pp. 111-127.
- [12] Zadeh LA (1999) From computing with numbers to computing with words - from manipulation of measurements to manipulation of perceptions," *IEEE Trans. on Circuits and Systems - 1: Fundamental Theory and Applications*, vol. 45, pp. 105-119.

Generating Fuzzy Rules for Financial Time Series by NN with Supervised Competitive Learning Techniques

Dušan MARČEK

The Faculty of Management Science and Informatics, University of Žilina,
Moyzesova 20, 010 26 Žilina, Slovakia, e-mail: marcek@fria.utc.sk

Abstract: Traditionally, fundamental Box-Jenkins analysis [1] and ARCH-GARCH techniques have been the mainstream methodologies used to develop time series models in finance. A dynamical process with linguistic values of the actual inflation observations is studied. To generate fuzzy rules, neural networks with SCL-based product-space clustering is used. Finally, we present our preliminary results and some further experiments that we have performed.

1 Introduction

Economic and statistical time series analysis is concerned with estimation of relationships among groups of variables, each of which is observed at a number of consecutive points in time. The relationships among these variables may be complicated. In particular, the value of each variable may depend on the values taken by many others in several previous time periods. Very often it is difficult to express exactly these dependencies, or there is not known hypothesis for that. Very frequently, in such cases more sophisticated approaches are considered. These approaches are based on the human experience knowledge and consist of series linguistic expressions each of which takes the form of an 'if ... then ...' fuzzy rule, and they are well known under the common name fuzzy controllers [3, 7].

The goal of this paper is to illustrate that two distinct areas, i.e. fuzzy sets theory and computational networks, may be used to economic time series modelling. Statistical and fuzzy time series models are presented in Section 2. Direct determination of the fuzzy rules based on the SCL (Supervised Competitive Learning) technique is described in Section 3. Concluding remarks are offered in Section 4.

2 Time series and fuzzy time series modelling

To illustrate modelling approaches used in conventional and fuzzy time series, we consider the time readings of the inflation (the 514 of monthly inflation observations in the U. S. from February 1956 to November 1998) was published at <http://neatideas.com/data/inflatdata.htm>. Fig. 1 illustrates the time plot of this time series. This time series shows no apparent trend or periodic structure. To build a forecast model the sample period for analysis y_1, \dots, y_{344} was defined, i.e. the period over which the forecasting model can be developed and estimated, and the ex post forecast period (validation data set), y_{345}, \dots, y_{514} as the time period from the first observation after the end of the sample period to the most recent observation.

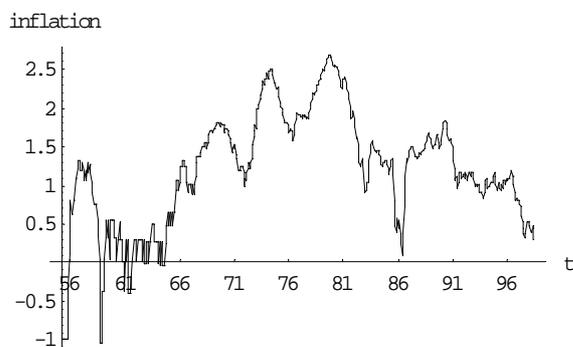


Fig. 1. Natural logarithm of monthly inflation from February 1956 to November 1998

Input selection and data preprocessing are of crucial importance to the development of time series models. Potential inputs (independent variables) were chosen based on traditional statistical tools. These include the autocorrelation function (ACF), the partial autocorrelation function (PACF) and the Akaike Information Criterion (AIC) [2]. Following this, at the starting point, we have formulated a model relating the value y_t of the series at time t that depends only on its previous value y_{t-1} and on the random disturbance ε_t , i.e.

$$y_t = \xi + \phi_1 y_{t-1} + \varepsilon_t, \quad (1)$$

where the variable y_t (in our case the first difference of inflation rate) is explained by only its previous value, Using Levinson-Durbin algorithm [2], [5] the model (1) is statistically fitted as

$$\hat{y}_t = -0,1248y_{t-1} \quad (2)$$

In contrast to the conventional time series, the observations of fuzzy time series are fuzzy sets (the observations of conventional time series are real numbers), the universes of discourse for the fuzzy sets are subsets of \mathfrak{R}^1 , where \mathfrak{R}^1 is the set of real numbers, either naturally or artificially defined. Song and Chisson [6] give a thorough treatment of these models. Let X_t , ($t = \dots, 1, 2, \dots$), a subset of \mathfrak{R}^1 , be the universe of discourse on which fuzzy sets y_t^i , ($i = 1, 2, \dots$) are defined and Y_t is the collection of y_t^i , ($i = 1, 2, \dots$). Then Y_t , ($t = \dots, 1, 2, \dots$) is called a fuzzy time series on X_t , ($t = \dots, 1, 2, \dots$).

The fuzzy time series modelling procedure consists of an implementation of several steps, usually as follows:

1. Define the input-output variables and the universes of discourse.
2. Define (collect) linguistic values and fuzzy sets on the universes of discourse.
3. Define (find) fuzzy relations (fuzzy rules).
4. Apply the input to the model and compute the output.
5. Defuzzify the output of the model.

Firstly, in the fuzzification process, we specified input and output variables. The input variable x_{t-1} is the lagged first difference of inflation values $\{y_t\}$ and is calculated as $x_{t-1} = y_{t-1} - y_{t-2}$, $t = 3, 4, \dots$. The output variable x_t is the first difference of inflation values $\{y_t\}$ and is calculated as $x_t = y_t - y_{t-1}$, $t = 2, 3, \dots$. The variable ranges are as follows:

$$-0,75 \leq x_t, x_{t-1} \leq 0,75 .$$

These ranges define the universe of discourse within which the data of x_{t-1} and x_t are, and on which the fuzzy sets have to be, specified. The universes of discourse were partitioned into the seven intervals.

Next, we specified the fuzzy-set values of the input and output fuzzy variables. The fuzzy sets numerically represented linguistic terms. Each fuzzy variable assumed seven fuzzy-set values as follows: NL: Negative Large, NM: Negative Medium, NS: Negative Small, Z: Zero, PS: Positive Small, PM: Positive Medium, PL: Positive Large.

Fuzzy sets contain elements with degrees of membership. Fuzzy membership functions can have different shapes. The triangular membership

functions were chosen. Fig. 2 shows membership function graphs of the fuzzy sets above.

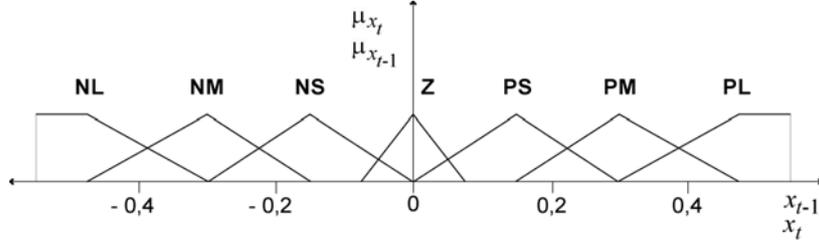


Fig. 2. The membership functions of fuzzy variables x_{t-1} and x_t

The input and output spaces were partitioned into the seven disjoint fuzzy sets. From membership function graphs μ_{t-1} , μ_t in Fig. 2 is shown that the seven intervals $[-0,75; -0,375]$, $[-0,375; -0,225]$, $[-0,225; -0,075]$, $[-0,075; 0,075]$, $[0,075; 0,225]$, $[0,225; 0,375]$, $[0,375; 0,75]$ correspond respectively to NL, NM, NS, Z, PS, PM, PL.

Next, we specified the fuzzy rule base or the bank of fuzzy relations. The appendix describes the neural network (see Fig. 3) which uses the SCL to derive fuzzy rules from data. As shown in Fig. 4(b) the bank contains the 5 fuzzy rules. For example the fuzzy rule of the 34th block corresponds to the following fuzzy relation

$$\text{IF } x_{t-1}^i = \text{PM THEN } x_t^j = \text{PS} \quad (3)$$

Finally, we determined the output action given the input conditions. We used the Mamdani's implication [4]. Each fuzzy rule produces the output fuzzy set clipped at the degree of membership determined by the input condition and the fuzzy rule. When the input value, say $x_{t-1}^i = x_{344}^i$, is applied to the time series model

$$x_t^j = x_{t-1}^i \circ R_{ij}(t, t-1), \quad (4)$$

where $x_t^j \in X_t$, $x_{t-1}^i \in X_{t-1}$, $i \in I$, $j \in J$, I and J are indices sets for X_t and X_{t-1} respectively, " \circ " is the sign for the *max-min* composition, $R_{ij}(t, t-1)$ is the fuzzy relation among the observations at t and $t-1$ times, the output value $x_t^j = x_{345}^j$ can be calculated. Since the Eq. (4) is equivalent to the linguistic conditional statement

$$\text{“if } x_{t-1}^i \text{ then } x_t^j \text{”}, \quad (5)$$

we have $R_{ij}(t, t-1) = x_{t-1}^i \times x_t^j$, where “ \times ” is the Cartesian product.

It is possible to compute the output fuzzy value x_t^j by the following simple procedure consisting of three steps:

- Compute the membership function values $\mu_{NL}(x_{t-1})$, $\mu_{NM}(x_{t-1})$, ..., $\mu_{PL}(x_{t-1})$ for the input x_{t-1} using the membership functions pictured in Fig. 2.
- Substitute the computed membership function values in fuzzy relations (4), (5).
- Apply the *max-min* composition to obtain the resulting value x_t^j of fuzzy relations.

Following the above principles, we have obtained the predicted fuzzy value for the inflation $x_t = x_{345}^j = 0,74933$.

To obtain a simple numerical value in the output universe of discourse, a conversion of the fuzzy output is needed. This step is called defuzzification. The simplest defuzzification scheme seeks for the value \hat{x}_t that is of middle membership in the output fuzzy set. Hence, this defuzzification method is called the Middle of Maxima, abbreviated MOM. Following this method, we have obtained the predicted value for the $\hat{x}_{345} = -0,15$. The remaining forecasts for ex post forecast period $t = 346, 347, \dots$ may be generated similarly.

3 Generating fuzzy rules with product-space clustering

The neural network pictured in Fig. 3 was used to generate structured knowledge of the form “if A, then B” from a set of numerical input-output data. In Section 2 we defined cell edges with the seven intervals of the fuzzy-set values in Fig. 2. The interval $-0,75 \leq x_t, x_{t-1} \leq 0,75$ was partitioned into seven nonuniform subintervals that represented the seven fuzzy-set values NL, NM, NS, Z, PS, PM, and PL assumed by fuzzy variables x_{t-1} and x_t . The Cartesian product of these subsets defines $7 \times 7 = 49$ fuzzy cells in the input-output product space \mathfrak{R}^2 .

We can represent all possible fuzzy rules as 7-by-7 linguistic matrix (see Fig. 4). The idea is to categorise a given set or distribution of input vectors

$\mathbf{x}_t = (x_{t-1}, x_t)$, $t = 1, 2, \dots, 344$ into $7 \times 7 = 49$ classes, and then represent any vector just by the class into which it falls.

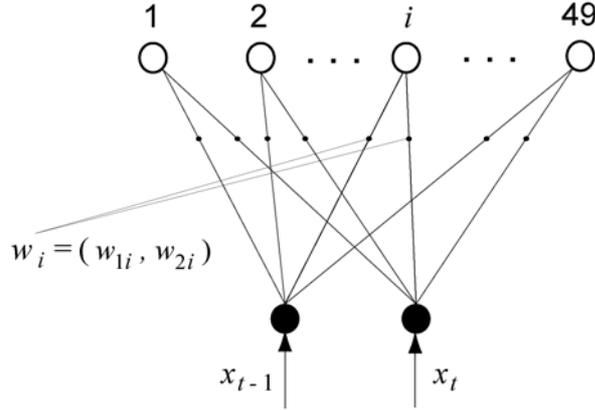


Fig. 3. The topology of the network for fuzzy rules generating by SCL-based product-space clustering

For each random input sample $\mathbf{x}_t = (x_{t-1}, x_t)$, the winning vector $\mathbf{w}_{i'} = (w_{1i'}, w_{2i'})$ was updated by the SCL algorithm according to

$$\left. \begin{array}{l} \tilde{w}_{1i'} \leftarrow \tilde{w}_{1i} + \eta (\tilde{x}_{1t} - \tilde{w}_{1i}) \\ \tilde{w}_{2i'} \leftarrow \tilde{w}_{2i} + \eta (\tilde{x}_{2t} - \tilde{w}_{2i}) \end{array} \right\} \text{if } i = i' ; \quad \left. \begin{array}{l} \tilde{w}_{1i'} \leftarrow \tilde{w}_{1i} - \eta (\tilde{x}_{1t} - \tilde{w}_{1i}) \\ \tilde{w}_{2i'} \leftarrow \tilde{w}_{2i} - \eta (\tilde{x}_{2t} - \tilde{w}_{2i}) \end{array} \right\} \text{if } i \neq i'$$

where i' is the winning unit defined $\|\tilde{\mathbf{w}}_{i'} - \tilde{\mathbf{x}}_t\| \leq \|\tilde{\mathbf{w}}_i - \tilde{\mathbf{x}}_t\|$ for all i , and where $\tilde{\mathbf{w}}_i$ and $\tilde{\mathbf{x}}_t$ is a normalized version of \mathbf{w}_i and \mathbf{x}_t respectively, η is the learning rate.

Supervised Competitive learning (SCL)-based product-space clustering classified each of the 344 input-output data vectors into 9 of the 49 cells as shown in Fig. 4(a). Fig. 4(b) shows the fuzzy rule bank. We added a rule to the rule bank if the count of input-output vectors in particular cells was larger than the value $0,05N$, where $N = 344$ is number of data pairs (x_{t-1}, x_t) , $t = 1, 2, \dots, N$ in the input and output series. For example the most frequent rule represents the cell 34. From most to least important (frequent) the fuzzy rules are (PM; PS), (PS; PL), (NL; NS), (PS; PL), and (PS; PS).

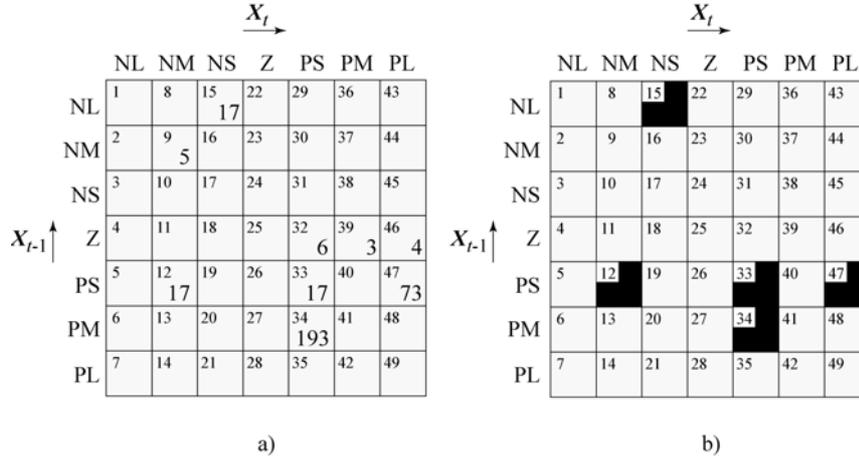


Fig. 4. Distribution of input-output data (x_{t-1}, x_t) in the input-output product space $X_{t-1} \times X_t$ (a). Bank of fuzzy rules of the time series modelling system (b).

4 Conclusion

In this paper, we have presented an application of the fuzzy time series model to forecast an autoregressive process. A formal framework for the definition of fuzzy rules has been presented. This framework is based on the simple competitive learning of networks. The neural network with the SCL clustering technique was used to determine the fuzzy relation (fuzzy rules) of first-order fuzzy time series models directly from data. The proposed method is also suitable for parameter estimations of econometric models, in applications of deterministic non-linear dynamics and chaos theory in contemporary economics and finance.

As a final point, let us examine what has been gained by use of a fuzzy time series model over an ordinary AR(1) model for the output x_{345} . For this purpose, we have computed prediction limits on the one-step-ahead forecast from the AR(1) model, and fuzzy time series model. The 95 percent interval around the actual inflation value based on the statistical theory is $\hat{x}_{345} \mp u_{1-\alpha/2} \hat{\sigma}_\varepsilon (1 + \phi_1^2)^{1/2} = (-0,0442; 0,05043)$, where $\hat{x}_{345} = 0,00312$ represents the forecast for period $t = 345$ made at origin $t = 344$, $u_{1-\alpha/2}$ is a $100(1 - \alpha/2)$ percentile of the standard normal distribution, and

$\hat{\sigma}_\varepsilon$ an estimate of the standard deviation of the noise. An intuitive method for constructing confidence intervals for fuzzy time series model is simply the defuzzification method First of Maxima and First of Minima to obtain prediction limits on the one-step-ahead forecast. In our example, the „confidence“ interval for fuzzy time series value $\hat{x}_{345} = -0,15$ is (-0,30256 to 0,3088). The sign of the actual value for the AR(1) model is opposite to the forecast value sign.

The method may be of real usefulness in practical applications where usually the expert can not explain linguistically, what control actions the process takes or there is no knowledge of the process.

Because the results were based of chosen inflation rates and data set, they were difficult to generalize in others situations. Yet, the results certainly provide a rational way for improvement of forecasting ability in chaotic economic systems.

Acknowledgement

This work is supported by Slovak grant foundation under the grant No. 1/9183/02.

References

1. Box, G. E. P., and G. M. Jenkins: Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, CA, 1970
2. Brockwell, P.J. and Davis, R.A.: Time Series: Theory and Methods. Springer-Verlag, New York, 1987
3. Kosko, B.: Neural networks and fuzzy systems - a dynamical systems approach to machine intelligence. Prentice-Hall Int., Inc. 1992
4. Mamdani, E., H.: Application of a fuzzy logic to approximate reasoning using linguistic synthesis. IEEE Trans. Comput. 26 (1977), pp. 1182-1191
5. Morettin, P., A.: The Levinson algorithm and its applications in time series analysis. International Statistical Review, 52, 1984, pp. 83-92
6. Song, Q., Chisson, B., S.: Fuzzy time series and its models. Fuzzy Sets and Systems, 54, 1993, pp. 269-277
7. Sugeno, M., and Yasukawa, T.: A fuzzy-logic-based approach to quantitative modelling. IEEE Trans. Fuzzy System. Vol. 1, No. 1, Feb. 1993, pp. 7-31

Pattern Recognition through Perceptually Important Points in Financial Time Series

Gul Zaib, Uzair Ahmed & Arshad Ali

National University of Sciences and Technology (NUST), NUST Institute of Information Technology, Chaklala Scheme III, Rawalpindi, Pakistan Phone: +92-51-9280658, Fax: +92-51-9280782
Email: {gul.zaib,uzair.ahmed,arshad.ali}@niit.edu.pk

Abstract

Technical Analysis is a financial risk management practice that has been in use since the advent of stock market and Pattern Recognition is an indivisible part of it. There has been a lot of research in to pattern recognition in time series. Existing pattern recognition techniques lacks dynamic extensibility. They do not provide any interfaces in order to include new patterns for recognition dynamically. This limits the operability of these techniques to a particular domain. This research devises a new technique for domain independent pattern recognition while giving sufficient speed and accuracy. This enables it to be used by critical Decision Support Systems for time series of different domains. The system emulates human visual cognition process by implementing the concept of Perceptually Important Points Identification (PIPI). Perceptually Important Points (PIP) represents minimal set of data points which are necessary to form a pattern. For dynamic inclusion of patterns a Pattern definition Language (PDL) has been conceptualized for defining patterns in time series by using declarative programming paradigm. This also results in domain independent pattern recognition without needing any modification.

Keywords: stock pattern mining, pattern recognition, financial time series, soft computing, declarative programming, time series analysis, perceptual modelling, computer vision, non-linear pattern recognition, modelling human cognition

Introduction

Historic time series data is one of the most precious assets of modern business and scientific research entities. Modern decision support systems (DSS) are expected to execute important policies based upon patterns occurring in time series. This makes pattern recognition an integral part of any future DSS.

Pattern recognition is primarily a human cognitive faculty. Depending upon an individual's ability faces severe limitations. Its expertise varies among individuals, thus it can not provide a basis for the decisions involving a high degree of risk. Secondly scanning huge amounts of data to find a single pattern or even recursively scanning it for multiple patterns is a humanly impossible task [1]. Still human cognitive process is indispensable as all the small nonlinearities and variations in patterns can not be captured through traditional parametric approaches. So we develop a system which utilizes the strengths of human cognitive process while removing its above-discussed weaknesses.

We develop a system that closely imitates human faculty of pattern recognition and is called Perceptually Important Point Identification (PIPI).

Briefly, it first divides the time series being scanned into segments. Once segmentation is completed the whole system explores the time series to discover PIPs. Once PIPs in a specific formation are found, the system classifies that segment of time series in which PIPs are found as making the desired pattern. To extend the system the user simply has to define the pattern in our specified PDL (Pattern Definition Language). The system would then be ready to start looking for that new pattern in large volumes of Time Series.

Rest of the paper is structured as follows. First the status of current research in the field of pattern recognition is explained. A PIP (Perceptually Important Point) will be defined; also the advantages gained by using perceptually important points instead of traditional parametric techniques are given after that. Then the whole PIPI is explained along with an example of stock time series. The structure of PDL and the architecture of its parser and interpreter will be given after that. The results of two performance gauging experiments done on PIPI are given and in the end there is a conclusion.

Previous work

[8] Have used hierarchal pattern analysis for pattern recognition in time series. Hierarchical pattern analysis The problems with this technique are a) a pattern must be seen at least once to be recognized, b) The user can not him self add patterns to it.

[7] Applied self organizing maps to the process of pattern discovery for making neural network training for pattern discovery efficient and accurate. [5] Has presented his technique of SNN (Single nearest neighbor) for pattern detection in long memory systems. This approach is extremely slow and further becomes slower as the time series or size of the segment becomes larger.

[6] pattern matching utilizes distance measure by storing the Euclidian distance between fixed points, and then compares that distance with the previous time series to locate a pattern.

Most of the research has been mainly focused on two attributes of pattern recognition i.e. efficiency and accuracy. There is still a need for a pattern recognition technique which should be able to work across domains, should be extensible enough, meaning that once it has been trained to recognize certain patterns at development time, it can be extended later to add more patterns. The features of extensibility and domain independence should not be achieved at the cost of efficiency or accuracy i.e. it should be efficient enough to be used in interactive system, should be accurate enough so that people could rely on it for critical decisions.

P.I.P.I Perceptually Important Point Identification:

P.I.P (Perceptually Important Points) definition

Time series are made up of finite but very large number of data points if we consider a pattern on a time series it will be subset of the points of making the whole time series, but even all those points are not required for pattern recognition. If we concentrate on even a smaller set of point we are still able to recognize pattern. These points which are necessary for pattern recognition are termed in this paper as Perceptually Important points or PIPS.

Challenges requiring the use of PIPS

- **Domain Independent:** The pattern recognition should work across domains the technique which should be able to ignore differences in the properties of time series. It should be flexible to work across domains
- **Time Unit Independent:** Time series can be on varying time scales within and across domains so there should be such a technique which is able to work no matter the time scale is hours, years or any thing in between as long as it is constant within a pattern.
- **Highly Non Linear Patterns:** Most of the patterns can not be defined strictly in terms of shape. Small non linearities will always exist in time and value.
- **Missing Data:** Most of the time several data values between 2 ranges of time series are missing. Thus we must also implement some method to overcome the problem of missing data
- **Need for a smoothing function:** Time series having a lot of small non linearities confuse the traditional techniques, so the data is passed through filters in order to smooth it.

Advantages of using PIP

1. **Time Unit Independence:** When using PIPS we get independent from time unit, it does not matter what unit of time each data point of time series enclose as long as all the PIPs in the pattern belong to same time unit.
2. **No need of a smoothing function:** When using PIPs there is no need of a smoothing function as highly noisy data does not affect accuracy. This is because we do not need to feed whole time series to the technique. Only we need the data related to PIPs.
3. **Missing data do not effect accuracy:** We do not need any filter to fill in missing data as long as the data points that will form the PIPs of the pattern exist, on these PIPs will be sufficient to do the pattern recognition
4. **Pattern Definition Feasible:** Further more when using PIPs, pattern definition is feasible as we only need some way to represent the PIPs in the time series. It also represents their relative position

in the pattern in order to make our algorithm domain independent and evolvable.

PIP identification algorithm

1. Find Minimum and Maximum value of time series for this time series. For this time series min and max values are 2.5 and 20.5 respectively.
2. Divide whole time series in to segments with a segmentation factor. The segmentation factor is fixed for time series belonging to each domain for example it is 1.5 % for stock data. However it can too be varied to increase the accuracy of recognition algorithm for each domain. The density of points lower as we go up, so the band size must increase in order to allow for lowering density. For the sample time series the number of bands will be 24
3. Start PIP Identification Process once the time series has been segmented. For this we first; arrange the segments in order by their volume, because almost all the pattern starts from the segment having high density of data points. In this sample series this segment is the segment 1.
4. Find Neighboring Pips that lie in the bottom layer, these are marked with red arrows, check if they satisfy the pattern i.e. in same band number is greater than four then we move on to upper layer so that the shoulder should be completed, now we will find the points in the upper layer at the mean of point already found, but not at exact mean, give or take a single time unit
5. Remember the matching points found, if they become equal to a specified number then the pattern is successfully detected e.g. for Head and Shoulder Pattern there must exist at least eight points.

Pattern Definition Language

PDL requirements

PDL (Pattern Definition Language) makes the system extensible, so it should be capable of modeling different aspects of the pattern. Each PIPs has two properties, the value on x axis which is always

some unit of time and the value on y axis which is different for each time series. To define the relation ship between PIPs we will also take, the distance of one PIP from another PIP in time and the distance of one PIP from another PIP in value

One of the strengths of human cognition is its ability to approximate patterns. Approximation means that humans have the capability to ignore small difference and take a look at the bigger picture to recognize a pattern. So the PDL designed should be capable of:

1. Defining the actual change in value that must be present between one PIP and the next.
2. The variation that can take place in the value change, i.e. the number of value units around the original where a PIP can exist if it's not found at the original change position.
3. As it is quite possible that the next PIP is not exactly at the specified time interval so we must also specify a variation for time and for value.
4. The variation or interval may be defined with respect to any previous point found. For example if the last point we found is "n". A requirement comes to find point that is 3 time units a head of point n-3 i.e. a point found three points back. So the PDL should be capable of doing this
5. Moreover the variation can be till the end of time series so there must be some way to specify the end of time series along X axis or Y axis.

PDL structure

Table 1, shows different key symbols.

Symbol	Usage
N	Specify variation to the end of time series
+,-	Define the direction of finding next PIP: +: Move forward in time or

	move higher in value -: Move previous in time or lower in value
value	Any value written between indicate the reference to any previous point i.e. 2 means move to point found 2 points back.
/	The line separator used in PDL, just like the use of “;”.
+	Used to show the start of a pattern i.e. to indicate that system should pick up the first point to start looking for new pattern

Conclusion

PIPI is a simple yet powerful technique to recognize patterns. It improves over previous work in this area in the following aspects

- It provides a high level abstraction of time series patterns, rather than treating pattern discovery process as a black box – something that the user may be uncomfortable with
- Its specification is close to how human keep patterns in memory, where as its working closely imitates the human way of pattern recognition
- PDL let us define patterns based upon the relationships between perceptually important points. It makes the system extensible in two aspects. Enabling dynamic inclusion of pattern and also the technique lends it self to be used across the domains without any modification.

A PIPI prototype implementation exhibits very impressive resource utilization during the detection process of the pattern and also shows highly optimized time series search. The experimental data was of year 1995 for share prices of CNN. Besides we also used a hypothetical data of 87 years for stress tests. A graphical user inter-

face is the future direction of this. This shall enable visual usage of PDL. It will help the user to define a pattern by marking the segment of time series in which the pattern occurs.

REFERENCES

- [1] Hedge, Human Information Processing – Pattern Recognition, Decisions And Attention,. Cornell University 1999.
- [2] Frietman, XCG-NEUROPC: Senter R & D program, Faculty of Applied Physics, Delft University of Technology, Delft, Netherlands 1994.
- [3] Identifying Seismicity Patterns Leading Flank Eruptions At Mt. Etna Volcano During 1981-1996, Center for Theoretical Physics, Laboratory for Nuclear Sciences and Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 1999
- [4] Tarr, Visual Pattern Recognition, Brown University 1996
- [5] Singh,A Long Memory Pattern Modelling And Recognition System For Financial Time-Series Forecasting, University of Exeter Department of Computer Science Prince of Wales Road Exeter 2001
- [6]Pratt LOCATING PATTERNS IN DISCRETE TIME-SERIES Department of Computer Science and Engineering College of Engineering University of South Florida 2001
- [7]Fu, Pattern Discovery From Stock Time Series Using Self-Organizing Maps, KDD2001 Workshop on Temporal Data Mining,
- [8] Ge, Pattern Matching in Financial Time Series Data, third conference on Knowledge Discovery in Databases and Data Mining 2000

Soft Clustering for Funds Management Style Analysis: Out-of-sample Predictability

Paul Lajbcygier and Asjad Yahya

School of Business Systems,

Monash University, Clayton, Australia.

Email:Paul.Lajbcygier@infotech.monash.edu.au

Abstract

Mutual funds are generally grouped on the basis of their investment objectives. Investors use these groupings, or 'styles', to make investment decisions and compare the performance of fund managers against appropriate indexes. Various authors have criticized the traditional classification approaches because they are made on the basis of funds' stated objectives, not the actual fund styles and are therefore often misclassified. [6] introduced the General Style Classification (GSC) approach as an alternative to the traditional technique which uses statistical (rather than stated) objectives. Results of their study suggest that investment style boundaries are continuous rather than "hard". In this work, we classify mutual funds using a soft clustering technique (Fuzzy-C-Means) and compare it with a hard clustering technique (GSC). This comparison demonstrates soft clustering can predict mutual fund performance better out-of-sample.

INTRODUCTION

The mutual funds industry has grown considerably over the years. Between 1995 and 2000, the number of mutual funds worldwide nearly doubled from 34,992 to 53,450 and assets managed by these funds nearly tripled from US \$5.386 trillion to US \$12.152 trillion [1]. In the United States alone, 47.0 million households had a stake in mutual funds in 2002 [2].

The fact that independent sources like Morningstar publish their own styles, despite officially stated objectives by the funds, highlights a problem faced by the mutual fund industry: misclassification of funds.

To address the problem of fund misclassification, [6] developed an alternative to the traditional techniques. The Generalized Style Classification (GSC) approach uses past returns to group funds into a pre-determined number of styles. The underlying concept is that funds following similar

investment patterns, irrespective of their objectives, will have similar returns and hence will be placed in one group when statistically clustered on the basis of past performance. The use of past returns has the added advantage that GSC is not fooled by window-dressing¹. Empirical evidence suggests that this technique is superior to traditional industry classifications in predicting cross-sectional future and past performances [6-7].

Results of the study in [6] showed that the investment actions of the mutual fund managers in America could be explained by eight different investment styles. Each style was formed using the GSC technology, and is comprised of a combination of a number of traditional style definitions. For example, the ‘Glamour’ style defined in [6] consists primarily of Morningstar ‘growth’ and ‘small company’ funds. This example illustrates that in-practice fund managers do not restrict themselves solely to one investment style and suggests that the style categories have continuous rather than “hard” boundaries. If this holds true, the analysis of mutual funds through soft clustering should yield greater insight into the actions of fund managers as opposed to an analysis through the use of a hard clustering approach (i.e. the GSC). The main problem with the latter is that, despite the fact that the results suggest the existence of continuous style boundaries, the algorithm places each fund exclusively into only one cluster. Soft clustering, on the other hand, ‘divides’ the fund among a number of clusters by assigning membership values to each of the pre-defined styles. This permits fund managers to belong to multiple investment styles.

The main aim of this paper is to establish a rationale for switching to the use of soft clustering from hard clustering for mutual funds analysis. This is achieved by illustrating the advantages of analyzing a dataset of American mutual fund returns using a soft clustering approach (Fuzzy-C-Means) as compared to an analysis using a hard clustering approach (GSC) and, in particular, comparing their out-of-sample mutual fund return predictions.

METHODOLOGY AND DATA

We have shown ([14]) that the theoretical returns generating process is consistent with the use of k-means clustering. This motivates the use of the quasi k-means clustering approaches used in this study — the ‘hard’ clustering GSC, and its soft clustering analog — Fuzzy-C-Means. The de-

¹ At the end of each period, fund managers often sell off poor performers in the portfolio and/or change the apparent strategy of the fund. This practice is known as ‘Window dressing’.

tails of their derivation (and tweaking) are omitted here due to space constraints, however they are discussed in [14].

Fuzzy-C-Means, which is a generalization of the “hard” k-means [8] clustering approach, was chosen because: it and the GSC are similar to k-means clustering and therefore can be justified based on the theoretical return generating process. In particular, mutual fund return predictions are compared using the GSC and Fuzzy-C-Means technology. We find that the Fuzzy-C-Means approach yields more accurate out-of-sample predictions.

The dataset was made up of 42 monthly returns, from 31 January 1997 to 30 June 2000, for 7594 U.S. equity funds. As is common with most mutual funds datasets, a number of values were missing².

RESULTS

This section discusses the results comparing GSC and Fuzzy-C-Means. It does this by determining which approach — GSC or Fuzzy-C-Means — provides more accurate out-of-sample predictions of funds management performance.

When attempting to compare different approaches for the estimation of funds management style we want to know how one approach can be said to be better than another. A key issue is that “there are no generally accepted standards for comparing style classifications” [6]. So, [6] borrowed “a natural measure from the asset pricing model” and used regression to measure the performance of GSC against other classification techniques, including the traditional methodology. Fund returns were cross-sectionally regressed on a matrix of dummy variables, which indicated whether a fund belonged to a particular style. This exercise was carried out for both GSC and industry-based classifications. A comparison of adjusted R^2 s helped identify the classification technique with superior predic-

² Missing returns were represented by ‘-95’ in the dataset. The GSC algorithm was designed to handle missing values, but results for Fuzzy-C-Means would have been distorted if unfiltered data had been used. Three methods were considered to overcome this problem: eliminating time series and/or funds with missing values beyond an acceptable threshold level; replacing missing returns with average returns; replacing missing values with zeros. After careful consideration of the effects these techniques would have on the final results, missing values were handled the way the GSC ([6]) algorithm handles them also by replacing missing returns with zeros. This ensured consistency between the two techniques.

tive ability. “This procedure resembles classical time-series, cross-section tests of pricing models, except that the cross-sectional regressors are not loadings but a matrix of dummy variables” [6].

The methodology used in this study to compare Fuzzy-C-Means with GSC is a variation of the technique described above. The following section presents the results of the regression analysis, along with details of deviations from the technique originally used by [6].

[6] used a twenty-four month rolling period of fund returns, instead of the entire dataset, for the cross-sectional regression. The use of a rolling window, which was moved forward twelve months at a time, relaxed the assumption that funds belonged to the same style over the entire period [6]. This exercise was carried out for both GSC and industry classifications. Adjusted R^2 s were compared to identify the classification technique that did a better job of explaining cross-sectional variance.

The regression analysis carried out in this study differs from the [6] methodology in two key ways. Firstly, instead of a twenty-four month rolling window, a thirty-six month period was used for the regression³. Secondly, due to the shortage of available data, the returns window was rolled on a month-by-month basis instead of a yearly basis. [6] found that the industry-based styles significantly under-performed other techniques when they used rolling month-by-month returns. This was due to the fact that the other techniques used data subsequent to the publication date of the industry styles to classify funds. Thus, in order to overcome this problem, [6] used a one-year test period. However, this study is not constrained by this limitation: both GSC and Fuzzy-C-Means use the same data to classify funds.

Table 1 reports the results of out-of-sample cross-sectional regressions for GSC and Fuzzy-C-Means. Since both GSC and Fuzzy-C-Means classify funds into the same number of styles, simple R^2 s were used instead of adjusted R^2 s to compare their predictive ability. The results appear to indicate that Fuzzy-C-Means does a better job of classifying funds than the GSC. Apart from the first two test periods, the R^2 s for the Fuzzy styles are significantly higher than those for GSC. Thus, the results seem to imply that the use of soft clustering in general, and Fuzzy-C-Means in particular, can provide more accurate prediction than hard-clustering.

³ [6] found that the use of rolling month-by-month returns for 24, 36, 48 and 60 months to classify funds yielded qualitatively similar results. Thus, due to the length of time it takes GSC to classify funds for each set of returns, a thirty-six month period was used.

Table 1 Regression Results

	R-squares	
	GSC	Fuzzy-C-Means
Jan 2000	0.112	0.1
Feb 2000	0.5504	0.5576
Mar 2000	0.4647	0.5415
Apr 2000	0.4232	0.5238
May 2000	0.4529	0.5557
Jun 2000	0.4983	0.5952
Mean	0.4169	0.479
Median	0.4588	0.5486
Std. Dev	0.1556	0.1871

A key consideration regarding this analysis is that, while the results seem to support the theoretical justification presented earlier in the study to switch from the a hard to a soft clustering approach, a number of caveats must kept in mind while interpreting them. Firstly, the test was conducted over a very short period: as mentioned earlier, due to the amount of time it takes GSC to classify each set of funds, a larger returns window was chosen to decrease the computation time, which led to fewer test periods. The GSC styles perform slightly better than their Fuzzy counterparts in the first test period. It is possible that GSC does a superior job of classifying funds prior to this period. Secondly, due to lack of availability of data, the returns window was rolled on a monthly basis. The use of such a small rolling period, combined with the small number of test periods, could have resulted in spurious results. Lastly, the presence of missing values in the test periods decreases the reliability of the results. Thus, although these results warrant further analysis into the use of soft clustering for style analysis, they should not be used as the sole basis for the replacement of GSC with Fuzzy-C-Means.

CONCLUSION

In this work we have introduced Fuzzy clustering to funds management style analysis and have found it to be a very powerful diagnostic tool. The technology appears to permit us to predict funds management performance out-of-sample more accurately than with the GSC technology.

References

1. Investment Company Institute (2001) Mutual Fund Fact book
2. Investment Company Institute (2002) Equity Ownership in America
3. Kim, M., Shukla, R. and Tomas, M.: Journal of Economics and Business 52(4) (2000) 309-323
4. Wolper, Gregg (2002): Our style box should help you, not bind you
<http://news.morningstar.com/doc/article/0,1,13564,00.html>
5. Costello M.: The new growth (or old value) <http://money.cnn.com/>
6. Brown, S. J. and Goetzmann, W. N.: Mutual Fund Styles. Journal of Financial Economics 43 (1997) 373-399
7. Lajbcygier, P. and Ong, M. Y.: Estimating the Number of Mutual Fund Styles Using the Generalized Style Classification Approach and the Gap Statistic. Proceedings of Computational Intelligence in Finance (2003) IEEE Press
8. Hartigan, J. A.: Clustering Algorithms, Wiley, New York (1975)
9. Bezdek, J. C.: Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York (1981)
10. Klir, G. J. and Yuan B.: Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice Hall PTR, New Jersey (1995)
11. Quandt, R.: Tests of the Hypothesis That a Linear Regression System Obeys Two Regimes. Journal of American Statistical Association 55 (1960) 324-330
12. Odeh, I. O. A., McBratney A. B. and Chittleborough D. J.: Soil Pattern Recognition with Fuzzy-means: Applications to Classification and Soil-Landform Interrelationships. Soil Science Society of America Journal 56 (1992) 505-516
13. Barès P. A., Gibson R. and Gyger S.: Style Consistency and Survival Probability in the Hedge Funds' Industry. Swiss Federal Institute of Technology (2001)
14. Yaha A., Lajbcygier P., A rationale for using soft clustering for funds management style analysis", pp.412-416, Proceedings of the International Fuzzy Systems Association World Congress, Okyay Kaynak (Ed.), Istanbul, Turkey, June 26-29 2003, Figur Grafik & Matbaacilik Sanayi Ticaret Ltd.

Tuning the Fuzzy Classification Models with Various Learning Criteria: the Case of Credit Data Classification

Serhiy SHTOVBA, Olga PANKEVICH, and Georgios DOUNIAS

Department of Computer Based Information & Management Systems,
Vinnitsa National Technical University, Vinnitsa, Ukraine
shtovba@ksu.vstu.vinnica.ua

Department of Financial and Management Engineering,
University of the Aegean, Chios, Greece, g.dounias@aegean.gr

Abstract. In this paper we study the efficiency of various learning criteria for the proper tuning of a fuzzy classifier. Different cases of crisp and noisy class borders are considered, and a specific credit-risk application is discussed.

1 Introduction

Fuzzy rule-based systems are powerful tools which perform adequately in classification tasks related to various financial and economic decision-making problems, such as customer segmentation, credit-risk prediction, project evaluation, fraud detection, etc. [3, 4]. Fuzzy classifiers usually provide a good balance between decision accuracy and model transparency. In this paper we study the efficiency of various learning criteria for the tuning process of a fuzzy classifier. Tuning corresponds to the search-process of weights of fuzzy if-then rules and parameters of the membership functions that minimize the difference between actual and inferred decisions. Cases of crisp and noisy class-separating curves with credit-risk assessing application are considered.

2 The Fuzzy Rule – Based Classifier

Let us consider a classifying system with n -inputs (x_1, x_2, \dots, x_n) and one output y . The classification can be considered as the mapping:

$$X = (x_1, x_2, \dots, x_n) \rightarrow y \in \{d_1, d_2, \dots, d_m\},$$

where d_1, d_2, \dots, d_m are decisions. The classification is performed with the aid of the following fuzzy knowledge base [1]:

$$\bigcup_{p=1}^{k_j} \left(\bigcap_{i=1}^n x_i = a_{i,jp} \text{ with weight } w_{jp} \right) \rightarrow y = d_j, \quad j = \overline{1, m}, \quad (2.1)$$

where \bigcap is logical operation AND, \bigcup is logical operation OR, $a_{i,jp}$ denotes a fuzzy term for the evaluation of input x_i in rule with number jp , $w_{jp} \in [0,1]$ is a subjective degree of the expert's confidence in rule with number jp , k_j is the number of rules corresponding to decision d_j .

The membership degrees of an object $X = (x_1, x_2, \dots, x_n)$ to decisions d_j ($j = \overline{1, m}$) are calculated as follows [1]:

$$\mu_{d_j}(X) = \bigvee_{p=1, k_j} w_{jp} \cdot \bigwedge_{i=1, n} [\mu_{jp}(x_i)], \quad j = \overline{1, m}, \quad (2.2)$$

where $\mu_{jp}(x_i)$ denotes the membership function of the fuzzy term $a_{i,jp}$, and \bigvee (\bigwedge) is the max (min) operation. The decision with the maximal fulfillment degree corresponds to object X :

$$y = \arg \max_{\{d_1, d_2, \dots, d_m\}} (\mu_{d_1}(X), \mu_{d_2}(X), \dots, \mu_{d_m}(X)). \quad (2.3)$$

3 Learning Criteria for Tuning the Fuzzy Classifier

Let us denote a fuzzy classifier by

$$y = F(X, P, W) \quad (3.1)$$

where X is an input vector, P is a vector of the membership functions' parameters in the knowledge base (2.1), W is a vector of the rule-weights in (2.1), and F is an input-output operator corresponding to (2.2) - (2.3).

We denote the training set by

$$(X_r, y_r), \quad r = \overline{1, M}, \quad (3.2)$$

where $X_r = (x_{r,1}, x_{r,2}, \dots, x_{r,n})$ and y_r are the input vector and its corresponding output for the input-output pair with number r .

The tuning corresponds to searching a vector (P, W) that minimizes the difference between actual (3.2) and inferred (3.1) decisions. This difference may be defined in various ways.

Criterion 1. The percentage of misclassification is widely used as a learning criterion for diverse pattern recognition tasks. For this case, the tuning is equivalent to the following minimization:

$$\frac{1}{M} \sum_{r=1, M} \Delta_r \rightarrow \min, \quad \text{where } \Delta_r = \begin{cases} 1, & y_r \neq F(X_r, P, W) \\ 0, & y_r = F(X_r, P, W) \end{cases} \quad (3.3)$$

An advantage of this criterion is its simplicity and the clear interpretation of values. The main drawback is the optimization difficulty related with plateau-shaped objective functions. It is often very hard to guess the suitable parameters of gradient optimizing routines, for example, the change in variables for finite difference gradient calculation.

Criterion 2. Let us apply fuzzification of the output variable in the training set (3.2) as follows [1]:

$$\left. \begin{aligned} \tilde{y} &= (1/d_1, 0/d_2, \dots, 0/d_m), & \text{if } y = d_1 \\ \tilde{y} &= (0/d_1, 1/d_2, \dots, 0/d_m), & \text{if } y = d_2 \\ & \dots \\ \tilde{y} &= (0/d_1, 0/d_2, \dots, 1/d_m), & \text{if } y = d_m \end{aligned} \right\} \quad (3.4)$$

The desirable values of the inferred membership grades (2.2) are equal to (3.4). Hence, the tuning problem may be now formulated as in the following minimization [1]:

$$\frac{1}{M} \cdot \sum_{r=1}^M \sum_{j=1}^m \left[\mu_{d_j}(y^r) - \mu_{d_j}(X^r, P, W) \right]^2 \rightarrow \min, \quad (3.5)$$

where $\mu_{d_j}(y^r)$ denotes the desirable membership degree according to (3.4), and $\mu_{d_j}(X^r, P, W)$ denotes the inferred membership degree of object X^r by formula (2.2).

The objective function in (3.5) does not have large plateaus, allowing the use of gradient-based optimization methods. However, the optimal vector for (3.5), sometimes does not obtain a minimal misclassification level

as well, due to the presence of objects, laying close to the class-separating curves, which almost equally contribute to correct and error classification.

Criterion 3. A combination of the advantages of the abovementioned criteria is proposed below. The main idea is to increase the contributions into (3.5), for the misclassified objects. The fuzzy classifier tuning process, is now formulated according to the following minimization problem:

$$\frac{1}{M} \cdot \sum_{r=1}^M \left((\Delta_r \cdot R + 1) \cdot \sum_{j=1}^m \left[\mu_{d_j}(y^r) - \mu_{d_j}(X^r, P, W) \right]^2 \right) \rightarrow \min, \quad (3.6)$$

where $R > 0$ is a penalty value.

Let us refer to (3.3) as criterion I, (3.5) as criterion II, and (3.6) as criterion III. The efficiency of the described learning criteria is studied below.

4 Experiment for the Case of Crisp Separating Curves

We consider a classification task with 2 inputs ($x_1, x_2 \in [0,1]$) and 3 decisions (d_1, d_2, d_3). Fig. 1 shows the data sets and class separating curves. The training set consists of 80 objects and the test set consists of 5000 objects. The input data in the sets were generated randomly.

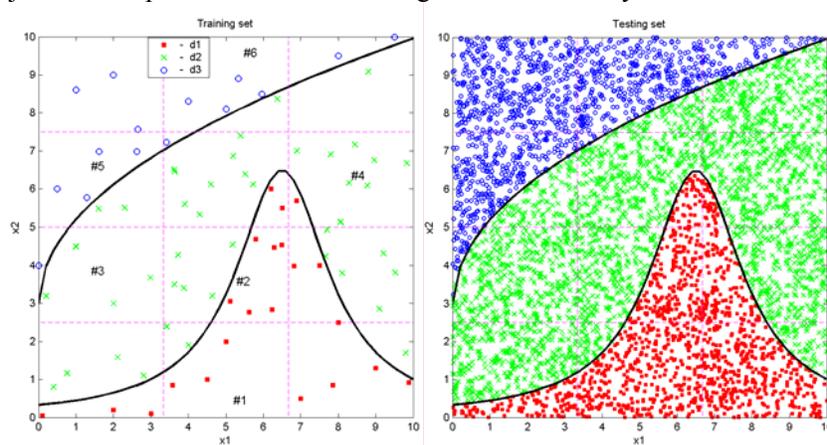


Fig. 1. Training and test sets

Table 1 shows the expert fuzzy knowledge base. The source and optimal (by various criteria) membership functions are shown on Fig. 2.

Table 1. Fuzzy knowledge base

x ₁	x ₂	y	w			
			source classifier	classifier I	classifier II	classifier III
average	low	d ₁	1	0.62	0.75	0.71
average	below average	d ₁	1	0.41	0.39	0.49
low	below average	d ₂	1	0.81	1	0.90
high	higher average	d ₂	1	0.46	1	0.71
low	higher average	d ₃	1	0.66	0.49	0.65
average	high	d ₃	1	0.02	0.02	0.91

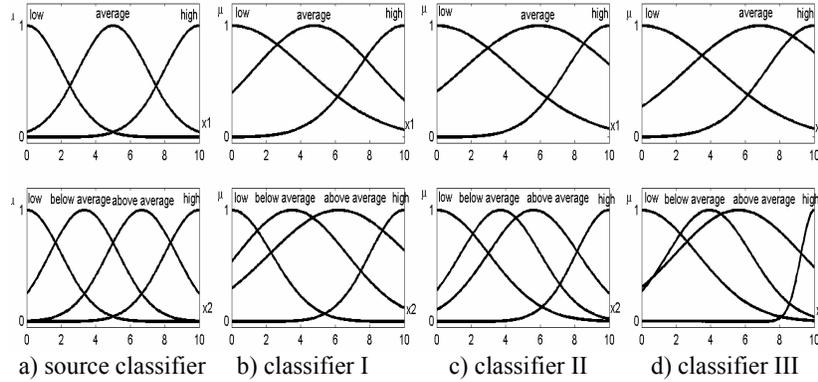


Fig. 2. Membership functions

Table 2 compares the results achieved by the tuning process according to various learning criteria. As an alternative classification tool we used the following decision tree (stated below as a set of equivalent decision if-then rules):

if ((x ₁ >1.2929) & (x ₂ ≤1))	then y=d ₁ ,
if ((x ₁ >4.6335) & (x ₂ >1) & (x ₁ ≤7.5) & (x ₂ ≤6))	then y=d ₁ ,
if ((x ₂ >8.3607) & (x ₁ >5.3301))	then y=d ₃ ,
if ((x ₁ ≤5.3301) & (x ₂ >6.9107))	then y=d ₃ ,
if ((x ₁ ≤1.2929) & (x ₂ >3.4988) & (x ₂ ≤6.9107))	then y=d ₃ ,
otherwise,	y=d ₂ .

Table 2 shows that the usage of criterion III provides the best classification accuracy (error of 9.28%). Criterion II performs worse of all criteria. In fact, there is a small difference between the accuracy levels of Classifiers I and III but the matching of the feasible parameters for the gradient-based optimization according to criterion I, is often a time-consuming process.

Table 2. Testing the classification models

Model	Criterion I	Criterion II	Criterion III (R=9)	Misclassification on test set
Source classifier	32.5%	0.53	3.99	25.92%
Classifier I	6.25%	0.52	1.03	9.78%
Classifier II	18.75%	0.44	1.88	16.42%
Classifier III	6.25%	0.46	0.92	9.28%
Decision tree	7.5%	n/a	n/a	15.24%

5 An Experiment with Noisy Separating Curves: Credit Risk Assessment

The task of credit-risk assessment corresponds to the differential decision-making process, for the acceptance or rejection of customers’ request for issuing a credit card, based on 15 customer parameters $x_1 \dots x_{15}$. The data sets are available upon request from [2].

Let us create a fuzzy classifier with three inputs: x_8 - mean time of occupation at a workplace, x_{11} - years of collaboration with the bank, and x_{15} - savings account balance. Fig. 3 shows the data sets for this case.

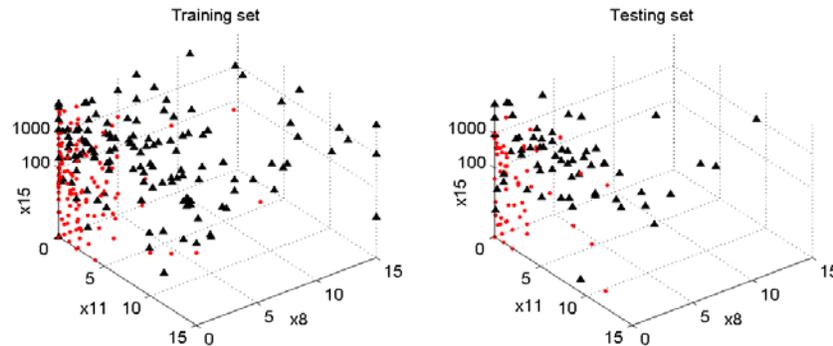


Fig. 3. Credit data sets (● - reject; ▲ - accept)

Table 3 shows the fuzzy knowledge base, created from the training data distribution with the aid of an expert. Source and optimal membership functions are shown in Fig. 4. Table 4 compares results achieved via tuning according to various learning criteria. As an alternative classification tool we used the following decision tree:

```

if      x11 > 2, then                y = "accept";
      elseif x15 ≤ 141, then          y = "reject";

```

```

else
    elseif x8>1.1,then
        y="accept";
        y="reject".
    
```

Table 3. Fuzzy knowledge base

x ₈	x ₁₁	x ₁₅	y	w			
				source classifier	classifier I	classifier II	classifier III
any	low	low	reject	1	1	0.63	1
low	low	any	reject	1	0.64	0.80	0.77
low	any	low	reject	1	1	0.50	0.63
any	high	any	accept	1	1	0.92	1
any	any	high	accept	1	1	0.61	1
high	average	any	accept	1	1	0.87	1
high	any	average	accept	1	0.38	0.05	0.12

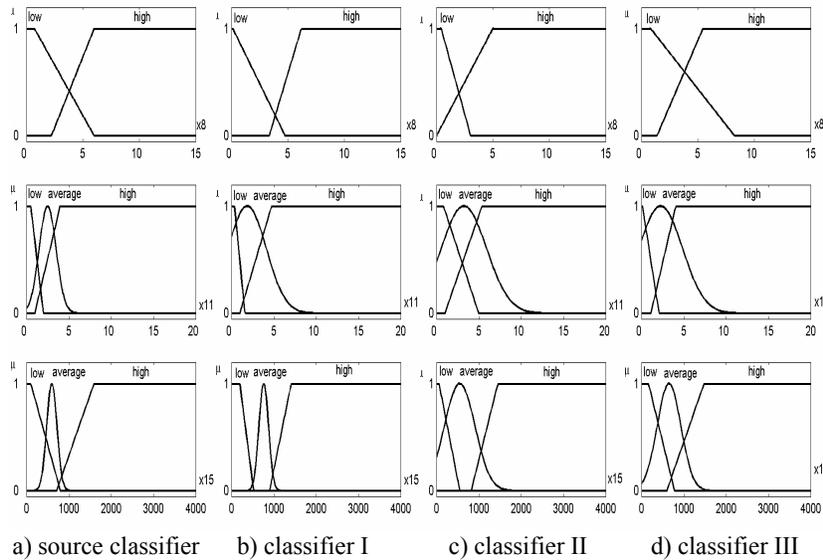


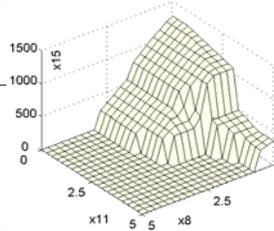
Fig. 4. Membership functions

Table 4 shows that the usage of criterion III provides the best classification accuracy (error of 21%). The difference of accuracy levels among all the classifiers is rather small, perhaps due to the noisy nature of the data.

Fig. 5 shows the minimal value of savings account balance, which provides a positive credit card issuing decision. The surface is created on the basis of fuzzy classifier III.

Table 4. Testing the classification models

Model	Crite- rion I	Crite- rion II	Crite- rion III (R=4)	Misclas- sification on test set
Source classifier	23.88%	0.46	2.34	23.5%
Classifier I	23.47%	0.43	2.16	22.0%
Classifier II	23.27%	0.35	1.49	22.5%
Classifier III	23.88%	0.42	2.00	21.0%
Decision tree	21.43%	n/a	n/a	22.5%

**Fig. 5.** Minimal value of the saving account for receiving a positive credit decision

5 Conclusion

We have studied three learning criteria for the tuning of a fuzzy classifier: 1) percentage of misclassification, 2) mean squared memberships' difference, and 3) penalized mean squared memberships' difference. The above-mentioned criteria were tested in 2 classification tasks: (a) a simple classification problem with nonlinear separating curves and (b) a real credit-risk assessment problem. The experiments suggest that the third criterion provides the best classification accuracy, especially in the case of crisp separating curves. This allows us to recommend the penalized mean squared memberships' difference, as a learning criterion of choice, for the proper tuning of fuzzy rule-based classification systems.

References

1. Rotshtein A (1998) Design and Tuning of Fuzzy Rule-Based Systems for Medical Diagnosis. In: Teodorescu NH, Kandel A, Jain LC (eds) Fuzzy and Neuro-Fuzzy Systems in Medicine. CRC-Press, Boca-Raton, pp 243-289
2. Sample Data Sets for the See5-Demo Product, Release 1.09a (2000). www.rulequest.com
3. Zimmermann H-J (1996) Fuzzy Set Theory and its Applications, 3rd ed.. Kluwer Academic Publishers, Norwell, Dordrecht
4. Zopounidis C, Pardalos PM, Baourakis G (eds., 2001) Fuzzy Sets in Management, Economics and Marketing. World Scientific, Singapore

prInvestor: Pattern Recognition based Financial Time Series Investment System

Dymitr Ruta¹

BT Exact Technologies, Adastral Park, Orion Building, 1st floor - pp12,
Martlesham Heath, Ipswich IP5 3RE, UK, dymitr.ruta@bt.com

ABSTRACT

Predictability of financial time series (FTS) is a well-known dilemma. A typical approach to this problem is to apply a regression model, built on the historical data and then further extend it into the future. If however the goal is to support or even make investment decisions, regression-based FTS predictions are inappropriate as on top of being uncertain and unnecessarily complex, they require further analysis to make an investment decision. Rather than precise FTS prediction, a busy investor may prefer a simple decision on the current day transaction: buy, wait, sell, that would maximise his return on investment. Based on such assumptions a classification model is proposed that learns the transaction patterns from optimally labelled historical data and accordingly gives the profit-driven decision for the current day transaction. Exploiting a stochastic nature of an investment cycle the model is locally reduced to a 2-class classification problem and is built on many features extracted from the share price and transaction volume time series. Simulation of the model over 20 years of NYSE:CSC share price history showed substantial improvement of the profit compared to a passive long-term investment.

KEY WORDS

Financial Time Series, Regression, Classification, Decision Support

1 Introduction

Prediction of the financial time series represents a very challenging signal processing problem. Many scientist consider FTS as very noisy, non-stationary and non-linear signal but believe that it is at least to a certain degree predictable [3], [6]. Other analysis suggest that a financial market is self-guarded against predictability as whenever it shows some signs of apparent predictability, investors immediately attempt to exploit the trading opportunities, thereby affecting the series and turning it unpredictable [2]. Stable

forecasting of FTS seems therefore unlikely to persist for longer periods of time and will self-destruct when discovered by a large number of investors. The only prediction model that could be successful and sustainable seems to be the one that exploits the supportive evidence either hidden to other investors or the evidence that is available but highly dispersed among many sources and therefore considered irrelevant, too difficult or too costly to be incorporated in the prediction model.

Irrespective of the above a number of techniques is being developed in an attempt to predict what seems unpredictable: tomorrow's share price based on historical data. Starting from simple linear Autoregressive Moving Average models (ARMA) [3] through conditional heteroscedastic models like ARCH or GARCH [3] up to the complex non-linear models [3], [4], the idea is similar: establish the regression-based description of the future samples based on the historical data series. More recently a number of machine learning techniques started to be applied to a financial forecasting and on a number of occasions showed considerable improvement compared to a traditional regression models [5], [6], [7]. Neural networks are shown to be particularly good at capturing complex non-linear characteristics of FTS [5], [6]. Support vector machines represent another powerful regression technique that immediately found applications in financial forecasting [8], [7].

While there is already extensive knowledge available in pattern recognition domain, it has been rarely used for FTS prediction. The major problem lies in the fact that classification model learns to categorise patterns into crisp classes, rather than numerical values of the series. A temporal classification models would have to provide a specific definition of classes or obtain it from the series by discretisation. Although some work has already been done in this field [10], [9], [11], [13] there is still lack of pattern recognition based models that would offer immediate investment applications surpassing in functionality and performance the traditional regression based models.

The proposed prInvestor model is a step forward towards a fully automated pattern recognition based investment system. Rather than predicting future series values it uses a classification model that learns from expandable historical evidence how to automatically categorise the future series into investment actions: buy, wait or sell. The prototype of prInvestor is tested on the 20-years of daily share price series and the results are analysed to give recommendations towards prospective fully automated platform development.

The remainder of the paper is organised as follows. Next section provides a detailed analysis of the proposed temporal classification with prInvestor, specifying the investment cycle, an algorithm for optimal labelling of training data, feature extraction process and the classification model used. Section 3 presents the results of extensive experiments with a real 20-year share price data series evaluating the performance of the prInvestor system. The concluding remarks and some suggestions for model refinement are shown in the closing Section 4.

2 Temporal classification with prInvestor

Classification represents a supervised learning technique that tries to correctly label the patterns based on multidimensional set of features [1]. The model is fully built in the training process carried out on the labelled dataset with a discriminative set of features. Based on the knowledge gained from the training process, a classifier assigns the label to a new previously unseen pattern.

Adapting the pattern recognition methodology, prInvestor would have to generate the action label: buy, sell or wait to the current day feature values based on the knowledge learnt from historical data. For that to be possible the training data have to be optimally labelled such that the investments corresponding to the sequence of labels generate the maximum return possible. Furthermore, to maximise the discrimination among classes, prInvestor should exploit the scalability of pattern recognition models and use as many relevant features as possible, far beyond just the historical share price series. All the properties mentioned above along with some mechanisms controlling model flexibility and adaptability are addressed in the presented prInvestor system.

2.1 Investment cycle

In the simplified investment cycle, considered in this paper, the investor is using all assets during each transaction which means he buys shares using all the available cash and sells all shares at once. Assuming this simplification, there are four different states an investor can fall into during the investment cycle. He enters the cycle at the state "wait with money" (WM), where the investor is assumed to possess financial assets and is holding on in the preparation for a "good buy". From there he can either wait at the same state WM or purchase the shares at the actual price of the day thereby entering the "buy" state (B). From state B investor can progress to two different states. Either he enters "wait with shares" state (WS) preparing for the good selling moment, or he sells immediately (the day after the purchase) all the shares transferring them back to money at the "Sell" (S) state. If the investor chooses to wait with shares (WS), he can stay at this state or may progress only to the state S. The sell state has to be followed by either the starting "wait with money" state WM or directly buy state (B) that both launch the new investment cycle. The complete investment cycle is summarised by the conceptual diagram and accompanied directed cyclic graph both shown in Figure 1.

Immediate consequence of the investment cycle is that the labelling sequence is highly constrained by the allowed transaction paths i.e. obeying the following sequentiality rules:

- WM may follow only with WM or B
- B may follow only with WS or S
- WS may follow only with WS or S
- S may follow only with WM or B

The above rules imply that for the current day sample the system has to pick only one out of two labels depending on the label from the previous sample. That way the 4-class problem is locally simplified to the 2-class classification problem. In a real-time scenario it means that to classify a sample, the model has to be trained on the dynamically filtered training data from only two valid classes at each sample. If the computational complexity is of concern, retraining at each sample can be replaced by 4 fixed models that can be built on the training data subsets corresponding to to 4 combinations of valid pairs of classes as stated in the above sequentiality rules.

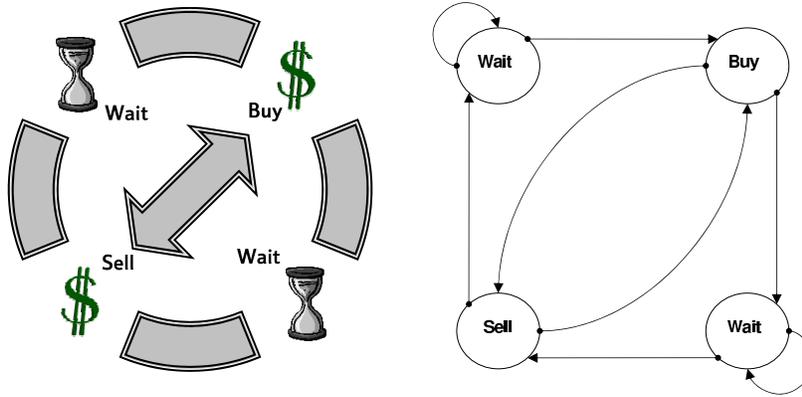


Fig. 1. Visualisation of the investment cycle applied in prInvestor. The wait state has been separated into 2 states as they have different preceding and following states.

2.2 Training set labelling

Classification, representing supervised learning model, requires labelled training data for model building and classifies incoming data using available class labels. In our share investment model, the training data initially represent unlabelled daily time series that has to be labelled using four available states WM, B, S, WS, subject to sequentiality rules. A sequence of labels generated that way determines the transaction history and allows for a calculation of the key performance measure - the profit.

In a realistic scenario each transaction is subjected to a commission charge, typically set as a fixed fraction of the transaction value. Let x_t for $t \in 0, 1, \dots, N$ be the original share price series and c stand for the transaction commission rate. Assuming that b and $s = b + p$, where $p \in 1, \dots, N - b$, denote the buying and selling indices, the relative capital after a single buy-sell investment cycle is defined by:

$$C_b^s = \frac{x_s}{x_b} \frac{1 - c}{1 + c} \quad (1)$$

Note that the same equation would hold if the relative capital was calculated in numbers of shares resulting from a sell-buy transaction. Assuming that there are T cycles in the series let $b(j)$ and $s(j)$ for $j = 1, \dots, T$ denote indices of buying and selling at j^{th} cycle such that $x_{b(j)}$ and $x_{s(j)}$ stand for buy and sell prices in j^{th} cycle. Then the relative capital after k cycles ($0 < k \leq T$) can be easily calculated by:

$$C_{b(1)}^{s(k)} = \prod_{j=1}^k C_{b(j)}^{s(j)} \quad (2)$$

The overall performance measure related to the whole series would be then the closing relative capital, which means the relative capital after T transactions:

$$C_T = C_{b(1)}^{s(T)} \quad (3)$$

Given C_T the absolute value of the closing profit can be calculated by:

$$P = C_0(C_T - 1) \quad (4)$$

where C_0 represents the absolute value of the starting capital. Finally, to be consistent with the investment terminology one can devise the return on investment performance measure, which is an annual average profit related to initial investment capital:

$$R = \frac{\overline{P}_{ANN}}{C_0} = \left[C_{b(1)}^{s(T)} \right]^{\frac{t_{ANN}}{s(T) - b(1)}} - 1 \quad (5)$$

where t_{ANN} stands for the average number of samples in 12 months.

The objective of the model is to deliver an optimal sequence of labels, which means the labels that through the corresponding investment cycles generate the highest possible profit. Such optimal labelling is only possible if at the actual sample to be labelled, the knowledge of the future samples (prices) is available. Although reasoning from the future events is forbidden in the realistic scenario, there is no harm of applying it to the training data series. The classification model would then have to try to learn the optimal labelling from historical data and use this knowledge for classification of new previously unseen samples.

An original optimal labelling algorithm is here proposed. The algorithm is scanning the sequence of prices and subsequently finds the best buy and sell indices labelling the corresponding samples with B and S labels respectively. Then all the samples in between of B and S labels are labelled with WS label and all the samples in between of S and B labels are labelled with WM label as required by the sequentiality rules. The following rules are used to determine whether the scanned sample is identified as optimal buy or sell label:

- Sample x_b is classified with the label B (optimal buy) if for all samples x_t ($b < t < s$) between sample x_b and the nearest future sample x_s $b, s \in$

- 1, ..., $N \cap b < s$, at which the shares would be sold with a profit ($C_b^s > 1$), the capital $C_t^s < C_b^s$.
- Sample x_s is classified with the label S (optimal sell) if for all samples x_t ($s < t < b$) between sample x_s and the nearest future sample x_b , $s \in 1, \dots, N \cap s < b$, at which the shares would be bought increasing the original number of shares ($C_b^s > 1$), the capital $C_b^t < C_b^s$.

2.3 Feature extraction

The data in its original form represents only a share price time series. Extensive research dedicated to time series prediction [3] proves that building the model solely on the basis of historical data of the FTS is very uncertain as it exhibits considerable proportion of a random crawl. At the same time an attractive property of classification systems is that in most of the cases they are scalable, which means they can process large number of features in a non-conflicting complementary learning process [1]. Making use of these attributes, prInvestor takes the original share price series, the average transaction volume series as well as the label series as the basis for the feature generation process. Details of the family of features used in prInvestor model are listed in Table 1. Apart from a typical moving average, differencing features, there are new features (*plf*, *ppf*) that exploit the labels of past samples in their definition. The use of labels as features might draw some controversies as it imposes that current model outputs depend on its previous outputs. This is however truly a reflection of the fact that investment actions strongly depend on the previous actions as for example if a *good buy* moment was missed the following *good sell* point could no longer be *good*. Incorporation of the dependency on previous system outputs (labels) injects also a needed element of the flexibility to the model such that after a wrong decision, it could quickly recover rather than make further losses. Another consequence of using past labels as features is the high non-linearity and indeterminism of the model and hence its limited predictability.

It is important to note that the features proposed in prototype of the prInvestor model are just a proposition of simple, cheap and available features which by no means form the optimal set of features. In fact as the series is time related, countless number of features starting from the company's P/E ratio or economy strength indicators up to type of weather outside or the investment mood could be incorporated. The problem of generation and selection of the most efficient features related to the prInvestor model is by far open and will be considered in more detail in the later version of prInvestor.

2.4 Classification model

Given the set of features, labelled training set and the performance measure, the model needs only a relevant classifier that could learn to label the samples based on optimal labels and the corresponding historical features available in

Name	Description
prc	Average daily share price
vol	Daily transactions volume
$mva_i(x)$	Moving average - mean from i last samples of the series x
$atd_i(x)$	Average difference between the current value of x and $mva_i(x)$
$dif_i(x)$	Series x differenced at i^{th} order
plf_i	The <i>past label</i> of the sample taken i steps before the current sample.
ppf	Difference between the current price and the price at B or S labels

Table 1. A list of features used in the prInvestor model.

the training set. Before the decision on the classifier is made, it is reasonable to consider the complexity and adaptability issues related to prInvestor working in a real-time mode. Depending on the choice of training data there are three different modes prInvestor can operate on. In the most complex *complete mode* the model is always trained on all available data to date. At each new day the previous day would have to be added to the training set and the model retrained on typically immense dataset covering all available historical evidence. In the *fixed mode* the model is trained only once and then used in such fixed form day by day without retraining that could incorporate the new data. Finally in the *window mode* each day the model is retrained on the same number of past samples. Undoubtedly the model is fastest in *fixed mode*, which could be a good short-term solution particularly if complexity is of concern. *Complete mode* offers the most comprehensive training, however at huge computational costs and poor adaptability capabilities. The most suitable seems to be the *window mode* in which the model is fully adaptable and its complexity can be controlled by the window width.

Given relatively large datasets and the necessity of retraining, it seems reasonable to select a rather simple easily scalable classifier that would not be severely affected by the increase in sizes of both feature and data sets. The simple quadratic discriminant analysis (QDA) classifier seems to be a good choice that accommodates the above properties while being still capable of capturing some non-linearities. Details of quadratic discriminant classifier can be found in [1]. It is important to note that given a day lag of the series, there is plenty of time for retraining even using large datasets. The model is therefore open for more complex classifiers or even the mixture of classifiers that could potentially improve its performance. The QDA classifier used in this work, due to its simplicity is particularly useful in this early prototyping stage where the experimentation comprehensiveness is the top priority rather than maximum possible performance. Moreover, simple classifiers are also preferred for shorter-lag series where the retraining might be necessary every hour or minute.

3 Experiments

Extensive experimentation work has been carried out to evaluate prInvestor. Specifically prInvestor was assessed in terms of the relative closing capital compared to the relative closing capital of the passive investment strategy of buying the shares at the beginning and selling at the end of the experimental series. Rather than a large number of various datasets only one dataset has been used but covering almost 20 years of daily average price and volume information. The dataset represents Computer Sciences Corporation average daily share price and volume series since 1964 to 1984, available at the corporation website (www.csc.com).

Initially the dataset has been optimally labelled for many different commission rates, just to investigate what is the level of maximum possible oracle-type return from the share market investment. The experimental results shown in Figure 2(b) reveal surprisingly large double-figure return for small commission charges ($< 1\%$) falling sharply to around 0.5 (50%) for high commission charges ($> 10\%$). Relating this information to the plot of the original series shown in Figure 2(a) it is clear that to generate such a huge return the algorithm has to exploit all possible price rises that exceed the profitability threshold determined by the commission rate. It also indicates that the most of the profit is generated on small but frequent price variations ($\pm 2\%$), which in real-life could be considered as noise and may not be possible to predict from the historical data. The maximum possible annual return or the profile presented in Figure 2(b) can be also considered as a measure describing potential speculative investment attractiveness of the corresponding company.

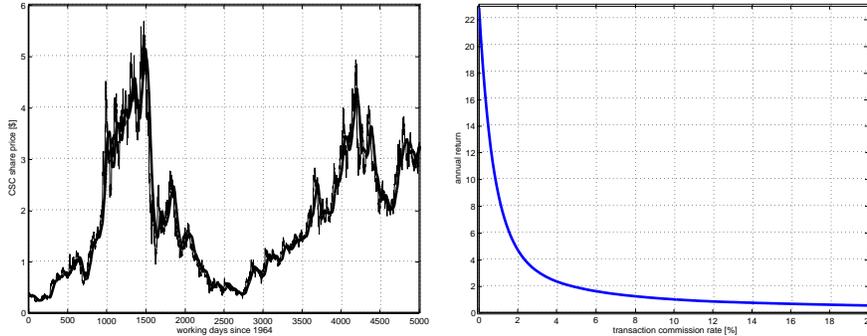
Due to many varieties of prInvestor setup and a lack of presentation space in this paper the experiments have been carried out in the moving-window mode as a balanced option featuring reasonable flexibility and adaptability mechanisms. A large pool of around 40 features have been generated from the family of features presented in Table 1. Then, for each of the window width fixed at: 6 months, 1, 2 and 3 years, prInvestor has been run 100 times using QDC classifier and a random subset of features. The performances obtained in a form of closing capital related to the capital of passive strategy *rcp* indicated that the width of 2 years (around 500 days) is optimal.

Having decided on the 2 years moving window mode prInvestor was further tuned by selection of the most relevant features. A simple evaluative search based on probability based incremental learning [12] has been used to find the best possible subset of features. As a result 21 features listed in Table 2 have been selected. The model was then trained on the first 500 days (2 years window) that have been optimally labelled. In the next step the investment simulation was launched, in which at each next day the model generated the transaction label based on the training on the preceding 500 optimally labelled samples. The resulted sequence of transaction labels represents complete output of the model. Figure 3(a) illustrates the transactions generated by the prInvestor model while Table 3 shows the performance results. Important

point is that most of investment cycles generated by prInvestor were profitable. Moreover the occasional loss cycles occur mostly during *bessa* and are relatively short in duration. The model is quite eager to invest during *hossa*, which seemed to be at least partially picked from the historical data. Numerical evaluation of the prInvestor depicted in Figure 3(b), shows more than 5 times higher closing capital than for the case of passive investor who buys at the beginning and sells at the end of 20 years period. Such remarkable results correspond on average to almost 20% annual return from investment and give a good prospect for the development of the complete investment platform with a number of carefully developed features and the data incoming automatically to the system for daily processing resulting in a final investment decision.

<i>prc</i>	<i>vol</i>	$mva_i(x)$	$atd_i(x)$	$dif_i(x)$	plf_i	ppf
<i>prc</i>	<i>vol</i>	$mva_1(vol)$	$atd_1(prc)$	$dif_1(p)$	plf_1	ppf
		$mva_5(prc)$	$atd_1(vol)$	$dif_1[mva_{20}(prc)]$		
		$mva_{20}(vol)$	$atd_5(prc)$	$dif_1[mva_{20}(vol)]$		
		$mva_{50}(prc)$	$atd_5(vol)$	$dif_2(prc)$		
		$mva_{50}(vol)$	$atd_{20}(prc)$	$dif_2(vol)$		
			$atd_{20}(vol)$	$dif_2[mva_{20}(prc)]$		
				$dif_2[mva_{20}(vol)]$		

Table 2. Features selected by the probability based incremental learning algorithm [12] from the pool of around 40 featured generated from the family of features introduced in Table 1. The selected features are grouped by their families.



(a) CSC daily share price since 1964 (b) Annual return as a function of transaction commission rate

Fig. 2. Visualisation of the optimal labelling algorithm capability.

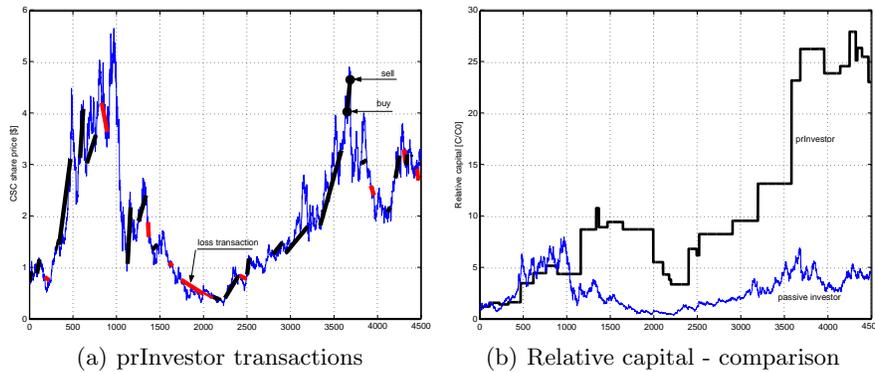


Fig. 3. prInvestor in action: the transactions returned as a result of learning to invest with profit from historical data. 3(a) Visualisation of the transactions generated by prInvestor. 3(b) Comparison of the relative capital evolutions of the prInvestor and passive investor always keeping shares.

Investor	Annual return [%]	Relative closing capital (rcc) [%]
Passive investor	8.9	463.8
prInvestor	19.1	2310.7

Table 3. Performance comparison between passive investment strategy and prInvestor obtained over investment simulation on 20 years of CSC share price history.

4 Conclusions

prInvestor is a proposition of the intelligent share investment system. Based on extensive historical evidence it uses pattern recognition mechanisms to generate a transaction decision for the current day: *buy*, *sell* or *wait*. The advantage of this model over the existing techniques is that it is capable of incorporating in a complementary, non-conflicting manner various types of evidence beyond just the historical share price data. The proposed system benefits further from optimal labelling algorithm developed to assigns the transaction labels to the training series such that the maximum possible return on investment is achieved. The model features basic flexibility and adaptability mechanisms such that it can quickly recover from bad decisions and adapt to the novel trend behaviour that may suddenly start to appear. The robustness of the prInvestor is demonstrated on just a the few simple features generated upon the share price and volume series. With such a simple setup the model hugely outperformed the passive investment strategy of buying at the beginning and selling at the end of the series, resulting on average in 20% annual return from investment. Despite tremendous average results the model is not always consistent and occasionally generates losses. Full understanding of this phe-

nomenon requires deeper analysis of the role of each individual feature in the decision process. In addition there is plenty of unknowns relating to the choice of features, classifier and the real-time classification mode. All these problems and doubts will be the subject of further investigation towards fully automated and robust investment system that could potentially find some commercial applications. For that to happen the system would have to meet very restrictive reliability requirements confirmed by the extensive testing across different company shares, markets, and time.

References

1. Duda RO, Hart PE, Stork DG (2001) Pattern classification. John Wiley & Sons. New York
2. Timmermann A, Granger CWJ (2004) Efficient market hypothesis and forecasting. *International Journal of Forecasting* 20(1): 15-27(13)
3. Tsai RS (2002) Analysis of financial time series. John Wiley & Sons
4. Casdagli M (1989) Nonlinear prediction of chaotic time series. *Physica* 35: 335-356
5. Vojinovic Z, Kecman V, Seidel R (2001) A data mining approach to financial time series modelling and forecasting. *International Journal of Intelligent Systems in Accounting, Finance & Management* 10(4): 225-239(15)
6. Giles CL, Lawrence S, Tsoi AC.(2001) Noisy time series prediction using recurrent neural networks and grammatical inference. *Machine Learning* 44(1/2): 161-183(23)
7. Tay FEH, Cao LJ (2002) Modified support vector machines in financial time series forecasting. *Neurocomputing* 48(1): 847-861(15)
8. Mukherjee S, Osuna E, Girosi F (1997). Nonlinear prediction of chaotic time series using support vector machines. In *Proceedings of the 7th IEEE Workshop on Neural Networks for Signal Proc.*, IEEE Press
9. Singh S (2000) Noisy time-series prediction using pattern recognition techniques. *Computational Intelligence* 16(1): 114-133(20)
10. Morik K, Rueping S (2002) A multistrategy approach to the classification of phases in business cycles. In Taprio E, Mannila H, Toivonen H (eds) *Lecture Notes in Artificial Intelligence (Machine Learning ECML 2002)*, pp. 307-318, Springer-Verlag
11. Wu B, Hung S-L (1999) A fuzzy identification procedure for nonlinear time series: With example on ARCH and bilinear models. *Fuzzy Sets and Systems* 108(3): 275-287(13)
12. Ruta D, Gabrys B. Application of the evolutionary algorithms for classifier selection in multiple classifier systems with majority voting. In *Proceedings of the 2nd International Workshop on Multiple Classifier Systems, LNCS 2096*, pp 399-408 Springer Verlag
13. Siekmann S, Neuneier R, Zimmermann H-G, Kruse R (1999) Neuro fuzzy systems for data analysis. In *Computing with Words in Zadeh LA, Kacprzyk J (eds) Information/Intelligent Systems Applications (Studies in Fuzziness & Soft Computing, Springer-Verlag*

On general scheme of invariant clustering procedures based on fuzzy similarity relation

Batyrshin^{1,2} I.Z., Rudas³ T. , Klimova⁴ A.

¹Mexican Petroleum Institute, Mexico

²Institute of Problems of Informatics, Academy of Sciences of Tatarstan, Russia

³TARKI, Social Research Center, Hungary

⁴Kazan State Technological University, Russia
batyr@imp.mx, batyr1@hotmail.ru

Abstract. We study the properties of general scheme of parametric invariant clustering procedures based on transformation of proximity function into fuzzy equivalence relation. The scheme gives possibility to build clustering procedures invariant to numeration of objects and to monotone transformation of proximity values between objects. The work of clustering procedure is demonstrated on example of clusterization of world stock markets.

1. Introduction

The requirements of invariance of clustering procedures under monotone transformations of similarity values between objects and invariance under permutation of objects are considered in cluster analysis as most important requirements on clustering algorithms [1, 6, 7]. The first requirement is necessary if similarity values are measured by experts or may be evaluated only in ordinal scale. This requirement is desirable also for insensitivity of clustering results to a choice of similarity or dissimilarity measure. The second requirement is necessary if we want to get as result of clustering procedure an intrinsic similarity structure of data independent on numeration of objects.

Different approaches to construction of invariant hierarchical clustering algorithms were considered in literature but only a few known algorithms fulfill to both requirements. One of such algorithms is single linkage algorithm discussed in many papers. This algorithm builds chains of clusters and for this reason reflects only specific point of view on “cluster” which is not always

acceptable. Nevertheless, this algorithm is considered as one of the most important clustering algorithm [7].

The scheme of clustering procedures discussed in this paper gives possibility to construct parametric clustering procedures invariant with respect to both invariance properties considered above.

In Section 2 we discuss the structure of clustering scheme and study its properties. In Section 3 we demonstrate the work of proposed clustering procedure on example of clusterization of world stock markets. In Conclusion we discuss obtained results.

2. General scheme of clustering algorithms

A valued relation on the set of objects X is a function $S: X \times X \rightarrow R$, where $R = [0, I]$ and I is some positive real value [5,10]. S called a similarity relation if it is reflexive: $S(x, x) = I$ and symmetric: $S(x, y) = S(y, x)$ on X . Linear ordering relation \leq on R defines operations *min* and *max* denoted as \wedge and \vee respectively: $a \dot{\cup} b = a$ if $a \leq b$ and $a \dot{\cup} b = b$ if $b \leq a$; $a \dot{\cap} b = b$ if $a \leq b$ and $a \dot{\cap} b = a$ if $b \leq a$. Similarity relation satisfying on X the property of (\vee, \wedge) -transitivity: $S(x, y) \geq \min\{S(x, z), S(z, y)\}$ will be called a valued equivalence relation.

For any value (level) a from R a valued relation S defines a usual relation $S_{[a]}$ and valued relation S_a as follows: $S_{[a]} = \{(x, y) \in X \times X \mid S(x, y) \geq a\}$; $S_a(x, y) = 1$, if $S(x, y) \geq a$ and $S_a(x, y) = 0$, if $S(x, y) < a$. Valued relation S_a may be considered as characteristic function of usual relation $S_{[a]}$. From $a < b$ it follows that $S_{[b]} \dot{\cap} S_{[a]}$ and $S_b \dot{\cap} S_a$. From reflexivity and symmetry of S it follows that for all $a \in R$ the usual relations $S_{[a]}$ also will be reflexive and symmetric. If S is $(\dot{\cup}, \dot{\cap})$ -transitive then all $S_{[a]}$ will be transitive relations [10]. As result, valued equivalence relation defines the nested set of usual equivalence relations and hence the nested partition of X on equivalence classes.

The subset A of X will be called a similarity class of similarity relation S on X if $S(x, y) > S(x, z)$ for all $x, y \in A$ and all $z \notin A$. A similarity class A may be considered as natural cluster in the set X .

Proposition 1. *A set of similarity classes of a valued equivalence relation S coincides with a set of equivalence classes of relations $S_{[a]}$, $a \in R$.*

The following scheme of clustering algorithms was proposed in [1]:

$$E = Q(S) = TC(F(S)) = \hat{F}(S),$$

where F is some “correction” of given similarity relation S such that $F(S) \hat{I} S$ and TC denotes the procedure of transitive closure \wedge of valued similarity relations [8, 10]. The procedure of transitive closure may be realized by single linkage algorithm [4]. This procedure has both types of invariance. Hence, if correction procedure F also have both types of invariance then the clustering procedure Q will satisfy to desirable properties of invariance.

The correction procedure may be based on the following idea. Two objects x and y are considered as identical in S if $S(x,y) = I$ and $S(x,z) = S(y,z)$ for all objects z from $X \setminus \{x,y\}$. More generally, two objects x and y will be called indistinguishable on the level $a \hat{I} R$ if $S(x,y) \geq a$ and for any $z \hat{I} X$ it is fulfilled $S(x,z) \geq a$ if and only if $S(y,z) \geq a$. It is clear that two objects indistinguishable on some level a will be identical in similarity relation S_a . It is clear also that all objects are indistinguishable on the minimal possible level 0 and the maximal possible level of indistinguishability of two objects x and y in relation S is equal to $a = S(x,y)$. Two objects x and y indistinguishable on the level $a = S(x,y)$ will be called indistinguishable in S .

Proposition 2. *A similarity relation S defined on X will be a valued equivalence relation if and only if all objects of X are indistinguishable in S .*

From the properties of the procedure of transitive closure TC it follows that it transforms any similarity relation S into valued equivalence relation E such that $S \subseteq E$ and E is the minimal valued equivalence relation including S . Hence transitive closure procedure produces minimal increase of values $S(x,y)$ for transformation S into valued equivalence relation E . From Proposition 2 we can conclude that this procedure transforms non-indistinguishable pairs of objects into indistinguishable. Hence we can suppose that the total value of transformation of S into E produced by TC depends on the number of non-indistinguishable pairs of elements in S and on the “degree of indistinguishability” of these elements, if we can measure this degree. Hence the correction procedure F decreasing similarity values $S(x,y)$ should produce such minimal corrections of these values which will increase the number of indistinguishable pairs of objects or increase the “degree of indistinguishability” of pairs of objects. In this case the transformation $TC(F(S))$ produced by procedure of transitive closure will be small.

For construction of suitable correction procedure it is desirable to decide: for what pairs of objects (x,y) the similarity values $S(x,y)$ should be corrected and how these values should be decreased. For these purposes the following evaluation of indistinguishability may be used. We will say that two objects x

and y are indistinguishable with respect to object z when $S(x,z) \approx S(x,y)$ if and only if $S(y,z) \approx S(x,y)$. In this case we will say that object z “support” similarity value $S(x,y)$. The more the objects in X supporting similarity value $S(x,y)$ the more the degree of indistinguishability of x and y . Our goal is to change the value of $S(x,y)$ such that the number of objects supporting similarity between x and y and hence the degree of indistinguishability will increase. As result we can say that if the objects x and y are indistinguishable only with respect to small part of objects and hence they show different behavior on large part of objects then the similarity value $S(x,y)$ does not confirmed or supported by objects of the set X and as result, the similarity value $S(x,y)$ may be corrected (decreased).

The correction procedure based on these ideas will depends on the following sets and parameters:

$$V_y(x) = \{z \in X \setminus \{x,y\} \mid S(x,z) \approx f_1(S(x,y))\},$$

$$V_x(y) = \{z \in X \setminus \{x,y\} \mid S(y,z) \approx f_1(S(x,y))\},$$

where $f_1: \mathcal{R} @ \mathcal{R}$ is some function. The sets $V_y(x)$ and $V_x(y)$ denote the sets of objects “similar” to x and to y respectively when the value $f_1(S(x,y))$ serves as criteria of this similarity. The set $V(x,y) = \{z \in X \setminus \{x,y\} \mid \max\{S(x,z), S(y,z)\} \approx f_2(S(x,y))\}$, contains the objects from X which are “similar” at least to one of the objects x and y . When $f_1 \equiv f_2$ we have $V(x,y) = V_y(x) \dot{\cup} V_x(y)$. This set will be considered as the set of “neighbors” of x and y . The “voices” of objects from $V(x,y)$ will be taken into account when decision about correction of the value $S(x,y)$ will be made. The following set $W(x,y) = \{z \in X \setminus \{x,y\} \mid \min\{S(x,z), S(y,z)\} \approx f_3(S(x,y))\}$ denotes the set of “strong” or “common” neighbors, i.e. objects which are “similar” to both objects x and y . The objects from $W(x,y)$ will “support” the value $S(x,y)$. When $f_1 \equiv f_3$ we have $W(x,y) = V_y(x) \cap V_x(y)$.

The decision about correction of the value $S(x,y)$ will depend on the respective part of objects “supporting” the similarity value $S(x,y)$ which may be considered as measure of indistinguishability of objects x and y . We can consider the following methods of calculating for each pair of objects x and y this respective part denoted as h_i :

$$h_1 = \frac{|W(x,y)|}{\min(|V_y(x)|, |V_x(y)|)}, \quad h_2 = \frac{|W(x,y)|}{\max(|V_y(x)|, |V_x(y)|)}, \quad h_3 = \frac{|W(x,y)|}{|V(x,y)|},$$

where we set $h_i = 1$ if denominator of h_i is equal to 0.

The correction procedure $F(S) = S_c$ of clustering procedure Q may be defined as follows:

$$S_c(x, y) = \begin{cases} S(x, y) & \text{if } h_i \geq p \\ F_j(x, y) & \text{else} \end{cases},$$

where $p \in [0, 1]$ is a parameter and $F_j(x, y)$ is a corrected value such that $F_j(x, y) \leq S(x, y)$. We will suppose that $F_j(x, y)$ depends on the values $S(x, z)$, $S(y, z)$ of objects z belonging to the sets of neighbors of x and y : $V_y(x)$, $V_x(y)$ and $V(x, y)$. It may be used some mean of this values. At least we require that $F_j(x, y) \leq \min_{z \in V} \{S(x, z), S(y, z)\}$, where $V = V_y(x) \cap V_x(y) \cap V(x, y)$. The following functions showed good result on many experimental data: $F_1 = \min$, $F_2 = \max$, $F_3 = \text{mean}$. When $p = 0$, from $h_i \geq 0$ it follows that $F(S(x, y)) = S_c(x, y) = S(x, y)$, i.e. for all x, y from X the values $S(x, y)$ will be not corrected and $Q(S) = TC(F(S)) = TC(S)$, i.e. clustering algorithm will coincide with single-linkage algorithm.

Proposition 3. For clustering procedures Q with identity functions $f_1 - f_3$ it is fulfilled $Q(S) = S$ if and only if S is a valued equivalence function.

One of the possible desirable properties of clustering algorithm is “to keep similarity classes”. Similarity classes satisfy to intuitively evident opinion about natural cluster. If such clusters exist in initially given similarity relation S then these clusters also should exist in clusterization obtained by clustering procedure. More generally we can require that if the subset A is a similarity class of any similarity relation S then A should be a similarity class of the valued equivalence relation E obtained by clustering procedure.

Proposition 4. Clustering procedure from proposed scheme “keeps similarity classes” if the functions f_1 and f_2 used in this procedure are identity functions.

Methods of clusterizations based on corrections procedures F_j with $j = 1, 2$ are invariant to monotone transformations of similarity values because they use only ordinal information about similarity values. Instead of method F_j with $j = 3$ it may be used its ordinal extension when instead of mean of similarity values $S(x, y), S(x, z)$ which are less than $S(x, y)$ the median of these values or “ k -th” similarity value chosen from ordered list of these values are used.

The experience of practical application of clustering procedures shows that these algorithms give good results for different data but for classification of some types of data these procedures may be modified. The use of non-identity functions $f_1 - f_3$ essentially extends the possibilities of proposed scheme. It may be used for construction clusters based on idea of “break bridges between clusters” [2,3]. This approach from some point of view is opposite to the approach “keep similarity classes” because this approach may

break the similarity clusters which may be considered as “bridges” between natural clusters.

Define dissimilarity measures as follows: $D(x,y)=I - S(x,y)$, $S(x,y)= I - D(x,y)$, where I is the maximal value of similarity or dissimilarity between object. Then the sets $V_y(x)$, $V_x(y)$, $V(x,y)$ and $W(x,y)$ will be defined as follows:

$$\begin{aligned} V_y(x) &= \{z \in \hat{I} X \setminus \{x,y\} \mid D(x,z) \leq f_1(D(x,y))\}, \\ V_x(y) &= \{z \in \hat{I} X \setminus \{x,y\} \mid D(y,z) \leq f_1(D(x,y))\}, \\ V(x,y) &= \{z \in \hat{I} X \setminus \{x,y\} \mid \min(D(x,z), D(y,z)) \leq f_2(D(x,y))\}, \\ W(x,y) &= \{z \in \hat{I} X \setminus \{x,y\} \mid \max(D(x,z), D(y,z)) \leq f_3(D(x,y))\} \end{aligned}$$

where $f_1, f_2, f_3: R @ R$ are some functions. When the dissimilarities are measured in quantitative scale we can define the functions f_i as follows: $f_i(D(x,y))=k_i \cdot D(x,y)$, for $i = 1,2,3$, where $k_i \in [0,1]$ are some parameters. If dissimilarities are measured in ordinal scale then the values of $f_k(D(x,y))$ may be chosen between the values of $D(u,v)$, $u,v \in \hat{I} X$. These sets have a natural interpretation as the sets of objects located from one of objects x, y or from both of them on the distance no greater than the distance $f_i(D(x,y))$.

The considered clustering procedures are based on the calculation of numbers of “strong neighbors” for all pairs of objects. These methods “keep the similarity values between indistinguishable objects”. But for non-indistinguishable pairs of objects even with minimal distances between them the strong neighbors may be absent and the similarity values between these objects will be not supported by another objects and hence may be corrected. Sometimes this situation is desirable but sometimes it may be better to keep these similarity values by extending the set of strong neighbors. All similarity values between these objects and the nearest objects will be corrected because they do not supported by another objects. In such cases the use of non-identity functions f_3 such that $D(x,y) \leq f_3(D(x,y))$ will give possibility to increase the number of strong neighbors for nearest objects and hence does not to correct the similarity values between them. These strong neighbors may be chosen between neighbors of x and y belonging to $V(x,y)$.

3. Example

To demonstrate the work of the clustering scheme it was applied to clusterization of World stock markets [9]. Distance measure was calculated based on information about these stock markets for 1996-2004 years. Table 1 shows the results of clusterization of these data by clustering procedure from

considered scheme. 24 stock markets were jointed in 4 clusters (A,B,C,D) and 11 constitute single objects (X) do not jointed in clusters with other objects. Cluster B contains only European stock markets, Cluster C contains only Asian stock markets, cluster D contain 2 country: Poland and Israel, cluster A is a mixture of stock markets from Asia and Europe.

Table 1. Clusterization of 35 world stock markets.

Class	Country	Stock Market	Class	Country	Stock Market
B	<u>Belgium</u>	BEL20	A	<u>Austria</u>	Trad
B	<u>Denmark</u>	KFX	A	<u>Norway</u>	Oslo All Share
B	<u>France</u>	Cac	A	<u>Czech Republic</u>	PX50
B	<u>Germany</u>	Dax	A	<u>India</u>	BSE Sens
B	<u>Ireland</u>	ISEQ	A	<u>Indonesia</u>	Jakarta Comp
B	<u>Italy</u>	Mibtel General	A	<u>Japan</u>	Second section
B	<u>Netherlands</u>	CBS All			
B	<u>Spain</u>	Madrid SE	X	<u>Finland</u>	HEX General
B	<u>Switzerland</u>	SMI	X	<u>Portugal</u>	PSI General
			X	<u>Sweden</u>	All Share
D	<u>Israel</u>	Tel Aviv-100	X	<u>Turkey</u>	IMKB Nat 100
D	<u>Poland</u>	WIG	X	<u>Hungary</u>	BUX
			X	<u>Russia</u>	RTS
C	<u>Hong Kong</u>	Hang Seng	X	<u>China</u>	CLSA China B
C	<u>Japan</u>	Nikkei 225	X	<u>Pakistan</u>	KSE-100
C	<u>Japan</u>	Topix	X	<u>South Korea</u>	KOSPI
C	<u>Malaysia</u>	KLSE Comp	X	<u>Sri Lanka</u>	CSE All Shr
C	<u>Philippines</u>	Manila Comp	X	<u>Thailand</u>	Bangkok SET
C	<u>Singapore</u>	Straits Times			
C	<u>Taiwan</u>	Weighted			

Conclusions

The properties of general scheme of invariant clustering procedures was studied in this work. It was shown that procedures from this scheme are rational, i.e. keep similarity classes etc for wide class of functions used in this scheme. But this scheme may be also extended for construction clusters based on the idea “break bridges between clusters”. The example of clusterization of worlds stock markets by clustering procedure from this scheme is considered.

Acknowledgements

This work was supported in part by Research Fellowship Program of Open Society Institute, by IMP, projects D.00006 and by RFBR grants 02-01-00092 and 03-01-96245.

References

- [1] Batyrshin I.Z., Shuster V.A. (1984). The structure of semantic spaces of verbal estimates of actions. Trans. AI, Principle Questions of Knowledge Theory, Acta et Comment. Univ. Tartuensis, Tartu, vol. 688, pp. 20-38 (in Russian).
- [2] Batyrshin I., Rudas T. (2000) Invariant clustering procedures based on corrections of similarities. In: Proceedings of East West Fuzzy Colloquium, Zittau, Germany, pp. 302-309.
- [3] Batyrshin I., Klimova A. (2002) New invariant relational clustering procedures, In: Proceedings of East West Fuzzy Colloquium, 0th Zittau Fuzzy Colloquium, Zittau, Germany, 2002, 264 – 269.
- [4] Dunn J.C. (1974). A graph-theoretic analysis of pattern classification via Tamura's fuzzy relation. IEEE Trans. SMC, vol. SMC-4, pp. 310-313.
- [5] Fodor J.C., Roubens M. (1995). Structure of transitive valued binary relations. Mathematical Social Sciences, vol. 30, pp. 71-94.
- [6] Hubert L.J. (1973). Monotone invariant clustering procedures. Psychometrika, vol. 38, pp. 47-62.
- [7] Jardine N. & Sibson R. (1971). Mathematical taxonomy. London: John Wiley & Sons.
- [8] Tamura S., Higuchi S. & Tanaka K. (1971). Pattern classification based on fuzzy relations. IEEE Trans. SMC, vol. SMC-1, pp. 61-66.
- [9] World stock market data base: <http://www.countrybriefings.com>
- [10] Zadeh L.A. (1973). Similarity relations and fuzzy orderings. Information Sciences, vol. 3, pp. 177-200.

Evolutionary procedures of visualization of multidimensional data

Angelica Klimova

Kazan State Technological University, e-mail: kangelica@front.ru

ABSTRACT. The methods of two dimensional (2D) and three dimensional (3D) visualization of multidimensional data are considered. They are both based on a minimization of the average alteration of all distances between objects. For each of procedures to visualization it was developed a genetic algorithm for minimization of correspondent distance alterations. The structures of these algorithms are discussed. The results of application of both methods are illustrated on the example of visualization of world stock markets.

1. Introduction

A visualization of data gives usually an important information for understanding a data structure. Such information about mutual spatial location of objects may be obtained by methods of data visualization representing initial data set in two dimensional (2D) or three dimensional (3D) spaces [3]. For these purposes the principal component analysis and the multidimensional scaling are often used. The first method is based on a sequential rotation of initial coordinate system such that the variance of data in the direction of each new axis becomes maximal. Data visualization is obtained as result of projection of data on hyper planes correspondent to the first two or three (principal) axis. But distances between objects in initial n -dimensional space of attributes may essentially differ from distances in 2D or 3D space of principal components. Multidimensional scaling gives possibility to decrease the space dimension with maximal preservation of initial distances between objects but a reduction of space dimension till 2 or 3 as usually causes valuable distance alteration.

The problem of data visualization may be considered as a problem of 2D (or 3D) representation of objects which minimizes alterations of initial distances between objects. Existing methods of optimization give as usually a local optimum for this problem. The search of a global optimum

may be based on the use of a genetic algorithm of optimization [2]. The structure of such algorithm is considered here.

Two genetic algorithms of data visualization are proposed in this paper. Both of algorithms minimize the average distance alteration between all objects. The proposed algorithms are illustrated on example. In Section 2 the genetic algorithm of 2D visualization of data is proposed. The genetic algorithm of 3D visualization of data is described in Section 3.

2. Genetic algorithm for 2D visualization of data

Suppose R is a distance function between objects from M in n -dimensional space of attributes, $n > 3$. We are finding 2D representation of these objects. First, the initial matrix of coordinates of objects in 2D space with axis X and Y is generated. Based on this matrix by standard optimization procedure it is determined a matrix of coordinates P of objects with the minimal error of approximation of initial matrix D by a distance matrix R calculated for matrix of coordinates P . Usually the obtained solution gives a local optimum. Then two objects a and b from M with the maximal value $R(a,b)$ are determined. The system of coordinate $\langle X, Y \rangle$ is moved and rotated so that its center is moved to the point a and the point b is located on axis X .

The rotation is realized by the following transformation. Suppose $\langle X, Y \rangle$ is the system of coordinate which we had before transformation but after moving and $\langle X', Y' \rangle$ is the “new” system of coordinate. The direction of the axis X' is coincide with the direction of the vector with the beginning in the point a and with the end in the point b . Calculate the angle between axis X and X' as following:

$$\varphi = \arcsin \left(\frac{R_b \cdot \sqrt{b_x^2 - \frac{R_b^2}{4}}}{b_x^2} \right)$$

where R_b is the distance between point b in the “old” system of coordinate after moving and the point b in the “new” system of coordinate, b_x is the coordinate of the point b in the “old” system of coordinate after moving. If the point b in the “old” system of coordinate after moving was in the I or III coordinate quarter then the angle of the rotation is equal $-\varphi$. If the point b in the “old” system of coordinate after moving was in the II or IV coordinate quarter then the angle of the rotation equals to φ .

Then define the matrix of the rotation like this:

$$M_{rot} = \begin{pmatrix} \cos \varphi & \cos\left(\frac{\pi}{2} + \varphi\right) \\ \cos\left(\frac{\pi}{2} - \varphi\right) & \cos \varphi \end{pmatrix}$$

And at last determinate the coordinates of objects in the “new ” system of coordinate $\langle X', Y' \rangle : P^* = M_{rot} P^T$.

Then it is defined an object c with the maximal absolute value y_c of coordinate Y in the matrix of coordinates obtained after rotation and displacement of coordinate system P . If $y_c < 0$ then the sign of Y coordinates of all objects is changing on opposite. The objects a, b, c will be reference elements for all future coordinate matrixes. The obtained matrix of coordinates P^* of objects from M is called a solution and an error of approximation of matrix D by matrix of distances R calculated from P^* is called an error of the solution.

Further it is randomly generated the set of m matrixes of initial coordinates of objects in 2D space and for each of them it is calculated the error of approximation. The obtained set of matrixes is called a population. The best q matrixes with minimal errors are selected from population and called elite.

For each matrix of coordinate from elite it is applied a shift and rotation of all coordinates in order to locate the reference elements a, b, c in similar positions as in the matrix P^* . The obtained solutions are used further for generation of new solutions called descendants. They are obtained as result of application of the following steps. Couples of solutions (“parents”) are randomly selected from elite and used further for construction of new solutions “descendants”) by means of recombination and mutation operations defined as follows.

Recombination operation: It is randomly selected an object x from M and all objects from one of parents with the number less than number of x receive the coordinates of the same objects from another parent.

Mutation operation: The matrix of coordinates of solution is added with the matrix of normally distributed increments multiplied by some small value Q .

A new population is obtained as a union of old elite, descendants obtained after recombination, mutated old elite and mutated descendants.

For new elements of elite defining the matrixes of coordinates of objects in 2D space they are calculated the correspondent distance matrixes and correspondent errors of approximation of initial matrix of distances in n -dimensional space.

From the obtained population new elite is selected as follows: half of the elite consists of the best solutions in population and half of the elite is ran-

domly selected from population. For new elite they are applied all the steps considered above: shift and rotation of all coordinates, selection of parents and generation of descendants etc.

The generation of new populations is repeated a given number of times or it is stopped if it will be found a solution with the error less than given value.

Proposed genetic algorithm was applied for visualization of 15 objects given in 8-dimensional space. The coordinates of objects in this space were from World Stock Market Data Base [7] and presented in Table 1. This table allows to compare the performance of 15 markets over the last 8 years with 15 key indices. Prices are quoted in local currency, the MSCI index is in US dollars. The % change is based on the opening level of the index on the first of January each year, and then on the first of each month.

Table 1.

Asia	2003	2002	2001	2000	1999	1998	1997	1996
1 China	10.25	-28.7	29.32	130.5	226.9	77.27	50.76	162.2
2 Hong Hong	34.92	10.34	-16.7	-25.9	25.15	17.28	-4.75	24.84
3 India	72.89	78.98	47.00	16.64	91.10	59.58	81.04	87.72
4 Indo- nesia	62.82	76.49	66.19	2.21	73.83	72.24	8.43	34.65
5 Japan	24.45	1.27	-22.6	-43.6	-22.9	-30	-44.9	-46.3
6 Japan	23.76	1.12	-18.7	-39.4	-3.98	-11.2	-29	-33.9
7 Japan	43.93	25.52	10.38	-18.1	81.23	78.74	17.16	3.65
8 Mala ysia	22.84	14.06	16.82	-2.26	35.45	33.56	-35.5	-20.2
9 Paki- stan	65.53	251.2	196.6	217.4	373.1	154.9	233.5	198.2
10 Phil- ippines	41.63	23.48	-3.49	-32.7	-26.7	-22.8	-54.3	-44.4
11 Sin- gapore	31.58	8.68	-8.42	-28.8	26.70	15.34	-20.4	-21.9
12 South Korea	29.19	16.87	60.66	-21.1	44.14	115.4	24.49	-8.18
13 Sri Lanka	30.34	71.07	137.4	85.58	77.86	51.29	76.18	67.65
14 Tai- wan	32.30	6.11	24.17	-30.3	-8.22	-28	-15	14.19
15 Thai- land	116.6	154.1	186.8	60.22	117	107.2	-3.86	-39.7

The optimal coordinates of objects in 2D space were found by standard procedure of approximation of nonlinear function by least squared method.

As initial values for these procedures we used 3 matrixes of 2D coordinates with minimal approximation errors selected from 100 randomly generated matrixes. After applying optimization procedure three locally optimal solutions were found. The best of them was used for comparison with the solution obtained by the genetic algorithm. The approximation error for this best locally optimal solution was equal to 16, 3453. The application of the considered above genetic algorithm gave the best solution with the approximation error 16, 3375. The genetic algorithm used 5000 generations of new populations. An initial population contained 100 matrixes. The best 14 matrixes were selected in elite.

From these elite matrixes 14 descendants were generated by means of the recombination operation. The increments of the mutation operation have been calculated as $d = 10000 * rndn / T$, where T is the number of current iteration and $rndn$ is a normally distributed standard random number.

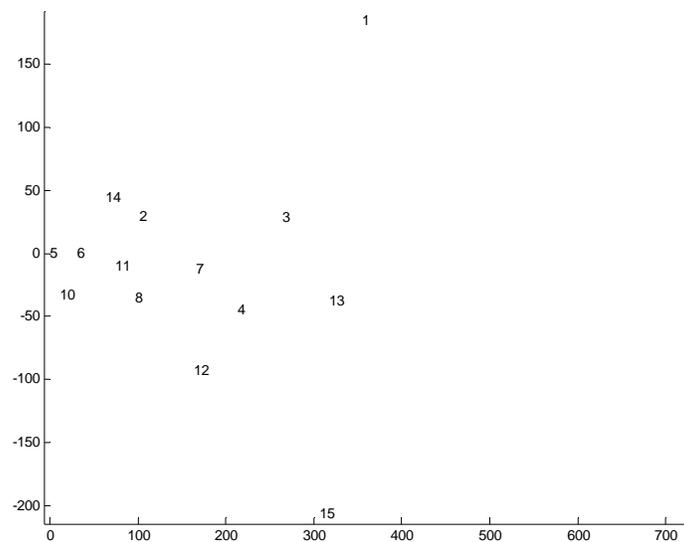


Fig. 1. 2D representation of 15 objects obtained by genetic algorithm.

3. Genetic algorithm for 3D visualization of data

Genetic algorithm for 3D visualization of multidimensional data is almost the same and the main difference from the genetic algorithm for 2D visualization is the rotation.

Suppose R is a distance function between objects from M in n -dimensional space of attributes, $n > 3$. We are finding 3D representation of these objects. First, the initial matrix of coordinates of objects in 3D space with axis X and Y is generated. Based on this matrix by standard optimization procedure it is determined a matrix of coordinates P of objects with the minimal error of approximation of initial matrix D by a distance matrix R calculated for matrix of coordinates P . Usually the obtained solution gives a local optimum. Then two objects a and b from M with the maximal value $R(a,b)$ are determined. The system of coordinate $\langle X, Y, Z \rangle$ is moved and rotated such that its center is moved to the point a and the point b is located on axis X .

The rotation is realized by the following transformation. Suppose $\langle X, Y \rangle$ is the system of coordinate which we had before transformation but after moving and $\langle X', Y' \rangle$ is the "new" coordinate system. The direction of the axis X' is coincided with the direction of the vector with the beginning in the point a and with the end in the point b . Assume (b_x, b_y, b_z) are coordinates of the point b before rotation and d is the distance between points a and b . In the system of coordinate $\langle X, Y \rangle$ defined coordinates of basis vectors of coordinate system $\langle X', Y' \rangle$ (the direction of the vector \vec{e}_1 is coincide with the direction of the vector with the beginning in the point a and with the end in the point b):

$$\vec{e}_1 = \left(\frac{b_x}{d}, \frac{b_y}{d}, \frac{b_z}{d} \right), \quad \vec{e}_2 = \left(\frac{b_y}{\sqrt{b_x^2 + b_y^2}}, \frac{-b_x}{\sqrt{b_x^2 + b_y^2}}, 0 \right),$$

$$\vec{e}_3 = \left(\frac{b_x \cdot b_z}{d \cdot \sqrt{b_x^2 + b_y^2}}, \frac{b_y \cdot b_z}{d \cdot \sqrt{b_x^2 + b_y^2}}, \frac{-\sqrt{b_x^2 + b_y^2}}{d} \right).$$

Coordinates of basis vectors of coordinate system $\langle X, Y \rangle$:

$$\vec{i} = (1, 0, 0), \quad \vec{j} = (0, 1, 0), \quad \vec{k} = (0, 0, 1).$$

The matrix of the rotation consists from the cosines of angles between vectors $\vec{e}_i (i=1,2,3)$ and $\vec{i}, \vec{j}, \vec{k}$:

$$M_{rot} = \begin{pmatrix} \frac{b_x}{d} & \frac{b_y}{\sqrt{b_x^2 + b_y^2}} & \frac{b_x \cdot b_z}{d \cdot \sqrt{b_x^2 + b_y^2}} \\ \frac{b_y}{d} & \frac{-b_x}{\sqrt{b_x^2 + b_y^2}} & \frac{b_y \cdot b_z}{d \cdot \sqrt{b_x^2 + b_y^2}} \\ \frac{b_z}{d} & 0 & \frac{-\sqrt{b_x^2 + b_y^2}}{d} \end{pmatrix}$$

Coordinates of objects are defined as following: $P^* = (M_{rot} P^T)^T$.

Then it is defined an object c with the maximal absolute value y_c of coordinate Y in the matrix of coordinates obtained after rotation and displacement of coordinate system P . If $y_c < 0$ then the sign of Y coordinates of all objects is changing on opposite. And then it is defined an object t with the maximal absolute value z_t of coordinate Z in the matrix of coordinates obtained after rotation and displacement of coordinate system P . If $z_t < 0$ then the sign of Z coordinates of all objects is changing on opposite. The objects a, b, c, t will be reference elements for all future coordinate matrixes. The obtained matrix of coordinates P^* of objects from M is called a solution and an error of approximation of matrix D by matrix of distances R calculated from P^* is called an error of the solution.

Further it is randomly generated the set of m matrixes of initial coordinates of objects in 3D space and for each of them it is calculated the error of approximation. The obtained set of matrixes is called a population. The best q matrixes with minimal errors are selected from population and called elite.

For each matrix of coordinates from elite it is applied a shift and rotation of all coordinates in order to locate the reference elements a, b, c, t in similar positions as in the matrix P^* . The obtained solutions are used further for generation of new solutions called descendants. They are obtained as result of application of the following steps. Couples of solutions (“parents”) are randomly selected from elite and used further for construction of new solutions “descendants”) by means of recombination and mutation operations defined as follows.

Recombination operation: it is randomly selected an object x from M and all objects from one of parents with the number less than number of x receive the coordinates of the same objects from another parent.

Mutation operation: the matrix of coordinates of solution is added with the matrix of normally distributed increments multiplied by some small value Q .

A new population is obtained as a union of old elite, descendants obtained after recombination, mutated old elite and mutated descendants.

For new elements of elite defining the matrixes of coordinates of objects in 3D space they are calculated the correspondent distance matrixes and correspondent errors of approximation of initial matrix of distances in n -dimensional space.

From the obtained population it is selected new elite as follows: half of elite consists of the best solutions in population and half of elite is randomly selected from population. For new elite they are applied all the steps considered above: shift and rotation of all coordinates, selection of parents and generation of descendants etc.

The generation of new populations is repeated a given number of times or it is stopped if it will be found a solution with the error less than given value.

Proposed genetic algorithm was applied for visualization of 15 objects given in 8-dimensional space. The coordinates of objects in this space were obtained from World Stock Market Data Base [7] and shown in Table 2. This table allows to compare the performance of 16 markets over the last 8 years. Prices are quoted in local currency, the MSCI index is in US dollars. The % change is based on the opening level of the index on the first of January each year, and then on the first of each month.

Table 2.

Europe	2003	2002	2001	2000	1999	1998	1997	1996
1 Austria	34.36	35.50	43.96	29.00	37.87	19.32	36.15	60.99
2 Belgium	10.82	-19.3	-25.8	-32.8	-36.2	-7.20	19.94	43.89
3 Denmark	22.49	-10.3	-22.5	-22.2	-4.44	11.40	16.05	79.50
4 Finland	4.44	-31.5	-53.7	-58.6	8.39	82.66	142.9	252.3
5 France	16.12	-23	-39.9	-40.3	-9.76	18.64	57.64	90.06
6 Germany	37.08	-23.2	-38.4	-43	-20.7	-6.70	39.19	75.93
7 Ireland	23.17	-13.3	-14	-1.93	-1.50	21.39	80.57	120.9
8 Israel	60.74	19.61	8.05	11.92	78.20	83.66	149.5	156.6
9 Italy	13.94	-12.8	-34.3	-31.3	-15.9	18.54	88.46	110.8
10 Netherlands	5.13	-31.4	-45.8	-47.9	-33.9	-21.5	13.24	51.10
11 Norway	47.96	10.64	-3.86	-1.62	46.61	5.42	36.14	75.60
12 Portugal	17.40	-6.85	-24.6	-30.8	-22	881.4	62.56	115.5
13 Spain	27.44	-1.99	-8.26	-19.9	-6.89	27.73	85.94	152.4
14 Sweden	29.82	-16.1	-95.9	-96.5	-94.1	-93.5	-91.8	-88.8
15 Switzerland	18.51	-14.5	-32.5	-27.5	-23.4	-12.4	39.21	66.41

The optimal coordinates of objects in 3D space were found by standard procedure of approximation of nonlinear function by least squared method. As initial values for these procedures we used 3 matrixes of 3D coordi-

nates with minimal approximation errors selected from 100 randomly generated matrixes. After applying optimization procedure three locally optimal solutions were found. The best of them was used for comparison with the solution obtained by the genetic algorithm. The approximation error for this best locally optimal solution was equal to 4, 0815. The application of the considered above genetic algorithm gave the best solution with the approximation error 4, 0313. The genetic algorithm used 7000 generations of new populations. An initial population contained 100 matrixes. The best 14 matrixes were selected in elite.

From these elite matrixes 14 descendants were generated by means of the recombination operation. The increments of the mutation operation have been calculated as $d = 10000 * rndn / T$, where T is the number of current iteration and $rndn$ is a normally distributed standard random number.

14

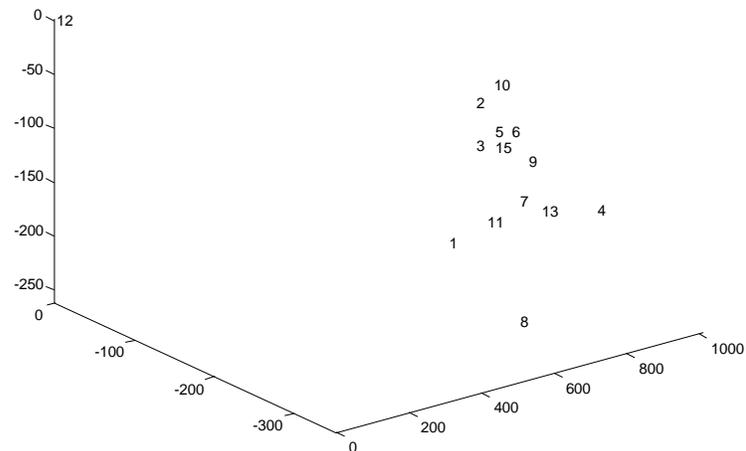


Fig. 2. 3D representation of 15 objects obtained by genetic algorithm.

4. Conclusions

Genetic algorithms for 2D visualization and 3D visualization of data were proposed. These algorithms minimize average alterations of distances

initially given in n -dimensional attribute space. Application of this algorithm gives possibility to obtain usually the better results than by standard optimization procedure. A visualization of multivariate data as usually results in distance alteration. For this reason additional methods of data analysis such as cluster analysis may be used independently from data visualization. Application of these methods to the same data gives possibility to supplement the classification of data with visualization of data [6]. In the work [4] it was proposed the method of multistage visualization of data based on hierarchical clustering of data. In this approach the results of cluster analysis are supplemented by visual information about distances between objects and mutual location of clusters. On the other hand a representation of clusters of similar objects in 2D space gives possibility to “compensate” the distance alterations in 2D representation of data. The method of 3D visualization of data proposed here gives possibility to extend on 3D case the method of hierarchical 2D visualization of data developed in [4]. It is also supposed in the future work to extend the evolutionary method of 3D visualization considered here on non-metric data visualization preserving the ordering of distances between objects [3].

Acknowledgment

The work is supported by RFFI grant 03-01-96245.

References

- [1] R. Duda and P. Hart. *Pattern Classification and Scene Analysis*, New York: Jon Wiley and Sons, 1973.
- [2] E. Goldberg. *Genetic algorithm in search Optimization and machine learning*, Addison-Wesley, MA, 1989.
- [3] A. Ju. Terekhina. *Data analysis by multidimensional scaling methods*, Moscow: Nauka, 1986 (in Russian).
- [4] Batyrshin I., Klimova A., Sheremetov L. On two dimensional visualization of hierarchical clustering. In: *IEEE Int. Conf. Comp. Cybern, ICCS 2003, Siofok, Hungary, 2003*, 181–184.
- [5] V. Vasiliev, B. Iliasov. *Intellectual systems of control with the use the genetic algorithms*, Ufa, 1999 (in Russian).
- [6] J. B. Kruskal. The relationship between multidimensional scaling and clustering, in: *Classification and clustering* (Ed. John Van Ryzin) New York: Acad. Press, 1977, 17-44 (in Russian).
- [7] World Stock Market Data Base: www.countrybriefings.com

**FUZZY GAMES, DECISIONS AND
EXPERT SYSTEMS**

Plenary Report

Vague Utilities in Cooperative Market

Mareš Milan

Fuzzy Games and Decision Making

The Shapley Value for Games on Lattices (L-Fuzzy Games)

Grabisch Michel

Optimal Strategies of Some Symmetric Matrix Games

De Schuymer Bart, De Meyer Hans and De Baets Bernard

On strict Monotonic t -norms and t -conorms on Ordinal Scales

Batyrshin I.Z. and Batyrshin I.I.

Belief Functions and the Imprecise Dirichlet Model

Utkin Lev V.

Fuzzy Decision Making Using the Imprecise Dirichlet Model

Utkin Lev V. and Augustin Thomas

Fuzzy Expert Systems

A Fuzzy Expert System for Predicting the Effect of Socio-Economic Status on
Noise-Induced Annoyance

Zaheeruddin and Jain V. K.

The Knowledge Representation Model in Fuzzy Expert Systems Provided by
Computing with Words and PRUF Language

Glova V.I., Anikin I.V., Katasev A.S. and Pheoktistov O.N

Decision Trees and Optimization

Construction of Incremental Fuzzy Decision Tree

Borisov Arkady and Bikesheva Gulnara

Fuzzy Binary Tree Model for European-Style Vanilla Options

Muzzioli Silvia and Reynaerts Huguette

Tree-Structured Smooth Transition Regression

Correa da Rosa Joel, Veiga Alvaro, Medeiros Marcelo

Discrete Dynamic Programming With Outcomes In Fuzzy Ordered Structures

Trzaskalik Tadeusz and Sitarz Sebastian

Vague Utilities in Cooperative Market

Milan MAREŠ

*Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
P.O.Box 18, 182 08 Praha 8, Czech Republic
e-mail: mares@utia.cas.cz*

Abstract: The Walras model of free exchange market and its relation to the cooperative games theory is well known from the literature. Here, we formulate its modification in which some of its components, namely the utilities and prices, are vague, represented by fuzzy quantities. This vagueness appears natural if we consider the subjective character of utilities, and the variability of prices in free market. We discuss the relation of the presented model to the classical deterministic one. Some of its elementary properties are derived, too.

1 Introduction

There exist several economic equilibrium models focused on particular phenomena. One of them is the free exchange market model including no production or consumption and analyzing the exchange of goods among agents aiming to satisfy their individual demands under natural market constraints. These constraints follow from the existence of limited amounts of particular goods in the market, and from the necessity to respect some price relations. The optimal satisfaction of such agents' needs is connected with the term of Walras equilibrium which is dealt in numerous works.

Here, we are interested in those which point at the close relation between such market model and some kind of coalitional games. It is presented, e.g., in [13], [11] and [12], and thoroughly analyzed in [4] where its asymptotic properties are considered, too. Papers [5], [6] and [7] have contributed to this model by including the possibility to form „strongly motivated“ coalitions of player whose members have a possibility to proceed cooperatively, to enter into the market with common property and to distribute the achieved wealth due to the needs of their participants without the limitations given by the existence of prices.

Both, the well known basic model as well as its modification mentioned above, are based on the assumption that all their elements are

exactly known. Nevertheless, in most of real markets this is not the truth. At least the subjective utilities of the goods may depend on the vagueness of human preferences, and also the prices vary in some interval with different degrees of possibility.

In the formulation of the fuzzy cooperative market model, we can be inspired by the concepts developed in the fuzzy cooperative games theory. Let us limit our attention on the fuzzification of the cooperative games with transferable utility (even if the fuzzy cooperation in games with non-transferable utility is studied in the literature, as well). First, the models of fuzzy cooperative games can be based either on the fuzzification of the coalitions (see, e.g., [1], [2], [3]) or on the fuzzification of the expected pay-offs (see, e.g., [9], [10]) with wide application of fuzzy quantities theory [8]. The vagueness of utilities and prices implies rather the vagueness of the expected pay-offs (or profits) and, consequently, we focus our attention of the model of markets where individual utilities are fuzzy functions (i.e., function with values in fuzzy quantities) and prices are fuzzy quantities. The game theoretical counterparts of such markets are the games described in [9].

In the following section we briefly recall the deterministic model of free exchange market, introduce the notations and concepts for the handling fuzzy quantities, formulate the fuzzified alternative of the market and its equilibrium, and show some results describing its basic properties.

2 Preliminary Concepts

In this section, we briefly recollect some notions important for the formulation of the suggested model. Namely, we introduce the original deterministic free exchange market, and the elements of the fuzzy quantities theory.

In the whole paper, we denote by I the (non-empty and finite) set of *agents*; to simplify the notations, we put $I = \{1, 2, \dots, n\}$. Every set of agents $K \subset I$ is called a *coalition*. By R we denote the set of real numbers, R_+ is the set of non-negative real numbers, and R^k , R_+^k are set of k -dimensional vectors for positive integer k .

2.1 Coalitional Game

If v is a mapping connecting each coalition K with a real number $v(K)$, such that $v(\emptyset)=0$ then the pair (I, v) is a *coalitional game*. We say that (I, v) is *superadditive* iff for any pair of disjoint coalitions K, L

$$v(K \cup L) \geq v(K) + v(L).$$

Every real-valued vector $\mathbf{x} = (x_i)_{i \in I}$ is called an imputation. A set of imputations

$$C = \left\{ \mathbf{x} = (x_i)_{i \in I} : \sum_{i \in I} x_i \leq v(I), \forall_{K \subset I}, \sum_{i \in K} x_i \geq v(K) \right\}$$

is called a *core* of the game in question. If $\mathbf{M} \subset 2^I$ is a class of coalitions such that for every $i \in I$ there exists $K \in \mathbf{M}$ such that $i \in K$, then the set

$$C_{\mathbf{M}} = \left\{ \mathbf{x} \in R^n : \sum_{i \in I} x_i \leq v(I) \text{ and for any } K \in \mathbf{M}, \sum_{i \in K} x_i \geq v(K) \right\}.$$

2.2 Cooperative Market

Let us consider n agents who share m types of *goods*. The symbol $x_j^i, i \in I, j = 1, \dots, m$ denotes the amount of the j -th good owned by agent i . We suppose that $x_j^i \geq 0$ for all i, j , denote $\mathbf{X} = (x_j^i)_{i \in I}$, $x^i = (x_j^i)_{j=1, \dots, m}, i \in I$. The structure \mathbf{X} of vectors $x^i \in R^m$ is called an *allocation of goods*. There exists exactly one allocation of goods, we denote it $\mathbf{a} = (a^i)_{i \in I}$, $a^i \in R_+^m$, representing the *initial allocation*. By the cartesian product $\mathbf{X} = \prod_{i \in I} X^i$, where

$X^i = \{x^i \in R_+^m : \forall_{j=1, \dots, m}, \sum_{i \in I} x_j^i \leq \sum_{i \in I} a_j^i\}, i \in I$, we denote the set of all theoretically possible allocations.

Let us suppose, further, that each agent $i \in I$ evaluates the realized allocation by means of his individual preferences represented by *utility function* $u_i : \mathbf{X} \rightarrow R$. The quadruple

$$\mathbf{m} = (I, \mathbf{X}, (u_i)_{i \in I}, (a^i)_{i \in I})$$

whose components are specified above, is called a *market*.

Finally, the goods are connected with some *prices*. The price vector is denoted $p = (p_j)_{j=1, \dots, m}$, where $p_j > 0$ for all $j=1, \dots, m$, and P is the set of all admissible price vectors. We suppose that $p \in P$ are row-vectors, meanwhile $x^i \in X^i$ are columns, so that the scalar products $p \cdot x^i$ have sense. Each pair (\mathbf{x}, p) formed by allocation and prices $\mathbf{x} \in \mathbf{X}, p \in P$, is called *state of market*. For every agent $i \in I$ we define the set

$$B_p^i = \left\{ \mathbf{x} \in \mathbf{X} : p \cdot x^i \leq p \cdot a^i \right\},$$

called his *budget set*.

A state of market (\bar{x}, \bar{p}) is called the (Walras) *equilibrium* iff

$$\bar{x} \in \mathbf{B}_{\bar{p}}^i \text{ for all } i \in I,$$

$$\text{if } y \in \mathbf{B}_{\bar{p}}^i \text{ then } u_i(\bar{x}) \geq u_i(y).$$

The above equilibrium concept can be extended into a cooperative one. It is done, e.g., in [11] or [12]. Here, we recollect its modification [5], [6]. Let us consider a non-empty coalition $K \subset I$, and denote for a price-vector p and allocation $\mathbf{x} \in \mathbf{X}$ the set

$$\mathbf{B}_p^K = \{\mathbf{x} \in \mathbf{X} : \sum_{i \in K} p \cdot y^i \leq \sum_{i \in K} p \cdot a^i\},$$

which we call the budget set of coalition K and the function $u_K : \mathbf{X} \rightarrow R$ such that

$$u_K(\mathbf{x}) = \sum_{i \in K} u_i(\mathbf{x}), \mathbf{x} \in \mathbf{X}.$$

If $\mathbf{M} \subset 2^I$ is a set of coalitions such that for each $i \in K$, there exists a coalition $K \in \mathbf{M}$ such that $i \in K$, then we say that a state of market (\bar{x}, \bar{p}) is an \mathbf{M} -*equilibrium* iff

$$\bar{x} \in \mathbf{B}_{\bar{p}}^K \text{ for all } K \in \mathbf{M},$$

$$u_K(\bar{x}) = \max\{u_K(\mathbf{x}) : \mathbf{x} \in \mathbf{X}, \sum_{i \in K} x_j^i = \sum_{i \in K} a_j^i \text{ for all } j = 1, \dots, m\}.$$

The cooperative definition of the market introduced above, suggested in [5] and [6], and modified in [7], follows from the latent assumption that there exist groups of agents being strongly motivated for cooperation reaching beyond the standard market principles (e.g., families, politicaly or nationally motivated allies, etc.). Their members aim to maximize their individual profit. But they are able to do so via maximizing the common profit of the coalition and dividing it among its members due to some „just“ rules of the cooperative game theory.

If $K \subset I$ then we define a value $v(K) \in R$ such that

$$v_m(K) = \max\{u_K(\mathbf{x}) : \mathbf{x} \in \mathbf{X}, \sum_{i \in K} x_j^i = \sum_{i \in K} a_j^i \text{ for all } j = 1, \dots, m\}.$$

The pair (I, v_m) , where $v : 2^I \rightarrow R$ and $v(\emptyset) = 0$, is said the *market game* of the market \mathbf{m} . The relation between the market equilibria and the core of the market game is treated by the main results of the referred market theory.

2.3 Fuzzy Quantity

The main fuzzy set theoretical tool used below is the concept of *fuzzy quantity* [8]. We use this term for any fuzzy subset a of the real line, with *membership function* $\mu_a : R \rightarrow [0,1]$. We suppose that $\mu_a(x) = 1$ for

some $x \in R$, and the support set of μ_a is limited. The fuzzy quantities can be algebraically processed by means of so called extension principle. Here, we need the operations of addition and product with crisp number. If a, b are fuzzy quantities with μ_a, μ_b , and $r \in R$ is crisp real number, then

$$a \oplus b \text{ and } r.a \text{ are fuzzy quantities, too, and}$$

$$\mu_{a \oplus b}(x) = \sup_{y \in R} [\min(\mu_a(x), \mu_b(x - y))]$$

$$\mu_{r.a}(x) = \mu_a(x/r) \text{ if } r \neq 0, \quad \mu_{0.a}(0) = 1, \mu_{0.a}(x) = 0 \text{ if } x \neq 0.$$

There exist various definitions of the ordering relation between fuzzy quantities. Here, we use the one due to which such ordering \geq is a fuzzy relation, too. It is defined by membership function $\nu(.,.)$ connecting each ordered pair of fuzzy quantities a, b with a real number

$$\nu(a, b) = \sup [\min(\mu_a(x), \mu_b(y)) : x, y \in R, x \geq y].$$

The value $\mu(a, b)$ evaluates the possibility of $a \geq b$ in the fuzzy ordering \geq .

3 Fuzzy Market

The original market model $\mathbf{m} = (I, X, (u_i)_{i \in I}, (a^i)_{i \in I})$ and its equilibrium are constructed as deterministic concepts. In the everyday reality, there is lot of vagueness hidden in the market situation. We focus our attention on the cases in which the numerical evaluations of allocations, i.e., the utility functions and prices, are subjective or destabilized.

In the next sections, we suppose that for each allocation $\mathbf{x} \in X$ and each agent $i \in I$, the utility value is a fuzzy quantity, let us denote it $u_i^F(\mathbf{x})$. If $u_i(\mathbf{x})$ is the deterministic utility value in the above market \mathbf{m} , if $\mu_{i,x}(\cdot) : R \rightarrow [0,1]$ is the membership function of $u_i^F(\mathbf{x})$, and if $\mu_{i,x}(u_i(\mathbf{x})) = 1$ for all $i \in I, \mathbf{x} \in X$, then we say that u_i^F is a *fuzzy extension of the utility function* u_i .

Similarly, we suppose that the prices $p_j, j=1, \dots, m$ are vague. Each of them is represented by fuzzy quantity p_j^F with membership function $\pi_j : R \rightarrow [0,1]$ such that $\pi_j(\xi) = 0$ for $\xi \leq 0$, and, moreover, the vector $(\pi_j)_{j=1, \dots, m}$ of membership functions is such that $\pi_1(\xi_1) \cdot \pi_2(\xi_2) \cdot \dots \cdot \pi_m(\xi_m) = 0$ if $(\xi_1, \xi_2, \dots, \xi_m) \notin P$. If $p \in P$ is a crisp

price vector used in the above market \mathbf{m} , if $\pi_j(p_j) = 1, j = 1, \dots, m$, then we say that $p^F = (p_j^F)_{j=1, \dots, m}$ is *fuzzy extension of prices* p .

The quadruple $\mathbf{m}^F = (I, \mathbf{X}, (u_i^F)_{i \in I}, (a^i)_{i \in I})$, is called *fuzzy market*. If the fuzzy utilities u_i^F are fuzzy extensions of $u_i, i \in I$, then \mathbf{m}^F is called *fuzzy extension of the market* \mathbf{m} .

Remark 1. If $P = R_+^m$ then for every vector of fuzzy prices $p^F = (p_j^F)_{j=1, \dots, m}$ there exists $p \in P$ such that p^F is fuzzy extension of p .

Remark 2. If $P = \{p \in R_+^m : p_1 + p_2 + \dots + p_m = 1\}$ and if the vector p^F is such that

$$\pi_1(\xi_1) \dots \pi_m(\xi_m) = 0 \text{ if } \xi_1 + \dots + \xi_m \neq 0$$

then there exists $p \in P$ such that p^F is fuzzy extension of p .

Let us consider a non-empty coalition K . Let $u_K^F : \mathbf{X} \rightarrow R$ be a fuzzy function such that for any $\mathbf{x} \in \mathbf{X}$, $u_K^F(\mathbf{x})$ is a sum of $u_i^F(\mathbf{x}), i \in K$, in the sense of Subsection 2.3. Then $u_K^F(\mathbf{x})$ is a fuzzy quantity with membership function $\mu_{K, \mathbf{x}} : R \rightarrow [0, 1]$.

Remark 3. If $K \subset I, K \neq \emptyset$, and if for all $i \in K, u_i^F$ is fuzzy extension of u_i then

$$\mu_{K, \mathbf{x}}(\sum_{i \in K} u_i(\mathbf{x})) = 1,$$

as follows from the properties of addition of fuzzy quantities (see [8]).

The vagueness of the utilities and prices brings its consequences in the vagueness of other components derived from them. It regards, especially, the budget sets and equilibria.

In the fuzzy market \mathbf{m}^F the product $p.x^i, x^i \in X^i, p \in P$, means a product of crisp vector x^i and vector of fuzzy quantities $p_j, j=1, \dots, m$, i.e.

$$p.x^i = p_1.x_1^i \oplus \dots \oplus p_m.x_m^i,$$

where each component of the sum is a product of fuzzy quantity p_j and crisp number x_j^i and the entire sum is a sum of fuzzy quantities. This is true even in the case of initial allocations $a^i, i \in I$. It means that the budget sets in the fuzzy market are based on the ordering relation over fuzzy quantities. We have accepted, in Subsection 2.3, the principle that

such relation is fuzzy, as well. Its membership function was denoted by ν . Hence, we define the fuzzy budget set ${}^F B_{p^F}^i$ for $i \in I$ and fuzzy prices p^F as fuzzy subset of X with membership function $\beta_{i,p^F} : X \rightarrow [0,1]$, where

$$\beta_{i,p^F}(\mathbf{x}) = \nu(p^F \cdot a^i, p^F \cdot x^i), \quad \mathbf{x} = (x^i)_{i \in I}.$$

Lemma 1. If the fuzzy market \mathbf{m}^F is fuzzy extension of market \mathbf{m} , if p^F is fuzzy extension of prices p , and if B_p^i is a budget set of \mathbf{m} for $i \in I$, then $\beta_{i,p^F}(\mathbf{x}) = 1$ for all $\mathbf{x} \in B_p^i$.

Proof. The statement follows from the previous definitions. If p^F is fuzzy extension of market \mathbf{m} , if p^F is fuzzy extension of p then $\pi_j(p_j) = 1$ for all $j=1,2,\dots,m$. Let us denote by μ_x the membership function of the fuzzy quantity $p_1^F \cdot x_1^i \oplus \dots \oplus p_m^F \cdot x_m^i$. Then $\mu_x(p \cdot x^i) = 1$. Similarly, let μ_a denote the membership function of $p_1^F \cdot a_1^i \oplus \dots \oplus p_m^F \cdot a_m^i$ for the initial allocation \mathbf{a} and $i \in I$. Then $\mu_a(p \cdot a^i) = 1$. Due to the definition of the membership ν , if $\mathbf{x} \in B_p^i$ then $p x^i \leq p a^i$ for all $i \in I$, and, consequently, $\nu(\mathbf{x}) = 1$. \square

Having introduced the fuzzy budget sets, we may formulate the concept of the *state of fuzzy market* \mathbf{m}^F as a pair (\mathbf{x}, p^F) , where $\mathbf{x} \in X$ and p^F is the vector of fuzzy prices, such that \mathbf{x} belongs to the fuzzy budget set and its fuzzy utility is maximal. This heuristic formulation has to be specified. The property of being equilibrium of a fuzzy market is a fuzzy property and the equilibria form a fuzzy subset of the class of all states of the fuzzy market. Let us denote the membership function of this fuzzy set by ρ , and for any state (\mathbf{x}, p^F) formed by (crisp) allocation and fuzzy prices the value $\rho(\mathbf{x}, p^F)$ denotes the possibility that (\mathbf{x}, p^F) may be an equilibrium of \mathbf{m}^F . By means of generalization of the deterministic equilibrium concept and using the above definitions we may define the values of ρ by the following procedure:

- Denote $\beta_{p^F}^i(\mathbf{x})$ the possibility that \mathbf{x} belongs to the fuzzy budget set, $i \in I$.
- Denote $\gamma_i(\mathbf{x}) = \min \left[\beta_{p^F}^i(\mathbf{y}), \nu(u_i^F(\mathbf{x}), u_i^F(\mathbf{y}))^{y \in X} : \mathbf{y} \in X \right]$, for all $i \in I$, the possibility that the fuzzy utility of allocation \mathbf{x} for agent i

is greater than the utility of any other allocation possibility belonging to the same fuzzy budget set.

- Then we put

$$\rho(\mathbf{x}, p^F) = \min(\beta_{p^F}^i(\mathbf{x}), \gamma_i(\mathbf{x}) : i \in I).$$

The equilibrium of fuzzy market \mathbf{m}^F reflects the basic structures of its fuzzy components. Namely, it extends the equilibrium of the deterministic market.

Theorem 1. Let \mathbf{m}^F be fuzzy extension of market \mathbf{m} , and let p^F be fuzzy extension of prices $\bar{p} \in P$. Let the state of market $(\bar{\mathbf{x}}, \bar{p})$ be an equilibrium of \mathbf{m} . Then $\rho(\bar{\mathbf{x}}, p^F) = 1$.

Proof. Using Lemma 1, we know that for all $i \in I$, $\beta_{p^F}^i(\bar{\mathbf{x}}) = 1$. The equilibrium assumption about $(\bar{\mathbf{x}}, \bar{p})$ immediately implies that $v(u_i^F(\bar{\mathbf{x}}), u_i^F(\mathbf{y})) = 1$ for all $\mathbf{y} \in \mathbf{X}$ such that $\beta_{p^F}^i(\mathbf{y}) = 1$, where the definition of u_i^F was used. These equations imply the statement. \square

The theorem means that the equilibrium of fuzzy market is a fuzzy extension of the deterministic equilibrium of the original deterministic market.

4. Cooperative Fuzzy Market

The above procedure can be repeated also for the concept of \mathbf{M} -equilibrium. Let \mathbf{m} be a market, let \mathbf{m}^F be its fuzzy extension, let p^F be vector of fuzzy prices which is a fuzzy extension of $\bar{p} \in P$. Let, finally, $\mathbf{M} \subset 2^I$ fulfil assumptions of Subsection 2.2.

If $K \subset I$ is a non-empty coalition then we introduce the fuzzy subset $B_{p^F}^K$ of \mathbf{X} with membership function $\beta_{p^F}^K$, where $\beta_{p^F}^K(\mathbf{x}) = v(p^F \cdot a^K, p^F \cdot x^K)$, $\mathbf{x} = (x^i)_{i \in I} \in \mathbf{X}$, where v represents the fuzzy ordering relation, and for any $\mathbf{x} \in \mathbf{X}$ the symbols $p^F \cdot x^K$ and $p^F \cdot a^K$ denote the sums of fuzzy quantities $p^F \cdot x^i$, or $p^F \cdot a^i$, summed over all

$i \in K$, respectively. The above fuzzy set is called *fuzzy budget set of coalition K*.

Analogously to the previous section, we define the M -equilibrium of \mathbf{m}^F as a fuzzy subset of the class of all states (\mathbf{x}, p^F) described by a membership function ρ_M , where

$$\rho_M(\mathbf{x}, p^F) = \min(\beta_{p^F}^K(\mathbf{x}), \gamma_K(\mathbf{x}) : K \in M)$$

where $\beta_{p^F}^K$ is defined above, and

$$\gamma_K(\mathbf{x}) = \min(\beta_{p^F}^K(\mathbf{y}), v(u_K^F(\mathbf{x}), u_K^F(\mathbf{y})) : \mathbf{y} \in \mathbf{X}).$$

Then the following statement can be derived analogously to the proof of the previous Theorem 1.

Theorem 2. Let \mathbf{m}^F be fuzzy extension of market \mathbf{m} , and let p^F be fuzzy extension of prices $\bar{p} \in P$. Let the state $(\bar{\mathbf{x}}, \bar{p})$ be an M -equilibrium of \mathbf{m} for some class of coalitions. Then

$$\rho_M(\bar{\mathbf{x}}, p^F) = 1.$$

The following two statements immediately follow from definitions.

Remark 4. If $M = \{\{1\}, \{2\}, \dots, \{n\}\}$ is the class of all one-agent coalitions then the M -equilibrium is identical with the equilibrium for any fuzzy market \mathbf{m}^F .

Lemma 2. If $M = \{I\}$ is the one-coalition class then for any fuzzy market \mathbf{m}^F and any fuzzy price vector p^F there exists an M -equilibrium $(\bar{\mathbf{x}}, p^F)$.

Proof. For this specific case of M the budget set $B_p^I = \{\mathbf{x} \in \mathbf{X} : \forall j = 1, \dots, m, \sum_{i \in I} x_j^i = \sum_{i \in I} a_j^i\}$ is independent on the price vector p . The existence of the maximal value of fuzzy coalitional utility function $u_K^F(\mathbf{x})$ for all $\mathbf{x} \in B_p$ follows from the definition of u_i^F and u_K^F , and from the properties of fuzzy ordering. \square

5. Conclusive Remarks

The above brief formulation of the fuzzy market and cooperative fuzzy market model covers only one of possible approaches to the fuzziness in market. Let us mention at least two alternatives. Both of them are inspired with fuzzy cooperative games theory.

The fuzziness of the market activities may regard not only utilities and prices (i.e., the quantitative values treated by the market model) but also the participation of agents in coalitions. It is possible (and realistic) to admit that each player may participate in more coalitions with different degree of participation (compare with [1], [2], [3]). This alternative approach opens qualitatively new aspects of the market model. The concept of \mathbf{M} -equilibrium (where the coalitions in \mathbf{M} need not be disjoint) offers a frame for the formal representation of such market.

On the other hand, the fuzziness of the quantitative aspects of market need not be represented by the fuzzification of its particular components like utilities and prices, but by analyzing a fuzzy subclass of the class of all deterministic markets with the set of agents I and initial allocation $\mathbf{a} = (a^i)_{i \in I} \in \mathbf{X}$. The difference of both approaches is analogous to the difference between the fuzzy cooperative game models given in [9] and [10]. Both of these models are mutually related – the memberships of fuzzy class of crisp markets can be derived from the memberships of fuzzy utilities.

The analysis of both these alternatives may be done in some following research.

References

- [1] J. P. Aubin: Cooperative fuzzy games. *Math. Oper. Res.* 6 (1981), 1-13.
- [2] D. Butnariu: Fuzzy games: a description of the concept. *Fuzzy Sets and Systems* 1 (1978), 181-192.
- [3] D. Butnariu, E. P. Klement: Triangular Norm – Based Measures and Games With Fuzzy Coalitions. Kluwer, Dordrecht 1993.
- [4] W. Hildenbrand: Core and Equilibria of a Large Economy. Princeton Univ. Press, Princeton 1974.
- [5] M. Mareš: Suggestion of a cooperative market model. *Kybernetika* 12 (1976), 6, 439-450.
- [6] M. Mareš: On cooperative games connected with markets. *Kybernetika* 12 (1976), 6, 451-461.
- [7] M. Mareš: General coalition games and markets. *Transactions of „Operations Research Tagung. Oberwolfach 1978“*, Springer-

- Verlag, Berlin 1979, 217-222.
- [8] M. Mareš: Computation Over Fuzzy Quantities. CRC-Press, Boca Raton 1994.
 - [9] M. Mareš: Fuzzy Cooperative Games. Physica-Verlag, Heidelberg 2001.
 - [1] M. Mareš, M. Vlach: Fuzzy classes of cooperative games with transferable utility. *Mathematica Japonica*. To appear in 2004.
 - [1] J. Rosenmüller: Kooperativne Spiele und Märkte. Springer-Verlag, Heidelberg 1971.
 - [1] J. Rosenmüller: The Theory of Games and Markets. North Holland, Amsterdam 1982.
 - [1] L. S. Shapley, M. Shubik: On market games. *Journal of Econ. Theory*. Vol. I (June 1969), No.1

The Shapley value for games on lattices (L-fuzzy games)

Michel GRABISCH

Université Paris I - Panthéon-Sorbonne ; LIP6

8, rue du Capitaine Scott 75015 Paris, France

email `Michel.Grabisch@lip6.fr`

Abstract

Several concepts of cooperative games (non monotonic fuzzy measures) have been proposed in the past, which can be all considered as games on lattices. For the case of distributive product lattices, we propose a general formulation of the Shapley value, together with an axiomatization. These games could be called *L*-fuzzy games. The case of global games proposed by Gilboa and Lehrer, where the underlying lattice is the partition lattice, is also cited.

1 Introduction

The field of cooperative game theory [1, 13] has been enriched these recent years by many new kinds of game, trying to model in a more accurate way the behaviour of players in a real situation. To the original notion of cooperative game (otherwise said, non-monotonic fuzzy measures), where to each coalition of players taking part into the game, an asset is associated, many variations have been added, let us cite for example bi-cooperative games proposed by Bilbao *et al.* [1] (see also Labreuche and Grabisch [12] for a slightly different view) where players are either defenders, defeaters or do not participate, ternary voting games of Felsenthal and Machover [5], where voters can choose between voting in favor or against, or abstain, multichoice games of Hsiao and Raghavan [11] where each player has several possible levels of participation to the game, games with r alternatives proposed by Bolger [3], fuzzy games [4] where each player has a degree of participation, global games of Gilboa and Lehrer [6], k -stage games, games defined on pairs (S, P) , where S is a set and P a partition containing it, etc.

Let us denote $N = \{1, \dots, n\}$ the finite set of players. As remarked by Gilboa and Lehrer [6], many of the above examples can be considered as particular cases of games on lattices, i.e. functions $v : (L, \leq) \rightarrow \mathbb{R}$ where (L, \leq) is a lattice, and such that $v(\perp) = 0$, \perp denoting the bottom element of L . This point of view has also been advocated by Grabisch and Labreuche, in a larger perspective of decision theory and data analysis [9, 10]. For example, classical cooperative games correspond to the case where (L, \leq) is the lattice $(2^N, \subseteq)$ of all coalitions of players ordered by inclusion,

bi-cooperative games and ternary voting games correspond to the case where $L = 3^N$, multichoice games correspond to $L = (m+1)^N$, with m the number of different levels of participation, fuzzy games to $L = [0, 1]^N$, and global games are built on the lattice of partitions of N . We remark that most of the examples of lattices are products of n lattices, all identical. This is rather natural since each dimension correspond to a player, and this was the motivation of Grabisch and Labreuche for considering only product lattices $L = L_1 \times \dots \times L_n$. Then interpreting the elements of L_i as level or *degree* of participation of player i to the game, such games could be called *L-fuzzy games*, in the same sense that Goguen [7] proposed *L-fuzzy sets*, i.e. valued on some lattice instead of $[0, 1]$. However in the sequel, we will keep the term “games on (product) lattices”, since this is more precise.

A central question in game theory is how to compute a value or solution concept for a game, i.e. how to individually reward players when the game has been played. For classical game, this is done by the well-known Shapley value [14]. For all the above cited examples, authors have tried to generalize or adapt the original Shapley value to their own framework. In this paper, we propose a general formulation for the case of product lattices, together with an axiomatization in the spirit of Weber [15]. The case of global games, which cannot be included into our framework, will be separately explained.

Similar work has been done by Bilbao [1] for games defined on other general structures (convex geometries and matroids).

To avoid heavy notations, we will often omit braces for singletons, and cardinality of sets S, T will be denoted by s, t . etc.

2 Classical cooperative games

Let $N := \{1, \dots, n\}$ be the finite set of players. A *game* on N is any function $v : 2^N \rightarrow \mathbb{R}$, such that $v(\emptyset) = 0$. We call *coalition* any subset of N . For any coalition $A \subseteq N$, $v(A)$ is the asset or income the coalition A will win if the game is played, i.e. if all players equally participate to the game, players in A^c playing against A (or do not participate). We denote $\mathcal{G}(2^N)$ the set of all games on N .

Player i is said to be *dummy* for v if $v(S \cup i) = v(S) + v(i)$, for all coalition S . Player i is *null* for v if $v(S \cup i) = v(S)$. Observe that a null player is dummy, and a dummy player with $v(i) = 0$ is null.

A *value* or *solution concept* is any function $\phi : \mathcal{G}(2^N) \rightarrow \mathbb{R}^N$, which represent an assignment of income to each player. Let us present examples of properties or axioms ϕ may fulfil.

- **linearity (l):** ϕ is linear over $\mathcal{G}(2^N)$.
- **dummy axiom (d):** if i is dummy for v , then $\phi_i(v) = v(i)$.
- **null axiom (n):** if i is null for v , then $\phi_i(v) = 0$.
- **symmetry (s):** ϕ does not depend on the labelling of the players.
- **efficiency (e):** $\sum_{i \in N} \phi_i(v) = v(N)$.

Note that the dummy axiom is stronger than the null axiom. The *Shapley value* [14] is the unique value satisfying axioms **I**, **n**, **s** and **e**, and is given by

$$\phi_i(v) := \sum_{S \subseteq N \setminus i} \frac{(n-s-1)!s!}{n!} [v(S \cup i) - v(S)]. \quad (1)$$

Weber [15] has shown how precisely each axiom contributes to the result. Since we will follow the same approach, we briefly present his results. Linearity implies the existence of 2^n real constants a_S^i such that $\phi_i(v) = \sum_{S \subseteq N} a_S^i v(S)$. If the null axiom is added, then these constants are pairwise opposite so that $\phi_i(v) = \sum_{S \subseteq N \setminus i} p_S^i [v(S \cup i) - v(S)]$, with $p_S^i := a_{S \cup i}^i$. If the dummy axiom replaces the null axiom, then in addition the 2^{n-1} coefficients p_S^i satisfy $\sum_{S \subseteq N \setminus i} p_S^i = 1$. If the symmetry axiom is added, then coefficients p_S^i depend only on the cardinality of S , so that only n coefficients p_0, \dots, p_{n-1} are needed ($n-1$ if the dummy axiom replaces the null axiom). Lastly, if efficiency is added, the unique possibility are the coefficients of the Shapley value in (1).

3 Games on product lattices

As remarked in the introduction, most of more general concepts of games can be embedded into a general notion of game on lattices, where the lattice is a product lattice $L := L_1 \times \dots \times L_n$. A general framework for these games has been proposed by Grabisch and Labreuche, aiming at defining interaction (a generalization of the Shapley value to coalitions). Since our aim is to focus on the Shapley value, we restrict here to necessary notions and avoid intricacies (see [9, 10] for details).

3.1 Preliminaries

We recall elementary definitions for lattices (see e.g. [2]), in a finite setting. A *lattice* is a set L endowed with a partial order \leq such that for any $x, y \in L$ their least upper bound $x \vee y$ and greatest lower bound $x \wedge y$ always exist. For finite lattices, the greatest element of L (denoted \top) and least element \perp always exist. x *covers* y (denoted $x \succ y$) if $x > y$ and there is no z such that $x > z > y$. The lattice is *distributive* if \vee, \wedge obey distributivity. An element $j \in L$ is *join-irreducible* if it cannot be expressed as a supremum of other elements. Equivalently j is join-irreducible if it covers only one element. Join-irreducible elements covering \perp are called *atoms*, and the lattice is *atomistic* if all join-irreducible elements are atoms. The set of all join-irreducible elements of L is denoted $\mathcal{J}(L)$.

An important property is that in a distributive lattice, any element x can be written as an irredundant supremum of join-irreducible elements in a unique way (this is called the *minimal decomposition* of x).

In a finite setting, *Boolean lattices* are of the type 2^N for some set N , i.e. they are isomorphic to the lattice of subsets of some set, ordered by inclusion. Boolean lattices are atomistic, and atoms corresponds to singletons. A *linear lattice* is such that \leq is a total order. All elements are join-irreducible, except \perp .

Given lattices $(L_1, \leq_1), \dots, (L_n, \leq_n)$, the product lattice $L = L_1 \times \dots \times L_n$ is endowed with the product order \leq of \leq_1, \dots, \leq_n in the usual sense. Elements of x can be written in their vector form (x_1, \dots, x_n) . We use the notation (x_A, y_{-A}) to indicate a vector z such that $z_i = x_i$ if $i \in A$, and $z_i = y_i$ otherwise. Similarly L_{-i} denotes $\prod_{j \neq i} L_j$. All join-irreducible elements of L are of the form $(\perp_1, \dots, \perp_{j-1}, i_0, \perp_{j+1}, \dots, \perp_n)$, for some j and some join-irreducible element i_0 of L_j . A *vertex* of L is any element whose components are either top or bottom. We denote $\Gamma(L)$ the set of vertices of L . Note that $\Gamma(L) = L$ iff L is Boolean.

Let (L, \leq) be some finite lattice, and consider $f : L \rightarrow \mathbb{R}$, and a join-irreducible element i of L . The *derivative* of f w.r.t i at point x is defined as [8]: $\Delta_i f(x) := f(x \vee i) - f(x)$. This definition can be considered as a *first-order* derivative. One can iterate the definition, taking several join-irreducible elements. If the lattice is distributive, since any element y can be decomposed in a minimal and unique way on join-irreducible elements, the derivative w.r.t. $y, \forall y \in L$, can be defined as well. The derivative is said to be *Boolean* if $x \vee i \succ x$ (or more generally if $[x, x \vee y]$ is a Boolean lattice).

3.2 The framework

We consider distributive finite lattices L_1, \dots, L_n and their product $L := L_1 \times \dots \times L_n$. A game on L is any function $v : L \rightarrow \mathbb{R}$ such that $v(\perp) = 0$. The set of such games is denoted $\mathcal{G}(L)$.

This can be interpreted as follows, in the setting of game theory (for other interpretations, see [9, 10]). Lattice L_i represents the (partially) ordered set of actions, choices, levels of participation of player i to the game. Each lattice may be different.

Let us give examples of this general framework. Let us consider that all lattices L_i 's are identical, so that the product lattice is of the form L^n . The simplest example is to take for L the smallest lattice $L := \{\perp, \top\}$, also denoted 2 since it has only 2 elements. Note that this is a linear lattice. Then we get 2^n , i.e. we recover classical games, defined on the power set of N ordered by inclusion, which is a Boolean lattice. Here \top has the meaning of participating to the game, and \perp means no participation (or playing against).

Bi-cooperative games and ternary voting games correspond to the case $L_1 = \dots = L_n = 3$, i.e. there are three possible actions, totally ordered¹.

Multichoice games correspond to all cases where $L_1 = \dots = L_n = m + 1$, i.e. a linear lattice of m actions, plus the \perp action which correspond to do nothing.

Fuzzy games correspond to the case $L_1 = \dots = L_n = [0, 1]$. Considering fuzzy sets valued on some lattice (the so-called L -fuzzy sets of Goguen [7]), it is clear that our proposal of games on product lattice may be called also *L-fuzzy games*.

Global games *do not* fit in our picture since the underlying lattice is the partition lattice, which is not a product lattice. We will consider this case later on.

¹This is true up to the fact that for bi-cooperative games as defined in [1, 12], the condition $v(\perp) = 0$ is replaced by $v(0, 0, \dots, 0) = 0$, if we denote all L_i by $\{\perp, 0, \top\}$. To fit with this definition, a slightly different framework has to be defined, which we do not address in this paper.

3.3 Value for games on lattices

In [9, 10] a general form for a solution concept was proposed for games on product lattices. We give its definition. We use the notation $L_{-j} := \prod_{i \neq j} L_i$.

Definition 1 Let $i = (\perp_1, \dots, \perp_{j-1}, i_0, \perp_{j+1}, \dots, \perp_n)$ be a join-irreducible element of L . The value w.r.t. i of v is any function of the form

$$\phi_i(v) := \sum_{x \in \Gamma(L_{-j}) \times \{i_0\}} \alpha_{h(x)}^1 \Delta_i v(x), \quad (2)$$

where i_0 is the (unique) element covered by i_0 in L_j , $h(x)$ is the number of components of x equal to \top_l , $l = 1, \dots, n$, and $\alpha_k^1 \in \mathbb{R}$ for any integer k .

Observe that the constants $\alpha_{h(x)}^1$ do not depend on i . Also, the derivative is Boolean.

Let us show that this definition encompasses the case of classical games, i.e. where L is the Boolean lattice 2^n . Join-irreducible elements of L are singletons, as remarked before all elements in L are vertices, and $h(x)$ is the cardinality of sets. Also for $S \subseteq N \setminus i$, $\Delta_i v(S) = v(S \cup i) - v(S)$, thus we get $\phi_i(v) = \sum_{S \subseteq N \setminus i} \alpha_{|S|}^1 [v(A \cup i) - v(S)]$ for any $i \in N$, as desired (see Section 2).

In [9, 10], Grabisch and Labreuche showed that if a condition similar to efficiency was added, specifically $\sum_{i \in \mathcal{J}(L)} I(i) = v(\top) - v(\perp)$, then the coefficients were determined, and equal to the Shapley coefficients $\alpha_k^1 = \frac{(n-1-k)!k!}{n!}$.

Let us show that under suitable axioms, one can recover this result in a unique way.

3.4 Axiomatization of the Shapley value for games on product lattices

We consider the linear case, where $L = \prod_{i=1}^n L_i$, with L_i a linear lattice denoted $L_i := \{0, 1, 2, \dots, l_i\}$. With some abuse of notation, for some $k \in L_i$ we write k_i for $(0_{-i}, k_i)$, i.e. $(0, \dots, 0, k, 0, \dots, 0)$, where k is at the i th position.

Our aim is to define $\phi_{k_i}(v)$, for any $i \in N$, any $k \in L_i, k \neq 0$ (i.e. k_i ranges over all join-irreducible elements of L).

For some $k \in L_i, k \neq 0$, player i is said to be k -dummy if $v(x, k_i) = v(x, (k-1)_i) + v(k_i)$, for any $x \in L_{-i}$.

Dummy axiom (D): $\forall v \in \mathcal{G}(L)$, for all join-irreducible k_i , $\phi_{k_i}(v) = v(k_i)$ if i is k -dummy.

For some $k \in L_i, k \neq 0$, player i is said to be k -null if $v(x, k_i) = v(x, (k-1)_i)$, for any $x \in L_{-i}$.

Null axiom (N): $\forall v \in \mathcal{G}(L)$, for all join-irreducible k_i , $\phi_{k_i}(v) = 0$ if i is k -null.

Remark that i is k -dummy and $v(k_i) = 0$ implies i is k -null, and i is 1-null implies i is 1-dummy.

Linear axiom (L): ϕ_v is linear on the set of games, i.e. for any join-irreducible k_i , $\phi_{k_i}(v) = \sum_{x \in L} a_x^{k_i} v(x)$, with $a_x^{k_i} \in \mathbb{R}$.

Proposition 1 Under (L) and (N), $\forall v \in \mathcal{G}(L)$, for all join-irreducible k_i , $\phi_{k_i}(v) = \sum_{x \in L_{-i}} p_x^{k_i} \Delta_{k_i} v(x, (k-1)_i)$, with $p_x^{k_i} \in \mathbb{R}$.

Recall that $\Delta_{k_i} v(x, (k-1)_i) = v(x, k_i) - v(x, (k-1)_i)$.

Let us see what we get if add the dummy axiom.

Proposition 2 Under (L), (D) and (N), $\forall v \in \mathcal{G}(L)$, for all join-irreducible k_i , $\phi_{k_i}(v) = \sum_{x \in L_{-i}} p_x^{k_i} \Delta_{k_i} v(x, (k-1)_i)$, with $p_x^{k_i} \in \mathbb{R}$, and $\sum_{x \in L_{-i}} p_x^{k_i} = 1$.

Let σ be a permutation on N . With some abuse of notation we write $\sigma(x) := (x_{\sigma(1)}, \dots, x_{\sigma(n)})$.

Symmetry axiom (S): $\phi_{\sigma(k_i)}(v \circ \sigma^{-1}) = \phi_{k_i}(v)$, for any game v , any join-irreducible k_i .

Proposition 3 Under (L), (N) and (S), $\phi_{k_i}(v) = \sum_{x \in L_{-i}} p_{n_1, \dots, n_l}^k \Delta_{k_i} v(x, (k-1)_i)$, where $l := \max(l_1, \dots, l_n)$, and n_j is the number of components of x being equal to j .

Difference axiom (Di) Let us consider v_1, v_2 on L such that for some $i \in N$ and some $k \geq 2$, $v_2(x, k_i) - v_2(x, (k-1)_i) = v_1(x, (k-1)_i) - v_1(x, (k-2)_i)$, $\forall x \in L_{-i}$. Then for any such games $\phi_{k_i}(v_2) = \phi_{(k-1)_i}(v_1)$.

The axiom says that when a game (v_2) is merely a shift of another game (v_1) concerning player i and level k , the Shapley values are the same. This implies that the way of computing ϕ_v does not depend on the level k , as shown in the next proposition.

Proposition 4 Under axioms (L), (N) and (Di), $p_x^{k_i} = p_x^{(k-1)_i}$, for all $x \in L_{-i}$, $\forall i \in N$, $\forall k \in L_i$, $k \neq 0$.

Recall that $\mathcal{J}(L)$ is the set of all join-irreducible elements of L .

Efficiency axiom (E): $\sum_{k_i \in \mathcal{J}(L)} \phi_{k_i}(v) = v(\top) - v(\perp)$.

Proposition 5 Suppose $L_1 = L_2 = \dots = L_n =: L$. Under axioms (L), (N), (S) and (E), the coefficients p_{n_1, \dots, n_l}^k satisfy $p_{0, \dots, 0, n-1}^1 = 1/n$, $p_{0, \dots, 0}^1 = 1/n$, and $n_l p_{n_1, \dots, n_{l-1}}^1 + \sum_{j=1}^{l-1} n_j (p_{n_1, \dots, n_{j-1}, \dots, n_l}^j - p_{n_1, \dots, n_{j-1}, \dots, n_l}^{j+1}) = (n - n_1 - \dots - n_l) p_{n_1, \dots, n_l}^1$.

The final result is the following.

Theorem 1 Suppose $L_1 = L_2 = \dots = L_n =: L$. Under axioms (L), (D), (S), (Di) and (E),

$$\phi_{k_i}(v) = \sum_{x \in \Gamma(L^{n-1})} \frac{(n - n_l - 1)! n_l!}{n!} [v(x, k_i) - v(x, (k-1)_i)],$$

where $\Gamma(L^{n-1})$ is the set of vertices of L^{n-1} .

It is remarkable that only vertices of L_{-i} remain in the summation, which justifies the general form given in (2).

Note that if axiom **(D)** is replaced by axiom **(N)**, then coefficients p_{n_1, \dots, n_l} with $n_1, n_2, \dots, n_{l-1} \neq 0$ are not determined, and so the summation is not restricted to vertices of L_{-i} .

4 Global games

We briefly describe global games as proposed by Gilboa and Lehrer [6]. In global games, all players tend to a global goal which will benefit to all of them if is attained (e.g. environment problems like water pollution, diminishing ozone layer, greenhouse effect, etc.), and a game v modelizes the global utility for a given profile of cooperation, represented as a partition of all players. Denoting $\Pi(N)$ the set of all partitions of N , we have $v : \Pi(N) \rightarrow \mathbb{R}$.

A partition π is *coarser* than a partition π' (denoted $\pi \supseteq \pi'$) if for every $A \in \pi$ there exists $B \in \pi'$ such that $A \supseteq B$. Then $(\Pi(N), \supseteq)$ is a lattice, known as the *partition lattice*, top being $\{N\}$ and bottom being $\{\{i\}\}_{i \in N}$.

The partition lattice is an example of a geometric lattice (atomistic and upper semi-modular). It is not distributive. This lattice being not a product lattice, our framework does not apply. Gilboa and Lehrer have proposed axioms for defining the Shapley value, which we present below.

To each global game h , we assign a (classical) cooperative game v_h , called the *induced game*, as follows: $v_h(A) := h(\{\{A\}, \{\{i\}_{i \notin A}\}\})$, $\forall A \subset N, A \neq \emptyset$. A player i is *null*² for global game h if for all partition π , $h(\pi) = h(\pi \wedge \{\{N \setminus i\}, \{i\}\})$. Two players i, j are *interchangeable* for h if for all partition π , $h(\pi \wedge \{\{N \setminus i\}, \{i\}\}) = h(\pi \wedge \{\{N \setminus j\}, \{j\}\})$. Gilboa and Lehrer defined the following axioms for a value ϕ on $\mathcal{G}(\Pi(N))$:

Linearity (\mathbf{L}_{GL}): ϕ is linear on $\mathcal{G}(\Pi(N))$.

Null axiom (\mathbf{N}_{GL}): for all $h \in \mathcal{G}(\Pi(N))$, for all $i \in N$, if i is null for h , then $\phi_i(h) = 0$.

Interchangeability (\mathbf{I}_{GL}): for all $h \in \mathcal{G}(\Pi(N))$, for all interchangeable i, j , $\phi_i(h) = \phi_j(h)$.

Efficiency (\mathbf{E}_{GL}): $\sum_{i \in N} \phi_i(h) = h(\{N\})$.

Note that these axioms are also close to the original ones of Shapley, interchangeability playing the role of symmetry. They proved the following.

Theorem 2 *The unique value (called the Shapley value) satisfying axioms (\mathbf{L}_{GL}), (\mathbf{N}_{GL}), (\mathbf{I}_{GL}), and (\mathbf{E}_{GL}) is equal to the Shapley value of the induced game: $\phi_i(h) = \phi_i(v_h)$.*

²Gilboa and Lehrer called this axiom “dummy”, but it seems to correspond more to the null axiom.

As remarked by Gilboa and Lehrer, it is remarkable that all partitions $\pi \in \Pi(N)$ do not intervene in the Shapley value, but only the following ones $\{\{\{A\}, \{\{i\}\}_{i \notin A}\}\}_{A \subseteq N}$ are used. These partitions could be called *vertices partitions*, and we recover what we have found in our framework, that only vertices of L_{-i} are used. This fact suggests that a yet more general framework encompassing our approach and global games should exist.

References

- [1] J.M. Bilbao. *Cooperative games on combinatorial structures*. Kluwer Academic Publishers, 2000.
- [2] G. Birkhoff. *Lattice Theory*. American Mathematical Society, 3d edition, 1967.
- [3] E.M. Bolger. A consistent value for games with n players and r alternatives. *Int. J. of Game Theory*, 29:93–99, 2000.
- [4] D. Butnariu and E.P. Klement. *Triangular norm-based measures and games with fuzzy coalitions*. Kluwer Academic Publishers, 1993.
- [5] D. Felsenthal and M. Machover. Ternary voting games. *Int. J. of Game Theory*, 26:335–351, 1997.
- [6] I. Gilboa and E. Lehrer. Global games. *Int. J. of Game Theory*, 20:129–147, 1991.
- [7] J.A. Goguen. L -fuzzy sets. *J. Math. Anal. Appl.*, 8:145–174, 1967.
- [8] M. Grabisch and Ch. Labreuche. Capacities on lattices and k -ary capacities. In *3d Int. Conf. of the European Soc. for Fuzzy Logic and Technology (EUSFLAT 2003)*, pages 304–307, Zittau, Germany, September 2003.
- [9] M. Grabisch and Ch. Labreuche. Interaction between attributes in a general setting for knowledge discovery. In *4th Int. JIM Conf. (Journées de l'Informatique Messine) on Knowledge Discovery and Discrete Mathematics*, pages 215–222, Metz, France, September 2003.
- [10] M. Grabisch and Ch. Labreuche. Derivative of functions over lattices as a basis for the notion of interaction between attributes. *Discrete Applied Maths.*, submitted.
- [11] C.R. Hsiao and T.E.S. Raghavan. Shapley value for multichoice cooperative games, I. *Games and Economic Behavior*, 5:240–256, 1993.
- [12] Ch. Labreuche and M. Grabisch. Bi-cooperative games and their importance and interaction indices. In *14th Mini-EURO Conference on Human Centered Processes (HCP'2003)*, pages 287–291, Luxembourg, May 2003.
- [13] G. Owen. *Game Theory*. Academic Press, 3d edition, 1995.
- [14] L.S. Shapley. A value for n -person games. In H.W. Kuhn and A.W. Tucker, editors, *Contributions to the Theory of Games, Vol. II*, number 28 in Annals of Mathematics Studies, pages 307–317. Princeton University Press, 1953.
- [15] R.J. Weber. Probabilistic values for games. In A.E. Roth, editor, *The Shapley Value. Essays in Honor of Lloyd S. Shapley*, pages 101–119. Cambridge University Press, 1988.

Optimal Strategies of Some Symmetric Matrix Games

Bart De Schuymer^{†1}, Hans De Meyer¹ and Bernard De Baets²

¹ Department of Applied Mathematics and Computer Science
Ghent University, Krijgslaan 281 (S9), B-9000 Gent, Belgium

² Department of Applied Mathematics, Biometrics and Process Control
Ghent University, Coupure links 653, B-9000 Gent, Belgium

Summary. Recently, we have developed a method for comparing dependent or pairwise independent discrete or continuous random variables. In this paper, we show that the same method can be used to describe certain games. Three different classes of symmetric matrix games, which are characterized by the copula that binds the marginal distributions, are proposed. The necessary and sufficient conditions for such a game to possess an optimal strategy are laid bare and the corresponding optimal strategies are described.

1 Dice models

The basic constituents of a dice model are the dice which can be identified with integer multisets, in the sense that each face of a dice carries exactly one element (integer) of its associated multiset. By considering the winning probabilities among all pairs of dice, any collection of dice generates a probabilistic relation that constitutes the corner stone of the dice model. The dice model, originally introduced in [3] as a method for comparing multisets, has been extended later on in various ways, in particular for the comparison of discrete or continuous, independent [4] or dependent [2] random variables. In the present article, we will only consider discrete dice models consisting of uniformly distributed random variables that are either independent or in some specific way pairwise coupled.

A basic concept is that of a probabilistic relation, often also called reciprocal or ipsodual relation. Probabilistic relations serve as a popular representation of various relational preference models [1, 5, 9]. A probabilistic relation Q on a set of alternatives A is a mapping from A^2 to $[0, 1]$ such that for all $a, b \in A$ it holds that $Q(a, b) + Q(b, a) = 1$.

[†] Bart.DeSchuymer@ugent.be, tel.: 32-9-264.47.66 – fax: 3-9-264.49.95

Given a collection of dice, we will define the winning probabilities for each pair of dice and the set of dice together with the generated probabilistic relation will be called a dice model [3]. For any multiset $\{a_1, a_2, \dots, a_n\}$ considered in this article it holds that $a_i \in \mathbb{N}_0$ and that the elements are ordered non-decreasingly: $a_i < a_j \Rightarrow i < j$.

Definition 1. We define for any two multisets $A = \{a_1, a_2, \dots, a_{n_1}\}$ and $B = \{b_1, b_2, \dots, b_{n_2}\}$:

$$P(A, B) = \frac{1}{n_1 n_2} (\#\{(a, b) \in A \times B \mid a > b\}), \tag{1}$$

$$I(A, B) = \frac{1}{n_1 n_2} (\#\{(a, b) \in A \times B \mid a = b\}), \tag{2}$$

$$D(A, B) = P(A, B) + \frac{1}{2} I(A, B). \tag{3}$$

Definition 2. A finite collection of multisets M_i generates a probabilistic relation $Q = [q_{ij}]$, defined by $q_{ij} = D(M_i, M_j)$. This collection together with the probabilistic relation Q is called a dice model.

The dice model can be interpreted in a probabilistic way, in the sense that the dice or multisets A_i can be identified with independent discrete random variables X_i that possess a uniform probability distribution defined on these discrete number sets A_i , whereas the probabilistic relation Q may be regarded as a way of comparing pairwise these random variables [4]. The following definition is then a straightforward generalization in which random variables instead of dice are compared.

Definition 3. Let (X_1, X_2, \dots, X_m) denote a discrete random vector with joint probability mass function p_{X_1, X_2, \dots, X_m} . Then the relation $Q = [q_{ij}]$, with elements

$$\begin{aligned} q_{ij} &= \text{Prob}\{X_i > X_j\} + \frac{1}{2} \text{Prob}\{X_i = X_j\} \\ &= \sum_{k>l} p_{X_i, X_j}(k, l) + \frac{1}{2} \sum_k p_{X_i, X_j}(k, k), \end{aligned} \tag{4}$$

is a probabilistic relation, where p_{X_i, X_j} denotes the bivariate marginal probability mass function of (X_i, X_j) .

For the computation of Q only the two-dimensional joint cumulative distribution functions F_{X_i, X_j} are required, since the two-dimensional marginal probability mass functions p_{X_i, X_j} can be immediately derived from them. Sklar's theorem [7] tells us that if a joint cumulative distribution function F_{X_i, X_j} has marginals F_{X_i} and F_{X_j} , then there exists a copula C such that for all x, y :

$$F_{X_i, X_j}(x, y) = C(F_{X_i}(x), F_{X_j}(y)). \tag{5}$$

On the other hand, if C is a copula and F_{X_i} and F_{X_j} are distribution functions, then the function defined by (5) is a joint cumulative distribution function with marginals F_{X_i} and F_{X_j} .

It is well known that all copulas are situated between the Lukasiwicz copula and the minimum copula, which are denoted as $T_{\mathbf{L}}$ and $T_{\mathbf{M}}$:

$$T_{\mathbf{L}}(x, y) = \max(0, x + y - 1) \leq C(x, y) \leq \min(x, y) = T_{\mathbf{M}}(x, y). \quad (6)$$

In the next sections we will discuss three games that differ from each other by the copula used: more precisely, we will consider the product copula $T_{\mathbf{P}}$, which corresponds to independent random variables, the minimum copula $T_{\mathbf{M}}$ and the Lukasiwicz copula $T_{\mathbf{L}}$. The one-dimensional marginal cumulative distribution functions will be those corresponding to discrete uniformly distributed random variables over an integer multiset of cardinality n .

Note that the copulas $T_{\mathbf{L}}$, $T_{\mathbf{P}}$ and $T_{\mathbf{M}}$ are associative and can therefore also be regarded as triangular norms [6]. Originally used for modelling the triangle inequality in probabilistic metric spaces, triangular norms have been adopted later on by the field of fuzzy logic for modelling many-valued conjunction. The three copulas mentioned above are also the three most important continuous t-norms, as any other continuous t-norm can be obtained from them using a construction known as ordinal sum.

2 Game-theoretical concepts

In this section, we briefly introduce some concepts from game theory, see f.e. [8]. The games that we will discuss in this paper are so-called non-cooperative games. In these games the goal of each participant (player) is to achieve the largest possible individual profit (payoff). The process of the game consists of each one of the players choosing a certain strategy $s_i \in S_i$. Thus as a result of each “round” of the game, a system of strategies $(s_1, \dots, s_n) = s$ is put together. This system is called a situation.

A situation s is *admissible* for a player if by replacing her/his present strategy in this situation with some other strategy, the player is unable to increase her/his payoff. A situation s , which is admissible for all the players is called an *equilibrium situation*. An *equilibrium strategy* of a player in a non-cooperative game is a strategy that appears in at least one equilibrium situation of the game.

An *antagonistic game* is a game with two players only and the values of the payoff function for these players in each situation are the same in absolute value but of the opposite sign. Antagonistic games in which each player possesses a finite number of strategies are called *matrix games*.

An equilibrium situation for the particular case of an antagonistic game is called a *saddle point*. Equilibrium strategies of players in an antagonistic game are called their *optimal strategies*.

A matrix game is completely determined by its *payoff matrix*. Consider a matrix $A = [a_{ij}]$ in which the rows correspond to the strategies of the first player and the columns to the strategies of the second, and the elements of

the matrix located at the intersection of the rows and columns correspond to the situations of the game. If we place in each cell the payoff of the first player in the appropriate situation, we obtain the description of the game in the form of a matrix, called the *payoff matrix*.

As the value in the payoff matrix of a matrix game is the same for each saddle point, all equilibrium strategies are equivalent for the players, hence the term optimal strategy. A matrix game with a payoff matrix $A = [a_{ij}]$ for which $a_{ij} = -a_{ji}$, is called a symmetric game. The value in the payoff matrix of a saddle point in a symmetric game equals zero.

In order for a matrix game to possess at least one optimal strategy, it is necessary and sufficient that there exist i, j for which the following two minimaxes are equal:

$$\max_i \min_j a_{ij} = \min_j \max_i a_{ij}.$$

If none of the minimaxes of the elements of a matrix are equal to each other, then there are no optimal strategies. It is therefore natural that the players should seek in these cases additional strategic opportunities in order to assure for themselves the largest possible share of this difference. It turns out that it is desirable that they choose their strategies for this purpose randomly. We then speak of a *mixed strategy* instead of a *pure strategy*. We do not go into further detail, as we only investigate equilibrium situations consisting of pure strategies.

3 Description of the games

The games we introduce here can be best described in the framework of dice models, explained in Section 1. In what follows, for clarity, we drop the notion of dice and just consider multisets of strictly positive integers. The integers in the multisets are, as has already been mentioned, always ordered non-decreasingly. We will investigate the existence of optimal strategies in the following symmetric games, called (n, σ) -games: denote the collection of all multisets containing strictly positive integers, of cardinality n and sum of the integers contained in the multiset equal to σ , as $C_{(n, \sigma)}$. Two players each have their own copy of the collection $C_{(n, \sigma)}$ of multisets and they choose, independently, one such multiset from their collection. Player 1 chooses multiset M_1 , player 2 chooses multiset M_2 .

We consider three variants of the same game concept. In the first variant, two discrete independent random variables uniformly distributed over $\{1, \dots, n\}$ are used to obtain two random integers $i_1, i_2 \in \{1, \dots, n\}$. The values i_1 and i_2 determine which integers of both multisets will be compared: the i_1 -th element of multiset M_1 is compared to the i_2 -th element of multiset M_2 . In the last two variants, only one random variables uniformly distributed over $\{1, \dots, n\}$ is needed to obtain the random integer $i \in \{1, \dots, n\}$. In the

second variant, the i -th element of multiset M_1 is compared to the i -th element of multiset M_2 . In the third variant, the i -th element of multiset M_1 is compared to the $(n - i + 1)$ -th element of multiset M_2 .

The first (second, third) game variant will be denoted as an $(n, \sigma)_{\mathbf{P}}$ -game ($(n, \sigma)_{\mathbf{M}}$ -game, $(n, \sigma)_{\mathbf{L}}$ -game). Here, \mathbf{P} refers to the product copula, \mathbf{M} to the minimum copula and \mathbf{L} to the Łukasiewicz copula [7]. It can be shown that there is a direct connection between these copulas and the corresponding game variants [2].

Player 1 wins from (loses from, plays a draw with) player 2 if the element selected from multiset M_1 is strictly greater than (strictly smaller than, equal to) the element selected from multiset M_2 .

The set of strategies is therefore given by all multisets from the collection $C_{(n, \sigma)}$, the payoff matrix for player 1 is given by $A = [a_{ij}]$, where $a_{ij} = q_{ij} - \frac{1}{2}$ and $q_{ij} = \text{Prob}\{\text{player 1 wins from player 2}\} + \frac{1}{2} \text{Prob}\{\text{player 1 plays a draw with player 2}\}$, under the premise that player 1 chooses multiset M_i and player 2 multiset M_j .

The three game variants differ from each other in their definition of q_{ij} . For two (n, σ) -multisets $M_i = \{i_1, \dots, i_n\}$, $M_j = \{j_1, \dots, j_n\}$,

(i) the first game variant defines q_{ij} as

$$q_{ij}^{\mathbf{P}} = \frac{\#\{(k, l) \mid i_k > j_l\}}{n^2} + \frac{\#\{(k, l) \mid i_k = j_l\}}{2n^2}, \quad (7)$$

(ii) the second game variant defines q_{ij} as

$$q_{ij}^{\mathbf{M}} = \frac{\#\{k \mid i_k > j_k\}}{n} + \frac{\#\{k \mid i_k = j_k\}}{2n}, \quad (8)$$

(iii) and the third game variant defines q_{ij} as

$$q_{ij}^{\mathbf{L}} = \frac{\#\{k \mid i_k > j_{n-k+1}\}}{n} + \frac{\#\{k \mid i_k = j_{n-k+1}\}}{2n}. \quad (9)$$

Note that $Q = [q_{ij}]$ is, in all three game variants, a probabilistic relation.

We end this section with two definitions:

Definition 4. A multiset containing strictly positive integers, with cardinality n and sum of its elements equal to σ , is called an (n, σ) -multiset.

Definition 5. The tally-representation of an (n, σ) -multiset M is given by $(1^{t_1}2^{t_2}3^{t_3} \dots)$, for which it holds that the number of occurrences of i in M equals t_i . When $t_i = 0$ the term i^{t_i} can be omitted.

For the tally-representation $(1^{t_1}2^{t_2}3^{t_3} \dots)$ of a given (n, σ) -multiset it clearly holds that $0 \leq t_i \leq n$, $\sum_{i>0} t_i = n$ and $\sum_{i>0} i t_i = \sigma$.

4 Optimal strategies of the games

In this section, we will determine, for the three game variants, the necessary and sufficient conditions on n and σ under which the game has at least one optimal strategy. We will also determine all of these optimal strategies. The class of $(n, \sigma)_{\mathbf{P}}$ -games and $(n, \sigma)_{\mathbf{M}}$ -games with optimal strategies will prove to be very limited, while all $(n, \sigma)_{\mathbf{L}}$ -games have at least one optimal strategy.

4.1 Optimal strategies of $(n, \sigma)_{\mathbf{P}}$ -games

The first game variant, connected to the $T_{\mathbf{P}}$ -copula, can be seen as a game in which two players each throw a fair dice with n faces, where the sum of the integers on the faces of each dice equals σ . The player whose bottom face contains the highest integer wins that round, they play a draw if both integers are equal.

This game variant does not often have an optimal strategy, as can be deduced from the following theorem which formulates the necessary and sufficient conditions for an $(n, \sigma)_{\mathbf{P}}$ -game to have at least one optimal strategy.

Theorem 1. *An $(n, \sigma)_{\mathbf{P}}$ -game has at least one optimal strategy if and only if one of the following five mutually exclusive conditions is satisfied:*

- (i) $n \leq 2$
- (ii) $(n, \sigma) = (3, 7)$
- (iii) $(n, \sigma) = (3, 8)$
- (iv) $(n, \sigma) = (2l, 4l + 1), l > 1$
- (v) $n > 2$ and there exist $a, b, k \in \mathbb{N}$ such that

$$\begin{cases} n = (a + b)k - b \\ \sigma = nk \end{cases} \quad (10)$$

- (vi) $n > 2$ and there exist $a, b, k \in \mathbb{N}$ such that

$$\begin{cases} n = (a + b)k \\ \sigma = (n + b)k \\ a \neq 0 \wedge b \neq 0 \end{cases} \quad (11)$$

The propositions below state the number of optimal strategies and how they look. We begin by handling the special cases in the first proposition.

Proposition 1.

1. The $(1, \sigma)_{\mathbf{P}}$ -game: the unique strategy (σ^1) is optimal.
2. The $(2, \sigma)_{\mathbf{P}}$ -game: all $\lfloor \frac{\sigma}{2} \rfloor$ strategies are optimal.
3. The $(3, 7)_{\mathbf{P}}$ -game: $(1^1 3^2)$ is the only optimal strategy.
4. The $(3, 8)_{\mathbf{P}}$ -game: $(1^1 3^1 4^1)$ is the only optimal strategy.
5. The $(n, n)_{\mathbf{P}}$ -game: the unique strategy (1^n) is optimal.
6. The $(2n, 4n + 1)$ -game, $n > 1$: $(1^{n-1} 2^1 3^n)$ is the only optimal strategy.

Proposition 2. All $(n, \sigma)_{\mathbf{P}}$ -games, with $n \neq \sigma$, satisfying (10) have exactly $\lfloor a/(k-1) \rfloor + \lfloor b/k \rfloor + 1$ optimal strategies and their tally-representation is given by $(1^a 2^b 3^a 4^b \dots (2k-2)^b (2k-1)^a)$, where a, b are different for each optimal strategy but k is the same.

Note that the above proposition implies that $\lfloor a/(k-1) \rfloor + \lfloor b/k \rfloor$ is an invariant of the solution space of system (10), for given values n and σ .

Proposition 3. All $(n, \sigma)_{\mathbf{P}}$ -games satisfying (11) have exactly one optimal strategy $(1^a 2^b 3^a 4^b \dots (2k-1)^a (2k)^b)$.

Note that Proposition 3 implies that system (11) has at most one solution, for given values n and σ .

4.2 Optimal strategies of $(n, \sigma)_{\mathbf{M}}$ -games

While an $(n, \sigma)_{\mathbf{P}}$ -game can be regarded as a game where dice are thrown, the $(n, \sigma)_{\mathbf{M}}$ - and $(n, \sigma)_{\mathbf{L}}$ -games are better regarded as games with ordered multisets of integers.

Lemma 1. The only optimal strategy M in an $(n, \sigma)_{\mathbf{M}}$ -game, with $n \geq 3$, for which the highest integer is strictly greater than 5 is $M = \{2, 4, 6\}$.

The above lemma, along with other independent reasonings, imply the following theorem which formulates the necessary and sufficient condition for an $(n, \sigma)_{\mathbf{M}}$ -game to have at least one optimal strategy.

Theorem 2. An $(n, \sigma)_{\mathbf{M}}$ -game has at least one optimal strategy if and only if one of the following three mutually exclusive conditions is satisfied:

- (i) $n \leq 2$
- (ii) $(n, \sigma) = (3, 12)$
- (iii) $n > 2$ and there exist $t_1, \dots, t_5 \in \mathbb{N}$ such that

$$\begin{cases} t_1 + t_2 + t_3 + t_4 + t_5 = n \\ t_3 > 0 \Rightarrow t_2 + 2 > (t_3 - 1) + t_4 + t_5 \\ t_4 > 0 \Rightarrow t_3 + 2 > t_1 + (t_4 - 1) + t_5 \\ t_5 > 0 \Rightarrow t_4 + 2 > t_1 + t_2 + (t_5 - 1) \end{cases} \quad (12)$$

As before, we are able to describe the optimal strategies of the $(n, \sigma)_{\mathbf{M}}$ -games. However, a closed formula for computing the number of optimal strategies of an arbitrary $(n, \sigma)_{\mathbf{M}}$ -game has not yet been found.

Proposition 4.

1. The $(1, \sigma)_{\mathbf{M}}$ -game: the unique strategy (σ^1) is optimal.
2. The $(2, \sigma)_{\mathbf{M}}$ -game: all $\lfloor \frac{\sigma}{2} \rfloor$ strategies are optimal.
3. The $(3, 12)_{\mathbf{M}}$ -game: $(2^1 4^1 6^1)$ is the only optimal strategy.

Proposition 5. All optimal strategies of $(n, \sigma)_{\mathbf{M}}$ -games not introduced in Proposition 4 have tally-representation $(1^{t_1} 2^{t_2} 3^{t_3} 4^{t_4} 5^{t_5})$, with (t_1, \dots, t_5) a solution of (12).

4.3 Optimal strategies of $(n, \sigma)_{\mathbf{L}}$ -games

While for the two previous types of games not all games have an optimal strategy, it turns out that all $(n, \sigma)_{\mathbf{L}}$ -games have at least one optimal strategy.

Theorem 3. *All $(n, \sigma)_{\mathbf{L}}$ -games have at least one optimal strategy.*

The exact characterization of these optimal strategies is given by the following proposition.

Proposition 6. *Consider an (n, σ) -multiset $M = \{i_1, i_2, \dots, i_n\}$ and let $a = \lfloor \frac{n}{2} \rfloor + 1$, $b = \lfloor \frac{\sigma - n + a}{a} \rfloor$, $c = n + 1 - \lfloor \frac{\sigma - n}{b - 1} \rfloor$. The (n, σ) -multiset M is an optimal strategy of the corresponding $(n, \sigma)_{\mathbf{L}}$ -game if and only if one of the following mutually exclusive conditions holds:*

- (i) $n = 1$.
- (ii) $(n, \sigma) = (n, 2n)$, $n > 1$ and:
 - $M = (1^l 2^{(n-2l)} 3^l)$, $l \in \{0, 1, \dots, \lfloor \frac{n}{2} \rfloor\}$
- (iii) $(n, \sigma) = (2l, \sigma)$, $l > 0$, $\sigma \neq 4l$ and:
 - $(i_c = b \wedge \sigma \neq l(b+2) + b - 1)$, or
 - $i_{l+1} \geq b + 1$, or
 - $M = (1^{l-1} b^2 (b+1)^{l-1})$
- (iv) $(n, \sigma) = (2l + 1, \sigma)$, $l > 0$, $\sigma \neq 4l + 2$ and:
 - $i_c = b$, or
 - $M = (1^l b^1 (b+1)^l)$

References

1. H.A. David, *The Method of Paired Comparisons*, volume 12 of *Griffin's Statistical Monographs & Courses*, Charles Griffin & D. Ltd., London, 1963.
2. H. De Meyer, B. De Baets and B. De Schuymer, Extreme copulas and the comparison of ordered lists, submitted.
3. B. De Schuymer, H. De Meyer, B. De Baets and S. Jenei, On the cycle-transitivity of the dice model, *Theory and Decision* 54: 261–285, 2003.
4. B. De Schuymer, H. De Meyer and B. De Baets, Cycle-transitive comparison of independent random variables, submitted.
5. P.C. Fishburn, Binary choice probabilities: on the varieties of stochastic transitivity, *Journal of Mathematical Psychology* 10: 327–352, 1973.
6. E. Klement, R. Mesiar and E. Pap, *Triangular Norms*, Trends in Logic, Studia Logica Library, Vol. 8, Kluwer Academic Publishers, Dordrecht, 2000.
7. R. Nelsen, *An Introduction to Copulas*, volume 139 of *Lecture Notes in Statistics*, Springer-Verlag, New York, 1998.
8. N.N. Vorob'eb, *Game Theory: Lectures for Economists and Systems Scientists*, Springer-Verlag, New York, 1977.
9. Z. Switalski, General transitivity conditions for fuzzy reciprocal preference matrices, *Fuzzy Sets and Systems*, 137:85–100, 2003.

On strict monotonic t -norms and t -conorms on ordinal scales

Batyrshin^{1,2} I.Z., Batyrshin³ I.I.

¹Mexican Petroleum Institute, Mexico

²Institute of Problems of Informatics, Academy of Sciences of Tatarstan, Russia

³Kazan State University, Russia

batyr@imp.mx, batyr1@hotmail.ru

Abstract. We introduce strict monotonic t -norms and t -conorms operating with expert evaluations measured in finite ordinal scale L . The solution of this problem achieved as result of extension of L on the set of multi-sets. Three such extensions are considered: containing only one strict monotonic operation, t -norm or t -conorm, and both strict monotonic operations. The possible applications of these operations in inference procedures of expert systems operating with expert evaluations of uncertainty are discussed.

1. Introduction

Expert evaluations of possibility, plausibility, true etc values as usually measured in ordinal scales. The following example of such scale may be considered: $L_p = \{Impossible < Slightly Possible < Average Possibility < Very Possible < Sure\}$. Denote the minimal and the maximal elements of such scale as $\mathbf{0}$ and $\mathbf{1}$ correspondingly. The grades of such scale may be represented by numbers with linear ordering corresponding to the ordering of L_p , e.g. $L_s = \{1, 2, 3, 4, 5\}$, $L_f = \{0, 0.2, 0.5, 0.8, 1\}$ etc. On the set of these numbers any strict monotonic transformations are admissible but numerical operations like addition, multiplication etc can not be used. min and max operations [10] defined by linear ordering of elements in scale are adequate conjunction T and disjunction S operations on ordinal scales but they are not strict monotonic operations, e.g. $min(a, x) = a$ for all $x \geq a$. For example, for $\wedge = min$ we have $1 \wedge 1 \wedge 1 \wedge 1 = 1$ and $1 \wedge 5 \wedge 5 \wedge 5 = 1$ but it seems evident that the result of the second conjunction should be more plausible in comparison with the first one. Such non-monotonicity of these operations does not give possibility to take

into account the change of plausibility of premises in expert systems rules which use qualitative expert evaluations. As result many conclusions on the output of inference procedure can obtain the equal plausibility evaluations in spite of having essentially different list of plausibility values of premises [3,4].

The solution of this problem was given in [2-4] by embedding the initial scale in the suitable set of strings where the result of operation $T(x,y)$ do not equal to one of the operands but to the string of operands ordered in a suitable way and called a lexicographic valuation of plausibility or uncertainty with memory. Several classes of such lexicographic valuations were introduced with strict monotonic conjunction and/or disjunction operations. These operations were realized in expert systems shell LEXICO [5].

The operations on the set of lexicographic valuations are based on manipulations with the strings. It caused some difficulties in realization and analysis of considered approach. Here we introduce a new more simple method of definition such operations on ordinal scale. This method is based on representation of results of operations by multi-sets [1, 8] which are simpler in realization and analysis than the method based on lexicographic valuations. These multi-sets may be considered as bags [9] studied in fuzzy theory. The properties of associative operations on ordinal scales are studied also in [6].

Our goal is to introduce strict monotonic conjunction T and disjunction S operations on ordinal scales satisfying the following conditions [3]:

- 1) $T(x,y) < T(x,z)$ if $y < z$ and $x \neq \mathbf{0}$ (strict monotonicity of T);
- 2) $S(x,y) < S(x,z)$ if $y < z$ and $x \neq \mathbf{1}$ (strict monotonicity of S).

Suppose $X_p = \{x_0, x_1, \dots, x_{n+1}\}$ is a linearly ordered scale of plausibility (true, possibility) values such that $x_i < x_j$ for all $i < j$. The elements x_0 and x_{n+1} will be denoted as $\mathbf{0}$ and $\mathbf{1}$ and will have interpretation as *Impossible* and *Sure* correspondingly. Denote $X = \{x_1, \dots, x_n\}$ the set of intermediate grades of plausibility. A multi-set \hat{A} over \tilde{O} is a string (a_1, a_2, \dots, a_n) , where a_i is a number of appearance of element x_i in A . If a_j equals zero then \hat{A} does not contain x_j . Denote F a set of all multi-sets over X and $A = (a_1, \dots, a_n)$, $B = (b_1, \dots, b_n)$ multi-sets from F . We will consider several ways of generation of multi-sets from elements of X corresponding to the classes of lexicographic valuations of plausibility [3].

In Section 2 and 3 we consider multi-sets corresponding to results of conjunction and disjunction operations. The corresponding classes of multi-sets will be considered as conjunctive and disjunctive classes of multi-sets and denoted as C -multi-sets and D -multi-sets. Each class will coincide with the set

of all multi-sets F but the comparison relations and operations on these classes of multi-sets will be defined by different way. We introduce conjunction and disjunction operations on these multi-sets which satisfy axioms of t -norms and t -conorms [7] but only one of these two operations will be strict monotonic. In Section 4 we introduce more complicated set of strings of multi-sets where both operations are strict monotonic. In Conclusion we discuss the obtained results and the possible applications of introduced operations.

2. Conjunctive multi-sets

Conjunctive multi-set consists of all elements from X participating in conjunction. On the set F introduce conjunction operation as follows:

$$(a_1, \dots, a_n) \wedge (b_1, \dots, b_n) = (a_1 + b_1, \dots, a_n + b_n).$$

The definition of the ordering of conjunctive (C -) multi-sets is based on the following property of conjunction operation: $A \wedge B \leq A$, i.e. an adding an elements of C -multi-set B to elements of C -multi-set A decrease the plausibility of resulting C -multi-set. An ordering relation on F is introduced as follows:

$$\begin{aligned} A = B, & \quad \text{if and only if } a_i = b_i \text{ for all } i = 1, \dots, n. \\ A < B, & \quad \text{if it exists } i, \text{ such that } a_i > b_i \text{ and } a_k = b_k \text{ for all } k < i. \end{aligned}$$

As it follows from the definition, the ordering of two C -multi-sets is defined by the minimal element where these C -multi-sets differ. C -multi-set may be considered as the result of application of conjunction operation generalizing \min operation. They are analogues of min-lexicographic valuations of plausibility considered in [3, 4].

Suppose $F_p = F \cup \{\mathbf{0}\} \cup \{\mathbf{1}\}$ is an extended set of plausibility values. The extension of ordering relation on F_p will be defined as follows: $\mathbf{0} < A < \mathbf{1}$, for all C -multi-sets A from F . In general case A and B will denote elements from F_p . Denote $A \leq B$, if $A = B$ or $A < B$. It is clear that \leq is a linear ordering on F_p i.e. it is transitive and for all A and B from F_p it is fulfilled $A \leq B$ or $B \leq A$.

Define an extension of conjunction operation on F_p as follows:

$$A \wedge \mathbf{0} = \mathbf{0} \wedge A = \mathbf{0}, \quad A \wedge \mathbf{1} = \mathbf{1} \wedge A = A.$$

Disjunction operation \vee on the set F_p is defined as follows:

$$\begin{aligned} A \vee \mathbf{0} &= \mathbf{0} \vee A = A, & A \vee \mathbf{1} &= \mathbf{1} \vee A = \mathbf{1}, \\ (a_1, \dots, a_n) \vee (b_1, \dots, b_n) &= \max\{(a_1, \dots, a_n), (b_1, \dots, b_n)\}, \end{aligned}$$

where \max is defined by the introduced linear ordering on F_p .

Theorem 1. Operations \wedge and \vee are t -norm and t -conorm [7] correspondingly on the set of C -multi-sets F_p , i.e. they satisfy conditions:

$$\begin{aligned} A \wedge \mathbf{1} &= A, & A \vee \mathbf{0} &= A, \\ A \wedge B \leq C \wedge D, \text{ if } A \leq C, B \leq D, & & A \vee B \leq C \vee D, \text{ if } A \leq C, B \leq D, \\ A \wedge B &= B \wedge A, & A \vee B &= B \vee A, \\ ((A \wedge B) \wedge C) &= (A \wedge (B \wedge C)), & ((A \vee B) \vee C) &= (A \vee (B \vee C)), \end{aligned}$$

and \wedge is a strict monotonic operation on F_p , i.e. it satisfies condition

$$A \wedge B < A \wedge D, \text{ if } B < D \text{ and } A \neq \mathbf{0}.$$

Note that disjunction operation does not strict monotonic on the set of C -multi-sets. Such operation will be introduced in the following section.

3. Disjunctive multi-sets

Disjunctive multi-set consists of all elements from X participating in disjunction. On the set F introduce disjunction operation \vee as follows:

$$(a_1, \dots, a_n) \vee (b_1, \dots, b_n) = (a_1 + b_1, \dots, a_n + b_n).$$

The definition of the ordering of disjunctive (D -) multi-sets is based on the following property of disjunction operation: $A \vee B \geq A$, i.e. an adding an elements of D -multi-set B to elements of D -multi-set A increase the plausibility of result. An ordering relation on F is introduced as follows:

$$\begin{aligned} A &= B, & \text{if and only if } a_i &= b_i \text{ for all } i = 1, \dots, n. \\ A &< B, & \text{if it exists } i, \text{ such that } a_i < b_i \text{ and } a_k &= b_k \text{ for all } k > i. \end{aligned}$$

As it follows from the definition, the ordering of two C -multi-sets is defined by the maximal element where these D -multi-sets differ. D -multi-set may be considered as the result of application of disjunction operation generalizing max operation. They are multi-set analogue of max-lexicographic valuations of plausibility considered in [3].

The extension of ordering relation on an extended set of plausibility values $F_p = F \cup \{\mathbf{0}\} \cup \{\mathbf{1}\}$ will be defined as follows: $\mathbf{0} < A < \mathbf{1}$, for all D -multi-sets A from F . In general case A and B will denote elements from F_p .

Denote $A \leq B$, if $A = B$ or $A < B$. It is clear that \leq is a linear ordering on F_p . Define an extension of disjunction operation on F_p :

$$A \vee \mathbf{0} = \mathbf{0} \vee A = A, \quad A \vee \mathbf{1} = \mathbf{1} \vee A = \mathbf{1}.$$

Conjunction operation \wedge on the set F_p is defined as follows:

$$A \wedge \mathbf{0} = \mathbf{0} \wedge A = \mathbf{0}, \quad A \wedge \mathbf{1} = \mathbf{1} \wedge A = A, \\ (a_1, \dots, a_n) \wedge (b_1, \dots, b_n) = \min\{(a_1, \dots, a_n), (b_1, \dots, b_n)\},$$

where \min is defined in correspondence with the linear ordering on F_p .

Theorem 2. Operations \wedge and \vee are t -norm and t -conorm correspondingly on the set of D -multi-sets F_p , and \vee is a strict monotonic operation on F_p , i.e. it satisfies

$$A \vee B < A \vee D, \text{ if } B < D \text{ and } A \neq \mathbf{1}.$$

Note that conjunction does not strict monotonic on the set of D -multi-sets.

4. DNF-multi-sets

In previous sections we considered multi-sets as analogues of conjunction or disjunction of elements of the scale X . But in each class of multi-sets only one of these operations is strict monotonic. Here we consider a set of ordered strings of multi-sets where both operations are strict monotonic. The approach is based on definition of max-min-lexicographic valuations [3].

DNF-multi-set is defined as an analogue of the DNF formula in Boolean logic which is a disjunction of conjunctions. As conjunctions we consider C -multi-sets and as disjunction we consider the list of C -multi-sets ordered in

descending order with the ordering relation defined on the set of C -multi-sets. DNF-multi-set may be represented generally as a column of C -multi-sets $A = [A_1, \dots, A_p]$ from F such that $A_k \geq A_{k+1}$:

$$A = \begin{bmatrix} A_1 \\ \dots \\ A_p \end{bmatrix} = \begin{bmatrix} (a_{11}, \dots, a_{1n}) \\ \dots \\ (a_{p1}, \dots, a_{pn}) \end{bmatrix}$$

We suppose that DNF-multi-set can contain repeated C -multi-sets. The set of DNF-multi-sets with C -multi-sets from F will be denoted as G . The number of C -multi-sets in DNF-multi-set $A = [A_1, \dots, A_p]$ will be called a length $|A|$ of A and we will write it also explicitly as $A(p) = [A_1, \dots, A_p]$. Suppose $A(p) = [A_1, \dots, A_p]$ and $B(q) = [B_1, \dots, B_q]$ are two DNF-multi-sets with length p and q respectively. The ordering of DNF-multi-sets from G is defined as follows:

$$A(p) \leq B(q) \quad \text{if } p \leq q \text{ and } A_i = B_i \text{ for all } i = 1, \dots, p \text{ or if } A_1 < B_1 \text{ or} \\ \text{if it exists } r > 1, r \leq \min(p, q), \text{ such that } A_r < B_r \text{ and} \\ A_i = B_i \text{ for all } i = 1, \dots, r-1.$$

This relation is based on the property of disjunction operation $A \leq A \vee B$ and determined by the maximal C -multi-set where two DNF-multi-sets differ.

As usually, we write $A = B$ if $A \leq B$ and $B \leq A$. We write $A < B$ if $A \leq B$, and $A \neq B$.

On the set G of DNF-multi-sets introduce disjunction operation \vee_G as follows. Suppose $A = [A_1, \dots, A_p]$ and $B = [B_1, \dots, B_q]$ are two DNF-multi-sets. The DNF-multi-set $C = A(p) \vee_G B(q)$ will contain all $p+q$ C -multi-sets from A and B sorted in descending order.

From the definition of disjunction operation we can write

$$A = [A_1, \dots, A_p] = [A_1] \vee_G \dots \vee_G [A_p].$$

The conjunction operation \wedge_G on the set G is defined as follows. If $A = [A_1] = [(a_{11}, \dots, a_{1n})]$ and $B = [B_1] = [(b_{11}, \dots, b_{1n})]$ then $A \wedge_G B = [A_1] \wedge_G [B_1] = [A_1 \wedge B_1] = [(a_{11} \wedge b_{11}, \dots, a_{1n} \wedge b_{1n})]$ where \wedge is the conjunction operation on F . Generally for $A = [A_1, \dots, A_p]$, $B = [B_1, \dots, B_p]$ a conjunction $A \wedge_G B$ is defined by the distributive law:

$$A \wedge_G B = ([A_1] \vee_G \dots \vee_G [A_p]) \wedge ([B_1] \vee_G \dots \vee_G [B_q]) = [A_1 \wedge B_1] \vee_G \dots \vee_G [A_1 \wedge B_q] \vee_G [A_2 \wedge B_1] \vee_G \dots \vee_G [A_2 \wedge B_q] \vee_G \dots \vee_G [A_p \wedge B_1] \vee_G \dots \vee_G [A_p \wedge B_q].$$

Finally $A \wedge_G B$ consists of $p \times q$ C -multi-sets $(a_{i1} + b_{j1}, \dots, a_{in} + b_{jn})$, $i = 1, \dots, p$, $j = 1, \dots, q$, ordered in descending order.

The extension of ordering relation on an extended set of plausibility values $G_p = G \cup \{\mathbf{0}\} \cup \{\mathbf{1}\}$ will be defined as follows: $\mathbf{0} < A < \mathbf{1}$, for all DNF -multi-set A from G . In general case A and B will denote elements from G_p .

Define an extension of disjunction operation on G_p :

$$A \vee_G \mathbf{0} = \mathbf{0} \vee_G A = A, \quad A \vee_G \mathbf{1} = \mathbf{1} \vee_G A = \mathbf{1}.$$

Conjunction operation \wedge on the set G_p is defined as follows:

$$A \wedge_G \mathbf{0} = \mathbf{0} \wedge_G A = \mathbf{0}, \quad A \wedge_G \mathbf{1} = \mathbf{1} \wedge_G A = A.$$

Theorem 3. Operations \wedge_G and \vee_G are strict monotonic t -norm and t -conorm correspondingly on the set of DNF -multi-sets G_p .

Conclusions

Several classes of strict monotonic conjunction and disjunction operations for processing expert valuations measured in ordinal scales were introduced. These operation are t -norms and t -conorms correspondingly. The approach is based on representation of results of these operations by multi-sets or by ordered list of multi-sets. In the first two classes of multi-sets only one of conjunction or disjunction operations is strict monotonic. In the last class of ordered lists of multi-sets both operations are strict monotonic. The approach to definition of strict monotonic conjunction and disjunction operations on ordinal scales is easier from practical point of view than the similar approach for defining strict monotonic operation based on lexicographic operations [3]. As result the realization of new approach based on multi-sets in inference procedures of expert systems will be simpler than for expert systems based on lexicographic operations [5]. The concept of multi-set is related with the concept of bag studied in fuzzy set theory [9] and hence the formalisms developed in this work may be integrated in the theory of bags.

Acknowledgements

The support for this research work has been provided by the IMP, projects D.00006 “Distributed Intelligent Computing” and by RFBR grants 02-01-00092 and 03-01-96245.

References

- [1] Aho A.V., Hopcroft J.E., Ulman J.D. (1976) *The Design and Analysis of Computer Algorithms*. Massachusetts: Addison-Wesley.
- [2] Batyrshin I.Z. (1993) Uncertainties with memory in decision-making and expert systems. In *Proceedings of the Fifth IFSA World Congress'93*. Seoul, Korea, pp. 737–740.
- [3] Batyrshin I.Z. (1994) Lexicographic estimates of the likelihood with universal bounds. I. *Technical Cybernetics, Izvestija Akademii Nauk*, vol. 5, pp. 28-45 (in Russian)
- [4] Batyrshin I. (1994) Modus ponens generating function in the class of \wedge -valuations of plausibility. In *Tenth Annual Conference on Uncertainty in Artificial Intelligence*, Seattle, Washington, pp. 55-59.
- [5] Batyrshin I., Zakuanov R., Bikushev G. (1994) Expert system based on algebra of uncertainties with memory in process optimization. In: Ruan D., D'hondt P., Govaerts P., Kerre E.E. (ed) *Fuzzy Logic and Intelligent Technologies in Nuclear Science*. World Scientific, Singapore, pp. 156-159.
- [6] Fodor J. Smooth associative operations on finite ordinal scales (2000) *IEEE Transactions on Fuzzy Systems*, vol. 8, 6, 791–795.
- [7] Klement EP, Mesiar R, Pap E (2000) *Triangular Norms*. Kluwer, Dordrecht
- [8] Reingold E.M., Nievergelt J., Deo N. (1977) *Combinatorial Algorithms. Theory and Practice*. New Jersey: Prentice-Hall.
- [9] Yager R.R. (1986) On the theory of bags. *Int. J. General Systems*, vol. 13, pp. 23-37.
- [10] Zadeh L.A. (1965) Fuzzy sets. *Information Control*, vol. 8, pp. 338-353.

Belief Functions and the Imprecise Dirichlet Model

Lev V. Utkin

Department of Computer Science, St. Petersburg Forest Technical Academy,
Russia

lvu@utkin.usr.etu.spb.ru

Summary. A belief function can be viewed as a generalized probability function and the belief and plausibility measures can be regarded as lower and upper bounds for the probability of an event. From this point of view, the common approach for computing belief and plausibility measures may be unreasonable in many real applications because there are cases where the probability function that governs the random process is not exactly known. In order to overcome this difficulty, Walley's imprecise Dirichlet model is used to extend the belief and plausibility measures. The obtained results are applied to aggregation of expert judgments. Simple numerical examples illustrate the proposed approach.

1 Belief Functions

Let U be a universal set under interest, usually referred to in evidence theory as the *frame of discernment*. Suppose N observations were made of a parameter $u \in U$, each of which resulted in an imprecise (non-specific) measurement given by a set A of values. Let c_i denote the number of occurrences of the set $A_i \subseteq U$, and $\mathcal{P}(U)$ the set of all subsets of U (power set of U). A frequency function m can be defined, called *basic probability assignment*, such that [4,7]:

$$m : \mathcal{P}(U) \rightarrow [0, 1], \quad m(\emptyset) = 1, \quad \sum_{A \in \mathcal{P}(U)} m(A) = 1.$$

Note that the domain of basic probability assignment, $\mathcal{P}(U)$, is different from the domain of a probability density function, which is U . According to [4], this function can be obtained as follows:

$$m(A_i) = c_i/N. \tag{1}$$

If $m(A_i) > 0$, i.e., A_i has occurred at least once, A_i is called a *focal element*.

According to [7], the *belief* $Bel(A)$ and *plausibility* $Pl(A)$ measures of an event $A \subseteq \Omega$ can be defined as

$$Bel(A) = \sum_{A_i : A_i \subseteq A} m(A_i), \quad Pl(A) = \sum_{A_i : A_i \cap A \neq \emptyset} m(A_i). \quad (2)$$

As pointed out in [6], a belief function can formally be defined as a function satisfying axioms which can be viewed as a weakening of the Kolmogorov axioms that characterize probability functions. Therefore, it seems reasonable to understand a belief function as a generalized probability function [4] and the belief $Bel(A)$ and plausibility $Pl(A)$ measures can be regarded as lower and upper bounds for the probability of A , i.e., $Bel(A) \leq Pr(A) \leq Pl(A)$. It is obvious that definition (1) can be used when N is rather large. However, this condition may be violated in many real applications. There are cases where the probability function that governs the random process is not exactly known. This difficulty has been investigated by Smets [8]. If N is small, inferences become too precise. I try to apply the imprecise Dirichlet model [10] to solving this problem.

Example 1. There are four firms ($U = \{1, 2, 3, 4\}$) which can be considered as candidates for investing. 5 experts ($N = 5$) try to make investment decision to maximize possible income. Two experts ($c_1 = 2$) suppose that the first firm is the best candidate ($A_1 = \{1\}$), one expert ($c_2 = 1$) supposes that the first or the second firm is preferred ($A_2 = \{1, 2\}$), two experts ($c_3 = 2$) suppose that the third firm will give maximal income ($A_3 = \{3\}$). Then there hold $m(A_1) = 0.4$, $m(A_2) = 0.2$, $m(A_3) = 0.4$. Let us find lower and upper bounds for the probability that the first firm is preferred ($A = \{1\}$). By using (2), we obtain

$$Bel(A) = m(A_1) = 0.4, \quad Pl(A) = m(A_1) + m(A_2) = 0.6.$$

It is obvious that the number of experts is rather small to consider the obtained belief and plausibility functions as bounds for the probability of A .

2 A Set of Multinomial Models

Suppose $U = \{u_1, \dots, u_L\}$ for simplicity and the set $A_i \subseteq U$ contains elements from U with indices J_i , i.e., $A_i = \{u_j : j \in J_i\}$. At that, the number of elements in J_i is l_i . Let n be the number of focal elements. Then $N = \sum_{i=1}^n c_i$. Let us calculate possible numbers of occurrences of every element of U . Associate the set A_i with an oblong box of size l_i with one open side and the set U with L small empty boxes of size 1. The i -th oblong box contains c_i balls which can move inside the box and we do not know location of balls in the i -th box because its open side is behind. Then we cover small boxes by the i -th oblong box and c_i balls enter in l_i small boxes with numbers from a set J_i . We do not know exact location of balls, but we know that they are in boxes with numbers from J_i . The same procedure is repeated n times. What can we say about possible numbers of balls in the small boxes now? It is obvious that there exist different combinations of numbers of balls except the case when

$l_i = 1$ for $i = 1, \dots, n$, i.e., all sets A_i consist of one element. Suppose that the number of the possible combinations is M . Denote the k -th possible vector of balls by $\mathbf{n}^{(k)} = (n_1^{(k)}, \dots, n_L^{(k)})$, $k = 1, \dots, M$. If to assume that the sets A_i occurred independently and a ball in the i -th small box has some unknown probability π_i , then every combination of balls in small boxes produces the *standard multinomial model*. M possible combinations of balls produce M equivalent standard multinomial models. The models are equivalent in the sense that we can not choose one of them as a more preferable case. For every model, the probability of an arbitrary event $A \subseteq U$ depends on $\mathbf{n}^{(k)}$, that is, we can find $P(A|\mathbf{n}^{(k)})$. So far as all the models are equivalent, even by precise probabilities of all categories only lower and upper probabilities of A can be computed

$$\underline{P}(A) = \min_{k=1, \dots, M} P(A|\mathbf{n}^{(k)}), \quad \overline{P}(A) = \max_{k=1, \dots, M} P(A|\mathbf{n}^{(k)}).$$

In particular, if all sets A_i consist of single elements, that is, all oblong boxes are of size 1, then $M = 1$ and $\underline{P}(A) = P(A|\mathbf{n}^{(k)})$, $\overline{P}(A) = P(A|\mathbf{n}^{(k)})$.

The following problem is to define $\mathbf{n}^{(k)}$ and $P(A|\mathbf{n}^{(k)})$. In the case of multinomial samples, the Dirichlet distribution is the traditional choice.

Remark 1. It is worth noticing that the Dirichlet distribution should be regarded as one of the possible multinomial models that can be applied here.

Remark 2. If U is some interval of real numbers, then we can always transform this universal set to a set with finite numbers of elements. Suppose that we have to find probabilities of an event A . Denote $A_{n+1} = A$. Let $\{\mathbf{i}\} = \{(i_1, \dots, i_n, i_{n+1})\}$ be a set of all binary vectors consisting of $n + 1$ components such that $i_j \in \{0, 1\}$. For every vector \mathbf{i} , we determine the interval B_k ($k = 1, \dots, 2^{n+1}$) as follows:

$$B_k = \left(\bigcap_{j: i_j=1} A_j \right) \cap \left(\bigcap_{j: i_j=0} A_j^c \right), \quad i_j \in \mathbf{i}.$$

As a result, we obtain a set of non-intersecting intervals B_k such that $B_1 \cup \dots \cup B_{2^{n+1}} = U$. Moreover, all intervals A_i can be represented as the union of a finite number of intervals B_k . This implies that every interval B_k corresponds to an element u_k of the transformed universal set U^* with the finite number (2^{n+1}) of elements.

3 Imprecise Dirichlet Model

The *Dirichlet* (s, α) *prior distribution* for π , where $\alpha = (\alpha_1, \dots, \alpha_L)$, has probability density function [3,11]

$$p(\pi) = C(s, \alpha) \cdot \prod_{j=1}^L \pi_j^{s\alpha_j - 1},$$

where $s > 0$, $0 < \alpha_j < 1$ for $j = 1, \dots, L$, $\alpha \in S(1, L)$, and the proportionality constant C is determined by the fact that the integral of $p(\pi)$ over the simplex of possible values of π is 1. Note that α_i is the mean of π_i under the Dirichlet prior and s determines the influence of the prior distribution on posterior probabilities. $S(1, L)$ denotes the interior of the unit simplex.

Walley [10] pointed out several reasons for using a set of Dirichlet distributions to model prior ignorance about probabilities π :

- 1) Dirichlet prior distributions are mathematically tractable because they generate Dirichlet posterior distributions;
- 2) sets of Dirichlet distributions are very rich, because they produce the same inferences as their convex hulls and any prior distribution can be approximated by a finite mixture of Dirichlet distributions;
- 3) the most common Bayesian models for prior ignorance about probabilities π are Dirichlet distributions.

The *imprecise Dirichlet model* is defined by Walley [10] as the set of all Dirichlet (s, α) distributions such that $\alpha \in S(1, L)$. For this model, the *hyperparameter* s determines how quickly upper and lower probabilities of events converge as statistical data accumulate. Walley [10] defined s as a number of observations needed to reduce the imprecision (difference between upper and lower probabilities) to half its initial value. Smaller values of s produce faster convergence and stronger conclusions, whereas large values of s produce more cautious inferences. At the same time, the value of s must not depend on L or a number of observations. The detailed discussion concerning the parameter s and the imprecise Dirichlet model can be found in [2,10]. By returning to the multinomial models considered in the example with boxes and balls and assuming that probabilities of balls are governed by the Dirichlet distribution, we can write the lower $\underline{P}(A, s)$ and upper $\overline{P}(A, s)$ probabilities of an event A , whose elements have indices from a set J , as follows:

$$\underline{P}(A, s) = \min_{k=1, \dots, M} \inf_{\alpha \in S(1, L)} \frac{n^{(k)}(A) + s\alpha(A)}{N + s},$$

$$\overline{P}(A, s) = \max_{k=1, \dots, M} \sup_{\alpha \in S(1, L)} \frac{n^{(k)}(A) + s\alpha(A)}{N + s},$$

where $\alpha(A) = \sum_{j \in J} \alpha_j$, $n^{(k)}(A) = \sum_{j \in J} n_j^{(k)}$.

4 Probability Bounds Induced by Belief and Plausibility Functions

Now we have to find $n^{(k)}(A)$ and $\alpha(A)$. The lower and upper probabilities $\underline{P}(A, s)$ and $\overline{P}(A, s)$ can be rewritten as

$$\underline{P}(A, s) = \frac{\min_{k=1, \dots, M} n^{(k)}(A) + s \cdot \inf_{\alpha \in S(1, L)} \alpha(A)}{N + s}, \quad (3)$$

$$\overline{P}(A, s) = \frac{\max_{k=1, \dots, M} n^{(k)}(A) + s \cdot \sup_{\alpha \in S(1, L)} \alpha(A)}{N + s}. \quad (4)$$

Note that $\inf_{\alpha \in S(1, L)} \alpha(A)$ is achieved at $\alpha(A) = 0$ and $\sup_{\alpha \in S(1, L)} \alpha(A)$ is achieved at $\alpha(A) = 1$ except a case when $A = U$. If $A = U$, then $\alpha(A) = 1$ for the minimum and maximum. In order to find the minimum and maximum of $n^{(k)}(A)$ we consider three sets A_1, A_2, A_3 such that $A_1 \subseteq A, A_2 \cap A = \emptyset, A_3 \cap A \neq \emptyset$ and $A_3 \subsetneq A$. Numbers of their occurrences are c_1, c_2, c_3 , respectively. It is obvious that all balls (c_1) corresponding to the set A_1 belong to the set A and $n^{(k)}(A)$ can not be less than c_1 . On the other hand, all balls (c_2) corresponding to the set A_2 do not belong to A . This implies that $n^{(k)}(A)$ can not be greater than $N - c_2$. A part of balls corresponding to A_3 may belong to A , but it is not necessary. Therefore, $\min n^{(k)}(A) = c_1$ and $\max n^{(k)}(A) = N - c_2$. By extending this reasoning on an arbitrary set of A_i , we get the minimal, denoted $L_1(A)$, and maximal, denoted $L_2(A)$, values of $n^{(k)}(A)$:

$$L_1(A) = \min_{k=1, \dots, M} n^{(k)}(A) = \sum_{i: A_i \subseteq A} c_i,$$

$$L_2(A) = \max_{k=1, \dots, M} n^{(k)}(A) = N - \sum_{i: A_i \cap A = \emptyset} c_i = \sum_{i: A_i \cap A \neq \emptyset} c_i.$$

It can be seen from (1) that

$$L_1(A)/N = \sum_{i: A_i \subseteq A} c_i/N = \sum_{A_i: A_i \subseteq A} m(A_i),$$

$$L_2(A)/N = \sum_{i: A_i \cap A \neq \emptyset} c_i/N = \sum_{A_i: A_i \cap A \neq \emptyset} m(A_i).$$

This implies that $L_1(A)/N$ and $L_2(A)/N$ (see (2)) are none other than the belief $Bel(A)$ and plausibility $Pl(A)$ measures of the set $A \subseteq U$. Then there hold

$$\underline{P}(A, s) = \frac{N \cdot Bel(A)}{N + s}, \quad \overline{P}(A, s) = \frac{N \cdot Pl(A) + s}{N + s}. \quad (5)$$

It is obvious from the above expressions that $\underline{P}(A, s) \leq Bel(A)$ and $\overline{P}(A, s) \geq Pl(A)$, i.e., $\underline{P}(A, s)$ and $\overline{P}(A, s)$ are extensions of the belief and plausibility measures of A taking into account the fact that N is restricted.

Example 2. Let us return to Example 1. By assuming $s = 1$, we obtain new bounds for the probability of the event $A = \{1\}$:

$$\underline{P}(A, 1) = 5 \cdot 0.4 / (5 + 1) = 0.33, \quad \overline{P}(A, 1) = (5 \cdot 0.6 + 1) / (5 + 1) = 0.67.$$

If $s = 2$, then $\underline{P}(A, 2) = 0.286, \overline{P}(A, 2) = 0.714$.

Let us consider some special cases and properties of the proposed extension of belief and plausibility measures.

1) In the case of total ignorance about elements of U , i.e., before making any

observations, it can be stated $c_i = N = 0$. Then for any $s \geq 0$, there hold $\underline{P}(A, s) = Bel(A) = 0$, $\overline{P}(A, s) = Pl(A) = 1$.

2) In the case $N \rightarrow \infty$, i.e., basic probability assignment coincides with probability in its classical sense, it can be stated for any s : $\underline{P}(A, s) = Bel(A)$, $Pl(A) = \overline{P}(A, s)$. Moreover, if $N \rightarrow \infty$, all the sets A_i are singletons and $A_1 \cup \dots \cup A_L = U$, then $\underline{P}(A, s) = Bel(A) = Pl(A) = \overline{P}(A, s)$.

3) If $s = 0$, then $\underline{P}(A, 0) = Bel(A)$, $\overline{P}(A, 0) = Pl(A)$.

4) If $s_1 \leq s_2$, then $[\underline{P}(A, s_1), \overline{P}(A, s_1)] \subseteq [\underline{P}(A, s_2), \overline{P}(A, s_2)]$.

5) If A^c is the complement of A , then $\overline{P}(A, s) = 1 - \underline{P}(A^c, s)$, $\underline{P}(A, s) = 1 - \overline{P}(A^c, s)$. The first equality follows from

$$\underline{P}(A^c, s) = \frac{N \cdot Bel(A^c)}{N + s} = \frac{N \cdot (1 - Pl(A))}{N + s} = 1 - \overline{P}(A, s).$$

The second equality is similarly proved.

6) $\underline{P}(A, s)$ and $\overline{P}(A, s)$ satisfy the coherence axioms [9].

5 Aggregation of Expert Judgments

Suppose that U is the real line restricted by some values $\inf U$ and $\sup U$. Then we can define lower and upper cumulative probability distribution functions of a random variable X , about which we have data in the form of intervals A_i , $i = 1, \dots, n$. By using (5) and taking into account the fact that $\alpha(A) = 1$ in (3) by $A = U$, we get

$$\underline{F}(t, s) = \underline{P}(\{u \leq t\}, s) = \begin{cases} \frac{N}{N+s} \sum_{A_i: \sup A_i \leq t} m(A_i), & t < \sup U \\ 1, & t = \sup U \end{cases},$$

$$\overline{F}(t, s) = \overline{P}(\{u \leq t\}, s) = \begin{cases} \frac{1}{N+s} \left(s + N \sum_{A_i: \inf A_i \leq t} m(A_i) \right), & t > \inf U \\ 0, & t = \inf U \end{cases}.$$

These functions are an envelope of all the possible cumulative distribution functions compatible with the data and they allow us to calculate the lower and upper expectations of X as follows:

$$\begin{aligned} \underline{\mathbb{E}}_s X &= \frac{s}{N+s} \int_U u \delta(u - \inf U) du + \frac{N}{N+s} \sum_{i=1}^n \int_U u \cdot m(A_i) \delta(u - \inf A_i) du \\ &= \frac{s}{N+s} \inf U + \frac{N}{N+s} \sum_{i=1}^n (\inf A_i) m(A_i), \\ \overline{\mathbb{E}}_s X &= \frac{s}{N+s} \sup U + \frac{N}{N+s} \sum_{i=1}^n (\sup A_i) m(A_i). \end{aligned}$$

Here $\delta(u - a)$ is the Dirac function which has unit area concentrated in the immediate vicinity of point a . It is worth noticing that the above expressions

can be viewed as a tool for aggregation of unreliable expert judgments [5]. Indeed, if experts provide judgments in the form of intervals, then the most common approach for their aggregation is to find some “average” judgments. At that, there exist a lot of averaging operators with different features of their application. Since judgments of reliable experts should be more important than those of unreliable ones, then most of the averaging operators use different types of weights characterizing the quality of experts. However, this way meets several difficulties. First, the behavior of experts is unstable, i.e., exact judgments related to some economical problem and elicited from an expert do not mean that this expert will provide results of the same quality for new problems and situations. Second, when experts provide imprecise values of a quantity, the weighted rules can lead to controversial results. For instance, if an expert with a small weight, say 0.1, provides a very large interval, say $[0, 10]$, for a quantity (covering its sample space), it is obvious that this expert is too cautious and the interval he supplies is non-informative, although this interval covers a true value of the quantity. On the other hand, if an expert with a large weight, say 0.9, supplies a very narrow interval, say $[5, 5.01]$, the probability that true value of the quantity lies in this interval is rather small. We can see that the values of weights contradict with the probabilities of provided intervals. It should be noted that sometimes we do not know anything about quality of experts or assignment of weights meets some ethical difficulties. This implies that weights of experts as measures of their quality can not be measures of the quality of provided opinions. The standard belief and plausibility functions also can not measure the expert quality in many cases because a number of experts may be small. Therefore, the extended belief and plausibility functions is a way to overcome at least a part of the mentioned difficulties. In sum, the lower and upper expectations given above can be applied to situations when (i) there is no information about the quality of experts and the experts may be unreliable; (ii) judgments elicited from experts are imprecise, i.e., they are represented in the form of intervals; (iii) a number of experts or expert judgments is rather small.

Example 3. Suppose that there are six expert judgments ($N = 6$) about possible values of a random quantity X , say, shares of a firm in conventional units next day, which can be changed in the interval $U = [0, 10]$. Three experts ($c_1 = 3$) provide the interval $A_1 = [4, 5]$, two experts ($c_2 = 2$) give $A_2 = [2, 4]$, and one expert ($c_3 = 1$) gives $A_3 = [1, 5]$. Then bounds for the aggregated interval by $s = 1$ are

$$\begin{aligned}\underline{\mathbb{E}}_1 X &= \frac{1}{6+1} \cdot 0 + \frac{6}{6+1} \left(4 \cdot \frac{3}{6} + 2 \cdot \frac{2}{6} + 1 \cdot \frac{1}{6} \right) = 2.43, \\ \overline{\mathbb{E}}_1 X &= \frac{1}{6+1} \cdot 10 + \frac{6}{6+1} \left(5 \cdot \frac{3}{6} + 4 \cdot \frac{2}{6} + 5 \cdot \frac{1}{6} \right) = 5.43.\end{aligned}$$

Note that if $s = 0$, we get the following bounds computed by means of expressions from [4]:

$$\underline{\mathbb{E}}_0 X = 4 \cdot \frac{3}{6} + 2 \cdot \frac{2}{6} + 1 \cdot \frac{1}{6} = 2.83, \quad \overline{\mathbb{E}}_0 X = 5 \cdot \frac{3}{6} + 4 \cdot \frac{2}{6} + 5 \cdot \frac{1}{6} = 4.67.$$

6 Conclusion

The approach for extending belief and plausibility measures when there is little information about elements of the frame of discernment has been proposed in the paper. Let me pointed out main virtues and shortcomings of the approach. First, the extended belief and plausibility functions can be simply updated after observing new events or obtaining new expert judgments because Dirichlet prior distributions generate Dirichlet posterior distributions. Second, the extended functions may be more realistic in many applications and possible large imprecision of results reflects insufficiency of available information. At the same time, the main open question here is how to determine the value of the parameter s because different values of s lead to different results and to large imprecision. Moreover, the second question is how to operate with expert judgments different in kind. These are directions for further work, which, in my opinion, could be solved in the framework of imprecise probability theory [9].

References

1. Bernard J-M (2002) Implicative analysis for multivariate binary data using an imprecise Dirichlet model. *Journal of Statistical Planning and Inference* 105(1):83-104
2. Coolen FPA (1997) An imprecise Dirichlet model for Bayesian analysis of failure data including right-censored observations. *Reliability Engineering and System Safety* 56:61-68
3. DeGroot M (1970) *Optimal Statistical Decisions*. McGraw-Hill, New York
4. Dempster AP (1967) Upper and lower probabilities induced by a multi-valued mapping. *Annals of Mathematical Statistics* 38:325-339
5. Dubois D and Kalfsbeek H (1990) Elicitation, assessment and pooling of expert judgement using possibility theory. In: Manikopoulos CN (ed) *Proc. of the 8th Inter. Congress of Cybernetics and Systems*, pages 360-367, Newark, NJ, New Jersey Institute of Technology Press
6. Halpern JY and Fagin R (1992) Two views of belief: Belief as generalized probability and belief as evidence. *Artificial Intelligence* 54:275-317
7. Shafer G (1976) *A Mathematical Theory of Evidence*. Princeton University Press
8. Smets P (1994) Belief induced by the knowledge of the probabilities. Technical report 94-4.1, Institut de Recherches Interdisciplinaires et de Developpements en Intelligence Artificielle, Brussels
9. Walley P (1991) *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London
10. Walley P (1996) Inferences from multinomial data: learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B* 58:3-57
11. Wilks S (1962) *Mathematical Statistics*. Wiley and Sons, New York

Fuzzy Decision Making Using the Imprecise Dirichlet Model

Lev V. Utkin¹ and Thomas Augustin²

¹ Department of Computer Science, St. Petersburg Forest Technical Academy, Russia

lvu@utkin.usr.etu.spb.ru

² Department of Statistics, Ludwig-Maximilians-University of Munich, Germany

augustin@stat.uni-muenchen.de

Summary. In most applications, probabilities of states of nature in decision making are not known exactly due to a lack of complete information. If available information is represented by a small number of statistical data, Walley's imprecise Dirichlet model may be regarded as a tool for determining interval probabilities of states of nature. It turns out that the resulting expected utilities constitute fuzzy sets and the initial problem is reduced to a fuzzy decision problem. A numerical example illustrates the proposed approach to solving decision problems under scarce information about states of nature. This approach also can be applied to construction of fuzzy sets on the basis of statistical observations.

1 Introduction

Consider the basic model of decision theory: One has to choose an *action* from a non-empty, finite set $\mathbb{A} = \{a_1, \dots, a_n\}$ of possible actions. The consequences of every action depend on the true, but unknown *state of nature* $\vartheta \in \Theta = \{\vartheta_1, \dots, \vartheta_m\}$. The corresponding outcome is evaluated by the *utility function*

$$u : (\mathbb{A} \times \Theta) \rightarrow \mathbb{R} \\ (a, \vartheta) \mapsto u(a, \vartheta),$$

and by the associated random variable $\mathbf{u}(\mathbf{a})$ on $(\Theta, \mathcal{P}o(\Theta))$ taking the values $u(a, \vartheta)$. Alternatively a loss function $l(a, \vartheta)$ is assigned, which can be embedded into the framework proposed by setting $u(a, \vartheta) = -l(a, \vartheta)$.

This model contains the essentials of every (formalized) decision situation under uncertainty and is applied in a huge variety of disciplines. If the states of nature are produced by a perfect random mechanism (e.g. an ideal lottery), and the corresponding probability measure $\pi(\cdot)$ on $((\Theta, \mathcal{P}o(\Theta)))$ is completely known, the Bernoulli principle is nearly unanimously favored. One chooses that action a_k which maximizes the expected utility $\mathbb{E}_\pi \mathbf{u}_k :=$

$\sum_{j=1}^m (u(a_k, \vartheta_j) \cdot \pi(\vartheta_j))$. For simplicity, we will denote $\pi_j = \pi(\vartheta_j)$ and $u_{rj} = u(a_r, \vartheta_j)$.

In most practical applications, however, the true state of nature can not be understood as arising from an ideal random mechanism. And even if so, the corresponding probability distribution will be not known exactly. Obviously, decision making model in this case depends on initial information about states of nature (more precisely about probability distribution of states of nature). An efficient approach for solving the decision problem under initial data in the form of lower and upper probabilities (expectations) in the framework of imprecise probability theory (Kuznetsov [5], Walley [9], Weichselberger [11]) has been proposed by Augustin [1,2]. A similar approach for solving the decision problem taking into account the second-order probabilities [7] has been discussed in [8]. However, available information may be represented in the form of a number of statistical observations. In this case, the proposed models do not work because the number of observations may be very small and it is very difficult to infer statistical characteristics describing states of nature. One of the most appropriate approaches in this case is to use the Dirichlet model. The Dirichlet model has been widely adopted to many application due to interesting properties of this model, in particular, due to the important fact that the Dirichlet density functions constitute a conjugate family of density functions. A very promising generalization of the existent Dirichlet models is Walley's imprecise Dirichlet model [10] taking into account incompleteness of statistical data. We show in this paper how to reduce the initial decision problem to a fuzzy decision problem by a small number of statistical observations. One should note explicitly that throughout the paper we assume the precise utilities are given.

2 Imprecise Dirichlet Model

The *Dirichlet* (s, α) *prior distribution* for π , where $\alpha = (\alpha_1, \dots, \alpha_m)$, has probability density function [12]

$$p(\pi) = C \cdot \prod_{j=1}^m \pi_j^{s\alpha_j - 1},$$

where $\alpha \in S(1, m)$, and the proportionality constant C is determined by the fact that the integral of p over the simplex of possible values of π is 1; α_i is the mean of π_i under the Dirichlet prior; $s > 0$ determines the influence of the prior distribution on posterior probabilities; $S(1, m)$ denotes the interior of the unit simplex.

The *imprecise Dirichlet model* is defined by Walley [10] as the set of all Dirichlet (s, α) distributions such that $\alpha \in S(1, m)$. For the imprecise Dirichlet model, the *hyperparameter* s determines how quickly upper and lower probabilities of events converge as statistical data accumulate. Smaller values of

s produce faster convergence and stronger conclusions, whereas large values of s produce more cautious inferences. The detailed discussion concerning the parameter s can be found in [10]. The application of the imprecise Dirichlet model in game theory for choosing a strategy in a two-player game has been studied by Quaeghebeur and de Cooman [6].

3 The Expected Utility by the Imprecise Dirichlet Model

Let us consider an approach for decision making under condition that π has the precise Dirichlet distribution. In this case, the expected utility $\mathbb{E}_\pi \mathbf{u}_r$ can be considered as a conditional expectation. Denote the dependency of $\mathbb{E}_\pi \mathbf{u}_r$ on the parameter s by $\mathbb{E}_\pi^{(s)} \mathbf{u}_r$. Then by using the Bayesian model, we find the unconditional expected utility $\mathbb{E}^{(s)} \mathbf{u}_r$ as

$$\mathbb{E}^{(s)} \mathbf{u}_r = \int_{\Theta} \sum_{i=1}^m u_{ri} \pi_i p(\pi) d\pi = \sum_{i=1}^m u_{ri} \int_{\Theta} \pi_i C \prod_{j=1}^m \pi_j^{s\alpha_j - 1} d\pi, \quad \Theta = [0, 1]^m.$$

The integral in the above expression is none other than the expectation $\mathbb{E}_p^{(s)} \pi_i$ of π_i with respect to $p(\pi)$. This implies that

$$\mathbb{E}^{(s)} \mathbf{u}_r = \sum_{i=1}^m u_{ri} \cdot \mathbb{E}_p^{(s)} \pi_i.$$

According to [10], there holds $\mathbb{E}_p^{(s)} \pi_i = (n_i + s\alpha_i)/(N + s)$, where n_i denotes the number of observations of the i -th state of nature in N trials. Then

$$\mathbb{E}^{(s)} \mathbf{u}_r = \sum_{i=1}^m u_{ri} \cdot \mathbb{E}_p^{(s)} \pi_i = \sum_{i=1}^m u_{ri} \cdot \frac{n_i + s\alpha_i}{N + s}.$$

By using the imprecise Dirichlet model, we can write the lower and upper expected utilities as the following optimization problems:

$$\underline{\mathbb{E}}^{(s)} \mathbf{u}_r = \inf_{\alpha \in S(1, m)} \mathbb{E}^{(s)} \mathbf{u}_r, \quad \overline{\mathbb{E}}^{(s)} \mathbf{u}_r = \sup_{\alpha \in S(1, m)} \mathbb{E}^{(s)} \mathbf{u}_r.$$

In other words, in order to find $\underline{\mathbb{E}}^{(s)} \mathbf{u}_r$ and $\overline{\mathbb{E}}^{(s)} \mathbf{u}_r$, it is necessary to solve two linear programming problems. Let us rewrite $\mathbb{E}^{(s)} \mathbf{u}_r$ as follows:

$$\mathbb{E}^{(s)} \mathbf{u}_r = \frac{1}{N + s} \sum_{i=1}^m u_{ri} n_i + \frac{s}{N + s} \sum_{i=1}^m u_{ri} \alpha_i.$$

Since extreme points of the simplex $S(1, m)$ are $(1, 0, \dots, 0)$, $(0, 1, \dots, 0), \dots, (0, 0, \dots, 1)$, then the optimal solutions can be found by substituting these points into the objective functions. As a result, we have

$$\underline{\mathbb{E}}^{(s)} \mathbf{u}_r = \frac{1}{N+s} \sum_{i=1}^m u_{ri} n_i + \frac{s}{N+s} \min_{i=1, \dots, m} u_{ri}, \tag{1}$$

$$\overline{\mathbb{E}}^{(s)} \mathbf{u}_r = \frac{1}{N+s} \sum_{i=1}^m u_{ri} n_i + \frac{s}{N+s} \max_{i=1, \dots, m} u_{ri}. \tag{2}$$

In particular, if $N = 0$, then there hold $\underline{\mathbb{E}}^{(s)} \mathbf{u}_r = \min_{i=1, \dots, m} u_{ri}$, $\overline{\mathbb{E}}^{(s)} \mathbf{u}_r = \max_{i=1, \dots, m} u_{ri}$. In this case, lower and upper expected utilities do not depend on s . If $N \rightarrow \infty$, $n_i = \pi_i N$, $i \leq m$, then $\underline{\mathbb{E}}^{(s)} \mathbf{u}_r = \overline{\mathbb{E}}^{(s)} \mathbf{u}_r = \sum_{i=1}^m u_{ri} \pi_i$.

It is worth noticing that we obtain intervals of expected utilities whose bounds $\underline{\mathbb{E}}^{(s)} \mathbf{u}_r$ and $\overline{\mathbb{E}}^{(s)} \mathbf{u}_r$ depend on the parameter s . In spite of various rules [10] for choosing s , it is difficult to assign a certain value for s in many cases. Moreover, the sense of s is not so clear for practitioners and different values of s lead to different optimal actions. Therefore, it is necessary to develop a more convenient way for decision making.

4 Interval Expected Utilities as a Fuzzy Set

Note that intervals $[\underline{\mathbb{E}}^{(s)} \mathbf{u}_r, \overline{\mathbb{E}}^{(s)} \mathbf{u}_r]$ by different values of s produce a set of nested intervals. At that, if $s \rightarrow 0$, then $\underline{\mathbb{E}}^{(0)} \mathbf{u}_r = \overline{\mathbb{E}}^{(0)} \mathbf{u}_r$. If $s \rightarrow \infty$, then $\underline{\mathbb{E}}^{(s)} \mathbf{u}_r = \min_{i=1, \dots, m} u_{ri}$, $\overline{\mathbb{E}}^{(s)} \mathbf{u}_r = \max_{i=1, \dots, m} u_{ri}$. Let us introduce a function $\mu(s)$ with the following properties: $\mu(0) = 1$, $\mu(\infty) = 0$, and $\mu(s)$ is non-increasing with s . Examples of such functions are $\mu(s) = \exp(-s)$, $\mu(s) = (1+s)^{-1}$. Thus, we have a set of nested intervals characterized by real numbers $\mu \in [0, 1]$. This set can be viewed as a fuzzy set $\tilde{\mathbb{E}} \mathbf{u}_r$ of the expected utility with the membership function μ . Now we can write a new criterion of decision making. An action a_k is optimal iff for all $r = 1, \dots, n$, $\tilde{\mathbb{E}} \mathbf{u}_k \geq \tilde{\mathbb{E}} \mathbf{u}_r$.

It can be seen from the above reasoning that we get the fuzzy decision problem. Now another question arises. How to order the fuzzy sets? This is one of the most controversial matters in fuzzy literature. Most ranking methods are based on transforming a fuzzy set into a real number called by the *ranking index*. Here we use the index proposed by Campos and Conzalez [3], which can be written in terms of the considered decision problem as

$$F_r = \int_0^1 \left(\eta \underline{\mathbb{E}}^{(\mu)} \mathbf{u}_r + (1-\eta) \overline{\mathbb{E}}^{(\mu)} \mathbf{u}_r \right) d\mu. \tag{3}$$

Here $\eta \in [0, 1]$ is a parameter of pessimism. Then we write the following criterion of decision making. An action a_k is optimal iff for all $r = 1, \dots, n$, $F_k \geq F_r$. It is worth noticing that the index F_r can be interpreted in another way. The branches of the membership function produce a set of probability distributions [4] with lower $\underline{H}(x)$ and upper $\overline{H}(x)$ survival functions of a random variable X determined as

$$\underline{H}(x) = \begin{cases} 1 - \mu, & x = \underline{\mathbb{E}}^{(\mu)} \mathbf{u}_r \\ 0, & x = \overline{\mathbb{E}}^{(\mu)} \mathbf{u}_r \end{cases}, \quad \overline{H}(x) = \begin{cases} 1, & x = \underline{\mathbb{E}}^{(\mu)} \mathbf{u}_r \\ \mu, & x = \overline{\mathbb{E}}^{(\mu)} \mathbf{u}_r \end{cases}.$$

For the given lower and upper survival functions, lower $\underline{\mathbb{E}}X$ and upper $\overline{\mathbb{E}}X$ expectations can be found as follows:

$$\underline{\mathbb{E}}X = \int_{-\infty}^{\infty} \underline{H}(x) dx, \quad \overline{\mathbb{E}}X = \int_{-\infty}^{\infty} \overline{H}(x) dx.$$

But these expectations can be represented as the following integrals:

$$\underline{\mathbb{E}}X = \int_0^1 \underline{\mathbb{E}}^{(\mu)} \mathbf{u}_r d\mu, \quad \overline{\mathbb{E}}X = \int_0^1 \overline{\mathbb{E}}^{(\mu)} \mathbf{u}_r d\mu.$$

This implies that $F_r = \eta \underline{\mathbb{E}}X + (1 - \eta) \overline{\mathbb{E}}X$. From this point of view, the parameter η can be regarded as a caution parameter [1] and the decision problem can be regarded as a decision problem with imprecise information about states of nature [2].

Let us try to simplify (3). Denote $\Psi(s) = \eta \underline{\mathbb{E}}^{(s)} \mathbf{u}_r + (1 - \eta) \overline{\mathbb{E}}^{(s)} \mathbf{u}_r$. Then

$$F_r = - \int_0^{\infty} \Psi(s) d\mu(s) = \Psi(0) + \int_0^{\infty} \mu(s) d\Psi(s).$$

By substituting (1)-(2) into the above expression and simplifying it, we get

$$F_r = A_r + ND_r \int_0^{\infty} \mu(s)(N + s)^{-2} ds, \quad (4)$$

where $D_r = -A_r + G_r$,

$$A_r = N^{-1} \sum_{i=1}^m u_{ri} n_i, \quad G_r = \eta \min_{i=1, \dots, m} u_{ri} + (1 - \eta) \max_{i=1, \dots, m} u_{ri}.$$

Here A_r is the expected utility of the r -th action under the given observations n_1, \dots, n_m ; G_r is the expected utility in accordance with Hurwicz criterion with optimism parameter $1 - \eta$; D_r can be regarded as some correction for G_r taking into account available statistical data. By using the mean value of an integral, we rewrite (4) as

$$F_r = A_r + ND_r \mu(\gamma) \int_0^{\infty} (N + s)^{-2} ds = A_r + D_r \mu(\gamma), \quad \gamma \in [0, \infty).$$

Hence $F_r = A_r \cdot (1 - \mu(\gamma)) + G_r \cdot \mu(\gamma)$. In other words, index F_r is a convex combination of the Hurwicz criterion and the expected utility.

Let us take the function $\mu(s) = (1 + s)^{-1}$. Then $s = \mu^{-1} - 1$. After substituting this function into (4) and integrating, we obtain

Table 1. Values of the utility function u_{rj}

	S1	S2	S3	S4
1	30	19	14	10
2	21	18	16	21
3	5	14	20	25

$$F_r = \frac{-\ln N + N - 1}{(N - 1)^2} \cdot A_r N + \frac{-N + 1 + N \ln N}{(N - 1)^2} \cdot G_r.$$

In sum, we have a very simple expression for F_r .

Let us consider a case of complete ignorance about states of nature. Before making any observations, it can be stated $n_i = N = 0$. Then

$$\lim_{N \rightarrow 0, n_i \rightarrow 0} \frac{-\ln N + N - 1}{(N - 1)^2} \sum_{i=1}^m u_{ri} n_i = 0, \quad \lim_{N \rightarrow 0} \frac{-N + 1 + N \ln N}{(N - 1)^2} = 1.$$

This implies that $F_r = G_r$. In other words, the comparison index F_r corresponds to the Hurwicz criterion with optimism parameter $1 - \eta$.

Let us consider a case when $N \rightarrow \infty$, i.e., there is a very large sample. Since n_i/N by $N \rightarrow \infty$ is none other than the precise probability π_i of the i -th state and it is supposed that $\pi_i > 0$, then there holds $n_i \rightarrow \infty$ for all $i = 1, \dots, m$, and $n_i = \pi_i N$. Hence

$$F_r = \lim_{N \rightarrow \infty} \left(\frac{-\ln N + N - 1}{(N - 1)^2} N A_r + \frac{-N + 1 + N \ln N}{(N - 1)^2} G_r \right) = \sum_{i=1}^m u_{ri} \pi_i.$$

5 Numerical Example

Let us consider a problem related to a number of repairmen for repairing equipment. There are available 40 units of equipment which can fail. Four states of nature are defined by the number of failed units during a certain time period: (S1) 0, . . . , 10 units fail; (S2) 11, . . . , 20 units fail; (S3) 21, . . . , 30 units fail; (S4) 31, . . . , 40 units fail. The maximal number of repairmen is 3. The utility function is given in Table 1. Let us study different cases of N . At that, numbers $n_i, i = 1, \dots, 4$, are generated by the random-number generator in accordance with the precise distribution $\pi: \pi_1 = 0.22, \pi_2 = 0.43, \pi_3 = 0.29, \pi_4 = 0.06$. We also use this distribution to consider the case $N \rightarrow \infty$. Membership functions of fuzzy sets $\tilde{\mathbb{E}}\mathbf{u}_1$ (curve 1), $\tilde{\mathbb{E}}\mathbf{u}_2$ (curve 2), and $\tilde{\mathbb{E}}\mathbf{u}_3$ (curve 3) are shown in Fig.1 (a) ($N = 3$), (b) ($N = 5$), (c) ($N = 10$), (d) ($N = 30$). It can be seen from the pictures that the fuzziness of $\tilde{\mathbb{E}}\mathbf{u}_1, \tilde{\mathbb{E}}\mathbf{u}_2, \tilde{\mathbb{E}}\mathbf{u}_3$ decreases as the number N increases.

Table 2 shows how the ranking indices F_r depend on the caution parameter η . It is obvious that F_r does not depend on η and the optimal action is a_1

Table 2. Fuzzy sets $\tilde{\mathbf{E}}\mathbf{u}_1, \tilde{\mathbf{E}}\mathbf{u}_2, \tilde{\mathbf{E}}\mathbf{u}_3$ and ranking indices F_1, F_2, F_3 as functions of η by various N

N	(n_1, n_2, n_3, n_4)	$\tilde{\mathbf{E}}\mathbf{u}_i, i = 1, 2, 3$	F_1	F_2	F_3
0	(0, 0, 0, 0)		$30 - 20\eta$	$21 - 5\eta$	$25 - 20\eta$
3	(0, 2, 1, 0)	Fig.1 a)	$21.4 - 6.5\eta$	$18.5 - 1.6\eta$	$18.9 - 6.5\eta$
5	(2, 2, 1, 0)	Fig.1 b)	$24.3 - 5.1\eta$	$19.3 - 1.3\eta$	$15.0 - 5.1\eta$
10	(2, 5, 2, 1)	Fig.1 c)	$21.2 - 3.5\eta$	$18.9 - 0.9\eta$	$16.3 - 3.5\eta$
30	(6, 14, 9, 1)	Fig.1 d)	$20.3 - 1.7\eta$	$18.3 - 0.4\eta$	$15.3 - 1.7\eta$
∞			19.43	18.26	14.42

Table 3. Optimal actions by different values of η and N

N	η				
	0	0.25	0.5	0.75	1
0	a_1	a_1	a_1	a_2	a_3
3	a_1	a_1	a_1	a_2	a_2
5	a_1	a_1	a_1	a_1	a_1
10	a_1	a_1	a_1	a_1	a_2
30	a_1	a_1	a_1	a_1	a_1
∞	a_1	a_1	a_1	a_1	a_1

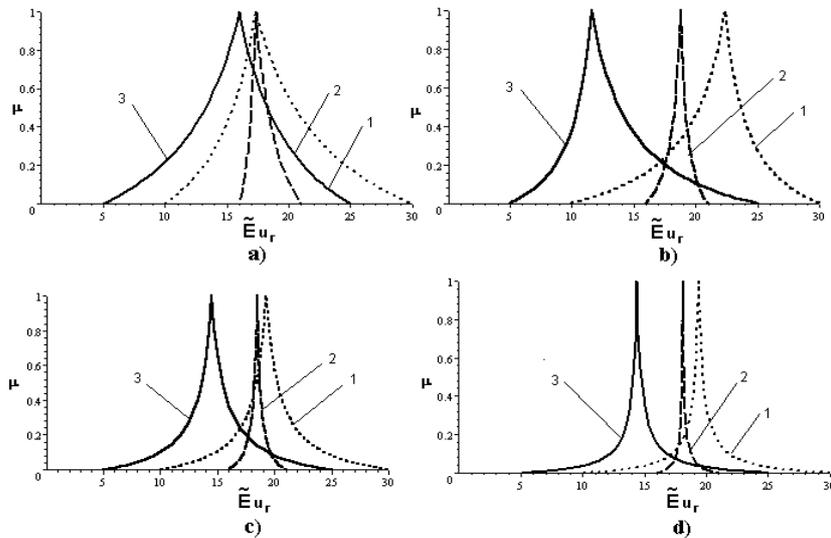


Fig. 1. Fuzzy sets $\tilde{\mathbf{E}}\mathbf{u}_1$ (curve 1), $\tilde{\mathbf{E}}\mathbf{u}_2$ (curve 2), and $\tilde{\mathbf{E}}\mathbf{u}_3$ (curve 3) by various N

in the case of complete information ($N \rightarrow \infty$). Optimal actions by different values of η and N are shown in Table 3.

6 Conclusion

The simplest decision problem has been studied in the paper. However, the fundamental ideas of this paper should be also applicable to more complex decision problems, like multi-criteria decision making, decision problems with combined objective and subjective initial information about states of nature, and decision problems with imprecise utilities. Another topic of further research is to extend the results obtained here to other optimality criteria, for example, Walley's maximality criterion. The problem can be also extended to a case of the randomized strategy instead of pure one considered in the paper. The proposed approach also can be applied to construction of fuzzy sets on the basis of statistical observations.

References

1. Augustin T (2001) On decision making under ambiguous prior and sampling information. In: de Cooman G, Fine TL, Seidenfeld T (eds) *Imprecise Probabilities and Their Applications*. Proc. of the 2nd Int. Symposium ISIPTA'01, pages 9-16, Ithaca, USA, June
2. Augustin T (2002) Expected utility within a generalized concept of probability - a comprehensive framework for decision making under ambiguity. *Statistical Papers*, 43:5-22
3. Campos LM and Gonzalez A (1989) A subjective approach for ranking fuzzy numbers. *Fuzzy Sets and Systems*, 29:145-153
4. Dubois D and Prade H (1988) *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Plenum Press, New York
5. Kuznetsov VP (1991) *Interval Statistical Models*. Radio and Communication, Moscow. in Russian
6. Quaeghebeur E and de Cooman G (2003) Game-theoretic learning using the imprecise Dirichlet model. In: Bernard J-M, Seidenfeld T, Zaffalon M (eds) *Proc. of the 3rd Int. Symposium on Imprecise Probabilities and Their Applications, ISIPTA'03*, pages 450-464, Lugano, Switzerland, July. Carleton Scientific
7. Utkin LV (2003) Imprecise second-order hierarchical uncertainty model. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 11(3):301-317
8. Utkin LV and Augustin T (2003) Decision making with imprecise second-order probabilities. In: Bernard J-M, Seidenfeld T, and Zaffalon M (eds) *Proc. of the 3rd Int. Symposium on Imprecise Probabilities and Their Applications, ISIPTA'03*, pages 545-559, Lugano, Switzerland, July. Carleton Scientific
9. Walley P (1991) *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London
10. Walley P (1996) Inferences from multinomial data: learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B*, 58:3-57
11. Weichselberger K (2001) *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung, vol. I Intervallwahrscheinlichkeit als umfassendes Konzept*. Physika, Heidelberg
12. Wilks S (1962) *Mathematical Statistics*. Wiley and Sons, New York

A Fuzzy Expert System for Predicting the Effect of Socio-economic Status on Noise-Induced Annoyance

Zaheeruddin¹ and V.K Jain²

¹ Department of Electrical Engineering, Faculty of Engineering and Technology,
Jamia Millia Islamia (A Central University),
New Delhi-110025, India
zaheer_2k@hotmail.com

² School of Environmental Sciences, Jawaharlal Nehru University,
New Mehrauli Road, New Delhi-110067, India
vkj0400@mail-jnu.ac.in

Abstract. This paper proposes a novel approach for modelling noise-induced annoyance using fuzzy technique. The fuzzy approach offers a convenient way of representing the relationships between inputs and the outputs of a system in the form of simple IF-THEN rules. The annoyance in the present model is considered as a function of noise level, its duration of occurrence, and socio-economic status of a person. The model has been implemented on Fuzzy Logic Toolbox of MATLAB using both Mamdani and Sugeno inference techniques. The results of the model are found to be in good agreement with the findings of World Health Organization (WHO) on noise-induced annoyance.

1 Introduction

The prominent adverse effects of noise pollution on human beings include loss of hearing, sleep disturbance, speech interference, reduction in work efficiency, and annoyance. Among these annoyance has been the most common term used to describe the community's collective response to noise [1,2]. It should be clear, however, that annoyance is not merely a slight irritation, but is a manifestation of a significant degradation in the quality of life. The established definition by the World Health Organization (WHO) says that health is "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" [3]. Hence, according to this definition, annoyance is basically a health problem and may be defined as a "feeling of displeasure associated with any agent or condition known or believed by an individual or a group to be adversely affecting them" [3].

A number of studies have been conducted to examine the noise-annoyance relationship [4-12]. Most of these studies have generally considered noise level and its duration of occurrence. However, social and economic factors, which may also have an important bearing on annoyance, have been neglected. It is, therefore, imperative that annoyance should be considered as a function of several parameters. The traditional quantitative techniques have been the most effective tools for modelling the noise-annoyance relationship [13,14]. These techniques of systems modelling have

significant limitations. In general, it is quite difficult to properly describe the behaviour of complex and non-linear systems by mathematical models. Also, all the real world's systems possess uncertainties to a large extent and there are other unknown phenomena that cannot be modelled mathematically at all [15]. The most important limitation of conventional techniques is that the solutions are found for the problems, which are very precise and certain. Consequently, they have a high computational cost. There are a number of cases in which excessive precision and certainty is not required and therefore computational cost can be reduced by incorporating some imprecision and uncertainty in the solution of the problem. In order to incorporate the imprecisions and uncertainties in complex and ill-defined systems, Zadeh [16] proposed a linguistic approach for modelling such type of system. Inspired by this idea of Zadeh, Mamdani and his associates [17] combined the ideas of rule-based systems with the use of fuzzy parameters to construct a controller, modelling the performance of a human operator.

It is in this context that an attempt has been made in this paper to examine the noise-annoyance relationships using fuzzy techniques because the parameters involved in the present study are fuzzy in nature. Few research papers in this area can be found in [18-20]. A brief introduction of fuzzy expert systems is given in Section 2. The methodology for the development of noise-annoyance fuzzy system is presented in Section 3. Finally the results are discussed in Section 4 followed by conclusions in Section 5.

2 Fuzzy Expert Systems

Fuzzy expert systems are known by various names viz., fuzzy rule-based systems, fuzzy inference systems, fuzzy knowledge-based systems, fuzzy models, and fuzzy controllers. Fig.1 shows the basic configuration of such systems which consist of (A) fuzzifier, (B) knowledge base, (C) inference engine, and (D) defuzzifier.

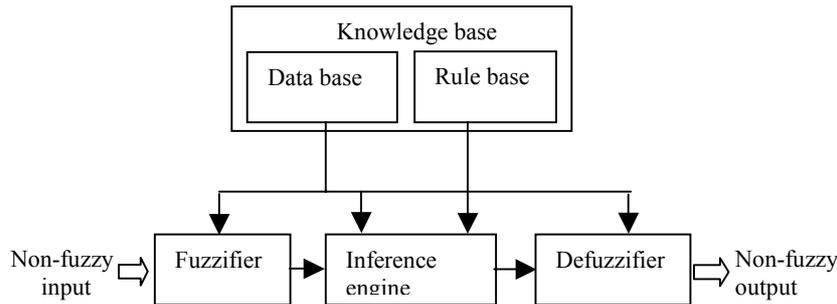


Fig. 1. Structure of a fuzzy expert system

Fuzzifier converts real numbers of input into fuzzy sets and this process is called fuzzification. Let U be a collection of objects denoted by $\{u\}$. U is called the universe of discourse and u is a generic element of U . A fuzzy set F in a universe of discourse U is characterised by a membership function μ_F in the interval $I = [0, 1]$.

The knowledge base comprises a data base and a rule base. Membership functions of the fuzzy sets are contained in the data base. The rule base is a set of linguistic statements in the form of IF-THEN rules with antecedents and consequents respectively connected by AND operator. In general, a fuzzy rule-based system with multi-inputs single-output (MISO) can be represented in the following manner [21]:

IF X_1 is B_{11} **AND** X_2 is B_{12} **AND ... AND** X_r is B_{1r} **THEN** Y_1 is D_1
ALSO

ALSO
IF X_1 is B_{i1} **AND** X_2 is B_{i2} **AND ... AND** X_r is B_{ir} **THEN** Y_s is D_s ,

where X_1, \dots, X_r are the input and Y_1, \dots, Y_s are the output variables, B_{ij} ($i = 1, \dots, m, j = 1, \dots, r$) and D_{ik} ($i = 1, \dots, m, k = 1, \dots, s$) are fuzzy subsets of the universes of discourse U_1, \dots, U_r , and V_1, \dots, V_s of X_1, \dots, X_r and Y_1, \dots, Y_s respectively.

The inference engine is the kernel of a fuzzy rule based system, which employs IF-THEN rules from the rule base to infer the output by a fuzzy reasoning method. Zadeh [22-24] proposed the concept of possibility theory as the theoretical foundation for inference mechanism to deal with the uncertainty and imprecision in rule based systems. The defuzzifier converts the fuzzy output obtained by inference engine into a non-fuzzy output real number domain and this process is called defuzzification. Though defuzzification seems to be the inverse of fuzzification but actually the maps need not be inverse of one another [25]. Among the several methods of defuzzification, the Centre of Area (COA) method is the most widely used.

3 Methodology

Noise-induced annoyance as a MISO system is shown in Fig. 2. It has three inputs and one output. Only those inputs, which affect the output to a large extent, have been selected. The most important input variables are noise level, duration of occurrence of noise, and the socio-economic status of the person. There is enough empirical evidence to suggest that a strong correlation exists between the socio-economic status of the person and noise-induced annoyance. Other factors such as gender, race, age, and motor-sidedness etc., which may influence the annoyance partially or marginally, have not been included. Inclusion of more inputs will increase the number of rules many fold and hence the complexity of the system increases.

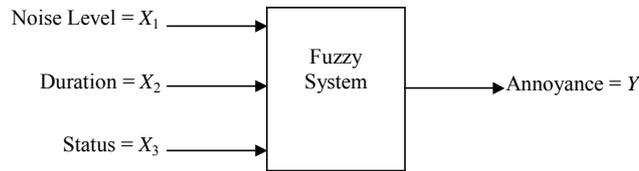


Fig. 2. Inputs and outputs of fuzzy system

The methodology for the development of the noise-annoyance fuzzy rule-based system involves the four steps: (1) Identification of the system’s variables; (2) Determination of the ranges of input and output variables; (3) Selection of the membership functions for various input and output variables; and (4) Formation of the set of linguistic rules.

The first and most important step in modelling is the identification of system’s input and output variables. The relationships between them will specify the objective of the model. The second step is to determine the ranges of the input and output variables. These variables in fuzzy modelling are defined as linguistic variables whose linguistic values are words or sentences in a natural or synthetic language. Table 1 shows the linguistic variables, their linguistic values and associated fuzzy intervals.

Table 1. Relations between inputs and outputs with their associated fuzzy values

S. No.	System’s Variables	Linguistic Variables	Linguistic Values	Fuzzy Intervals
1.	Inputs	Noise Level	Low	30-60 dB(A)
			Medium	50-80 dB(A)
			High	70-100 dB(A)
			Very High	90-120 dB(A)
2.		Duration	Very Short	0-20 Seconds
			Short	10-40 Seconds
			Medium	30-70 Seconds
			Long	60-100 Seconds
3.		Status	Very Poor	0-30 Point Scale
			Poor	20-50 Point Scale
			Middle Class	40-70 Point Scale
			Upper Class	60-90 Point Scale
	Creamy Layer		80-100 Point Scale	
4.	Output	Annoyance	Negligible	0-20 %
			Less	10-40 %
			Moderate	30-60 %
			High	50-80 %
			Extreme	70-100 %

The next step is to express linguistic values in the form of fuzzy sets, which are represented by its membership functions. The membership functions are constructed from several basic functions such as piecewise linear functions, Gaussian distribution function, sigmoid curve, quadratic and cubic polynomial curves. The triangular membership function is the simplest one and is commonly used due to its computational efficiency [26]. Membership functions for one input are shown in Fig. 3.

Finally, the relationships between inputs and output are represented in the form of IF-THEN rules. Two types of inference mechanisms are commonly used in fuzzy rule-based systems. The first was proposed by Mamdani [17]. In this case, both antecedents and consequent predicates are fuzzy and IF-THEN rules are generally provided by an expert. The second was proposed by Takagi-Sugeno-Kang (TSK) and the mod-

els employing this mechanism are commonly called TSK model [27]. TSK model assumes the inputs to be fuzzy sets and the outputs to be constants. Generally, a single spike is used to represent this constant output, also known as a singleton output membership function.

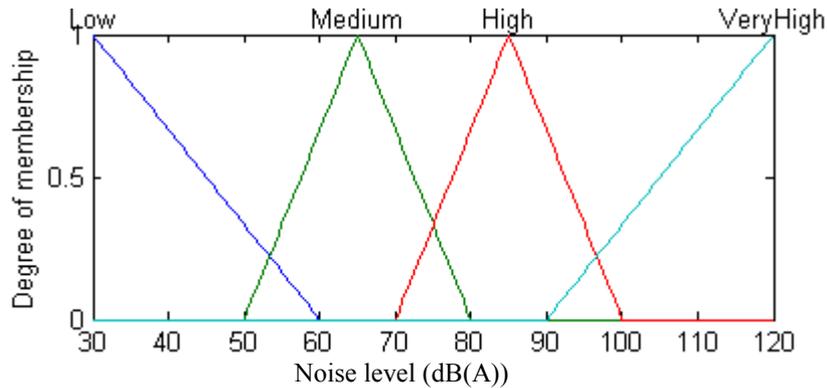


Fig. 3. Membership function for input (noise level)

4 Results and Discussion

The model for noise-induced annoyance has been implemented on Fuzzy Logic Toolbox of MATLAB® [28]. The model employs both Mamdani and Sugeno fuzzy inference methods. The results can be obtained either in 3-D or 2-D forms. Since the output is a function of three variables, hence there are three possible plots in 3-D and nine possible graphs in 2-D forms. The number of graphs in 2-D form depends upon the number of membership functions of input variables. As an example, only one result of 2-D form is depicted in Fig. 4.

Fig. 4 shows the plot of annoyance versus noise levels for long duration with socio-economic status as parameter. It is inferred from the figure that even for 'medium' noise level, the annoyance is 'high' in 'upper class' and 'creamy layer', 'moderate' in 'middle class', 'less' among 'poor' and 'negligible' among 'very poor'. At 'high' noise level, the reaction from 'upper class' and 'creamy layer' is 'extreme', 'high' from 'middle class', 'moderate' from 'poor' and 'less' from 'very poor'. At 'very high' noise level, even the 'very poor' category shows 'moderate' annoyance followed by 'high' annoyance among 'poor' sections. Not surprisingly, the other three classes, viz., 'creamy', 'upper', and 'middle' respond with 'extreme' annoyance at 'very high' noise level.

To validate our model, a comparison has been made in Table 2 between the model predictions and the findings from field surveys carried out in different parts of the world [3]. Table 2 shows the percentage of highly annoyed population as a function of low, medium, high, and very high noise levels. The population distribution based on per capita per month consumption expenditure in urban areas of India is given in Table 3 [29]. Further, the status wise distribution has been defined by us in the last row

of Table 3 using a realistic economic-criteria. An examination of Table 2 clearly confirms the adequacy of model in predicting the percentage of highly annoyed population for low, medium, high, and very high noise levels of long duration as the predicted values are well within the range of affected population.

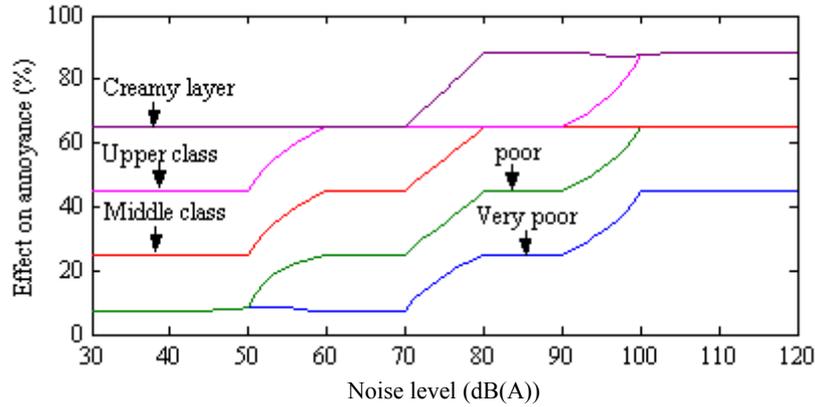


Fig. 4. Annoyance as a function of noise level for long duration with status as parameter

Table 2. Comparison between model predictions and findings of surveys

Noise Level	Population (highly annoyed)	
	Model predictions	Findings of the surveys*
Low (≤ 50 dB(A))	3.5 %	0-4 %
Medium (55-70 dB(A))	15.8 %	1-30 %
High (75-90 dB(A))	64.1 %	26-70 %
Very high (≥ 95 dB(A))	92.5 %	No survey conducted

* Source: Reference [3]

Table 3. Distribution of per capita consumption expenditure per month in a population sample of 1000 for urban India⁺

Consumption Expenditure (in Rupees)	300	350	425	500	575	665	775	915	1120	1500	1975	>1975
Population	77	62	107	113	103	105	98	89	88	82	41	35
Status	Very poor	Poor		Middle Class				Upper Class		Creamy Layer		

⁺ Source: Reference [29]

5 Conclusion

The main thrust of the present work has been to develop a fuzzy rule-based system for the prediction of noise-induced annoyance. The annoyance has been modelled as a function of noise levels, their durations of occurrence, and socio-economic status of the population. The fuzzy model has been implemented using both Mamdani and Sugeno inference techniques. The results with the proposed model are in good agreement with the findings of field surveys conducted in different parts of the world. The model clearly brings out the salient feature of the surveys concerning the variation of annoyance with noise levels while taking into account the status as one of the most important parameters.

References

1. Suter, A.H.: Noise and its effects. This is available online at <http://www.nonoise.org/library/suter/suter.html>. (1991)
2. EPA (ed.): Noise Effects Handbook. [online]. Available: <http://www.nonoise.org/library/handbook/handbook.html>. U.S. Environmental Protection Agency, Florida (1981)
3. Berglund, B., Lindvall, T., (eds.): Community Noise. [online]. Available: <http://www.nonoise.org/library/whonoise.html>. Document prepared for WHO, Stockholm, Sweden (1995)
4. Miedema, H.M.E.: Annoyance caused by two noise sources. *J. Sound and Vibration*. 98 (1985) 592-595
5. Fidell, S., Barber, D.S. Barber, Schultz, T.J.: Updating a dosage-effect relationship for the prevalence of annoyance due to general transportation noise. *J. Acoust. Soc. Am.* 89 (1991) 221-223
6. Miedema, H.M.E., Vos, H.: Exposure response relationship for transportation noise. *J. Acoust. Soc. Am.* 104 (1998) 3432-3445
7. Buchta, E., Vos, J.: A field survey on the annoyance caused by sounds from large firearms and road traffic. *J. Acoust. Soc. Am.* 104 (1998) 2890-2902
8. Green, D.M., Fidell, S.: Variability in the criterion for reporting annoyance in community noise surveys. *J. Acoust. Soc. Am.* 89 (1991) 234-243
9. Hall, F., Birnie, S., Taylor, S.M., Palmer, J.: Direct comparison of community response to road traffic noise and to aircraft noise. *J. Acoust. Soc. Am.* 70 (1981) 1690-1698
10. Fields, J.M.: A quantitative summary of non-acoustical variables effects on reactions to environmental noise. In: Busch-Vishniac, I. (ed.): National Conference on Noise Control Engineering: Reducing the Annoyance of Noise, University of Texas (1990)
11. U. S. Environmental Protection Agency. An Assessment of Noise Concern in Other Nations. Rep. No. NTID 300.6, Vol. 1 (1971)
12. Denver Department of Environmental Health: Denver Noise Survey. [online]. Available: <http://www.nonoise.org/library/denver/survey.html>. (1997)
13. Arana, M., Garcia, A.: A social survey on the effects of environmental noise on the residents of Pamplona, Spain. *Applied Acoustics* 53 (1998) 245-253
14. Pandya, G.H.: Urban noise—A need for acoustic planning. *Environmental Monitoring and Assessment*. 67 (2001) 379-388

15. Emami, M.R., Turksen, I.B., Goldenberg, A.A.: Development of a Systematic Methodology of Fuzzy Logic Modelling. *IEEE Transactions on Fuzzy Systems* 6(1998) 346-361
16. Zadeh, L.A.: Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Syst., Man, and Cyber. SMC-3* (1973) 28-44
17. Mamdani, E.H., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. *Intern. J. of Man-Machine Studies* 7 (1975) 1-13
18. Zaheeruddin, Singh, G.V., Jain, V.K.: Fuzzy Modelling of Human Work Efficiency in Noisy Environment. *Proc. The IEEE Internat. Conf. Fuzzy Systems, Vol. 1* (2003) 120-124
19. Zaheeruddin, Jain, V.K.: A Fuzzy Approach for Modelling the Effects of Noise Pollution on Human Performance. *Internat. Journal of Advance Computational Intelligence and Intelligent Informatics*. 8(2004) to appear
20. Zaheeruddin, Jain V.K.: An Intelligent System for Noise-induced Hearing Loss. *Proc. of Intern. Conf. Intelligent Sensing and Information Processing (ICISIP – 2004), Chennai, India, January 4-7, (2004)* 379-384
21. Yager, R.R., Filev, D.P.: *Essentials of Fuzzy Modeling and Control*. John Wiley & Sons, Inc. New York (1994)
22. Zadeh, L.A.: Fuzzy logic and approximate reasoning. *Synthese*. 30 (1975) 407-428
23. Zadeh, L.A.: The concept of linguistic variable and its application to approximate reasoning. *Information Sciences*. 1,2,3 (1975) 199-249, 301-357, 43-80
24. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*. 1 (1978) 3-28
25. Lygeros, J.: A formal approach to fuzzy modeling. *IEEE Transactions on Fuzzy Systems*. 5 (1997) 317-327
26. Arslan, A., Kaya, M.: Determination of fuzzy logic membership functions using genetic algorithms. *Fuzzy Sets and Systems*. 118 (2001) 297-306
27. Takagi, T., Sugeno, M: Fuzzy Identification of Systems and its Applications to Modelling and Control. *IEEE Trans. Syst., Man, and Cyber*. 15 (1985) 116-132
28. *Fuzzy Logic Toolbox for use with MATLAB®*. The MathWorks Inc., USA (2000)
29. Department of Statistics (Government of India). *National Sample Survey: Level and Pattern of Consumer Expenditure in India, NSS Report No. 457, 1999-2000*

The knowledge representation model in fuzzy expert systems provided by computing with words and PRUF language

V.I. Glova, Dr.Sci., I.V. Anikin, PhD, A.S. Katasev, p.-gr.st., O.N. Pheoktistov, p.-gr.st.

Department CSIS, Kazan State Technical University named after A.N. Tupolev, e-mail: glova@kai.ru, AnikinIgor@evm.kstu-kai.ru

ABSTRACT. We considered the problem of knowledge representation and processing in fuzzy expert systems, also their application in economics. We developed the representation and processing model of expert knowledge in expert systems provided by computing with words and PRUF language. Authors introduced the architecture of PRUF language translator, which is developed as one of the expert system modules. Provided by extended Backus-Naur forms we described the basic PRUF language structures.

Introduction

The solving of many problems of optimization, diagnostics, prediction and functional business component modelling is reviewed in a new fashion at present because of rapid development of intelligent information technologies. More often we must solve these problems in conditions of fuzziness, uncertainty and incompleteness of source information. In this situation there is a critical question of using such methods of information processing, which are suitable for functioning in these conditions.

The fuzzy sets theory offers one of possible methods of linguistic uncertainty processing, qualitative conceptions, which are used by people in everyday life [1],[2]. Using the fuzzy sets theory is more efficiently in case of need of attraction to task solution process sufficiently large volume of subjective knowledge, which can describe facts, goals and heuristic devices of tasks solution. This situation arises very often during solving of financial analysis tasks, investment planning, technical analysis and stock market prediction, dynamic scheduling and business-process modelling [3],[4].

One of more perspective directions of solving these tasks is attraction of computer technology for this process in form of decision-support systems (DSS) and expert systems.

The production models of knowledge representation in expert systems

Expert system (ES) is a computer system designed for solving of difficult formalized tasks on basis of knowledge base which reflects the collective operational experience of experts in research area, and for solving process explanation of task for expert system user.

Expert system is an instrument which accelerates expert's intelligent capacities and must execute next functions [5]:

- consultant for novice and unprofessional users;
- assistant because of necessity of analysis by expert different variants of decision-making;
- expert's partner on questions, which concern to knowledge sources from adjacent areas.

Of one's own structure any ES consists of two builders [5]:

- knowledge base (KB), which stores knowledge about data domain;
- software component of access and knowledge processing, which consists of inferential mechanism, knowledge acquisition mechanism, explanation of receiving results and intelligent interface.

The central components of any ES are knowledge base and inference engine. Therefore, functioning of ES as a whole depends on formalization, representation and processing expert's knowledge about data domain.

The production models and their generalizations are sufficiently popular models of procedural knowledge representation in ES as the most convenient for creating output mechanisms [6]. The declarative knowledge about problem area in ES are better represented with help of semantic networks [7]. View the production rules as a model of knowledge formalization in ES.

In general, the production rules have next form:

IF A_1 AND/OR A_2 ... AND/OR A_n THEN B_1, B_2, B_3 ... B_m

where A_1, \dots, A_n – premises set, and B_1, \dots, B_m – conclusions set, which take place on condition that executed premises set.

The completeness and consistency of these rules is achieved at the expense of the ES adjustment on the simulation systems together with experts.

For conclusion on knowledge, which are presented in form of production rules, the resolution principle is used [8].

The fuzziness and qualitative of initial data, informal methods of tasks solution and attainable aims results in necessity of using fuzzy production rules, which are represented in form of Mamdani and Sugeno models as well as their special cases – Cykamoto and Larsen models [9],[10].

In Mamdani model the production rules have next form:

$$\begin{aligned} \text{IF } A_1 \text{ is } \tilde{A}_1 \text{ AND/OR } A_2 \text{ is } \tilde{A}_2 \text{ ... AND/OR } A_n \text{ is } \tilde{A}_n \\ \text{THEN } B_1 \text{ is } \tilde{B}_1, \dots, B_m \text{ is } \tilde{B}_m \end{aligned}$$

where $\tilde{A}_1, \dots, \tilde{A}_n, \tilde{B}_1, \dots, \tilde{B}_m$ – fuzzy restrictions on values of variables $A_1, \dots, A_n, B_1, \dots, B_m$, which are specified as fuzzy variables – values of respective linguistic variables.

In Sugeno model the production rules have next form:

$$\begin{aligned} \text{IF } A_1 \text{ is } \tilde{A}_1 \text{ AND/OR } A_2 \text{ is } \tilde{A}_2 \text{ ... AND/OR } A_n \text{ is } \tilde{A}_n \\ \text{THEN } z = f(A_1, \dots, A_n) \end{aligned}$$

where f – some function of arguments A_1, \dots, A_n .

The conclusion on Mamdani and Sugeno fuzzy models is used in fuzzy expert systems and fuzzy control systems, including in expert systems in different economics areas [11],[12],[13]. Using the fuzzy control systems allows considerably enhance stability of their functioning under existing conditions (noises, fuzziness, lack of initial information, fuzziness of attainable aims), reduce the computational complexity of obtained decision, more softly adapt for environment fluctuation.

The representation and processing of expert knowledge provided by computing with words and PRUF language

The tasks solution efficacy, which solve the expert systems, is defined by expert's level of proficiency by quality of ES finalizing on the systems of simulation as well as capability of using model for knowledge representation in knowledge base of expert system on expert's knowledge formalization.

At present the most prevailing knowledge representation models can not often satisfy all expert's requirements. His knowledge passes through the prism of those restrictions, which are in concrete model. As a result the question of building the model, which would allow an expert to formulate one's own knowledge in convenient form, on the near natural language, is very actual.

A number of preference on expert's knowledge formalization, in comparison with existent, offer the models of computing with words technology and PRUF language [14],[15]. The main their virtue is a capability of

expert knowledge representation on the near natural language as a canonical forms, in which is made processing. The main purpose of PRUF is a truth definition of natural language expressions in desired conditions.

Provided by PRUF language L.A. Zade introduces the correct language constructions, which can be represented and processed [16]. The basic constituent of these constructions are statement sets and restrictions, which are imposed upon these statements. The canonical form of writing such expert expression is represented in form:

$$X \text{ is } R$$

where X is a statement, R is a restriction, which is laid on possible values of X , is is a variable-bunch, which defines the type of restriction.

Provided by PRUF language, which is under considered by L.A. Zade, are used such constituent language structures as

1. Fuzzy and linguistic variables, permitting to formalize the fuzzy conceptions and linguistic categories, which are used by expert during statements forming. For example «FINANCIAL PERFORMANCE» = {POOR, SATISFACTORY, GOOD}.
2. Fuzzy modifiers, permitting to modify the semantics of entered restrictions by given way and generate the extended term-set of linguistic variable values from base terms set (VERY POOR FINANCIAL PERFORMANCE, NOT VERY LOW NET PROFIT RATIO and so on).
3. Fuzzyfiers, permitting to fuzzyfy crisp and fuzzy given restrictions, covered on controlled variables (MORE-LESS SATISFACTORY FINANCIAL PERFORMANCE, APPROXIMATELY ZERO NET PROFIT RATIO and so on).
4. Limiters (possible, probabilistic and so on).
5. Bunchs (AND, OR)

The base types of restrictions, which are viewed by L.A. Zade, are next:

1. Possible value (is).
2. Probabilistic restriction (isp).
3. Restriction by random set (isrs).
4. Restriction given in form fuzzy graph (isfg).
5. Restriction by linguistic probability (ispv).
6. Restriction by linguistic truth (istv).
7. Restriction of commonness (isu).

Provided by these restrictions expert can form the statements, which is referred to problem area as well as appreciate the truth degree of natural language expressions in the given situation. In practice, these expressions can join with each other controlled variables and laid as data domain description to knowledge base of expert system. Using this approach will al-

low a great extent increase the expert's possibility on forming of expert system knowledge base.

Review the examples of expert's statements, which can be formalized provided by PRUF language.

Example 1. «Fully probably that in the near future oil price will appreciably rise».

In the terms of PRUF language this expression can be written in the following way:

LV_Probability {LV_Space_of_time(O_Space_of_time) is
FV_Near_future and_p LV_Change_of_oil_price
(O_Change_of_oil_price) is M_Appreciably(FV_Rise)} isp
M_Fully(FV_Probably)

Here LV_Probability, LV_Space_of_time, LV_Change_of_oil_price – linguistic variables, FV_Probably, FV_Near_future, FV_Rise – their values (fuzzy variables), M_Appreciably, M_Fully – modifiers, O_Space_of_time, O_Change_of_oil_price – objects.

There is the hidden linguistic variable LV_Probability in this expression and her value FV_Probably (fuzzy variable).

Example 2. «If demand for commodity will be low and expenses will be small, then in the near future the price of commodity will fall».

This expression can be written in the following way:

IF LV_Demand(O_Commodity) is FV_Low
AND LV_Expenses(O_Commodity) is FV_Small
THEN LV_Period_of_price_change(O_Commodity) is FV_Near_future
AND LV_Change_of_price(O_Commodity) is FV_Fall

In this case there are four linguistic variables in expression – LV_Demand, LV_Expenses, LV_Period_of_price_change, LV_Change_of_price, four fuzzy variables – FV_Low, FV_Small, FV_Near_future, FV_Fall, as well as object O_Commodity, his characteristic we must evaluate.

The semantics of conceptions, which are used in expressions, is defined in EDF-tables in form of membership functions.

The creation of PRUF language translator for expert systems

We developed the PRUF language translator for his using in inference engine of expert systems, including in expert systems in different economics fields.

The vocabulary and syntax of more widespread language structures of PRUF is described by authors provided by extended Backus-Naur forms,

on basis of which was created the finite state models of lexical, syntactical and semantic analyzer of quasinatural language constructions.

The process of expert's expression analysis, which is written provided by PRUF language, passes series of phases, which are named as translate phases: lexical analysis of expression, syntactic analysis and semantic interpretation of expression. The functional scheme of developed PRUF language translator is on the Figure.

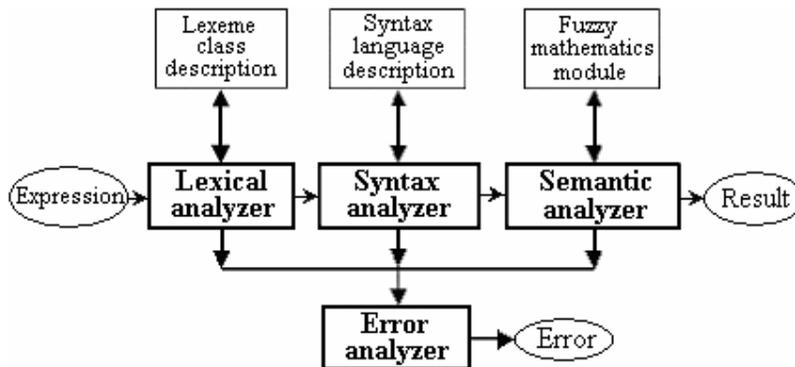


Fig. The functional scheme of PRUF language translator

The work of translator consists in semantic evaluation of entered expression, which is written in canonical form of PRUF language. This expression is entered the entrance of lexical analyzer, disassembled on independent lexical units (lexemes), as a result is formed descriptor text, which is entered the entrance of syntactic analyzer. The goal of syntactic analyzer is a syntactic correctness evaluation of entered expression, his sentence analysis and forming the input data for semantic interpretation module in form of inverse polish record. The semantic analyzer forms the result of semantic expression diagnosis with help of fuzzy mathematics module and refined EDF- frames, which are defined in expert system.

As an example consider the forming process of semantic evaluation of natural language expression, which is written in canonical form of PRUF language, in terms of developed translator.

Let have next expression:

The oil price is very high

Represent this expression in canonical form:

LV_Price (O_Oil) is M_Very FV_High

This expression is entered the entrance of lexical analyzer, which analyzes his on individual lexemes: linguistic variable LV_Price, object O_Oil, modifier M_Very and fuzzy variable FV_High as well as possible restric-

tion *is* and brackets. The descriptor tables are formed from these lexemes. Each lexeme is written as descriptor as xx xxx. First two numbers are number of descriptor table, which stores the lexeme. Last three numbers are order number of lexeme, which is stored in this table.

The set of these descriptor lexemes are descriptor text, which is entered the entrance of syntax analyzer. Then this text is transformed to more convenient form for working. In our example we have:

<lv>'(<o>')<r><m><fv>

The syntax analyzer solves such tasks as:

defines the syntactical correctness of expression on basis of well-known syntactically correct templates:

transforms this expression into inverse polish record as

<o><lv><fv><m><r>

which is entered the entrance of semantic analyzer.

The analysis of given set in semantic analyzer allows identically define his numerical semantic evaluation of truth on condition of formalization of semantic meaning for all composite elements of given construction. This formalization is realized provided by developed fuzzy mathematics library.

The translator realization represents the special-purpose library, which will be connected up to external with respect to this library expert system.

Using of this translator will allow to experts form more full description of knowledge domain on solving of difficult formalized tasks.

References

- [1] Altunin A.E., Semuhin M.V. The models and algorithms of decision-making in fuzzy conditions: Monograph. Tyumen: Publisher by Tyumen State University, 2000. – 352 p.
- [2] Frolov U.V. The intellectual systems and management decisions. - Moscow: MSPU, 2000. – 294 p.
- [3] The decision-making models in terms of linguistic variable / A.N. Borisov and others. – Riga: Zinatne, 1982.
- [4] Nedosekin A.O. The methodological foundations for modeling of financing activities with help of fuzzy-multiple descriptions. – Saint-Petersburg, 2003.
- [5] Telnov U.F. The intellectual information systems in economics: Tutorial. – M.: SYNTEG, 1999. – 216 p.
- [6] Gavrilova T.A., Horoshevsky V.F. The knowledge bases of intellectual systems. – St. Petersburg.: Piter, 2001. – 384 p.

- [7] Marenko V.A. The methods of knowledge representation in expert systems //The mathematical structures and modelling. – 2001. – №8. – P.34–39
- [8] P. Jackson. Introduction to expert systems. M: Williams, 2001.
- [9] V.V. Kruglov, M.I. Dli, R.U.Golunov. Fuzzy logic and artificial neuron networks. M.: Physmatlit, 2001.
- [10] Averkin A.N., Batyrshin I.Z., Blishun A.F. and others. Fuzzy sets in management models and artificial intelligence. // Under edit. Pospelov D.A. M.: Science, 1986. 311 p.
- [11] Dourra H., Siy P. Investment using technical analysis and fuzzy logic. // Fuzzy Sets and Systems, – № 127, 2002.
- [12] Nedovic L., Devedgic V. Expert system in finance - a cross-section of the field. // Expert Systems with Applications, – № 23, 2002.
- [13] Zopounidis C., Doumpos M., Matsatsinis M. On the use of knowledge-based decisions support systems in financial management: A survey. // Decision Support Systems, – № 20, 1997.
- [14] L.A. Zadeh. Fuzzy Logic = Computing with words // IEEE Transactions on fuzzy systems, Vol. 4, N. 2, 1996.
- [15] L.A. Zadeh. PRUF – A meaning representation language for natural languages//Int. J. Man-Mach. Studies, vol. 10, 1978.
- [16] L.A. Zadeh. Test-score semantics for natural languages and meaning representation via PRUF//Empirical Semantics, B. Reiger, Ed., W.Germany: Brockmeyer, 1981.

Construction of Incremental Fuzzy Decision Tree

Arkady Borisov* and Gulnara Bikesheva**

*Riga Technical University”, 1Meza Str., Professor, Dr.habil.sci.comp.

E-mail: aborisov@egle.cs.rtu.lv

**Joint-stock company DATI, Ganibu-dambis Str. 17a, Programmer, Master of Computer Science

E-mail: Gulnara.Bikesheva@dati.lv

Abstract

This paper considers algorithms for incremental induction of fuzzy decision trees. Incremental learning ability is very important for solving real-world problems for the following reasons. When new examples appear, one can make some revision on the trained system and incorporate new knowledge. Modifying a trained system takes much less time than building a new system. The algorithm presented combines three properties: it generates decision trees which are able to classify and generalize the data; it utilizes fuzzy logic that allows it to operate with indistinct concepts characteristic to human reasoning; and it is incremental, which enables it to use new data without building a new tree.

The paper also describes incremental algorithms for fuzzy induction of decision trees using classical logic. Then this algorithm is modified for use in indistinct logic. It has a variety of modifications. It allows working with both numeric and symbolic information, processing both inconsistent and incomplete data. An example of construction of an incremental fuzzy decision tree is considered in detail to illustrate the new algorithms potential.

Key words: fuzzy decision tree, classification, incremental learning, non-incremental learning, tree induction, fuzzy logic

1. Introduction

Decision-tree algorithms provide one of the most popular methodologies for symbolic knowledge acquisition. The popularity of decision trees is based on such advantages, as speed of construction and ease of usage at

classification. Because of the natural interpretation of knowledge, symbolic decision trees can easily be translated into a set of rules suitable for use in rule-based systems.

Today, a great number of algorithms for automatic construction of decision trees are in existence. Most of them are based on well-known algorithms ID3 and CART which allow working with both symbolic and continuous-valued attributes. In many practical applications, the data used are inherently of imprecise and subjective nature. A popular approach to capture this vagueness of information is the use of fuzzy logic, going back to Zadeh [7]. To combine the advantages of decision trees and fuzzy logic [8], the concept of a decision tree has been supplemented with fuzzy logic, and there is a set of algorithms creating such trees [6]. Unlike approaches based on classical logic, a fuzzy decision tree is able to deal with a degree of truth in classification of new examples.

To construct a fuzzy decision tree, several algorithms can be used. They differ in the search criteria employed to find the most important attribute. Say, Zeidler [1] applies traditional entropy measure borrowed from the papers of Quinlan [12]. Wang [9] describes maximum information gain that is based on linguistic information. In [10] fuzzy entropy measure introduced by Kosko [11] is used to evaluate the effectiveness of attributes. All the aforementioned algorithms suffer from one drawback: they are unable to work in the online mode.

Moreover, the present algorithms for induction of fuzzy decision trees are non-incremental. These algorithms do not allow adopting the received results to the new data. The incremental algorithms allow re-structuring the existing decision tree for incorporation of new data, which enables one to essentially reduce calculation costs.

This paper presents a new classification algorithm based on Zeidler's works. The algorithm combines three properties: it generates decision trees which have the ability to classify and generalize the data; it utilizes fuzzy logic which enables us to deal with vague concepts characteristic of human reasoning; and it is incremental which allows one to use new data, without building a new tree.

This paper consists of the following parts:

- necessary definitions from the fuzzy logic theory and decision trees;
- the concept of fuzzy decision tree based on ID3;
- description of incremental induction, based on Utgoff [3,4,5];
- an algorithm that combines advantages of fuzzy decision tree and incremental approach;
- an illustrative example of fuzzy decision tree construction and restructuring.

The paper ends with conclusions and references.

2. Fuzzy Sets and Fuzzy Linguistic Variables

Set and element are two basic notions of the set theory. In the classical set theory, an element either belongs to a certain set or it doesn't. Therefore, a set can be defined by a two-valued characteristic function $U \rightarrow \{0,1\}$, where U is the universe of discourse. In the fuzzy set theory, a fuzzy subset of the universe of discourse U is described by a membership function $\mu_v(V): U \rightarrow [0,1]$ which represents a degree by which $u \in U$ belongs to set v .

The concept of the fuzzy linguistic variable is used to describe objects and phenomena using fuzzy sets. The meaning of the linguistic term is defined by a fuzzy set, which in its turn is defined by its membership function.

An atomic fuzzy proposition (often called a fuzzy restriction) is defined as $[V \text{ is } A]$, where V is a fuzzy variable and v is a fuzzy term from its linguistic domain. The classical rule **If then** in fuzzy logic may be represented as

$$R = \text{if } V_1 \text{ is } A_1 \text{ and } \dots \text{ } V_j \text{ is } A_j \text{ then } D \text{ is } A_D$$

A fuzzy rule-based system is a set of certain fuzzy rules along with special inference procedures.

3. Decision Trees

A decision tree is a representation of rules in a hierarchical, consecutive structure where each object has its corresponding unique node giving a decision. Several algorithms have been developed in order to ensure the construction of decision trees and their use for the classification task. The *ID3* and *CART* are the two most important discriminative learning algorithms based on recursive partitioning.

An important property of these algorithms is that they attempt to minimize the size of the tree by optimizing some quality measure. *ID3* assumes discrete domains with small cardinalities. *CART* enables one to use continuous attributes. Now we will consider the *ID3* algorithm, which will later be modified to a fuzzy decision tree.

1. Compute the information content at node N given by

$$Info^N = - \sum_{i=1}^{|C|} (p_i \cdot \log p_i), \quad (1)$$

where C is the set of decisions, and p_i is the probability that a training example in the node represents class i .

2. For each attribute a_i not appearing on the path to N and for each of its domain values a_{ij} , compute the information content $Info^{N|a_{ij}}$ in a child node restricted by the additional condition $a_i = a_{ij}$.

3. Select the attribute a_i maximizing the information gain

$$Gain = Info^N - \sum_j^{|D_i|} (w_j \cdot Info^{N|a_{ij}}) \quad (2)$$

where w_j is the relative weight of examples in the child node following a_{ij} condition to the total number of examples in N , and D_i is the symbolic domain of the attribute.

4. Split node N using the selected attribute.

4. Fuzzy Decision Trees

Fuzzy decision trees differ from the traditional ID3 algorithm in the following points [1]:

1. There is a membership grade μ ($0 \leq \mu \leq 1$) for all input examples
2. The algorithm does not only create a leaf node if all data belong to the same class but also in the following cases:
 - if the proportion of a data set of class C_k is greater than or equal to an assigned threshold (prepruning of subsequent nodes because “nearly all” data belong to the same class),
 - if the number of elements in a data set is less than an assigned threshold (prepruning because of “numerical tininess” of the set) or
 - if there are no more attributes for classification (in ID3 there is a null class for this leaf node)
3. More than one class name may be assigned to the leaf node (a real advantage of a fuzzy decision tree).
4. Each attribute is considered as a linguistic variable. The membership function for each attribute is user-defined or can be calculated from the boundary points of the interval using the following algorithm:
 - Values of the attribute are divided into intervals and cut points are defined.
 - For representing membership functions we choose the trapezoidal one. Storing these functions is simple, because four corner points (A, B, C, D) are sufficient description.

- Start and end points for each range where the membership grade equals 1 (B and C)
- Start and end points for each range where the membership grade equals 0 (A and D)

For fuzzy decision trees we use $\frac{freq(C_j, T)}{|T|}$ instead of p_i in calculating the information content of the node in formulas (1) and (2), where $|T|$ is the sum of membership functions of all classes of training set T and $freq(C_j, T)$ is the sum of membership functions of class C_j of training set T .

5. Incremental Induction

Suppose we have a tree t constructed from a training set λ . Now we get some new information and build a new tree t' that is based on a new training set λ' containing λ and a couple of new training examples. In comparing the two trees, two things may occur: first, the annotated class distribution may change, and second, the structure of the tree may change.

To be able to change the test that is installed in a decision node, one needs to maintain information in that node to provide the basis for evaluating the quality of each possible test. For each possible test, the frequency counts for each outcome-class combination are kept and updated as necessary. For example, the test (color =blue) contains 5 examples from the training set: two of them belong to class 1, three – to class 2. Hence, in this node we keep two frequencies $2/5$ and $3/5$.

As the first step in reconstructing of the tree, we add a new example without changing the structure of the tree. When an example is to be incorporated into an empty tree, the tree is replaced by a “leaf” node that indicates the class of the leaf, and the example is saved in the leaf node. Whenever a new example is to be incorporated, the branches of the tree are followed as far as possible according to the values in the example. If the example belongs to the same class as the leaf, the example is simply added to the set of examples saved in the leaf node. If the example has a class label different from the leaf, the algorithm attempts to turn the leaf into a decision node, picking the best test according to the test-selection metric. The examples saved in the node that has just been converted from a leaf node into a decision node are then incorporated recursively by sending each one down its proper branch according to the new test. Thus, an example can be added to a node, and it will work its way down the tree to a leaf, possibly sprouting branches at leaves as it moves down the tree.

After adding new examples to the tree the tests in the nodes may be no longer optimal, hence the tree needs to be revised. To ensure that every decision node has the desired test installed, according to the attribute selection metric, one needs to visit the subtrees recursively. In each decision node that requires that a different test be installed we use a tree transposition operator.

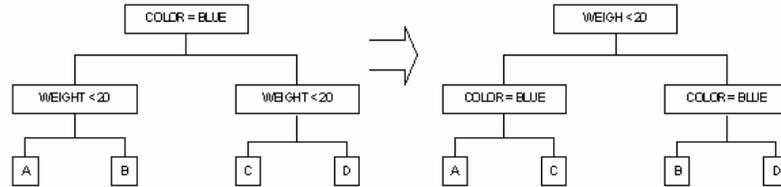


Fig. 1. Tree Transposition Operator (left on TRUE, right on FALSE)

Let us consider one of the base cases, as illustrated in Figure 1. The goal is to transpose the left-hand tree into the right-hand tree. This can be accomplished easily by rearranging the pointer because the test (Weight < 20) that is to be installed at the root is already installed at each of the children of the root. It should be noted that subtrees A, B, C, and D are simply reattached, and are not revisited for this case. Similarly, the set of examples corresponding to the root does not change either. Only the sets of the examples at the two new children of the root, now with test (Color = Blue), change.

The recursive tree transposition operator provides the ability to restructure a given tree into another that has the desired test installed at the root. To ensure this goal, we need to visit all the subtrees recursively. Any operation that changes the set of the examples on which a subtree is based marks the test as stale. This can happen either when incorporating an example or when transposing a tree. Whenever a desired test has been identified and, if necessary, installed, one removes the test's stale mark. To ensure that each decision node has the desired test installed, one proceeds recursively in the following manner: at the root, identify the desired test and install it via recursive transposition: for each subtree, if it is marked stale, recursively identify its desired test and install it.

6. Incremental Induction of Fuzzy Decision Trees

The algorithm of incremental induction of a fuzzy decision tree consists of several stages. For the first step we add some new examples without changing the structure of the tree. For the second step we check whether any structural changes in the tree are necessary and perform the

necessary changes in the tree. This procedure determines the best tests for decision nodes. We use a transposition operator for swapping nodes in a subtree so as to get the desired conditions on the top node of the subtree. Additionally we use the procedure for classification of new examples, which do not enter into the tree.

This algorithm can have several modifications depending on the criteria of estimation of quality in the decision nodes. In our case we use information Gain. The best test is one with the maximal value of information Gain. If it is necessary to add some new examples in the tree, the algorithm can restructure the tree after each example. If there are quite a few new examples, it is expedient to restructure the tree after adding all the new examples or part of them. Upon presentation of a new training example, one can decide whether or not to incorporate the training example into the tree. We can use the ability of the already existing tree to correctly classify the new example for making the decision.

7. An Example of Creating the Incremental Fuzzy Decision Tree

Our algorithm is illustrated with data from [1,2].

We have the following attributes: height; weight; hair color (blond, black, and red) and two classes C1 and C2.

μ ($0 \leq \mu \leq 1$) – is the class membership grade of an example.

Data T with μ are given in Table 1.

Table 1. Data T

number	1	2	3	4	5	6	7	8
height	16	18	17	1	16	1	16	180
weight	60	80	75	6	75	6	60	70
hair color	blond	black	black	red	Black	red	Black	blond
class	C1	C2	C2	C	C2	C	C2	C1
μ	1	0,8	0,2	0,7	1	0,3	1	0,5

We only use the first four examples for constructing a fuzzy decision tree on the basis of the fuzzy ID3 algorithm suggested.

First, the attribute values height and weight are arranged in the ascending order and the cut points are determined. In our example we have one or two possible cut points between data of different classes.

Table 2. Data sorted by Height

Cut points:	1		2	
number	1	3	4	2
height	1	1	17	18
	60	70	5	0
class	C	C	C1	C2
	1	2		
μ	1	0,	0,7	0,8
		2		

Table 3. Data sorted by weight

Cut points:	1			
number	1	4	3	2
weight	6	6	75	80
	0	0		
class	C	C	C2	C2
	1	1		
μ	1	0,	0,2	0,8
		7		

To construct the fuzzy decision tree, we do not calculate the number of examples for class 1 and class 2, we count the sum of class membership grades of the examples:

$$\text{freq}(C1, T) = 1 + 0.7 = 1.7 \quad \text{freq}(C2, T) = 0.8 + 0.2 = 1$$

$$||T|| = 1 + 0.8 + 0.2 + 0.7 = 2.7$$

$$\text{Info}(T) = -1.7/2.7 * \log_2(1.7/2.7) - 1/2.7 * \log_2(1/2.7) = 0.95$$

For the *hair_color* attribute we obtain this:

$$\text{Info}(\text{Hair_color}, T) = 0$$

$$\text{Gain}(\text{Hair_color}, T) = \text{Info}(T) - \text{Info}(\text{Hair_color}, T) = 0.95 - 0 = 0.95$$

For the height and weight attributes, we find discretization with the candidate cut points and calculate the information gains:

$$\text{Info}(\text{Height}, T, \text{Cut_point1}) = 0.843 \quad \text{Gain}(\text{Height}, T, \text{Cut_point1}) = \mathbf{0.128}$$

$$\text{Info}(\text{Height}, T, \text{Cut_point2}) = \mathbf{0.342} \quad \text{Gain}(\text{Height}, T, \text{Cut_point2}) = \mathbf{0.629}$$

$$\text{Info}(\text{Height}, T, \text{Cut_point1_2}) = \mathbf{0.289} \quad \text{Gain}(\text{Height}, T, \text{Cut_point1_2}) = \mathbf{0.682}$$

$$\text{Info}(\text{Weight}, T) = \mathbf{0} \quad \text{Gain}(\text{Weight}, T, \text{Cut_point1}) = \mathbf{0.971}$$

We select the discretization with maximum information gain among all the candidate cut points. They are attributes *Weight* and *Color*. Attribute *weight* is chosen as the testing attribute in the root node. The selected cut point is between 60 and 75. As a result of splitting we have received two subsets which are leaves of the tree. Then we added example 5 into the tree based on the algorithm described above. The structure of the tree did not change. Example 5 was simply added in the right leaf node of the tree.

After the addition of the other examples, the tree is transformed as follows (Fig. 2). The resulting tree may not be optimal. For each test the information gain should be calculated.

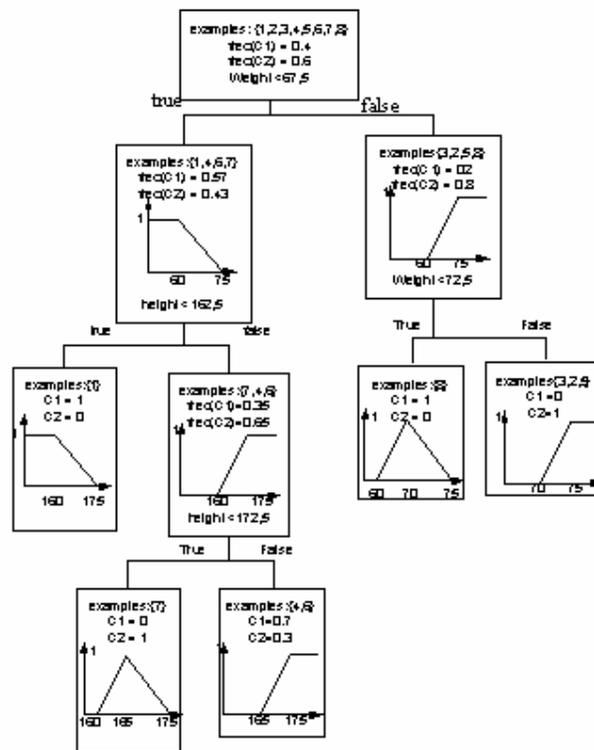


Fig. 2. Fuzzy decision tree after adding all examples

Table 4. Data sorted by height

cut points:	1	2
number	1 5 7	3 4 6 8 2

heig	1	1	1	1	1	1	1	1	1
ht	60	60	65	70	75	75	80	80	
clas	C	C	C	C	C	C	C	C	C
s	1	2	2	2	1	2	1	2	
μ	1	1	1	0	0	0	0,5	0,8	

Table 5. Data sorted by weight

cut points:	1)				2)			
number	1	4	6	7	8	3	5	2
weight	6	6	6	6	7	7	7	8
height	0	0	0	0	0	5	5	0
class	C	C	C	C	C	C	C	C
s	1	1	2	2	1	2	2	2
μ	1	0,7	0,3	1	0	0,5	0,2	1,8

Attribute *Weight* with cut point 1 has the maximal information gain if we have binary splitting. If we use two cut points, then attribute *Weight* has better information gain. Hence, if we deal with binary tree we do not need any transposition of this tree. If we do not want to be limited to a binary tree, we can transpose our tree as is shown in Fig. 3.

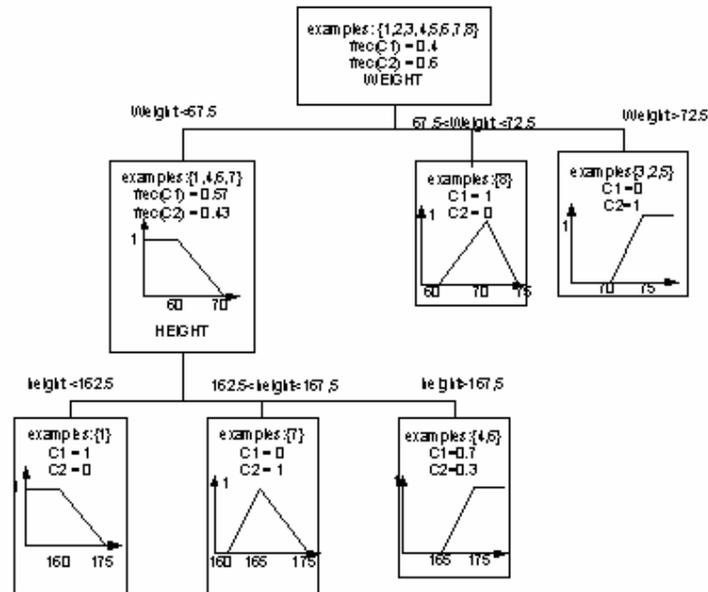


Fig. 3. Fuzzy decision tree after transposition

8. Conclusions

In this paper we have considered the algorithm for incremental induction of fuzzy decision trees. An incremental algorithm is suitable for an application that is based on a stream of examples. Employing an incremental method allows online tree updating.

An incremental fuzzy decision tree has a number of advantages as compared with classical decision trees. Due to the fuzzy approach, such a tree can deal with uncertain and inconsistent information. When using the incremental algorithm, we do not need to build a new tree each time new data appear. New examples are incorporated into the tree instead of creating a new tree, which essentially reduces calculation cost. An incremental fuzzy decision tree enables one to work with both numerical and symbolical data, and use examples with the values of attributes missing.

This paper considers an example of construction of an incremental fuzzy decision tree created on the basis of data from work [1,2]. The resulting fuzzy tree is equivalent to the one described in paper [2]. Hence, we can draw the conclusion that the incremental approach enables one to

obtain a decision tree of the same quality as if it were built from scratch. Further research will be devoted to the possibility of using the incremental approach in constructing other kinds of fuzzy decision trees described in [6].

9. References

1. J. Zeidler, Michael Schlosser. Fuzzy Handling of Continuous-Valued Attributes in Decision Trees // In: Proc. ECML-95 Mnet Familiarization Workshop "Statistics, Machine Learning and Knowledge Discovery in Databases", Heraklion, Crete, Greece, April 28-29, 1995, 41-46.
2. J. Zeidler, Michael Schlosser. Continuous-Valued Attributes in Decision Trees // In: Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'96), Granada, Spain, 1-5 July 1996, Proceedings, Volume I, S. 395-400.
3. Utgoff P. E., Berkmann N. C., Clouse J. A. Decision Tree Induction Based on Efficient Tree Restructuring // In: Machine Learning, 29, 1997, pp. 5-44.
4. Utgoff P. E. Incremental induction of decision trees // In: Machine Learning 4, 1989, p. 161-186.
5. Utgoff P.E. An improved algorithm for incremental induction of decision trees // In: Machine Learning: Proceedings of the Eleventh International Conference, 1994, p. 318-325.
6. Bikesheva G., Borisov A. Fuzzy Decision Trees: Methods and Problems // In: Proceedings of Riga Technical University, Information technology and management science, Riga 2002, Vol. 5, Issue 10, pp. 10-19.
7. Zadeh L. A. Fuzzy sets // In: Information and Control, 8, 1965, p. 407-428.
8. Borisov A., Alexeyev A. et al. *Fuzzy Information Processing in Decision Making Systems*, Moscow, Radio i Sviaz Publishers, 1989 (In Russian).
9. Wang Xizhao and Jiarong Hong. On the handling of fuzziness for continuous-valued attributes in decision tree generation. *Fuzzy Sets and Systems*, 99 (1998), p.283-290.
10. Yufei Yuan and Michael J. Shaw. Induction of fuzzy decision trees. *Fuzzy Sets and Systems*, 69 (1995), p.125-139.
11. Kosko, B. *Fuzzy Engineering*, Prentice Hall, Inc. 1997.
12. Quinlan, J.R. *C4.5: Programs for Machine Learning*. Morgan Kauffman. 1993.

Fuzzy Binary Tree Model for European-Style Vanilla Options

Silvia Muzzioli

Department of Economics, University of Modena and Reggio Emilia,
Italy

`muzzioli.silvia@unimore.it`

Huguette Reynaerts

Department of Applied Mathematics and Computer Science, Ghent
University, Belgium

`huguette.reynaerts@UGent.be`

Abstract The derivation of the risk neutral probabilities in a binary tree, in the presence of uncertainty on the underlying asset moves, boils down to the solution of dual fuzzy linear systems. The issue has previously been addressed and different solutions to the dual systems have been found. The aim of this paper is to apply a methodology which leads to a unique solution for the dual systems.

Keywords Dual fuzzy linear systems, vanilla option, binary tree.

1 Introduction

The plan of the paper is the following. In section 2 we recall the financial problem and we introduce the vector solution to dual fuzzy linear systems. In section 3 we highlight that the system, connected to the financial problem, has no solution if one applies standard fuzzy arithmetic. In section 4 we propose the vector solution to that system and in section 5 we show the algorithm in order to compute the vector solution. The last section concludes.

2 European-style plain vanilla options

A (European-style Plain Vanilla) option is a financial security that provides its holder, in exchange for the payment of a premium, the right but not the obligation to either buy (call option) or sell (put option) a certain underlying asset at a certain date in the future for a prespecified price K . The binary tree model of Cox-Ross-Rubinstein (1979) can be considered as a discrete-time version of the Black & Scholes (1973) model. The following assumptions are made: (A1) the

markets have no transaction costs, no taxes, no restrictions on short sales, and assets are infinitely divisible; (A2) the lifetime T of the option is divided into N time steps of length T/N ; (A3) the market is complete; (A4) no arbitrage opportunities are allowed, which implies for the risk-free interest factor, $1+r$, over one step of length T/N , that $d < 1+r < u$, where u is the up factor and d is the down factor. The European call option price at time zero, has a well-known formula in this model:

$$EC(K, T) = \frac{1}{(1+r)^N} \sum_{j=0}^N \binom{N}{j} p_u^j p_d^{N-j} (S(0)u^j d^{N-j} - K)_+,$$

where K is the exercise price, $S(0)$ is the price of the underlying asset at time the contract begins. p_u and p_d are the resp. up and down risk-neutral probabilities which are solutions to the system:

$$\begin{cases} p_u + p_d = 1 \\ up_u + dp_d = 1 + r. \end{cases} \tag{1}$$

The standard methodology (see Cox et al. (1979)) leads to set $u = e^{\sigma\sqrt{T/N}}$ and $d = e^{-\sigma\sqrt{T/N}}$, where σ is the volatility of the underlying asset. If there is some uncertainty about the value of the volatility, then it is also impossible to precisely estimate the up and down factors. Muzzioli and Torricelli (2001, 2004) suggest to model the up and down jump factors by triangular fuzzy numbers.

A triangular real fuzzy number is uniquely defined by three numbers (f_1, f_2, f_3) or can be written in terms of its α -cuts, $f(\alpha)$, α in $[0, 1]$:

$$f(\alpha) = [\underline{f}(\alpha), \bar{f}(\alpha)] = [\underline{f}, \bar{f}] = [f_1 + \alpha(f_2 - f_1), f_3 - \alpha(f_3 - f_2)].$$

Since the α -cuts of a triangular fuzzy number are compact intervals of the set of real numbers, the interval calculus of Moore (1966) can be applied on them.

In this setting the up and down factors are represented by the triangular fuzzy numbers: $u = (u_1, u_2, u_3)$ and $d = (d_1, d_2, d_3)$. Assumptions (A1), (A2), (A3) are still valid, while assumption (A4) changes as follows: $d_1 \leq d_2 \leq d_3 < 1+r < u_1 \leq u_2 \leq u_3$.

Note that this condition guarantees that the resulting fuzzy matrix:

$$\begin{bmatrix} 1 & 1 \\ (d_1, d_2, d_3) & (u_1, u_2, u_3) \end{bmatrix}$$

has always full rank for all $d \in (d_1, d_2, d_3)$ and for all $u \in (u_1, u_2, u_3)$. There is no fuzziness in the risk free rate of interest, since it is given at time zero.

A fuzzy version of the two equations of the system (1) should now be introduced. This can be done (for each equation) in two different ways, since for an arbitrary fuzzy number f there exists no fuzzy number g such that $f + g = 0$ and for all non-crisp fuzzy numbers $f + (-f) \neq 0$:

$$p_u + p_d = (1, 1, 1), \quad p_u = (1, 1, 1) - p_d$$

respectively

$$up_u + dp_d = (1 + r, 1 + r, 1 + r) \quad up_u = (1 + r, 1 + r, 1 + r) - dp_d$$

where $p_u, p_u = ((p_u)_1, (p_u)_2, (p_u)_3)$, and $p_d, p_d = ((p_d)_1, (p_d)_2, (p_d)_3)$, are the fuzzy up and down probabilities.

Therefore the linear system (1) can be rewritten in four different ways:

$$\begin{cases} p_u + p_d = 1 \\ up_u + dp_d = 1 + r \end{cases} \quad \begin{cases} p_u = 1 - p_d \\ up_u + dp_d = 1 + r \end{cases}$$

$$\begin{cases} p_u = 1 - p_d \\ dp_d = (1 + r) - up_u \end{cases} \quad \begin{cases} p_u + p_d = 1 \\ dp_d = (1 + r) - up_u \end{cases}$$

Those systems are called dual linear systems in the terminology of Ming Ma, et al. (2000). Buckley et al. (1991, 2002) propose the following procedure to solve the fuzzy matrix equation $Ax = b$, where the elements, a_{ij} , of the $n \times n$ -matrix A and the elements, b_i , of the $n \times 1$ -vector b are triangular fuzzy numbers: (1) solve the linear system by using fuzzy number arithmetic; (2) if no such solution exists use the vector solution X_J , with

$$X_J(\alpha) = \{x \mid A_\alpha x = b_\alpha, (A_\alpha)_{ij} \in a_{ij}(\alpha), (b_\alpha)_i \in b_i(\alpha)\}.$$

Muzzioli and Reynaerts (2004) prove that the same procedure can be applied to dual fuzzy linear systems $A_1x + b_1 = A_2x + b_2$, where the elements of the $n \times n$ -matrices A_1 and A_2 and the elements of the $n \times 1$ -matrices b_1 and b_2 are fuzzy numbers. Indeed, the dual fuzzy system $A_1x + b_1 = A_2x + b_2$ has a vector solution X_J^* , with α -cuts

$$\begin{aligned} X_J^*(\alpha) = & \{x \mid A_{1\alpha}x + b_{1\alpha} = A_{2\alpha}x + b_{2\alpha}, \quad (A_{1\alpha})_{ij} \in a_{1,ij}(\alpha), \\ & (A_{2\alpha})_{ij} \in a_{2,ij}(\alpha), (b_{1\alpha})_i \in b_{1,i}(\alpha), (b_{2\alpha})_i \in b_{2,i}(\alpha)\}, \end{aligned}$$

if all matrices $A_{1,0} - A_{2,0} = [a_{1,ij}^0 - a_{2,ij}^0]$, with $a_{1,ij} \in a_{1,ij}(0)$ and $a_{2,ij} \in a_{2,ij}(0)$, are nonsingular. They also prove that the linear systems $Ax = b$ and $A_1x + b_1 = A_2x + b_2$ have the same vector solutions if $A = A_1 - A_2$ and $b = b_2 - b_1$.

3 Solution using fuzzy arithmetic

As we already mentioned, the first step when solving fuzzy linear systems, is to use fuzzy arithmetic. Therefore the systems are expressed in α -cuts. This method is applied for each system, keeping in mind that the coefficients as well as the unknowns are positive:

System 1:

$$\begin{aligned} [\underline{p}_u, \overline{p}_u] + [\underline{p}_d, \overline{p}_d] &= [1, 1] \\ [\underline{u}, \overline{u}][\underline{p}_u, \overline{p}_u] + [\underline{d}, \overline{d}][\underline{p}_d, \overline{p}_d] &= [1 + r, 1 + r]. \end{aligned}$$

leads to the following solution :

$$\begin{aligned} \underline{p}_u &= \frac{(1+r) - \underline{d}}{\underline{u} - \underline{d}} & \overline{p}_u &= \frac{(1+r) - \overline{d}}{\overline{u} - \overline{d}} \\ \underline{p}_d &= \frac{\underline{u} - (1+r)}{\underline{u} - \underline{d}} & \overline{p}_d &= \frac{\overline{u} - (1+r)}{\overline{u} - \overline{d}}. \end{aligned}$$

Clearly, $\overline{p}_u < \underline{p}_u$. The conclusion is that the system has no solution in the set of the fuzzy numbers if one applies fuzzy arithmetic. Note that if one poses $\underline{p}_u^1 = \overline{p}_u$ and $\overline{p}_u^1 = \underline{p}_u$, one obtains the solution given in Muzzioli and Torricelli (2001) and (2004), i.e.:

$$p_u^1 = \left[\frac{(1+r) - \overline{d}}{\overline{u} - \overline{d}}, \frac{(1+r) - \underline{d}}{\underline{u} - \underline{d}} \right] \quad p_d^1 = \left[\frac{\underline{u} - (1+r)}{\underline{u} - \underline{d}}, \frac{\overline{u} - (1+r)}{\overline{u} - \overline{d}} \right].$$

System 2:

$$\begin{aligned} [\underline{p}_d, \overline{p}_d] &= [1, 1] - [\underline{p}_u, \overline{p}_u] \\ [\underline{u}, \overline{u}][\underline{p}_u, \overline{p}_u] + [\underline{d}, \overline{d}][\underline{p}_d, \overline{p}_d] &= [1 + r, 1 + r]. \end{aligned}$$

leads to the solution proposed by Reynaerts and Vanmaele (2003):

$$\begin{aligned} \underline{p}_u &= \frac{(1+r)(\overline{u} + \underline{d}) - \underline{d}(\overline{d} + \overline{u})}{\underline{u}\overline{u} - \underline{d}\overline{d}} & \overline{p}_d &= 1 - \underline{p}_u \\ \overline{p}_u &= \frac{(1+r)(\underline{u} + \overline{d}) - \overline{d}(\underline{d} + \underline{u})}{\underline{u}\overline{u} - \underline{d}\overline{d}} & \underline{p}_d &= 1 - \overline{p}_u, \end{aligned}$$

which is different from what was found for System 1. One can show that $\overline{p_u} < \underline{p_u}$ (and thus $\overline{p_d} < \underline{p_d}$). The conclusion is that the system has no solution in the set of the fuzzy numbers if one applies fuzzy arithmetic.

System 3:

$$\begin{aligned} [\underline{p_d}, \overline{p_d}] &= [1, 1] - [\underline{p_u}, \overline{p_u}] \\ [\underline{d}, \overline{d}] [\underline{p_d}, \overline{p_d}] &= [1 + r, 1 + r] - [\underline{u}, \overline{u}] [\underline{p_u}, \overline{p_u}]. \end{aligned}$$

leads to the following solution :

$$\begin{aligned} \underline{p_u} &= \frac{(1+r) - \overline{d}}{\underline{u} - \overline{d}} & \overline{p_u} &= \frac{(1+r) - \underline{d}}{\overline{u} - \underline{d}} \\ \underline{p_d} &= 1 - \overline{p_u} & \overline{p_d} &= 1 - \underline{p_u}. \end{aligned}$$

Remark that, if one considers the solutions for $\alpha = 0$ and $\alpha = 1$, one can conclude that $(p_u)_2 < (p_u)_1$. Thus the system has no solution in the set of the fuzzy numbers if one applies fuzzy arithmetic.

System 4:

$$\begin{aligned} [\underline{p_d}, \overline{p_d}] + [\underline{p_u}, \overline{p_u}] &= [1, 1] \\ [\underline{d}, \overline{d}] [\underline{p_d}, \overline{p_d}] &= [1 + r, 1 + r] - [\underline{u}, \overline{u}] [\underline{p_u}, \overline{p_u}]. \end{aligned}$$

leads to the following solution :

$$\begin{aligned} \underline{p_u} &= \frac{(1+r)(\overline{u} + \overline{d}) - \overline{d}(\underline{d} + \overline{u})}{\underline{u}\overline{u} - \underline{d}\overline{d}} & \underline{p_d} &= 1 - \underline{p_u} \\ \overline{p_u} &= \frac{(1+r)(\underline{u} + \underline{d}) - \underline{d}(\underline{u} + \overline{d})}{\underline{u}\overline{u} - \underline{d}\overline{d}} & \overline{p_d} &= 1 - \overline{p_u}. \end{aligned}$$

Depending on the values of u and d , $\overline{p_u} < \underline{p_u}$ or $\overline{p_d} < \underline{p_d}$. Thus the system has no solution in the set of the fuzzy numbers if one applies fuzzy arithmetic.

In sum, none of the four dual fuzzy linear systems has a solution if one applies fuzzy arithmetic. More details on the properties of the solutions to the four systems can be found in Muzzioli (2004).

4 Vector solution

Since the four dual fuzzy linear systems have no solution if one applies fuzzy arithmetic, we now investigate the vector solution. This solution

of all four dual linear systems is obtained by solving System 1:

$$\begin{aligned} (A_\alpha)_{1,1} &= (A_\alpha)_{1,2} = (b_\alpha)_1 = 1 & (b_\alpha)_2 &= 1 + r \\ (A_\alpha)_{2,1} &= u_1 + \alpha(u_2 - u_1) + \lambda_1(u_3 - u_1 - \alpha(u_3 - u_1)), & \lambda_1 &\in [0, 1] \\ (A_\alpha)_{2,2} &= d_1 + \alpha(d_2 - d_1) + \lambda_2(d_3 - d_1 - \alpha(d_3 - d_1)), & \lambda_2 &\in [0, 1]. \end{aligned}$$

The vector solution is:

$$\begin{aligned} p_u(\alpha) &= \frac{(1+r) - (d_1 + \alpha(d_2 - d_1) + \lambda_2(d_3 - d_1)(1 - \alpha))}{frac} \\ p_d(\alpha) &= \frac{u_1 + \alpha(u_2 - u_1) + \lambda_1(u_3 - u_1)(1 - \alpha) - (1+r)}{frac}, \\ &\lambda_1, \lambda_2 \in [0, 1], \\ frac &= u_1 - d_1 + \alpha((u_2 - u_1) - (d_2 - d_1)) + (1 - \alpha)(\lambda_1(u_3 - u_1) \\ &\quad - \lambda_2(d_3 - d_1)). \end{aligned}$$

The marginals are:

$$\begin{aligned} p_u(\alpha) &= \left[\frac{1+r-d_3+\alpha(d_3-d_2)}{u_3-\alpha(u_3-u_2)-d_3+\alpha(d_3-d_2)}, \right. \\ &\quad \left. \frac{1+r-d_1-\alpha(d_2-d_1)}{u_1+\alpha(u_2-u_1)-d_1-\alpha(d_2-d_1)} \right] \\ p_d(\alpha) &= \left[\frac{u_1+\alpha(u_2-u_1)-(1+r)}{u_1+\alpha(u_2-u_1)-d_1-\alpha(d_2-d_1)}, \right. \\ &\quad \left. \frac{u_3-\alpha(u_3-u_2)-(1+r)}{u_3-\alpha(u_3-u_2)-d_3+\alpha(d_3-d_2)} \right] \end{aligned}$$

Note that this solution is the same solution found in Muzzioli and Torricelli (2001,2004).

5 An algorithmic solution of the fuzzy system

Muzzioli et al. (2004) proposes a practical algorithm that finds directly the marginals for each unknown. Therefore one has to solve 2^k systems, where k is the number of fuzzy parameters in the original fuzzy system. Each element of the extended coefficient matrix of those systems is either the lower or the upper bound of the α -cut of the corresponding element of the original fuzzy extended coefficient matrix. The final solution is investigated by taking the minimum and the maximum of the solutions found in each system for each unknown. This procedure

ensures that all possible solutions, consistent with the parameters of the system, are taken, but it does not guarantee that the solutions are fuzzy numbers. Thus an ex-post check is needed to exclude the solutions that are not fuzzy numbers. If $x_j(1)$ does not belong to $[\underline{x}_j(0), \bar{x}_j(0)]$ for all j , then there is no solution to the system. If one applies this algorithm to the financial example, one should solve the following systems:

$$\begin{cases} p_u + p_d = 1 \\ \underline{u}p_u + \underline{d}p_d = 1 + r. \end{cases} \quad \begin{cases} p_u + p_d = 1 \\ \bar{u}p_u + \bar{d}p_d = 1 + r. \end{cases}$$

$$\begin{cases} p_u + p_d = 1 \\ \underline{u}p_u + \bar{d}p_d = 1 + r. \end{cases} \quad \begin{cases} p_u + p_d = 1 \\ \bar{u}p_u + \bar{d}p_d = 1 + r. \end{cases}$$

The solutions to those systems are resp.:

$$\begin{cases} p_u = \frac{(1+r)-\underline{d}}{\underline{u}-\underline{d}} \\ p_d = \frac{\underline{u}-(1+r)}{\underline{u}-\underline{d}}. \end{cases} \quad \begin{cases} p_u = \frac{(1+r)-\bar{d}}{\bar{u}-\bar{d}} \\ p_d = \frac{\bar{u}-(1+r)}{\bar{u}-\bar{d}}. \end{cases} \quad \begin{cases} p_u = \frac{(1+r)-\bar{d}}{\underline{u}-\bar{d}} \\ p_d = \frac{\underline{u}-(1+r)}{\underline{u}-\bar{d}}. \end{cases} \quad \begin{cases} p_u = \frac{(1+r)-\underline{d}}{\bar{u}-\underline{d}} \\ p_d = \frac{\bar{u}-(1+r)}{\bar{u}-\underline{d}}. \end{cases}$$

The final solution is obtained by taking the minimum and maximum for each unknown. Since $p_u(1)$ and $p_d(1)$ belong to the resp. solution intervals, the vector of fuzzy numbers,

$$\left(\left[\frac{(1+r)-\bar{d}}{\bar{u}-\bar{d}}, \frac{(1+r)-\underline{d}}{\underline{u}-\underline{d}} \right] \left[\frac{\underline{u}-(1+r)}{\underline{u}-\bar{d}}, \frac{\bar{u}-(1+r)}{\bar{u}-\underline{d}} \right] \right),$$

is a solution to the system. Remark that one obtains the same solution as the vector solution in section 4.

6 Conclusions

The derivation of the risk neutral probabilities in a lattice framework, in the presence of uncertainty on the underlying asset moves, boils down to the solution of a fuzzy linear system. In this paper we have investigated the solution of such a system by using the methodology proposed by Muzzioli and Reynaerts (2004). The solution, that is the same solution found in Muzzioli and Torricelli (2001, 2004), is here given in terms of vector solution. The properties of such a solution have been highlighted.

References

- [1] Black F, Scholes M (1973) The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*
- [2] Buckley JJ, Eslami E, Feuring T (2002) *Fuzzy mathematics in economics and engineering*, Studies in Fuzziness and soft computing, Physica Verlag
- [3] Buckley JJ, Qu Y (1991) Solving systems of linear fuzzy equations, *Fuzzy sets and systems* 43: 33-43
- [4] Cox J, Ross S, Rubinstein S (1979) Option pricing, a simplified approach, *Journal of Financial Economics* 7: 229-263
- [5] Friedman M, Ming M, Kandel A (2000) Duality in fuzzy linear systems, *Fuzzy sets and systems* 109: 55-58
- [6] Moore RE (1979) *Methods and applications of interval analysis*, SIAM Studies in Applied Mathematics
- [7] Muzzioli S, Torricelli C (2001) A multiperiod binomial model for pricing options in a vague world, *Proceedings of the second International Symposium on Imprecise Probabilities and Their Applications*: 255-264
- [8] Muzzioli S, Torricelli C (2004) A Multiperiod Binomial Model for Pricing Options in a Vague World, *Journal of Economic Dynamics and Control*, Special Issue on Financial Modelling 28: 861-887
- [9] Muzzioli S (2003) A note on fuzzy linear systems, Discussion paper n. 447, Economics Department, University of Modena and Reggio Emilia.
- [10] Muzzioli S, Reynaets H (2004) Dual fuzzy linear systems, submitted to *Fuzzy sets and systems*
- [11] Reynaerts H, Vanmaele M (2003), A sensitivity analysis for the pricing of European call options in a binary tree model, *Proceedings of the third International Symposium on Imprecise Probabilities and Their Applications*: 467-481

Tree-Structured Smooth Transition Regression

Joel Corrêa da Rosa¹, Alvaro Veiga², Marcelo Medeiros³

¹ Department of Statistics of Federal University of Paraná, e-mail: joel.rosa@est.ufpr.br

² Department of Electrical Engineering of Catholic University of Rio de Janeiro
e-mail: alvf@ele.puc-rio.br

³ Department of Economics of Catholic University of Rio de Janeiro e-mail:
mcm@econ.puc-rio.br

Abstract Over the recent years, many nonlinear models have been proposed, both in classical econometrics, where multiple regime linear models like STR (Smooth Transition Regression) are receiving much attention, and in machine learning theory, including recursive partitioning methods like regression trees, artificial neural networks (ANN) and nonparametric regression. In this work, we try to combine the advantages of two of these models: STR and CART. From the STR model we inherited statistical inferential tools available to specify, test and to provide diagnostic checks. The CART methodology offers an easy interpretation of the final model through a binary tree that can be stated as logical sentences. In this paper we present the methodology of a tree structured regression with smooth splits instead of the traditional CART partitioning of the explanatory variables space. The performance of this new approach is compared to CART one in a set of benchmark datasets.

1 Introduction

In recent years, much attention has been devoted to nonlinear modeling. Techniques such as artificial neural networks, nonparametric regression and recursive partitioning methods are frequently used to approximate unknown functional forms. In spite of their success in various applications, several times these approaches lack interpretability due to the complexity of the final model. Some cases in which the fitted model can be given a reasonable interpretation, there are no inferential procedures that guarantee the statistical significance of the parameters.

Send offprint requests to: joel.rosa@est.ufpr.br

The proposal of the present paper is the construction of a nonlinear regression model that combines aspects of two well-known methodologies: Classification and Regression Trees (CART) discussed in [1] and the Smooth Transition Regression (STR) (see [3] and [2]) by taking advantages of their main capabilities. Our model inherits from CART the simplicity and interpretability of the tree-based models while the STR framework provides tools for inference-based decisions.

The CART methodology represents a unification of all tree-based classification and prediction methods that have been developed since a first approach presented by [4]. It transformed the regression tree models in an important nonparametric alternative to the classical methods of regression. Since then, the attractiveness of this methodology has motivated many authors to create hybrid modeling strategies that merge tree techniques with known statistical methods.

This paper presents in Section 2 the basis of the proposal of a fuzzy regression tree called Tree-STR which can be found by applying sequential tests of hypotheses. Section 3 brings an application of Tree-STR model to the famous Boston Housing dataset. In Section 4, we compare our approach to CART one in a set of benchmark datasets by using in-sample and out-of-sample evaluation criteria. Finally, conclusions are presented in Section 5.

2 Tree-STR model a new approach

Given a set of observations taken from the pair (y_t, \mathbf{x}_t) in which y is a univariate response variable and \mathbf{x} is a vector of p explanatory variables, traditional regression techniques will try to explain the relationship between them by assuming that errors follow a probability distribution.

A regression tree is often used as a nonparametric model which looks for the best local prediction, or explanation, of a continuous response through the recursive partitioning of the predictor variables' space. The fitted model is usually displayed in a graph which has the format of a binary decision tree which grows from the root node to the terminal nodes, also called leaves.

The idea proposed here is the simple substitution of CART sharp splits by smooth splits in an operation that exchange step functions by continuous smooth functions assuming values between 0 and 1. By doing this, the simplest tree model, which contains 2 terminal nodes, could be mathematically written as:

$$y_t = \beta_1 G_1(\mathbf{x}_t; \Psi_0) + \beta_2 G_2(\mathbf{x}_t; \Psi_0) + \epsilon_t \quad (1)$$

where $G_2(\cdot) = 1 - G_1(\cdot)$.

Assuming that the error term follows a probability distribution, we can interpret the STR model in equation (1) as a simple regression tree in which the nodes contain fuzzy sets instead of binary partition of the dataset. Here, the commonest choice for function $G(\cdot)$ is the logistic function:

$$G_1(\mathbf{x}_t; \Psi_0) = \frac{e^{-\gamma_0(x_{s_0 t} - c_0)}}{1 + e^{-\gamma_0(x_{s_0 t} - c_0)}} \quad (2)$$

in which the vector Ψ_0 , associated to the partition of the root node (node 0), has three parameters γ_0 , c_0 and s_0 . The first one takes account of the smoothness of the transition function while the second would represent the CART threshold. If the split is smooth, c_0 , the location parameter, indicates the value which the observations have equal membership degrees to the created nodes¹. The parameter s_0 indicates the index of the splitting variable selected among the elements of the vector $\mathbf{x}_t = (x_1, x_2, \dots, x_p)$. As long as γ_0 approaches infinity, the CART simplest tree is nested by the Tree-STR model in (1).

The parameters β_1 and β_2 represent constant models to be fitted within the two terminal nodes.

A More complex tree model would correspond to a re-specification of (1) by splitting one of the two created nodes. If the first (node 1) is to be split, the model equation, by omitting the argument of transition functions, is updated to:

$$y_t = [\beta_3 G_3(\cdot) + \beta_4 G_4(\cdot)] G_1(\cdot) + \beta_2 G_2(\cdot) + \epsilon_t \quad (3)$$

where $G_4(\cdot) = 1 - G_3(\cdot)$.

The regression tree behind the model in equation (3) has 3 terminal nodes that contain fuzzy sets with membership functions $G_1(\cdot)G_3(\cdot)$, $G_1(\cdot)G_4(\cdot)$ and $G_2(\cdot)$.

2.1 Tree Growing Methodology

We adopt an inferential approach to grow the tree based on linearity tests of the kind proposed in [5]. In this framework, the Taylor expansion of the logistic functions creates an auxiliary regression that could be used to test the significance of splitting a node.

A problem that emerges from this procedure is the control of the overall error in this sequence of tests. Here, we adopt a strategy that reduces the significance level whenever the tree grows. By forcing the size of the test to be lowered during the tree construction, we also avoid the use of post-pruning techniques.

The tree growing strategy is based on the modelling cycle in [6]. The final decision for splitting a node will be taken after following the steps listed below:

- Step 1 Use the LM-type test to specify a node and a splitting variable.
- Step 2 Estimate model parameters by nonlinear least squares.
- Step 3 Conditionally on the nonlinear parameters, evaluate the significance of the constants within the nodes.

¹ The created nodes are labeled as node 1 and node 2

When the LM-Type test does not identify significant splits in the existent nodes, the algorithm is stopped and the tree-based model is specified.

It is important to enhance that the LM-type test in step 1 works only as a tool for ranking splitting variables and nodes to be split.

The estimation procedure corresponds to the maximization of the likelihood function since we assume that the errors are i.i.d and normally distributed.

Immediately after splitting a node its significance is tested by evaluating tests in the parameters of the conditional linear model. During the tree growing, the nonlinear parameters are kept fixed while the linear parameters are updated at each split.

3 Application to Boston Housing Dataset

We present in this section an application of the proposed methodology to the famous Boston Housing dataset which has been largely used in regression analysis literature. The main variable in this dataset is the median value of homes (MV) in Boston census tracts. Important explanatory variables are: crime rate (CRIM), lower status population (LSTAT), average number of rooms (RM), weighted distance to employment centers (DIS), percent built before 1940 (AGE), air pollution concentration (NOX) and pupil/teacher ratio (P/T).

This dataset was extensively used by the CART algorithm, thus it becomes easier to compare it with the Tree-STR results.

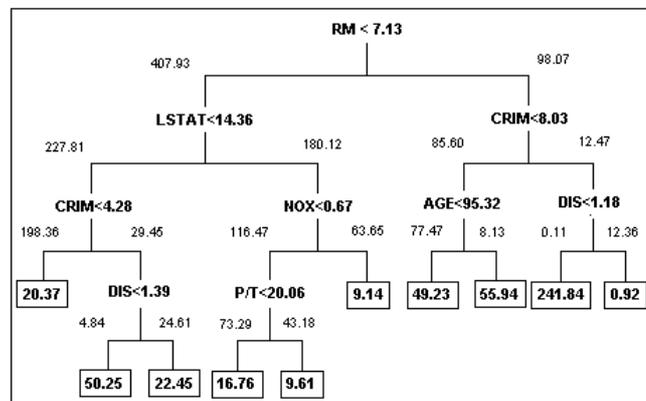


Fig. 1 Tree-STR model for Boston Housing Dataset

We specified the Tree-STR model to this dataset using the iterative modeling cycle described in Section 2 and reducing the significance level of hypotheses testing after each node splitting. The corresponding tree is shown in Figure 1.

Some aspects illustrated in Figure 1 are quite different from the traditional CART regression tree. First, instead of distributing the sample over the terminal nodes, from this methodology we distribute the membership degree of a observation over them. Second, besides the threshold (location) parameters we also estimate a set of smoothness parameters that are not present in Figure 1 but which are shown in Table 1.

Table 1 Parameters Estimates of Logistic Functions in the Tree-STR Model fitted to Boston Housing Dataset

Nonlinear Parameters	
Location	Smoothness
$\hat{c}_0 = 7.13_{(0.003)}$	$\hat{\gamma}_0 = 1.62$
$\hat{c}_1 = 14.36_{(1.166)}$	$\hat{\gamma}_1 = 1.93$
$\hat{c}_2 = 8.03_{(1.085)}$	$\hat{\gamma}_2 = 3.53$
$\hat{c}_3 = 4.28_{(1.086)}$	$\hat{\gamma}_3 = 65.13$
$\hat{c}_4 = 0.67_{(4 \times 10^{-4})}$	$\hat{\gamma}_4 = 10.63$
$\hat{c}_5 = 95.32_{(4.317)}$	$\hat{\gamma}_5 = 115.36$
$\hat{c}_6 = 1.18_{(0.001)}$	$\hat{\gamma}_6 = 103.41$
$\hat{c}_9 = 20.06_{(5 \times 10^{-4})}$	$\hat{\gamma}_8 = 245.24$
$\hat{c}_8 = 1.39_{(0.145)}$	$\hat{\gamma}_9 = 5.04$

(.) - s.e of estimates

In the Table 1, the parameters indexes, represent the location of the node in the tree display. Thus, c_0 is the location parameter at the root node (node 0).

Another difference from CART that should be outstanding is that the Tree-STR model can fit a constant to the node that does not belong to the range of the observed responses. In fact, the predicted value, unlike CART, is obtained from a weighted sum of the constants within the nodes in which the weights are given by the membership of that observation to a terminal node.

The sum of membership degrees within the terminal node is an important statistic to detect outliers candidates. A good example can be taken from Figure 1 where the lowest value (0.11) for this statistic is associated to node 10 that would contain an unlikely prediction rule which is stated in Figure 2.

With the use of the function *treefit* present in the MATLAB statistical toolbox, the tree grown by the CART algorithm for this data obtained initially 95 terminal nodes. By following the same post-pruning principle, we found 30 terminal nodes with the Tree-STR model. After re-evaluating the tree performance by 10-fold cross-validation to get honest estimates of the errors, the estimated proportion of variance explained by CART was

If		
Condition		
1	Number of Rooms greater than 7.13	<i>and</i>
2	Crime Rate greater than 8.03	<i>and</i>
3	Distance to employment center lower than 1.18	
Then		
Predicted Value is	241.84	

Fig. 2 Prediction Rule for Observations in Node 10

about 78 % and applying the same procedure to the Tree-STR model, this proportion raised to 85 %.

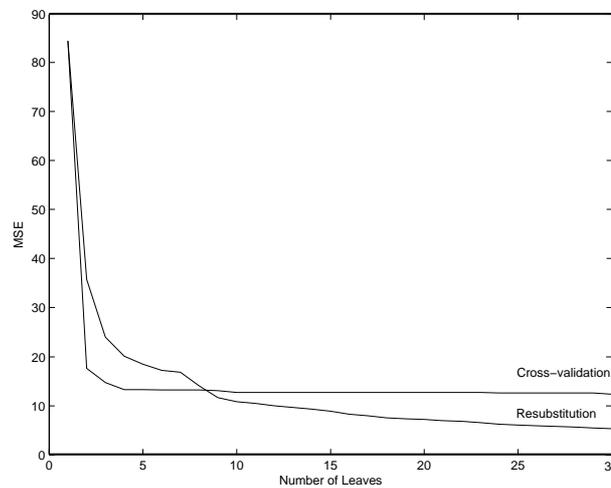


Fig. 3 Resubstitution and Cross-validation Estimates of Mean Square Error - Boston Housing Dataset

For the so-called resubstitution estimate, which means in-sample evaluation, the errors behaved quite differently in our model when compared to CART. We can see in Figure 3 that for small trees, the in-sample evaluation overestimates the mean square error and after reaching the supposed "right-sized" tree it starts to underestimate it.

4 Tree-STR vs CART Using Benchmark Datasets

To get a better evaluation of the Tree-STR performance when compared to CART algorithm, we selected 3 benchmark datasets to apply both methodologies. A brief description of them is given below.

1. CPUs - The CPUs data is present in [7]. The purpose of applying regression trees methodology to this dataset is to provide a model that explains the performance of 209 different types of CPUs by some hardware characteristics.
2. Cars - This data were taken from MASS library in *R* software and it describes the prices and other 25 variables of 93 new cars models for the 1993 year in the United States.
3. Auto imports - This dataset was taken from Ward's 1985 Automotive Yearbook and consists of 195 prices of cars followed by some features such as: fuel consumption, length, width, engine size, among others. The informations are similar to the previous dataset, but there are more continuous variables to be included in a regression model.

Table 2 brings general information about the models fitted by CART and Tree-STR respectively.

Table 2 Features of CART and Tree-STR Models fitted to Different Datasets

CART	Number of Nodes	Splitting Variables
Boston	9	X_6, X_5, X_8 and X_{13}
CPUs	5	X_2, X_3 and X_4
Cars	3	X_4
Auto Imports	9	X_3, X_5, X_6, X_7 and X_8
Tree-STR	Number of Nodes	Splitting Variables
Boston	10	$X_5, X_6, X_7, X_8, X_{11}$ and X_{13}
CPUs	4	X_3 and X_4
Cars	2	X_4
Auto Imports	9	X_3, X_4, X_5, X_6, X_8 and X_{11}

In Table 2, it should be noted that the tree sizes produced by each method are very close, but the Tree-STR model captures more information as long as it selects a greater number of splitting variables.

To evaluate the results in-sample and out-of-sample, we present in Table 3 the resubstitution and 10-fold cross-validation estimates for the mean square error (MSE).

Table 3 reveals that Tree-STR model has a better in-sample and out-of-sample performance on average if the mean square error measures the goodness-of-fit.

5 Conclusions

In this paper we present a new approach that provides an alternative non-linear model that can be applied to regression problems. This model can be interpreted as a multiple regime smooth transition regression, a neural network or a fuzzy regression.

Table 3 In-Sample and Out-of-Sample Evaluation of CART and Tree-STR Models in Benchmark Datasets

In-sample evaluation of MSE		
Dataset	CART	Tree-STR(LM)
Boston	12.56	11.55
CPUs	3024.84	1283.72
Cars	35.53	33.93
Auto Imports	3.26	7.23
Out-of-sample evaluation of MSE		
Dataset	CART	Tree-STR(LM)
Boston	22.85 _(4.07)	18.10 _(2.74)
CPUs	3772.91 _(969.08)	3672.12 _(1774.36)
Cars	60.34 _(18.38)	39.89 _(15.61)
Auto Imports	8.84 _(1.60)	10.77 _(1.58)

(.) - s.e. of MSE estimates in 10 samples.

Application of this model to benchmark datasets has shown that the fitted regression can be presented in the same fashion as CART, but with some useful additional informations such as the membership degrees of the observations to the terminal nodes and the possibility of predicting beyond the range of the dependent variable.

The overall results indicated that Tree-STR model yields parsimonious tree-based models that outperform CART trees if using the same number of nodes. Additionally, as shown in the Boston Housing application, the membership degrees are tools to identify outliers and influential observations.

References

1. L. Breiman and J. H. Friedman and R. A. Olshen and C. J. Stone, *Classification and Regression Trees* (Wadsworth Int. Group, New York 1984).
2. D. van Dijk, T. Tersvirta and P. H. Franses, *Econometric Reviews* **21** (2002) 1-47
3. C. W. J. Granger and T. Tersvirta, *Modelling Nonlinear Economic Relationships* (Oxford University Press, Oxford 1993).
4. J. Morgan and J. Sonquist, *Journal of the American Statistical Association* **58**, (1963) 415-434.
5. P. Saikkonen and R. Luukkonen, *Scandinavian Journal of Statistics* **15**, (1988) 55-68.
6. T. Tersvirta, *Kybernetika* **30**, (1994) 319-330.
7. B. D. Ripley, W. N. Venables *Modern Applied Statistics with S-Plus* (Springer Verlag, New York 2002).

Discrete Dynamic Programming With Outcomes In Fuzzy Ordered Structures

Tadeusz Trzaskalik¹, Sebastian Sitarz²

¹Department of Operations Research, The University of Economics,
ul. Bogucicka 14, 40-226 Katowice, Poland, ttrzaska@ae.katowice.pl

²Institute of Mathematics, The University of Silesia,
ul. Bankowa 14, 40-007 Katowice, Poland, ssitarz@ux2.math.us.edu.pl

Abstract. One way of generalization single criterion dynamic programming model is to consider partially ordered criteria structure. Generalization of Bellman's principle of optimality is employed to create solution of such problem. Ordered structures defined by fuzzy numbers and triangular norms are considered. Numerical illustration is given at the end of the paper.

Keywords. discrete dynamic programming, fuzzy numbers, triangular norms.

1. Introduction

Decision problems with conflicting objectives and multiple stages can be considered as multi-objective dynamic programming problems. A survey has been presented by Li and Haimes (1989), more recently by Trzaskalik (1998). Another way of generalization single-criterion dynamic programming models is to consider outcomes in partially ordered criteria set. Mitten (1974) described a method for solving a variety of multistage decisions in which the real value objective function is replaced by preference relation. Sobel (1975) extended Mitten's result to infinitive horizon for deterministic and stochastic problems. Preference order dynamic programming was described by Steinberg and Parks (1979). Henig (1985) defined a general sequential model with returns in partially ordered set. It is shown that Bellman's principle of optimality (Bellman(1957)) is valid with respect to maximal returns and leads to an algorithm to approximate these returns. Discrete dynamic programming with partially ordered criteria set was also considered by Trzaskalik and Sitarz (2002).

Application of fuzzy logic to control started with the work written by Bellman and Zadeh (1970). Many contemporary approaches in this field are presented in Kacprzyk (1997).

The aim of the paper is to describe finite discrete model and solving procedure and partially ordered fuzzy structures. In Section 2 discrete dynamic decision process is defined and solving procedure is described. In Section 3 two fuzzy ordered structures: one based on trapezoid fuzzy numbers and the second – on triangular norm are considered. Two illustrative examples are given in Section 4.

2. Dynamic programming in partially ordered structure

Discrete dynamic process P which consists of T periods is considered. Let us assume that for $t=1, \dots, T$, that:

Y_t is the set of all feasible state variables at the beginning of period t ,
 Y_{T+1} is the set of all states at the end of the process,
 $X_t(y_t)$ is the set of all feasible decision variables for period t and state $y_t \in Y_t$.

We assume that all above sets are finite. Now let us define:

$D_t = \{d_t = (y_t, x_t) : y_t \in Y_t, x_t \in X_t(y_t)\}$ - the set of all period realizations in period t ,

$W_t : D_t \rightarrow Y_{t+1}$ - transformations.

Process P is given if sets $Y_t, \forall t, Y_{T+1}, X_t(y_t), \forall t, X_T(y_T)$ and transformations $W_t, \forall t, W_T$ are identified.

Let us denote

$D = \{d = (d_t, \forall t) : \forall t \in \{1, \dots, T\} y_{t+1} = W_t(y_t, x_t) \text{ and } x_t \in X_t(y_t)\}$ - the set of all process realizations.

$D_t(y_t) = \{(y_t, x_t) : x_t \in X_t(y_t)\}$ - the set of all realizations in period t which begin at y_t .

$d(y_t) = (y_t, x_t, \forall t, y_{T+1})$ - the backward partial realization which begins at y_t .

$D(y_t) = \{d(y_t) : d \in D\}$ - the set of all backward partial realizations, which begin at y_t .

$D(Y_t) = \{D(y_t) : y_t \in Y_t\}$ - the set of all backward partial realizations for Y_t .

$d'(y_t) = (y_{t+1}, x_{t+1}, \forall t, y_t, x_t)$ - the forward partial realization which ends at $y_t = W_t(y_{t+1}, x_{t+1})$.

$D'(y_t) = \{d'(y_t): d\hat{\mathbf{I}}D\}$ - the set of all forward partial realizations, which end at y_t .

$D'(Y_t) = \{D'(y_t): y_t\hat{\mathbf{I}}Y_t\}$ - the set of all forward partial realizations for Y_t .

$(W, \mathbf{f}, \hat{\mathbf{I}})$ - ordered structure with binary relation \leq and binary operator fulfilling following conditions:

$$\begin{aligned} & "_{a\hat{\mathbf{I}}W} a\mathbf{f}a, \\ & "_{a,b\hat{\mathbf{I}}W} a\mathbf{f}b \hat{\mathbf{U}} b\mathbf{f}a \mathbf{P} a=b, \\ & "_{a,b,c\hat{\mathbf{I}}W} a\mathbf{f}b \hat{\mathbf{U}} b\mathbf{f}c \mathbf{P} a=c, \\ & "_{a,b,c\hat{\mathbf{I}}W} a (b c) = (a b) c, \\ & "_{a,b,c\hat{\mathbf{I}}W} a\mathbf{f}b \mathbf{P} a c\mathbf{f}b c \hat{\mathbf{U}} c a\mathbf{f}c b. \end{aligned}$$

Relation $<$ is defined as follows:

$$a < b \hat{\mathbf{U}} a\mathbf{f}b \hat{\mathbf{U}} a^1b.$$

Applying relation \mathbf{f} we define for each finite subset $A\hat{\mathbf{I}}W$ the set of maximal elements:

$$\max(A) = \{a^*\hat{\mathbf{I}}A: \sim\mathcal{S}_{a^*\hat{\mathbf{I}}A} a^* < a\}.$$

For $t=1, \dots, T$ let $(W, \mathbf{f}, \hat{\mathbf{I}}_t)$ be a sequence of ordered structures and $f_t: D_t \rightarrow W$ a sequence of period criteria functions. Applying period criteria functions f_t , we define functions $F_t: D(Y_t) \rightarrow W$ in the following way:

$$\begin{aligned} F_T &= f_T \\ F_t &= f_t \hat{\mathbf{I}}_t (f_{t+1} \hat{\mathbf{I}}_{t+1} (\dots \hat{\mathbf{I}}_{T-1} (f_T))), \quad t=T-1, \dots, 1. \end{aligned}$$

Let $F: D \rightarrow W$ be the function defined in one of the following ways:

i) $F = F_1$

ii) $F = F_T$

F is called the multi-period criteria function. Discrete dynamic decision process (P, F) is given if there are defined: discrete dynamic process P and multi-period criteria function F .

Realization $d\hat{\mathbf{I}}D$ is efficient, iff

$$F(d) \hat{\mathbf{I}} \max F(D).$$

Our problem is to find the set of all maximal values of the process i.e. set $\max F(D)$.

Theorem 1. Decision dynamic process (P, F) is given. For all $t=T-1, \dots, 1$ and all $y_t\hat{\mathbf{I}}Y_t$ holds:

- a) $\max \{F_t(D(y_t)) = \max \{f_t(d_t) \hat{\mathbf{I}}_t \max(F_{t+1}(d(W_t(d_t))))): d_t\hat{\mathbf{I}}D_t(y_t)\}.$
- b) $\max\{F(D)\} = \max \{\max F_t(d(y_t)): y_t\hat{\mathbf{I}}Y_t\}.$

Proof. Trzaskalik and Sitarz (2000), (Theorem 2, Theorem 3, p. 191).

Theorem 1 yields backward iterative computational method.

Step 1. Compute $\max\{F_t(D(y_t))\}$ for all states $y_t \in \hat{Y}_t$

Step t (for $t=T-1, \dots, 1$). Compute $\max\{F_t(D(y_t))\}$ for all states $y_t \in \hat{Y}_t$ by using theorem 1a).

Step T. Compute $\max\{F(D)\}$ by using theorem 1b).

3. Fuzzy ordered structures

3.1. Fuzzy numbers

We consider trapezoidal fuzzy numbers with membership function:

$$m(x) = \begin{cases} 0, & x \in \langle -\infty, m-a \rangle \cup \langle n+b, \infty \rangle \\ 1, & x \in \langle m, n \rangle \\ 1 - \frac{m-x}{a}, & x \in (m-a, m) \\ 1 - \frac{x-n}{b}, & x \in (n, n+b) \end{cases}$$

Shortly we denote such trapezoidal numbers as (m, n, α, β) , and the set of all trapezoidal fuzzy numbers as $T_{m,n,a,b}$. We use fuzzy max-order (Furu-kava (1994))

$$A \mathbf{f}_F B \hat{U} \text{ " }_{\alpha \in [0,1]} (\sup A_\alpha \mathbf{f} \sup B_\alpha) \hat{U} (\inf A_\alpha \mathbf{f} \inf B_\alpha)$$

where $A_\alpha = \{x \in \hat{I} : m_\alpha(x) \geq \alpha\}$, $B_\alpha = \{x \in \hat{I} : n_\alpha(x) \geq \alpha\}$ are α -cuts of fuzzy numbers A, B.

This relation in a class of trapezoidal fuzzy numbers has the following form:

$$(m_1, n_1, a_1, b_1) \mathbf{f}_F (m_2, n_2, a_2, b_2) \hat{U} (m_1 - a_1 \mathbf{f} m_2 - a_2) \hat{U} (m_1 \mathbf{f} m_2) \hat{U} (n_1 \mathbf{f} n_2) \hat{U} (n_1 - b_1 \mathbf{f} n_2 + b_2),$$

We add fuzzy numbers by adding parameters:

$$(m_b, n_b, a_b, b_b) + (m_2, n_2, a_2, b_2) \hat{U} (m_1+m_2, n_1+n_2, a_1+n_2, b_1+b_2)$$

One can easily prove that triple $(T_{m,a,b}, \mathcal{F}_F, +)$ is ordered structure. We can built another ordered structure of fuzzy numbers using L-fuzzy numbers presented by Furukava (1994).

3.2. Triangular norms

Let us consider following structure (W, \mathcal{F}, \cdot) consisting of:
 $W = [0, 1]$ (real numbers between 0 and 1),
 \mathcal{F} (standard relation of comparing real numbers),
 $\cdot = T$ (T-norm).

This triple creates ordered structures because every t-norm fulfills conditions:

$$\begin{aligned} & \text{" }_{a,b,c \in [0,1]} T(a,T(b,c)) = T(T(a,b),c) \\ & \text{" }_{a,b,c \in [0,1]} a \mathcal{F} b \Rightarrow T(a,c) \mathcal{F} T(b,c) \Rightarrow T(c,a) \mathcal{F} T(c,b) \end{aligned}$$

4. Numerical illustrations

We consider a dynamic process which consists of 4 periods $[T=4]$, in which:

$$\begin{aligned} & Y_t = \{0, 1\}, \text{ for } t \in \{1, 2, 3, 4\} \\ & X_t(y_t) = Y_{t+1} \text{ for } t \in \{1, 2, 3\} \text{ and } y_t \in Y_t \\ & W_t(y_t, x_t) = x_t \text{ for } y_t \in Y_t \text{ and } x_t \in X_t(y_t) \end{aligned}$$

Two examples presented below are considered over the same dynamic process. The difference lays in the structures of outcomes. Fuzzy numbers are used in example 4.1, and triangular norms are used in example 4.2.

Example 4.1 (fuzzy numbers)

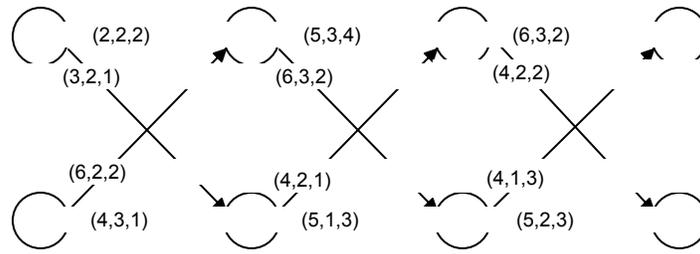
In the example we use special case of trapezoidal fuzzy numbers i.e. triangular fuzzy numbers. We denote them as $(m, \mathbf{a}, \mathbf{b})$, where m is the center and \mathbf{a}, \mathbf{b} are deviation parameters. The outcomes are described in ordered structure: $(T_{m,a,b}, \mathcal{F}_F, +)$. The values of period criteria functions are as follows

$$\begin{array}{lll} f_1(1,1) = (2, 2, 2); & f_2(1,1) = (5, 3, 4); & f_3(1,1) = (6, 3, 2); \\ f_1(1,0) = (3, 2, 1); & f_2(1,0) = (6, 3, 2); & f_3(1,0) = (4, 2, 2); \\ f_1(0,1) = (6, 2, 2); & f_2(0,1) = (4, 2, 1); & f_3(0,1) = (4, 1, 3); \end{array}$$

$$f_1(0,0) = (4, 3, 1); \quad f_2(0,0) = (5, 1, 3); \quad f_3(0,0) = (5, 2, 3).$$

The process with values of period criteria functions is presented in Fig 1.

Fig. 1. Graph of the process in example 1.



Applying the backward procedure we obtain results presented in Tab 1.

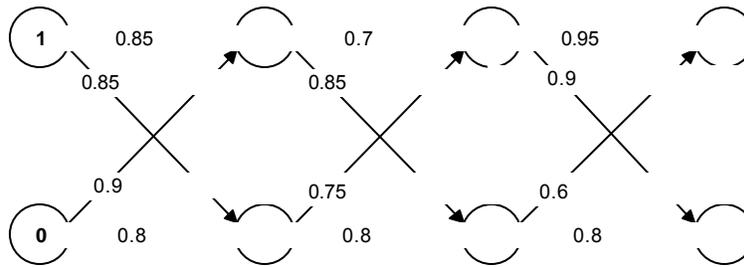
Table 1. Maximal values obtained in example 1.

3	(5, 2, 3)	(6, 3, 2)
2	(10, 3, 6)	(11, 6, 6), (11, 5, 5)
1	(17, 7, 7), (17, 8, 8)	(13, 5, 7), (13, 8, 8)
	max F(D)	
	(17, 7, 7), (17, 8, 8)	

Example 4.2. (T-norms)

We consider dynamic process with outcomes in structure $([0, 1], \leq, T)$, where T denotes T-norm. The values of period criteria functions are presented in graph of the process Fig 2.

Fig. 2. Graph of the process in example 2.



Applying the backward procedure we obtain results for chosen T norms presented in Tab 2.

Table 2. Results obtained in example 2.

$\min \{a, b\}$	minimum	0.8	(0,1, 1,0, 0,0)
$\max \{a+b-1, 0\}$	Lukasiewicz	0.55	(0,1, 1,0, 0,0), (1,0, 0,1, 1,1), (0,1, 1,1, 1,1)
ab	probablistic	0.621	(0,1, 1,0, 0,0)
$\begin{cases} \min \{a, b\} & \text{if } \max \{a, b\} = 1 \\ 0 & \text{otherwise} \end{cases}$	weak	0	(0,1, 1,0, 0,0), (1,0, 0,1, 1,1), (0,1, 1,1, 1,1)

5. Conclusions

Dynamic programming model with outcomes in partially ordered criteria set gives a good possibility of modeling decision maker's preference. Two examples of fuzzy structures were presented. The approach given in the paper can be easily extended and among dimensions defining criteria space real numbers, random variables, fuzzy numbers and other mathematical structures can be simultaneously considered.

References

- Bellman, R.E. (1957) *Dynamic Programming*. Princeton University Press, Princeton, N.J.
- Bellman, R.E., Zadeh, (1970) L.A. Decision making in fuzzy environment. *Management Science* 17 141-164.
- Furukava N. (1994) A parametric total order on fuzzy numbers and a fuzzy shortest route problem, *Optimization*, 30 (1994) 367-377.
- Henig, M.I. (1985) The principle of optimality in dynamic programming with returns in partially ordered sets". *Math. of Oper. Res.*, 10 3 462-470.
- Li, D., Haimes, Y.Y. (1989) "Multiobjective dynamic programming: the state of the art". *Control Theory and Advanced Technology*, 5 4 471-483.
- Kacprzyk J. (1997) *Multistage fuzzy control. A model-based approach to fuzzy control and decision making*. Willey.
- Mitten, L.G. (1974) "Preference order dynamic programming". *Management Science*, 21 1 43-46.
- Sobel, M.M.(1975): "Ordinal dynamic programming". *Management Science*, 21, 9, 967-975.
- Steinberg E., Parks, M.S. (1979) "A preference order dynamic program for a knapsack problem with stochastic reward". *Operational Research Society Journal* 30 2 141-147.
- Trzaskalik, T.: (1998) *Multiobjective Analysis in Dynamic Environment*, The Karol Adamiecki University of Economics in Katowice, 1998.
- Trzaskalik, T., Sitarz, S. (2000) "Dynamic discrete programming with partially ordered criteria set". In:Trzaskalik T., Michnik J. (eds.): *Multiple objective and Goal Programming. Recent Developments*. Physica-Verlag,186-195.

**FUZZY MATHEMATICAL
STRUCTURES AND GRAPH
THEORY**

Plenary Report

Fractal Methods of Graphs Partitioning
Kureichik V.V. and Kureichik V.M.

Fuzzy Mathematical Structures

A Method for Defining and Fuzzifying Mathematical Structures

Ionin V.K. and Plesniewicz G.S.

On Symmetric MV-Polynomials

Di Nola Antonio, Lettieri Ada and Belluce Peter

Lattice Products, Bilattices and some Extensions of Negations, Triangular Norms
and Triangular Conorms

Tarassov Valery B.

Soft Computing Methods in Graphs Theory and Applications

Definition of Optimum Allocation of the Service Centers

Bershtein Leonid, Bozhenyuk Alexander and Rozenberg Igor

Evolutionary Algorithm of Minimization of Intersections and Flat Piling of the
Graph Mathematic Models

Gladkov L.A. and Kureichik V.M.

Use of a Fuzzy Strategy to Create the Shortest Hamiltonian Cycle

Piech Henryk and Ptak Aleksandra

Fractal Methods of Graphs Partitioning*

V.V.Kureichik, V.M.Kureichik

Taganrog State University of Radio Engineering

vkur@tsure.ru; kur@tsure.ru

The problem of partitioning of the graph nodes on parts is considered. For its decision genetic algorithms and fractals aggregation methods are offered. It allows receiving partitioning with local optimum for polynomial time. Complexity of algorithms has the square-law order on average.

Introduction

In 1980 B.Mandelbrot has specified on fractal geometry of natural systems (NS) [1,3]. Fractals are invariant to analyzed object and are capable to similar duplication at various existential levels and to the transfer of the information to system about infringement of stability of the structural condition. Fractals have properties of adaptation to external influence by reorganization of fractal sets (FS). For artificial systems the similarity of structures represents the determining property of FS and is realized only on the limited scales. According to [1-3,4,8] evolving systems have a fractal nature and a presence of directed not spontaneous selection and the self-coordinated evolution.

The new type of genetic algorithms of graphs partitioning on parts with minimization of total number of external edges is offered. These methods can be used in economics and finance. For example, when deciding the problem of optimal resources distribution, goods distribution between bases, shares distribution for shareholders etc. Strategy of partitioning is based on aggregation of fractals and the use of evolutions ideas [5]. Running time of algorithm (RTA) allows

* Work is executed at financial support of the Russian Federal Property Fund (the project №03-01-00336)

receiving local optimum. It roughly makes $O(n^2) - O(n^3)$, where n – is a number of the graph nodes researched.

1. Problem statement. Fractal objects are similar, i.e. their kind does not undergo changes when changing scales. The sets having such structure are considered being ones possessing geometrical (scale) universality. Such sets of objects refer to fractal sets. Among these are Set of Cantor (SC) and Carpet of Serpinsky (CS) [1,2]. They possess a geometrical invariance and are called « sets of average thirds ». At SC construction a piece of individual length $[0,1]$ is divided into three equal parts and an average from them – an interval $[1/3, 2/3]$ is cut out. Further the same happens to each piece. We receive a sequence of pieces of the decreasing length. At the first stage it is one piece, on the second – two, on the third – 4, etc., on k -that – $2k$. At $k \rightarrow \infty$ we have a set of points named SC. The total length of all cut out pieces is equal to 1. The circuit of a snowflake can be an example of a fractal object.

New strategy of the decision of the important combinatory problem of graph partitioning concerning to a class of NP-complete problems is proposed. In this paper, new and modified genetic algorithms (GA) with fractals aggregation for graph partitioning are offered. They differ by application of non-conventional architectures of genetic search (GS), and new genetic operators (GO) focused on use of knowledge of decided problems.

In [9-12] statement of the graph partitioning problem on the set or any number of parts is given. Cost function (CF) for a graph G partitioning will be written as follows:

$$K = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n K_{i,j}, (i \neq j), \quad (1)$$

where $K_{i,j}$ – is a number of connections between parts B_i and B_j at partitioning graph G on parts; l – is an amount of parts in partitioning; K – is a total amount of edges at the graph partitioning. The standard problem of partitioning consists in minimization K ($K \rightarrow \min$).

It follows from the previously mentioned that the decision of a graph partitioning problem based on full search is inconvenient because of the process exponential complexity. Thereupon various

heuristics of the decision of the given problems are developed [9]. Some new heuristics using various incident methods recently have appeared. These are methods of evolution simulation, simulated annealing, fractals aggregation and their updatings. They are based on the analysis and search of alternative variants of decisions; however they differ on technology and principles of realization.

There is the mechanism of aggregation (DLA) describing fractals creation. According to DLA the certain version of fractals can be received during the disorder growth. For example, a cluster is given growing as follows: eventually the molecule joins it and sticks to it at once. This process is named aggregation. Let particles diffuse to growing cluster randomly. The aggregation of particles proceeding in conditions of random movement, it is DLA [5].

2. Fractal algorithm of the graph partitioning. Let us consider an algorithm of graph partitioning based on fractals aggregation [1,5,9-12]. The mechanism of the graph partitioning is based on seven principles:

- essential dependence on entry conditions;
- "Razor Okkama";
- use of various evolution models;
- clusters construction (the massifs of graph nodes definitely connected among themselves);
- clusters factorization, i.e. reduction of dimension the graph by representation of clusters as the new incorporated nodes of the graph;
- DLA, i.e. the aggregation proceeding in conditions of random, directed and combined connection of the graph nodes to clusters;
- hierarchy.

Such mechanism is open, i.e. the amount of used principles can be varied. We shall describe this process in detail. Process of clusters creation is based on the mechanism of construction of the minimal files in the graph or hypergraph [10]. Let us introduce a concept of cluster in graph $G = (X, U)$. Cluster is a part of graph $\Phi \subseteq G$, and $\Phi = (X_1, U_1)$, $X_1 \subseteq X$, $U_1 \subseteq U$. Nodes of the cluster are connected by internal edges to other nodes $X \setminus X_1$ of graph G . Let us name the capacity of a subset of cluster external edges as f . We shall name cluster Φ_i as minimal one, if for any another cluster Φ_j , $\Phi_j \subset \Phi_i$

the condition $f_i \leq f_j$ is satisfied. In other words, the removal of any nodes from $\Phi_i (\Phi_i \setminus X_m)$ results to new cluster with a big number of external edges. By definition we shall consider:

$$\begin{aligned} & (\forall \Phi_i \subseteq G) (\Phi_i \neq \emptyset), \\ & (\forall x_i \in X) (x_i - \text{minimal cluster}). \end{aligned} \quad (2)$$

It means that minimal cluster (CM) cannot be empty. Besides in a trivial special case each node of the graph forms CM. Construction of CM is connected to allocation of all subsets of set X, and it makes 2^n , if $|X| = n$. In this connection are developed heuristics of allocation of a clusters set and CM. These heuristics are based on theorems of the minimal files in graphs [5].

Theorem 1. *If Φ_i and Φ_j are CM and $\Phi_i \not\subset \Phi_j \wedge \Phi_j \not\subset \Phi_i$ then $\Phi_i \cap \Phi_j = \emptyset$.*

Theorem 2. *Let Φ_i, Φ_j – CM. If $\Phi_s = \Phi_i \cup \Phi_j \wedge f_s < f_i \wedge f_s < f_j$ then Φ_s is CM.*

Theorem 3. *Let $\Phi_1, \Phi_2, \dots, \Phi_L$ – the CM, and any association of a part from them is not CM. If $\Phi_i = \Phi_1 \cup \Phi_2 \cup \dots \cup \Phi_L \wedge f_i < f_1 \wedge f_i < f_2, \dots, \wedge f_i < f_L$ then Φ_i is CM.*

The proof of theorems 1,2,3 follows from the analysis of CM construction. On the basis of these theorems we shall describe heuristic algorithm of CM construction. From the theorem 1 it follows that generally at full search procedure of CM construction is converging. The decision is unique at the graph partitioning when the number of nodes of partitioning is not set. At the graph partitioning on the set number of nodes the amount of decisions generally represents some subset of alternatives. Theorems 2 and 3 allow aggregating (uniting) clusters.

Let us present heuristic indistinct algorithm of CM construction in the graph.

1. Let us order all local degrees of the graph nodes as trivial CM.
2. Starting by nodes with the greater local degree, we shall analyze all nodes of graph G in pairs. If a pair of nodes (x_i, x_j) for which it is defined:

$$f = \rho_{(x_i)} + \rho_{(x_j)} - 2r_{i,j}, \quad (3)$$

then $\Phi = (X_1, U_1)$, $X_1 = \{x_i, x_j\}$ and Φ is CM. Here $\rho_{(x_i)}, \rho_{(x_j)}$ – are local degrees of nodes x_i, x_j ; $r_{i,j}$ – is a number of the edges connecting nodes x_i, x_j among themselves, f – is a number of the edges

- connecting CM with other nodes of graph G , and $f < \rho(x_i)$, $f < \rho(x_j)$. Minimal cluster Φ is put in the special list, and nodes x_i, x_j are excluded from consideration. Transition to 2. If new CM was not formed, transition to 3.
3. We replace nodes x_i, x_j in CM by one node $x_{i,j}$. Thus the edges connecting x_i and x_j , are removed, and edges from nodes x_i, x_j to other nodes are attributed to the common node $x_{i,j}$. We have graph G' .
 4. In graph G' all nodes are analyzed. If there are three nodes x_a, x_b, x_c for which the following is fair:

$$f = \rho(x_a) + \rho(x_b) + \rho(x_c) - 2r_{a,y} - 2r_{b,c} - 2r_{a,c}, \quad (5)$$
 and $f < \rho(x_a)$, $f < \rho(x_b)$, $f < \rho(x_c)$ the top x_a, x_b, x_c forms CM according to the theorem 3. Transition to 2. Minimal cluster will be worn out in the list. If the CM is not present more, transition to 4.
 5. Let's increase parameter of CM capacity by unit and we pass to 3. If CM capacity is equal to number of nodes of the set part of partitioning or it is equal to number of nodes of the graph, transition to 5.
 6. The end of work of algorithm.

The given algorithm can be applied for construction of the set amount of the CM, the set capacity, and also it is used for construction of quasioptimal cluster, when $f_i - f_j \leq \varepsilon$. Here ε - is the least relative deviation from capacity of CM ($\varepsilon = 0, 1, 2, \dots$).

After construction of a set of CM if in graph G there were free nodes, it is necessary to carry out procedure of aggregation. It consists in a trial premise of nodes in CM with analysis of CF. The top is located in CM if its infringement occurs on the size which is not exceeding ε . If after the second step there were free nodes, procedures of joint evolution are realized.

Let us define labor input of the given algorithm. Let N_p - be a size of a population, N_s - be an amount of the received descendants, N_G - be an amount of GA generations. Then the common labor input of algorithm can be defined roughly:

$$T \approx t_f [(N_p t_p + N_s t_s + (N_p + N_s) t_c] N_G, \quad (4)$$

where t_f - is a labor input of minimal cluster construction; t_p - is a labor input of construction of one chromosome (the alternative decision) in a population (a set of decisions); t_s - is a labor input of generation of one descendant; t_c - is a cumulative labor input of se-

lection. Let us notice that labor input of construction of one chromosome can change from $O(n)$ up to $O(n^2)$, and labor input of generation of one descendant depends on complexity of the used GO and roughly changes from $O(n)$ up to $O(n \log n)$. Labor input of realization of various models of evolution can vary from $O(n)$ up to $O(n^2)$. RTA of the described algorithm roughly is $O(\lambda n^2)$ for one generation, here λ - is a factor determined by the amount of iterations.

At the decision of problems of graph partitioning there are problems of a grouping of the elements possessing identical properties frequently. For the decision of such problems we shall use ideas of construction of quasi-optimal clusters on the basis of fractals aggregation [5].

3. Results of experimental research. Let us construct the following table (tab. 1) in which we shall give results of the experiments carried out.

Table 1.

N	n	m	N_p	N_i	ρ_{ok}	ρ_{mok}	ρ_{oc}	S	t	CF
1	50	150	50	1000	0.8	0.9	0.1	⊖	6	0.54
2	100	300	50	1000	0.8	0.9	0.1	⊖	20	0.5
3	200	600	50	1000	0.8	0.9	0.1	⊖	30	0.48
4	300	900	50	1000	0.8	0.9	0.1	⊖	40	0.41
5	400	1200	50	1000	0.8	0.9	0.1	⊖	46	0.45
6	500	1500	40	500	0.8	0.9	0.1	⊖	57	0.47
7	600	1800	30	500	0.8	0.9	0.1	⊖	60	0.50
8	800	2400	20	500	0.8	0.9	0.1	⊖	80	0.52
9	1000	3000	10	500	0.8	0.9	0.1	⊖	160	0.54

The table presents results of experiments when partitioning not one graph is analyzed, but a set of test problems. The behavior of partitioning algorithm is investigated when using various modified GO. In the sixth, the seventh and the eighth columns of the table the sizes of probability of application of the modified operators are specified. In 11 column values of conditional CF are given. Conditional CF is defined by the relation of amount of external edges to amount of internal edges at partitioning. It is necessary to aspire to such partitioning, that conditional $CF \Rightarrow 0$. The elite selection is chosen. The size of a population is fixed constant. But during the decision after each generation the size of a population varied basi-

cally aside reduction, leaving the best from point of view CF of a chromosome. On fig.1 the schedule of dependence of time of reception of the best decision from number of nodes researched the graph is shown.

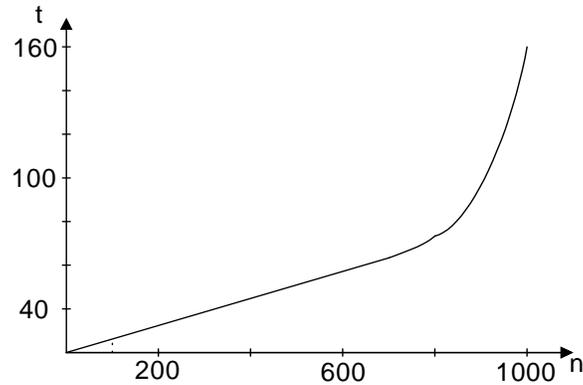


Fig. 1.

The conclusion

Realization of partitioning algorithms has shown the advantage of fractals aggregation ideas, uses of joint models of evolutions and non-standard architecture of search in comparison with algorithms of a pair exchange and random algorithms. Application of the modified genetic algorithms allows increasing the partitioning quality. From the resulted statistical data follows, that generally time of the decision linearly depends on amount of iterations and RTA on the basis of fractals aggregation confirms theoretical preconditions and $\approx O((n \log n) - (n^3))$. It is necessary to notice that when increasing the number of iterations in GA the time of the decision raises, but this increase is insignificant and is compensated by reception of a set of local - optimum decisions.

References

1. *Cronover R.M.* Fractal and chaos in dynamic systems. Bases of the theory. M.: the Postmarket, 2000.
2. *Pajtgen H.O., Rokhner P.H.* Beauty of fractals. M.: the World, 1990
3. *Prigozhin II., Stengers I.* Order from chaos. M.: Progress, 1986.
4. *Goldberg D.E.* Genetic Algorithms in Search, Optimization and Machine Learning. USA: Addison-Wesley Publishing Company, Inc., 1989.
5. *Emeljanov V.V., Kureichik V.V., Kureichik V.M.* Theory and practice of evolution simulation. – M.: Физматлит, 2003. – 432 p.
6. *Kureichik V.V.* Evolutionary, synergetic and homeostatic methods of decision-making. The monography. Taganrog: Publishing house TRTU, 2001.
7. *Batishchev D.A.* Genetic algorithms of the decision of extreme problems. Voronezh: Publishing house VGTU, 1995.
8. *Kureichik V.M.* Genetic algorithms: the Monography. Taganrog: Publishing house TRTU, 1998.
9. *Sherwani N.* Algorithms for VLSI Physical Design Automation, Kluwer Academic Publishers, Boston/Dordrecht/London, 1995.
10. *Morozov K.K., Melihov A.N., Kureichik V.M., etc.* Methods of partitioning of circuits PЭА on structurally finished parts. M.: the Soviet radio, 1978.
11. *Chandrasekharam R., Subhramanian and chadhurys.* Genetic algorithms for node partitionaly problem and application in VLSI design // IEEE Proc-E, v.140, N.5, September, 1993.
12. *Kureichik V.M., Kureichik V.V.* Genetic algorithm of partitioning the graph // Izvestija of The Russian Academy of Science, № 5, 1999.

A Method for Defining and Fuzzifying Mathematical Structures

V.K.Ionin¹ and G.S.Plesniewicz²

¹ Moscow Power Institute (technical university) evdionina@mtv-net.ru

² "MATT"-Russian University of Technology olples@mail.ru

Abstract. A new method for defining mathematical structures and the method for their fuzzifying is presented. In particular, the definition of fuzzy vector space over arbitrary field is given.

1 Introduction

In [1,2] one of the authors (V.K.Ionin) has proposed some very general method for defining mathematical structures. Briefly, the method consists in the following.

Let \mathbf{K} be any category, A, B be its objects, and $\Gamma \subseteq Hom(A, B)$ be some set of morphisms from A to B . Given object X let (X, F, Φ) be such that (a) $F \subseteq Hom(A, X)$, $\Phi \subseteq Hom(X, B)$; (b) $\Phi F \subseteq \Gamma$; (c) it is not possible to extend F or Φ so that (a) is satisfied. (Here ΦF is the set of all compositions φf of morphisms $\varphi \in \Phi$ and $f \in F$.)

Remark. Due to the property (b), if F (or Φ) is known then Φ (or F , correspondingly) is uniquely defined.

We say that (X, F, Φ) is a Γ -space and (Φ, F) is the Γ -structure. Thus, by mathematical structure we mean a Γ -structure. An object X with some Γ -structure can be considered as Γ -structure. One can consider the set Γ as *kind* of the structure (F, Φ) .

For example, consider one of the simplest mathematical structure - the structure of partial order. This structure is the pair (X, θ) where $\theta \subseteq X \times X$ is a binary relation such that: (1) $x \theta x$ for all $x \in X$; (2) if $x \theta y$ and $y \theta x$ then $x = y$ for all $x, y \in X$; (3) if $x \theta y$ and $y \theta z$ then $x \theta z$ for all $x, y, z \in X$.

It turns out that the structures of partial order can be represented as Γ -structures. Let $A = B = \{0, 1\}$ and Γ be the set of all mapping from A to B besides the mapping $\gamma: \gamma(0)=1, \gamma(1)=0$. For any Γ -space (X, F, Φ) we define: $x \theta y$ if and only if $\varphi(x) \leq \varphi(y)$ for all

$\varphi \in \Phi$. It is easy to prove that (X, θ) is a partial order structure. Thus, we have the transition

$$(X, F, \Phi) \rightarrow (X, \theta). \quad (1.1)$$

On the contrary, let (X, θ) be any partial order structure. For any pair $(a, b) \in X \times X$ we denote by f_{ab} the following mapping from A to X as follows: $f_{ab}(0) = a$, $f_{ab}(1) = b$. Then $F =_{def} \{f_{ab} | a\theta b\}$. Defining Φ , we firstly denote by Φ_{ab} the set of all mapping φ from X to B such that if $\varphi(a) = 1$ then $\varphi(b) = 1$. Then, by definition, Φ is intersection of all Φ_{ab} such that $a\theta b$. Thus, we have the inverse transition

$$(X, \theta) \rightarrow (X, F, \Phi). \quad (1.2)$$

Both transitions (1.1) and (1.2) define the correspondence

$$(X, F, \Phi) \leftrightarrow (X, \theta) \quad (1.3)$$

In this case we say that the Γ -structure (X, F, Φ) is equivalent to the structure (X, θ) .

One goal of our investigations is in establishing correspondences like (1.3) between Γ -spaces and usual mathematical structures.

In the above definition of Γ -structure we implicitly suppose that Γ is usual (i.e. crisp) subset of the set $Hom(A, B)$. But if we replace Γ by fuzzy subset of $Hom(A, B)$ then we come to the concept of fuzzy Γ -structure and fuzzy Γ -space. Thus, fuzzification of the concepts related to crisp mathematical structures consists in applying fuzzy sets of morphisms rather than crisp sets. "Higher degree" of fuzzification is obtained when we consider Hom as fuzzy sets.

Remark. Our method is different from well known category theory method for specifying fuzzy structures (originating by J.Goguen). See for instance [3].

2 Preliminary Information

2.1 Γ -spaces

Let Γ be a subset of the set $Hom(A, B)$ of all morphisms from A to B where A and B are objects of a category \mathbf{K} . A triple (X, F, Φ)

where $X \in \text{ob}\mathbf{K}$, is said to be the Γ -space if the following two axioms are satisfied.

Axiom 1. For any $F' \subseteq \text{Hom}(A, X)$ the following statements are equivalent: (a) $F' \subseteq F$; (b) $\varphi f \in \Gamma$ for any $\varphi \in \Phi$ and $f \in F'$.

Axiom 2. For any $\Phi' \subseteq \text{Hom}(X, B)$ the following statements are equivalent: (a) $F' \subseteq F$, (b) $\phi f \in \Gamma$ for any $\varphi \in \Phi$ and $f \in F'$ are equivalent: (c) $\Phi' \subseteq \Phi$: (d) $\varphi f \in \Gamma$ for all $\varphi \in \Phi'$ and $f \in F$.

It easy to note that these axioms are equivalent (together) to the following two statements: (i) if $f \in F$, $\varphi \in \Phi$ then $\varphi f \in \Gamma$, (ii) it is not possible to extend sets F and Φ without losing the statement (i).

If (X, F, Φ) is a Γ -space then we also say that X is the Γ -space with the Γ -structure (F, Φ)

2.2 Category $\mathbf{K}(\Gamma)$

Let objects $X, Y \in \mathbf{K}$ are considered as Γ -spaces with Γ -structures (F, Φ) and (G, Ψ) , correspondingly. Morphism $h \in \text{Hom}(X, Y)$ is said to be Γ -morphism if $\psi h f \in \Gamma$ for any $f \in F$ and $\psi \in \Psi$. It is clear that Γ -spaces and Γ -morphisms is a category.

2.3 Kind of structures

Denote by **BijSet** the category of sets and bijections and by **BijK** the category of \mathbf{K} 's objects and isomorphisms of \mathbf{K} . *Kind of structure* is a covariant functor T from **BijK** to **BijSet**. Elements of $T(X)$ are the structures of the object X . For example, the kind of the vector structure over an arbitrary field B is the covariant functor $T : \mathbf{BijSet} \rightarrow \mathbf{BijSet}$ which with every set X associates the set of pairs. Each pair contains in the first position a binary operation which specifies on X the structure of commutative group, and contains in the second position an operation - mapping from $B \times X$ to X . The first operation responds to vector addition and the second one responds to multiplication of vectors by scalars. Usual identities relates these operations. In section 3 we give an equivalent (though, strongly different in appearance) definition of kind of the vector structure. This allow us to give other definition of vector spaces. (Similar situation arises with other kinds of structures.)

2.4 Fuzzy Γ -spaces

Now let Γ be a fuzzy subset of $Hom(A, B)$, i.e. Γ is a mapping from $Hom(A, B)$ to the segment of reals $I = \{x \mid 0 \leq x \leq 1\}$. Let us take some function $\tau: I \times I \rightarrow I$ satisfying the following conditions:

$$\tau(0, 0) = 0, \tau(1, 1) = 1, \tau(x, \tau(y, z)) = \tau(\tau(x, y), z) \text{ for all } x, y, z \in I.$$

(Due associativeness of τ , it is convenient to write $x \tau y$ rather than $\tau(x, y)$.)

A triple (X, F, Φ) and a pair (F, Φ) are said to be the *fuzzy Γ -space* and *fuzzy Γ -structure* for the object $X \in Ob\mathbf{K}$, if F and Φ are fuzzy subsets of $Hom(A, X)$ and $Hom(X, B)$, correspondingly, and the following two axioms are satisfied.

Axiom 1 ° For any mapping $F': Hom(A, X) \rightarrow I$ the following two statements are equivalent: a) $F'(f) \leq F(f)$ for all $f \in Hom(A, X)$; b) $\Phi(\varphi)\tau F'(f) \leq \Gamma(\varphi f)$ for all $f \in Hom(A, X)$ and $\varphi \in h=Hom(X, B)$.

Axiom 2 ° . For any mapping $\Phi': Hom(X, B) \rightarrow I$ the following two statements are equivalent: c) $\Phi'(\varphi) \leq \Phi(\varphi)$ for all $\varphi \in Hom(X, B)$; d) $\Phi'(\varphi)\tau F(f) \leq \Gamma(\varphi f)$ for all $f \in Hom(A, X)$ and $\varphi \in Hom(X, B)$.

In the sequel for any Γ -space (X, F, Φ) we write (for brevity) γ , f and φ instead of $\Gamma(\gamma)$, $F(f)$ and $\Phi(\varphi)$, correspondingly.

2.5 Fuzzy sets of morphisms

We define fuzzy sets of morphisms as $Hom(X, Y) \rightarrow I$ for any Γ -spaces (X, F, Φ) and (Y, G, Ψ) . The fuzzy set of morphisms depends on the function τ and on the fuzzy set Γ . In the sequel we write h instead of $H(h)$. Suppose h the exact upper bound for λ such that $0 \leq \lambda \leq 1$, $\lambda\tau f \leq h$, $\psi\tau\lambda \leq \psi h$ for all $f \in Hom(A, X)$ and $\psi \in Hom(Y, B)$.

Let (X_i, F_i, Φ_i) , $i = 1, 2, 3$ be fuzzy Γ -spaces and $h_1 \in Hom(X_1, X_2)$, $h_2 \in Hom(X_2, X_3)$.

Lemma 1. If τ is a continuous function then the following inequality holds:

$$h_2\tau h_1 \leq h_2 h_1. \quad (2.1)$$

Remark. It is naturally to mean by a fuzzy subcategory of category \mathbf{K} relative a function τ such mapping H of morphisms into

reals that inequality (2.1) is satisfied for any morphisms h_1 and h_2 having the composition h_2h_1 . Lemma 1 states that mapping H defines a fuzzy subcategory of the category of all Γ -spaces and all morphisms for category \mathbf{K} . It is clear that fuzzy subcategories are natural generalizations of (crisp) subcategories.

3 Fuzzy Vector Spaces

In the sequel \mathbf{K} is the category of all sets and all mapping, i.e. $\mathbf{K} = \mathbf{Set}$.

3.1 Connectivity and Separability

Let $A, B \in \mathit{Ob}\mathbf{K}$ and $\Gamma : \mathit{Hom}(A, B) \rightarrow I$. A fuzzy Γ -space (X, F, Φ) is said to be λ -connected (where $\lambda \in I$) if for any points p and q (possibly, $p = q$) there exists $f \in \mathit{Hom}(A, X)$ that $p, q \in f(A)$ and $f \geq \lambda$. If $\lambda = 1$ then (X, F, Φ) is said to be a *strongly connected* fuzzy Γ -space. The space is to be λ -separable if for any different points p and q there exists $\varphi \in \mathit{Hom}(X, B)$ such that $\varphi(p) \neq \varphi(q)$ and $\varphi \geq \lambda$. If $\lambda = 1$ then the space is to be *strongly separable*.

3.2 Linear mapping

In the sequel B is any fixed field and $A = B \times B$. A mapping $\gamma : A \rightarrow B$ is *linear* if there are $a, b \in B$ that $\gamma(x, y) = ax + by$ for all $x, y \in B$. The set of all linear mapping from A to B is denoted by Λ .

3.3 Fuzzy vector space

Let $\Gamma : \mathit{Hom}(A, B) \rightarrow I$ be such that $\Lambda = \Gamma^{-1}(1) = \{\gamma \mid \Gamma(\gamma) = 1\}$. Then Γ -space (X, F, Φ) is said to be a *fuzzy vector Γ -space* if the following three axioms (besides axioms 1° and 2°) are satisfied.

Axiom 3°. X is not empty.

Axiom 4°. (X, F, Φ) is strongly connected Γ -space.

Axiom 5°. (X, F, Φ) is strongly separable Γ -space.

Let $\Gamma \subseteq \mathit{Hom}(A, B)$, $F \subseteq \mathit{Hom}(A, X)$, $\Phi \subseteq \mathit{Hom}(X, B)$ be injections defined by equalities $\gamma = 1$, $f = 1$, $\varphi = 1$.

The following two statements hold.

Lemma 3. A) If $f \in F, \varphi \in \Phi$ then $\varphi f \in \Gamma$.

B) If for a function $f : A \rightarrow X$ the condition $\varphi f \in \Gamma$ holds for all $\varphi \in \Phi$ then $f \in F$.

3.4 Vector addition

In this section, members of X and B are called vectors and scalars, correspondingly.

By definition, a vector r is said to be a sum of vectors p and q (notation: $r = p + q$) if there are a function $f \in F$ and scalars x_1, y_1, x_2, y_2 such that

$$p = f(x_1, y_1), q = f(x_2, y_2), r = f(x_1 + y_1, x_2 + y_2).$$

Lemma 4. There is one and only one operation of vector addition satisfying this definition.

3.5 Product of scalar and vector

By definition, a vector q is a product of a scalar λ and a vector p (notation: $q = \lambda p$) if there are scalars x, y and function $f \in F$ such that

$$p = f(x, y), q = f(\lambda x, \lambda y).$$

Lemma 5. There is one and only one operation of product of scalars and vectors satisfying this definition.

4 Γ -Structures vs Bourbaki's Mathematical Structures

A partial order structure (X, θ) is an example of mathematical structure according N.Bourbaki. In general, Bourbaki has given the following definition. Let X be any set (so-called the support of the structure) and $\tau_j = \tau_j(X)$ ($1 \leq j \leq n$) are terms obtained by applying usual set-theoretic operation (intersection, union, complement, Cartesian product, projects, and power set formation) starting X (considered as a variable). Then a *mathematical structure is* a tuple

$$(X, c_1, c_2, \dots, c_n) \quad \text{where} \quad c_j \in \tau_j(X) \quad (1 \leq j \leq n) \quad (4.1)$$

For example, if $\tau(X) = \wp(X \times X)$ and $\theta \in \wp(X \times X)$ then we have the structure of binary relation. In particular, if we choose θ to satisfy the conditions (1), (2) and (3) then (X, θ) will be the structure of partial order.

It is natural to consider term $\wp(X \times X)$ as the kind of binary relation structure. But how to define naturally a kind of the structure of partial order?

It is clear that properties (1), (2), (3) can be represented by three formulas:

$\Delta \subseteq \theta, \theta \cap \theta^\circ \subseteq \Delta, \theta\theta \subseteq \theta$ where Δ is the "diagonal" relation, θ° is the inverse relation, i.e.. $\theta^\circ = \{(x, y) \mid (y, x) \in \theta\}$.

Then it is natural to consider the expression

$\langle \wp(X), (\Delta \subseteq \theta) \ \& \ (\theta \cap \theta^\circ \subseteq \Delta) \ \& \ \theta\theta \subseteq \theta \rangle$

as the kind of partial order relation. (This expression is an example of so-called statement of Tarskian algebra of binary relations.)

In general, let $\omega = \omega(c_1, c_2, \dots, c_n)$ be a statement obtained from variable c_1, c_2, \dots, c_n by applying the set-theoretic operations, inclusion operation and propositional connectives. Then the expression $\sigma = \langle \tau_1, \tau_2, \dots, \tau_n \rangle$ is the kind of the Bourbaki's structure (4.1).

We propose the following conjecture. Let **FinSet** be the category whose objects are finite sets $\underline{n} = \{0, 1, 2, \dots, n\}$ and morphisms are mapping $\underline{n} \rightarrow \underline{m}$.

Conjecture. For each Bourbaki's structure (4.1) with the kind σ there is an equivalent Γ -structure relatively category **FinSet**. Furthermore, there is an algorithm which returns Γ after applying to $\langle \sigma, \omega \rangle$.

We have verified the conjecture in particular cases when τ_j ($1 \leq j \leq n$) are all the same $\wp(X \times X)$ and operation \wp is not used in ω . In other words, all c_j ($1 \leq j \leq n$) are binary relations and ω is a statement of Tarskian algebra.

So, the conjecture is true for the case of the structures represented by means of Tarski's formalism for binary relations.

References

1. V.K.Ionin. Structures generated by sets of morphisms (in Russian), Transactions of Siberian Institute of Mathematics "Research in geometry and mathematical analysis", Novosibirsk, 1987..

2. V.K.Ionin. Algebraic principles of building mathematical structures, American Mathematical Society Translations, 1995, vol. 163 (2).
3. L.A.Zadeh. Fuzzy sets , Information and Control. 1965, vol. 8 (1).
4. S.E.Rodabaugh et al. (eds). Application of Category Theory to Fuzzy Subsets, Kluwer Acad. Publishers, 1992.

On Symmetric MV - polynomials

Antonio Di Nola¹ Ada Lettieri² Peter Belluce³

1 Introduction

An *MV*-algebra is an algebraic structure $A = (A, \oplus, *, 0)$ of type $(2,1,0)$ satisfying the following axioms:

- (1) $(x \oplus y) \oplus z = x \oplus (y \oplus z)$;
- (2) $x \oplus y = y \oplus x$;
- (3) $x \oplus 0 = x$;
- (4) $(x^*)^* = x$;
- (5) $x \oplus 0^* = 0^*$;
- (6) $(x^* \oplus y)^* \oplus y = (y^* \oplus x)^* \oplus x$.

Therefore, if we define the constant 1 by $1 = 0^*$ and the operation \odot by $x \odot y = (x^* \oplus y^*)^*$, then from (4), we obtain $1^* = 0$. Moreover, setting $y = 1$ in (6), it follows $x^* \oplus x = 1$. On A two new operations \vee and \wedge are defined as follows: $x \vee y = (x^* \oplus y)^* \oplus y$ and $x \wedge y = (x^* \odot y)^* \odot y$. The structure $(A, \vee, \wedge, 0, 1)$ is a bounded distributive lattice. We shall write $x \leq y$ iff $x \wedge y = x$. A remarkable example is the *MV*-algebra having, as support, the real interval $[0, 1]$ and, as basic *MV*-algebraic operations on $[0, 1]$,

$$\begin{aligned}x \oplus y &= \min(1, x + y); \\ x^* &= 1 - x.\end{aligned}$$

We refer to this *MV*-algebra by $[0, 1]$. For each positive integer n , let L_n be the set $\{0, \frac{1}{n}, \dots, 1\}$ endowed with the following operations :

$$x \oplus y = \min(x + y, 1),$$

¹Dipartimento di Matematica e Informatica, Università di Salerno, Via S. Allende, Baronissi (Salerno).

e-mail: dinola@unisa.it

²Dipartimento di Costruzioni e Metodi Matematici in Architettura, Università di Napoli Federico II, via Monteoliveto 3, 80134 Napoli, Italy.

e-mail: lettieri@unina.it

³University of British Columbia, Dept. Math. , Vancouver, B.C. Canada

$$x^* = 1 - x.$$

For each positive integer n , the algebra L_n is a finite totally ordered MV -algebra (MV -chain) and every non-trivial finite MV -chain is isomorphic to one of them.

Let A be an MV -algebra, $x \in A$ and n a nonnegative integer. In the sequel we will denote by nx the element of A , inductively defined by $0x = 0$, $nx = (n-1)x \oplus x$. Analogously we will set $x^0 = 1$, $x^n = x \odot x^{n-1}$. Moreover we consider the $*$ operation more binding than any other operation, and \odot operation more binding than \oplus .

For all unexplained notions regarding MV -algebras we refer to [1] [2] [3] [4].

Let \mathcal{L} be the poset, under \subseteq , of subalgebras of the MV -algebra $[0, 1]$. \mathcal{L} then has a unique minimal element, $\{0, 1\}$, and a unique maximal element, $[0, 1]$.

\mathcal{L} also contains atoms, that is subalgebras $A \subseteq [0, 1]$ such that if $A' \subseteq A$, then $A' = \{0, 1\}$ or $A' = A$. The algebra $\{0, 1/2, 1\}$ is such an atom.

Since, for a maximal ideal M of an MV -algebra A , $A/M \in \mathcal{L}$, we have a method to refine the structure of the maximal ideal space $MaxA$. Heuristically, *smaller* is the quotient A/M , *larger* is the maximal ideal M . In effect this provides a pre-order on the set of maximal ideals.

In this work we shall study these ideas for the set of maximal ideals of *finite type*, that is maximal ideals M with A/M finite. We shall first look at *super maximal* ideals M , that is, those maximal ideals M such that A/M is a *small* as possible, namely $A/M = \{0, 1\}$. Next we shall look at some classes of *big maximal* ideals M , that is, those maximal ideals M which if not super maximal are such that A/M is an atom of \mathcal{L} .

Our study will use a class of MV -polynomials, that we will call *symmetric* which shall permit us to construct the appropriate MV -algebras.

The first part of this work concerns supermaximal ideals. In a boolean algebra all maximal ideals are supermaximal (considering the boolean algebra as an MV -algebra).

Given an MV -algebra, its set of idempotents, $B(A)$, is a subalgebra which is a boolean algebra. We shall examine extensions of $B(A)$ in A , that is subalgebras A' of A such that $B(A) \subseteq A' \subseteq A$, that have supermaximal ideals. We shall also study properties of the set of supermaximal ideals.

The second part of this work will take up the case of certain extensions of $B(A)$ which may have big maximal ideals, and we shall study some properties of these algebras.

Both of these parts will be presented as a special case of subalgebras determined by certain symmetric MV -polynomials.

Definition 1. Let A be an MV -algebra. An $M \in MaxA$ is called of *type n* , provided that $A/M \cong L_n$.

Definition 2. Let A be an MV -algebra. $M \in MaxA$ is called of *finite type*, if M is of type n for some integer n .

Definition 3. Given an MV -algebra A , an $M \in MaxA$ is called *supermaximal*, in symbols $SMax$ provided $A/M \cong \{0, 1\}$.

Definition 4. Given an MV -algebra A , we shall call $M \in MaxA$, *big-maximal* iff $A/M \cong S$ where S has not nontrivial MV -algebras.

We shall focus on big-maximal ideals of finite type.

Not every MV-algebra A has supermaximal ideals, for example $[0, 1]$, or less trivially, $[0, 1]^X$. We shall construct algebras which do have super maximal ideals, and some algebras where all maximal ideals are super maximal.

We shall look at the topological aspect of the subspace $SMaxA \subset MaxA$ of super maximal ideals.

We shall also look at some cases where A contains big maximals that are not super maximal.

2 Symmetric Polynomials

By an MV-polynomial (in one variable) we mean a polynomial $p(z)$ built from a symbol z and the symbols $\oplus, \odot, *, \vee, \wedge, 0, 1$. Given such a polynomial $p(z)$, we have an evident map on any given MV-algebra A , $p(z) : A \rightarrow A$, by evaluation, $p(a)$, $a \in A$.

We shall call $p(z)$ *symmetric* if $p(z) = p(z^*)$. We shall call $p(z)$ *ideal-uniform* if for any MV-algebra A and every ideal $I \subseteq A$, we have $p(0) \in I$ and if $p(a), p(b) \in I$, then $p(a \oplus b) \in I$.

We immediately have:

Proposition 5. *If $p(z)$ is symmetric and ideal-uniform, then for any MV-algebra A and any ideal $I \subseteq A$ we have that $Sym(p, I) = \{a \in A \mid p(a) \in I\}$ is a subalgebra of A .*

Observe that, since every ideal in an MV-algebra is semi-prime, it suffices to check the ideal-uniform condition only on the prime ideals.

Proposition 6. *Let $p(z)$ be symmetric and ideal-uniform. Suppose if A is a linearly ordered MV-algebra and $p(z) = 0$ on A , then $A \cong L_n$, for some positive integer n . Then for any MV-algebra A and any ideal $I \subseteq A$, the subalgebra $Sym(p, I)$ satisfies the following:*

- i) If Q is a prime ideal of $Sym(p, I)$ and $I \subseteq Q$, then $Sym(p, I)/Q \cong L_n$ or $Sym(p, I)/Q \cong \{0, 1\}$.*
- ii) If $I \subseteq J$, then $Sym(p, I) \subseteq S(p, J)$.*

We will apply this proposition to several different symmetric MV-polynomials.

$Sym(p, I)$ (or just $Sym(I)$ if p is understood) will be called the *p-symmetric subalgebra over I* . A will be called a *p-symmetric algebra* if $A = Sym(p, I)$ for some $I \subseteq A$ where $p(z) \not\equiv 0, I \neq A$.

A simple example of a non-trivial (that is, non-constant) symmetric polynomial is $p_1(z) = z \wedge \bar{z}$. With above notations we get:

Lemma 7. *p_1 is symmetric and ideal-uniform.*

We shall examine some consequences of the above lemma.

Proposition 8. *Let A be an MV-algebra. $B(A)$ is a p_1 -symmetric subalgebra of A .*

Proposition 9. *Let A be an MV-algebra, $I \subseteq A$ be an ideal and $Sym(I) = Sym(p_1, I)$. Then,*

- i) I is an ideal in $Sym(I)$.*
- ii) $B(A) \subseteq Sym(I)$.*
- iii) $Sym(I)/I$ is a Boolean algebra.*
- iv) $B(A/I) \cong Sym(I)/I$.*
- v) $Sym(I)$ is the largest subalgebra R of A for which R/I is a Boolean algebra.*
- vi) If A is α -complete and I is an α -complete ideal, then $Sym(I)$ is α -complete.*
- vii) If J is an ideal and $I \subseteq J$, then $Sym(I) \subseteq Sym(J)$.*

We can view, therefore, $Sym(I)$ as a generalization of $B(A)$. We shall compare some properties of $Sym(I)$ and $B(A)$.

Every ideal in a boolean algebra, if maximal, is supermaximal. On the other hand, $A = [0, 1]^X$, $X \neq \emptyset$ has no supermaximal ideals since the constant function $f(x) = 1/2$ satisfies $f \wedge \bar{f} = f$. As f has finite order it belongs to no ideal.

From Proposition 9 we have $B(A) \subseteq Sym(I) \subseteq A$ for any ideal I of A . Since every ideal I in an MV-algebra is contained in some prime ideal, it is evident that $Sym(I)$ always contains supermaximal ideals.

Let us now focus on an MV-algebra A that is p_1 -symmetric over I for some ideal $I \subseteq A$, $I \neq A$.

Set $M^* = \{x \in A : x^* \in M\}$. Then we get:

Proposition 10. *Let $M \in Max(A)$. The following are equivalent:*

- i) $M \in SMax(A)$.*
- ii) for all $x \in A$, $x \in M$ or $x^* \in M$.*
- iii) for all $x, y \in A$, $x \odot y \in M$ implies $x \in M$ or $y \in M$.*
- iv) $A = M \cup M^*$.*

Proposition 11. *$SMax(A)$ is a closed subspace of $Spec(A)$.*

Proposition 12. *$SMax(A)$ is a closed boolean subspace of $Spec(A)$.*

Call an MV-algebra A *Boolean-mixed*, if A is not Boolean and $A = A' \times B$ where B is a Boolean algebra and A' is non-boolean MV-algebra. We shall prove that every p_1 -symmetric algebra is a subdirect subalgebra of a Boolean-mixed algebra.

Denote by:

N_A the ideal of A , generated by the elements $\{x \wedge x^* : x \in A\}$ and

N_A^\perp , the set $\{x \in A : x \wedge a = 0, \text{ for every } a \in N_A\}$.

With above notations we get:

Proposition 13 (see [5]). *Let A be an MV-algebra. Then following are equivalent:*

- i) A is p_1 -symmetric.*
- ii) A has a supermaximal ideal.*
- iii) A has a boolean homomorphic image.*
- iv) $N_A \neq A$.*
- v) $A \in \mathbf{BP}$.*

Suppose then that A is a subdirect subalgebra of $A' \times B$ where A' is non boolean and B is boolean MV-algebra respectively. Then A has B as a homomorphic image and by the above proposition, A is p_1 -symmetric.

Suppose now that A is p_1 -symmetric. Consider the ideal $(N_A)^\perp$. We note that $(N_A)^\perp \subseteq B(A)$. For if $x \in (N_A)^\perp$, then $x \wedge (x \wedge x^*) = 0$ and so $x \wedge x^* = 0$. Therefore $x \in B(A)$. Since $N_A \cap (N_A)^\perp = 0$, we have the condition for a subdirect representation of A with A/N_A and $A/(N_A)^\perp$. We consider two different representations.

Consider first the map $A \rightarrow A \times A/N_A$ given by $x \rightarrow (x, x/N_A)$. This map is an injective morphism, thus if A is non-Boolean we have A as a subdirect subalgebra of a Boolean-mixed algebra. Moreover we have $A \rightarrow A \times A/N_A \rightarrow A$ given by $x \rightarrow (x, x/N_A) \rightarrow x$; therefore A is a retract of a Boolean-mixed MV-algebra. We have,

Proposition 14. *A is a non-boolean p_1 -symmetric MV-algebra iff A is a subdirect algebra of a boolean-mixed algebra.*

Theorem 15. *Suppose A is a retract of a boolean-mixed algebra. Then A is p_1 -symmetric.*

3 Other Symmetric Functions

Here we want to consider other symmetric functions and the type of subalgebras of $[0, 1]$ they determine. These functions generalize $p_1(z)$ and are defined as follows:

Set:

$$q_0(z) = z \wedge z^*$$

$$p_2(z) = (z^2 \vee (z^*)^2) \wedge q_0(z)$$

and

$$p_3(z) = (z^3 \vee 2z^* \odot z^*) \wedge ((z^*)^3 \vee (2z \odot z) \wedge q_0(z)).$$

Moreover, for $n = 5$ or $n = 7$,

$$p_n(z) = q_0(z) \wedge \bigwedge_{k=1}^{\frac{n-1}{2}} [z \odot (\frac{n-1}{k} z)^k \vee z^* \odot (k(z^*)^{\frac{n-1}{k}})] \wedge [z^* \odot (\frac{n-1}{k} z^*)^k \vee z \odot (kz^{\frac{n-1}{k}})].$$

With above notations we get:

Theorem 16. *Let A be an MV-algebra, I be an ideal of A and $n = 2, 3, 5, 7$. Then we have:*

1. $p_n(z)$ is symmetric and ideal-uniform.
2. $Sym(p_n, I)$ is a subalgebra of A .
3. I is an ideal of $Sym(p_n, I)$.
4. $B(A)$ is a subalgebra of $Sym(p_n, I)$.
5. If Q is a prime ideal in $Sym(p_n, I)$ and $I \subseteq Q$, then Q is supermaximal or big ideal; thus $Sym(p_n, I)/Q \cong \{0, 1\}$ or $Sym(p_n, I)/Q \cong \{0, \frac{1}{n}, \dots, \frac{n-1}{n}, 1\}$.

4 General Case p_n , n prime number and $n \geq 11$

Let n be a prime number and $n \geq 11$. Set:

$$q_{n,1}(z) = [z^n \vee z^* \odot ((n-1)z^*)] \wedge [(z^*)^n \vee z \odot ((n-1)z)]$$

$$q_{n,2}(z) = [z \odot (\frac{n-1}{2}z)^2 \vee z^* \odot (2(z^*)^{\frac{n-1}{2}})] \wedge [z^* \odot (\frac{n-1}{2}z^*)^2 \vee z \odot (2z^{\frac{n-1}{2}})]$$

$$q_{n, \frac{n-1}{2}}(z) = [z \odot (2z)^{\frac{n-1}{2}} \vee z^* \odot (\frac{n-1}{2}(z^*)^2)] \wedge [z^* \odot (2z^*)^{\frac{n-1}{2}} \vee z \odot (\frac{n-1}{2}z^2)]$$

To define $q_{n,k}(z)$, for $3 \leq k < \frac{n-1}{2}$, we need some preliminar considerations. Dividing the prime n by k yields a quotient d_0 and a remainder r_0 , in symbols

$$n = kd_0 + r_0, 0 < r_0 < k. \quad (1)$$

If $d_0 < r_0$, you have to apply a similar process to n and r_0 , obtaining

$$n = r_0d_1 + r_1, 0 < r_1 < r_0. \quad (2)$$

If $d_1 < r_1$, you have to repeat the division algorithm to n and r_1 and so forth.

$$\dots n = r_\ell d_{\ell+1} + r_{\ell+1}, 0 < r_{\ell+1} < r_\ell. \quad (3)$$

Since the finite sequences of positive intergers $(d_0, d_1, \dots, d_{\ell+1})$ and $(r_0, r_1, \dots, r_{\ell+1})$ are strictly increasing and decreasing respectively, there is $\min\{i \in N : r_i < d_i\}$. Denote such a minimum by i_k .

$$n = r_{i_k-1}d_{i_k} + r_{i_k}, 0 < r_{i_k} < r_{i_k-1} \text{ and } 0 < r_{i_k} < d_{i_k}.$$

Then we have:

$$q_{n,k}(z) = [(r_{i_k}t) \odot (d_{i_k}t)^{r_{i_k-1}} \vee (t^*)^{r_{i_k}} \odot (r_{i_k-1}(t^*)^{d_{i_k}})] \wedge [(r_{i_k}t^*) \odot (d_{i_k}t^*)^{r_{i_k-1}} \vee t^{r_{i_k}} \odot (r_{i_k-1}t^{d_{i_k}})]$$

where

$$t = (d_{i_k-2} \dots ((d_2(d_0z)^{d_1})^{d_3} \dots)^{d_{i_k-1}}), \text{ if } i_k \text{ is even;}$$

$$t = (d_{i_k-2} \dots ((d_2(d_1(z^*)^{d_0})^{d_2} \dots)^{d_{i_k-1}}), \text{ if } i_k \text{ is odd.}$$

An example: Let $n = 23$ and $k = 8$

Dividing we get:

$23 = 2 \times 8 + 7$; $d_0 = 2 < 7 = r_0$, then we proceed with n and r_0 ,

$23 = 3 \times 7 + 2$; $d_1 = 3 > 2 = r_1$, then we stop.

In this case $i_k = 1$ is odd. $d_0 = d_{i_k-1} = 2$, $r_{i_k} = r_1 = 2$, $d_{i_k} = d_1 = 3$,
 $r_0 = r_{i_k-1} = 7$

Then:

$$q_{23,8}(z) = [(2t) \odot (3t)^7 \vee (t^*)^2 \odot (7(t^*)^3)] \wedge [(2t^*) \odot (3t^*)^7 \vee t^2 \odot (7t^3)]$$

where

$$t = (z^*)^2 = (2z)^*.$$

Substituting $t = (2z)^*$, we get:

$$q_{23,8}(z) = [(2(2z)^*) \odot (3(2z)^*)^7 \vee (2z)^2 \odot (7(2z)^3)] \wedge [(2(2z)) \odot (3(2z))^7 \vee ((2z)^*)^2 \odot (7((2z)^*)^3)]$$

From above notations we have, for n prime and $n \geq 11$:

$$p_n(z) = q_0(z) \wedge q_{n,1}(z) \wedge q_{n,2}(z) \wedge q_{n,\frac{n-1}{2}}(z) \wedge \bigwedge_{k=4}^{\frac{n-3}{n}} q_{n,k}(z).$$

Conjecture: Theorem 16 holds in the general case for $p_n(z)$, n prime number and $n \geq 11$.

References

- [1] L.P. Belluce, *Semisimple algebras of infinite-valued logic and bold fuzzy set theory*, Canadian J. Math., **38** (1986), 1356–1379.
- [2] C.C. Chang, *Algebraic analysis of many valued logics*, Trans. Amer. Math. Soc. **88**, (1958), 467–490.
- [3] C.C. Chang, *A new proof of the completeness of the Lukasiewicz axioms*, Trans. Amer. Math. Soc. **93**, (1959), 74–80.
- [4] R.L.O. Cignoli, I.M.L. D’ottaviano, D. Mundici *Algebraic Foundations of Many-valued Reasoning*, Kluwer, Dordrecht. (Trends in Logic, Studia Logica Library) (2000)
- [5] A. Di Nola, F. Liguori, S. Sessa *Using Maximal Ideals in the Classification of MV-Algebras*, Portugaliae Mathematica vol.50 **1**, (1993), 87-102.

Lattice products, bilattices and some extensions of negations, triangular norms and triangular conorms

Valery B. Tarassov, Ph.D.

*CIM Department, Bauman Moscow State Technical University
e-mail: tarasov@rk9.bnstu.ru*

Key words: Fuzzy Set, Lattice Product, Bilattice, Negation, Meet, Join

ABSTRACT

Sources of vagueness, uncertainty and inconsistency in economics and finance are revealed. The need in their joint consideration is argued. In this context the advantages and restrictions of $[0,1]$ -fuzzy sets and lattice-based fuzzy sets together with appropriate logical prerequisites are discussed. Lattice products and bilattices are suggested as basic structures for introducing fuzziness, examples of their practical use are given. Some generalizations of negations, triangular norms, conorms, as well as some hybrid operations are proposed. Such bilattice structures induce the properties of two-fold fuzzy estimates on bipolar scales

1. Introduction. How to face the problem of information imperfection in economics and finance ?

One of the main features of human relations in economics, business and finance is a wide use of natural language descriptions to deal with ill-defined problems and complex context-dependent situations. A good example is financial forecasting problem concerning the prognosis of expected results in the company activities under dynamic and non-stable market environment. For instance, the preliminary estimation of any investment project aimed at increasing the value of firm supposes the analysis of costs and profits related to the future. In such situations there exist principal limitations on precision of specifying both quantitative economical indices (such as profitability index, internal rate of return, risks, etc.) and (in particular) qualitative economical factors (partner reliability, strategy "cost", etc.). These problems and situations make appeal to imperfect,

in particular, incomplete, contradictory, uncertain information and vague, linguistic estimates, preferences and beliefs, such as “high profitability”, “small return period”, “the investment alternative 1 is much preferred to the investment alternative 2”, “the cost of product supply by the company A is rather high, and the proposal of company B is a bit better”, and so on.

Thus, the concept of linguistic financial/economical indices based on Zadeh’s linguistic variables to develop computing with words and manipulations of perceptions [31] for business and economy seems to be crucial.

The canonical form of ill-defined linguistic financial estimation is: $M_{CF}(X \text{ is } mA)$, where A is an atomic linguistic value (for instance, “cheap”), m is a modifier (for instance, “very”), mA is a composite linguistic value (for instance, “very cheap”), CF is a certainty (uncertainty) factor (such as necessity, probability, belief), M is a qualifier (“quite”, “rather”, “weakly”), M_{CF} stands for linguistic certainty label (“quite probable”, “weakly believable”). The typical example of such plausible statement is “It is quite probable that tomorrow “Euro/Dollar” (€/\$) exchange rate will considerably increase”.

Let us note that in this statement two different types of information imperfection (or indetermination) appear: *vagueness* and *uncertainty* (their detailed comparison may be found in [9, 18]).

On the one hand, *uncertainty* stands for the lack of information about the occurrence of some event (here, the increase of Euro). As a rule the word “occurrence” inherently involves time, i.e. uncertainty is connected with the future event, or with the question whether the given event may be regarded within some period or not. Uncertainty may take the form of *randomness* that makes appeal to probabilistic models. Conventional probabilistic techniques occupy a rather important place in economics and finance, but they could not be considered as panacea in all situations. Their applicability is reduced to the analysis of exactly given, mass events and is related to the mean-values statistics. Hence, probabilistic approaches are not suitable in constructing “individual trajectory of company evolution”. Here uncertainty may be expressed in the form of *modal values* (for instance, alethic, deontic, belief modalities). Nowadays dual possibility/ necessity measures [9, 30] which provide a convenient tool to model individual events and construct extreme-values statistics become more and more popular in uncertainty modeling.

On the other hand, *vagueness* is usually understood as a lack of precise boundaries. It may be also interpreted within the process of grouping objects that possess a property A with some degrees. So the transition from having a property to not having it is smooth. Vagueness is generally attributed to linguistic labels. It is expressed by *fuzzy set* [29], where the transition from membership to non-membership is not abrupt but gradual. Here

membership function generalizes characteristic function and may be specified between 1 (full membership) and 0 (non-membership) $\mu_A: X \rightarrow [0, 1]$; it can be seen as a way of assessing degrees of satisfaction or subjective preferences for some different elements or interpretations. So the representation of vagueness is closely related to continuity: a small difference between objects (for instance, two investment alternatives) cannot lead to considerable changes in the decision (in particular, both of them may be seen as feasible).

Moreover, in financial management the estimations “cheap – expensive”, “poor-rich”, “low-high”, as well as other similar estimations on bipolar scales are often used. Here we may face such information imperfection factors as *inconsistency* (presence of *contradiction*) associated with simultaneous taking into account of both polar properties (that is typical for expert groups) or *total ignorance* due to the absolute lack of relevant knowledge. In fuzzy mathematics contradiction is studied in the form of *ambiguity*: following Batyrshin [3] and Yager [28], the ambiguity may be expressed as a distance between fuzzy set and its pseudocomplement.

The analysis made above shows that there exist various indetermination factors in economics and finance; most of them can be faced in the framework of fuzzy mathematics. Fuzzy sets express vagueness, possibility and necessity measures tightly connected with fuzzy sets (which can be seen as possibility distributions [30]) are suitable for uncertainty modeling, possibility/probability conversions may be performed too and the simultaneous consideration of fuzzy set and its pseudocomplement enables the analysis of contradictory information. Nevertheless, some restrictions of standard fuzzy techniques, as well as some proposals of generalized approaches taking into account various information/belief imperfections within unified models, will be discussed in this paper. In particular, we shall argue that special algebraic models called bilattices are of primary concern in modeling plausible reasoning with different indetermination factors.

2. Some restrictions of fuzzy logics and ways of their overcoming

Let us point out that Zadeh’s fuzzy sets theory is based upon some important assumptions about membership. Here key assumptions are *membership completeness* and *membership discernability*. *Membership completeness* expresses the total character of membership or non-membership consideration: the third is excluded. *Membership discernability* means that each two elements of a set may be distinguished on membership scale.

Zadeh's fuzzy set is a straightforward generalization of two-valued characteristic function $f \in \{0,1\}$. This function is supposed to be total. It means that from logical viewpoint (when membership value is identified to truth-value) the basic membership logic meets two main principles of classical logical semantics: 1) *bivalence principle* $\mathbf{I} \cup \mathbf{I}^c$ (each element is characterized either by its membership or by its non-membership to a set, so the *membership gaps* like $\neg(\mathbf{I} \wedge \mathbf{I}^c) = \neg(\mathbf{I} \cup \mathbf{I}^c)$ are not possible); 2) *singleton valuation principle* (only singletons $\mathbf{I} = \{i\}$, $\mathbf{I}^c = \{i^c\}$ are subject of membership valuation, composite values $\{\mathbf{I}, \mathbf{I}^c\}$ expressing membership and non-membership at the same time are prohibited; that may be written as $\neg(\mathbf{I}, \mathbf{I}^c)$). The enhanced form of this principle is *non-contradiction principle* $\neg(\mathbf{I} \wedge \mathbf{I}^c)$ considered as the absence of *membership gluts*.

Furthermore, membership and non-membership are tightly coupled. In particular, here membership/ non-membership mutual compensation principle holds: if the grade of membership increases, then the grade of non-membership decreases and vice versa. Often their links are represented by negation operation that justifies the use of only membership function.

The introduction of L-fuzzy sets by Goguen [16] mainly does not change the assumptions mentioned above. Indeed, he focused on complete lattices and respected classical semantics principles.

It may be argued that for many applications we ought to select another basic semantics to introduce fuzziness. For example, to fuzzify partial and paraconsistent logics we should start with three-valued semantics, and to fuzzify minimal hybrid logics [25] needed to express two various types of information imperfection already four-valued semantics are necessary.

This new challenge had been already perceived by fuzzy community. As a result, some non-traditional approaches in fuzzy sets theory appeared. An important intuitionism idea concerning the existence of rather loose relations between affirmations and negations was used by Atanassov to define intuitionistic fuzzy sets [2] described by both membership function and non-membership function. Here singleton valuation principle does not more hold. Similarly, Ren suggested generalized fuzzy sets [21] to deal with incomplete information. In their turn, Dubois and Prade [10] introduced two-fold fuzzy sets which also were given by two membership functions. They express respectively degrees of possibility and necessity of membership in an ill-known crisp set. In section 6 we shall also consider generalized fuzzy sets with two membership functions which extend the procedure of estimation on bipolar scales [20]. Nevertheless, general theoretical background to explain such generalizations is still lacking. Below, we present lattice products and bilattices as a natural tool for building multi-dimensional models of information/belief imperfection factors.

3. Lattice products and bilattices as possible background for some non-standard fuzzy and uncertainty models

Let L and K be lattices (see [6]). Let us take the set $L \times \hat{E}$ of ordered pairs (x, y) , where $x \in L, y \in \hat{E}$. The greatest lower bound \wedge and least upper bound \vee are defined coordinatewisely: $(x_1, y_1) \wedge (x_2, y_2) = (x_1 \wedge x_2, y_1 \wedge y_2)$, $(x_1, y_1) \vee (x_2, y_2) = (x_1 \vee x_2, y_1 \vee y_2)$.

As a result $L \times \hat{E}$ becomes a new lattice called direct lattice product of L and K . Here an important special case (for $L = K$) are lattice exponents $ML = L^m$ and, in particular, lattice squares L^2 .

Another important case of lattice-based structures is the concept of bilattice introduced in [13, 15]. Bilattice is a triple $BL = (X, \leq_1, \leq_2)$, where X is a non-empty set, \leq_1 and \leq_2 are two different ordering relations, (X, \leq_1) and (X, \leq_2) are complete lattices. In [15] bilattice is given in a more specific way as a quadruple $BL_G = (X, \leq_1, \leq_2, m)$, where m is a unary operation on X such that: i) if $x \leq_1 y$ then $m(x) \geq_1 m(y)$; ii) if $x \leq_2 y$ then $m(x) \leq_2 m(y)$. It is obvious that bilattice may be viewed as an algebraic structure containing two different meet and join operations: $BL = (X, \wedge_1, \vee_1, \wedge_2, \vee_2)$.

A well-known example of bilattice “four” semantics is Dunn-Belnap’s [5, 11] semantics with two orderings: information ordering and truth ordering. In case of aboutness valuation, apart standard atomic truth values – true $T = \{T\}$ and false $F = \{F\}$ – another two truth values $B = \{T, F\}$ and $N = \{\}$ are introduced; these B and N correspond to semantic gluts (contradiction) and semantic gaps (total ignorance). In this model the unary ϕ -operation m is a kind of pseudonegation: it makes inversion on the truth order but leaves the information order unchanged.

Let us recall that various ways of modeling fuzziness are obtained by using the following criteria [3]: a) general approach to representing fuzziness (fuzzy sets, probabilistic sets, families of ordinary sets, etc.); b) domain of membership function (standard fuzzy sets, level fuzzy sets, fuzzy numbers, etc.); c) co-domain of membership function (conventional $[0,1]$ -fuzzy sets, $[-1,+1]$ -fuzzy sets, L-fuzzy sets, fuzzy sets of type 2, etc.); d) the type of correspondence between the domain and co-domain (one-to-one, one-to-many). Lattice products are widely used to extend co-domains of membership functions. For instance, de Luca and Termini [8] defined $[0,1]^m$ -fuzzy sets to model the situations when the family of fuzzy sets $\{A\}_m$ corresponds to m properties of considered object and each element $x \in X$ is characterized by a vector of membership values $(\mu_1(x), \dots, \mu_m(x))$. Later on Koczy introduced the concept of vector-valued fuzzy set as a mapping $\mu: X \rightarrow F_1 \times F_m$, where $F_i, i = 1, \dots, m$ – are bounded linearly

ordered sets. A very interesting generalization of L-fuzzy set is Kaufmann's heterogeneous fuzzy set [17] where various elements of X may take values from different lattices.

Below we shall extend modern functional-axiomatic approach to the specification of basic logical operations to lattice square and bilattice case.

4. Negations on lattice squares and bilattices

The classes of negations on the real unit interval were investigated by many authors (e.g. by Alsina, Trillas and Valverde [1], Ovchinnikov [19], Weber [27], Fodor and Roubens [14]). Mainly involutions were thoroughly studied, whereas some interesting non-involutive negations (for instance, contracting and expanding negations) were introduced by Batyrshin [4]. Negation classes on lattices and posets were discussed by Yager [28] and De Cooman and Kerre [11]. In this section we introduce negations on lattice squares and pseudonegations on bilattices.

Classical negation is a unary operator that specifies the truth value of NOT p for a proposition p ; it can be given pointwisely. The basic structure to define negation is a bounded partially ordered set $C = (X, \leq)$. So a negation operation on $C = (X, \leq)$ is an order reversing permutation of (X, \leq) [7].

Let L be a lattice L with the least element 0 and the greatest element 1 . We recall (see, for instance [1, 3-4, 14, 18]) that the function $n: L \rightarrow L$ is called a *negation* on L if it satisfies the conditions: n1) $n(0) = 1; n(1) = 0$ (boundary condition); n2) $x \leq y \Rightarrow n(x) \geq n(y), \forall x, y \in L$ (antimonotonicity). Such negations that also meet the requirement n3) $n(n(x)) = x, \forall x \in L$ (involutivity) are called involutions.

Following [7], we also define the function $a: L \rightarrow L$ called an *affirmation* on L if it meets the following requirements: a1) $a(0) = 0; a(1) = 1$ (boundary condition); a2) $x \leq y \Rightarrow a(x) \leq a(y), \forall x, y \in L$ (monotonicity).

In case of lattice squares and bilattices the concept of negation can be naturally extended and three main classes of negations may be introduced: a) $n^{**}(p) = n^{**}(x, y) = (\text{NOT } \bar{o}, \text{NOT } y)$; b) $n_1^*(p) = n_1^*(x, y) = (\text{NOT } \bar{o}, y)$; c) $n_2^*(p) = n_2^*(x, y) = (\bar{o}, \text{NOT } y)$.

Let L^2 be a lattice square.

Proposition 1. Suppose n is a negation on L , then a function

$n^{**}: L^2 \rightarrow L^2$ that satisfies

$$n^{**}(x, y) = (n(x), n(y)), \forall (x, y) \in L^2 \quad (\text{nn-condition})$$

is a negation on L^2 . The negation n^{**} will be involution if n is involution.

Let us note that generally different classes of n may be used to specify n^{**} , $n^{**}(x, y) = (n_1(x), n_2(y)), \forall (x, y) \in L^2$ ((n_1, n_2) -condition).

For example, in case of $L = [0, 1]$ (NOT x , RATHER NOT y) may be interpreted as $(1-x, 1-y^k)$, where $k \geq 2$.

Example 1 (for simple lattice square “four”, $L^2 = \{0, 1\} \times \{0, 1\}$)

p	(1, 1)	(1, 0)	(0, 1)	(0, 0)
$n^{**}(p)$	(0, 0)	(0, 1)	(1, 0)	(1, 1)

Definition 1. A left negation on BL is a function $n^{1*}: BL \rightarrow BL$ such that $n^{1*}(x, y) = (n(x), a(y))$, $\forall (x, y) \in BL$ (na -condition), where n is a negation with respect to \leq_1 , and a is an affirmation with respect to \leq_2 .

Definition 1c A left negation n^{1*} on BL is called left involution if n is an involution.

Example 2 (for simple bilattice “four”, $BL_4 = \{0, 1\} \times \{0, 1\}$)

p	(1, 1)	(1, 0)	(0, 1)	(0, 0)
$n^{1*}(p)$	(0, 1)	(0, 0)	(1, 1)	(1, 0)

Definition 2. A right negation on BL is a function $n^{2*}: BL \rightarrow BL$ such that

$$n^{2*}(x, y) = (a(x), n(y)), \forall (x, y) \in BL \quad (an\text{-condition}),$$

where a is an affirmation with respect to \leq_1 , and n is a negation with respect to \leq_2 .

Definition 2c A right negation n^{2*} on BL is called right involution if n is an involution.

Example 3 (for simple bilattice “four”, $BL_4 = \{0, 1\} \times \{0, 1\}$)

p	(1, 1)	(1, 0)	(0, 1)	(0, 0)
$n^{2*}(p)$	(1, 0)	(1, 1)	(0, 0)	(0, 1)

Note. It is obvious that from formal viewpoint left negation n^{1*} and right negation n^{2*} are not negations. For instance, $n^{1*}(1, 1) = (0, 1)$, $n^{2*}(1, 1) = (1, 0)$, $n^{1*}(0, 0) = (1, 0)$, $n^{2*}(0, 0) = (0, 1)$. They may be seen as mixed negation-affirmation operations or pseudonegations. Hence, left involution and right involution are pseudoinvolutions.

Now let $a(x) = x$, $a(y) = y$. Then we have $n^1(x, y) = (n(x), y)$, $n^2(x, y) = (x, n(y))$. We shall call them ny -involution and xn -involution.

Theorem 1. The (nn) -involution n^{**} and identity operation (trivial involution) $I = (x, y)$ together with left (ny) involution $n^1(x, y)$ and right (xn) involution $n^2(x, y)$ form Klein group.

To show it, we need to verify the conditions of associativity, identity, inversion and composition. Let us depict in Figure 1

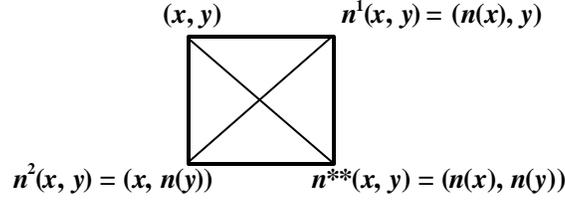


Figure 1. The illustration of relations between involutions and pseudoinvolutions

Let $n^{**}(x, y) = A$, $n^1(x, y) = B$, $n^2(x, y) = C$. Then we obtain

1. Associativity: $A(B(C(x, y))) = B(C(A(x, y))) = C(A(B(x, y)))$.
2. Identity condition: $B(A(C(x, y) = (x, y)$ (triple negation law),
 $I B = B$, $I A = A$, $I C = C$.
3. Inversion condition (homogeneous double negation law):
 $B(B(x, y)) = (x, y)$, $A(A(x, y) = (x, y)$, $C(C(x, y)) = (x, y)$.
4. Composition (heterogeneous double negation law):
 $C(x, y) = A(B(x, y))$, $A(x, y) = C(B(x, y))$, $B(x, y) = A(C(x, y))$.

5. Triangular norms and conorms on bilattices

Triangular norms and conorms (shortly referred as t-norms and t-conorms) were introduced by Menger in the context of stochastic geometry: his aim was to extend the triangle inequality. Later on they were elaborated by Schweizer and Sklar [22] within the framework of probabilistic metric spaces. In fuzzy set theory they were first introduced by Alsina, Trillas and Valverde [1]. Besides, the t-norms (t-conorms) generators and transforms were presented in [3, 14, 18].

Triangular seminorms and semiconorms on $[0, 1]$ as noteworthy generalizations were suggested by Suarez Garcia and Gil Alvarez [23] to deal with fuzzy integrals. In their turn, de Cooman and Kerre proposed an important extension of these concepts by defining t-norms and t-conorms on bounded posets [7].

It is evident that two orderings give rise to two classes of meet and join operations. Below we shall introduce t-norms and t-conorms on lattice squares (viewed as a simplified example of lattice products), as well as we shall discuss about some special hybrids on bilattices.

Let L be a lattice. A binary operation $T: L \times L \rightarrow L$ is called triangular seminorm if: 1) $T(0, 0) = 0$, $T(x, 1) = T(1, x) = x$, $\forall x \in L$ (boundary condition); 2) $x \leq u, y \leq v \Rightarrow T(x, y) \leq T(u, v)$, $\forall x, y, u, v \in L$ (monotonicity). A dual operation called triangular semiconorm is given by $S: L \times L \rightarrow L$,

where: 1') $S(I, I) = I, S(x, 0) = S(0, \bar{0}) = x, \forall x \in L$ (boundary condition);
 2') $x \leq u, y \leq v \Rightarrow S(x, y) \leq S(u, v), \forall x, y, u, v \in L$ (monotonicity).

A triangular norm on L is a t-seminorm that is commutative: 3) $T(x, y) = T(y, x), \forall x, y \in L$ and associative: 4) $T(T(x, y), z) = T(x, T(y, z)), \forall x, y, z \in L$. Similarly a triangular conorm on L is a t-semiconorm that is commutative: 3) $S(x, y) = S(y, x), \forall x, y \in L$ and associative: 4) $S(S(x, y), z) = S(x, S(y, z)), \forall x, y, z \in L$.

Now let L^2 be a lattice square.

Proposition 2. Suppose T is a t-norm on L , then a function $T^{**}: L^2 \rightarrow L^2 \otimes L^2$ such that

$T^{**}(x, y, u, v) = (T(x, y), T(u, v)), \forall (x, y), (u, v) \in L^2$ (TT-condition) is a triangular norm on L^2 .

Example 4. $p \otimes q = (x \wedge y, u \wedge v), \forall (x, y), (u, v) \in L^2$.

Let us note that this greatest lower bound is often referred as generalization operation.

Proposition 3. Suppose S is a t-conorm on L , then a function $S^{**}: L^2 \rightarrow L^2 \otimes L^2$ such that

$S^{**}(x, y, u, v) = (S(x, y), S(u, v)), \forall (x, y), (u, v) \in L^2$ (SS-condition) is a triangular conorm on L^2 .

Example 5. $p \oplus q = (x \vee y, u \vee v), \forall (x, y), (u, v) \in L^2$.

This least upper bound is often referred as unification operation.

The next definitions will be based on bilattices.

Definition 3. A TS operator on BL is a function $TS: BL \rightarrow BL \otimes BL$ such that

$TS(x, y, u, v) = (T(x, y), S(u, v))$ (TS-condition),

where T is triangular norm and S is triangular conorm.

Example 6. $p \wedge q = (x \wedge y, u \vee v), \forall (x, y), (u, v) \in BL$.

Let us note that this operation is often seen as a natural extension of classical conjunction to bilattices. For instance, in the context of estimation bilattice, conjunction supposes linking positive estimates by AND and negative estimates by OR.

Definition 4. A ST operator on BL is a function $BL: BL \rightarrow BL \otimes BL$ such that

$ST(x, y, u, v) = (S(x, y), T(u, v))$ (ST-condition),

where S is triangular conorm and T is triangular norm.

Example 7. $p \vee q = (x \vee y, u \wedge v), \forall (x, y), (u, v) \in BL$.

Let us note that this operation is often seen as a natural extension of classical disjunction to bilattices. For instance, in the context of estimation bilattice, disjunction supposes linking positive estimates by OR and negative estimates by AND.

Some other interesting extended operations can be easily obtained. For instance a class of Sheffer connectives (called generalized anticonjunctions) $Sh(x, y) = T(n(x), n(y)), \forall x, y \in L$ together with a class of Peirce connectives (generalized antidisjunctions) $Pe(x, y) = S(n(x), n(y)), \forall x, y \in L$ may be extended for lattice square and bilattice case.

6. Generalized fuzzy estimates on bipolar scales

Let $E = \{(x, y) \mid x \in L_1, y \in L_2\}$ be a composite estimate on bipolar scale. Here x is an exact estimate of the positive property A^+ and y is an exact estimate of the negative property A^- . Specifically, if the bipolar scale range is $[-1, +1]$, we have $L_1 = [0, +1]$ and $L_2 = [-1, 0]$ (see [26]).

Definition 5. A generalized fuzzy estimate on bipolar scale is a pair $\mathbf{\mu} = (\mathbf{\mu}_{A^+}, \mathbf{\mu}_{A^-})$, where $\mu_{A^+}(x) \in L_1, \mu_{A^-}(y) \in L_2$.

Let $\mu^1 = (\mu_{A^+}^1(x), \mu_{A^-}^1(y))$ and $\mu^2 = (\mu_{A^+}^2(x), \mu_{A^-}^2(y))$ be two generalized fuzzy sets defined on $L_1 \times L_2$. We shall specify the main operations over these fuzzy sets in a following way.

Inclusion: $\mu^1 \subset \mu^2 \Leftrightarrow (\mu_{A^+}^1(x) \leq \mu_{A^+}^2(x), \mu_{A^-}^1(y) \geq \mu_{A^-}^2(y)), \forall (x \in L_1, y \in L_2)$

Equality: $\mu^1 = \mu^2 \Leftrightarrow (\mu_{A^+}^1(x) = \mu_{A^+}^2(x), \mu_{A^-}^1(y) = \mu_{A^-}^2(y)), \forall (x \in L_1, y \in L_2)$

Negation: $-\mu \Leftrightarrow (\mu_{A^-}(y), \mu_{A^+}(x)), \forall (x \in L_1, y \in L_2)$

Intersection: $\mu^1 \wedge \mu^2 \Leftrightarrow (\mu_{A^+}^1(x) \wedge \mu_{A^+}^2(x), \mu_{A^-}^1(y) \vee \mu_{A^-}^2(y)), \forall (x \in L_1, y \in L_2)$

Union: $\mu^1 \vee \mu^2 \Leftrightarrow (\mu_{A^+}^1(x) \vee \mu_{A^+}^2(x), \mu_{A^-}^1(y) \wedge \mu_{A^-}^2(y)), \forall (x \in L_1, y \in L_2)$

It can be easily shown that these intersection and union operation satisfy commutative, associative, mutual distributive and idempotent laws. The negation is involutive. De Morgan laws hold too. So the algebra of generalized fuzzy estimates is a subclass of De Morgan algebras.

7. Conclusion

A rather general lattice product (and specifically, lattice square and bilattice) framework for building hybrid logics and composite fuzzy models has been introduced. Its further development may stimulate the generation of both new fuzzy theoretical approaches and useful fuzzy application techniques in many fields, including the area of economics and finance.

The author is grateful to I. Batyrshin for valuable suggestions and comments.

References

[1] Alsina C., Trillas E, Valverde L. On Some Logical Connectives for Fuzzy Sets Theory, *Journal of Mathematical Analysis and Applications*, 1983, vol.93, p.15-26.

- [2] Atanassov K.T. Intuitionistic Fuzzy Sets, *Fuzzy Sets and Systems*, 1986, vol.20, p.87-96.
- [3] Averkin A.N., Batyrshin I.Z., Blishun A.F., Sylov V.B., Tarassov V.B. Fuzzy Sets in the Control and Artificial Intelligence Models. Nauka, Moscow, 1986 (in Russian).
- [4] Batyrshin I.Z. On the Structure of Involutive, Contracting and Expanding Negations, *Fuzzy Sets and Systems*, 2003, vol.139, p.661-672.
- [5] Belnap N. A Useful Four-Valued Logic. In J.M.Dunn and G.Epstein (eds.) *Modern Uses of Multiple-Valued Logic*. Dordrecht, D.Reidel, 1977, p.837.
- [6] Birkhoff G. Lattice Theory. Providence, Rhode Island, 1967.
- [7] De Cooman G., Kerre E.E. Order Norms on Bounded Partially Ordered Sets, *The Journal of Fuzzy Mathematics*, 1994, vol.2, 12, p.281-310.
- [8] De Luca A., Termini S. Entropy of L-Fuzzy Sets, *Information and Control*, 1974, vol.24, p.55-73.
- [9] Dubois D., Prade Í. Theorie des possibilites. Masson, Paris, 1987.
- [10] Dubois D., Prade Í. Twofold Fuzzy Sets and Rough Sets, *Fuzzy Sets and Systems*, 1987, vol.23, p.3-18.
- [11] Dunn J.M. Partiality and its Dual, *Studia Logica*, 2000, vol.66, p.225-256.
- [12] Finn V.C. On a Version of Argumentation Logic, *Scientific and Technical Information. Series 2. Information Processes and Systems*, 1996, 156, p.3-19 (in Russian).
- [13] Fitting M. Bilattices and the Theory of Truth, *Journal of Philosophical Logic*, 1989, vol.18, p.225-256.
- [14] Fodor J., Roubens M. Fuzzy Preference Relations and Multi-Criteria Decision Support. Kluwer Academic Publishers, Dordrecht, 1994.
- [15] Ginsberg M. Multi-Valued Logics: a Uniform Approach to Reasoning in AI, *Computer Intelligence*, 1988, vol.4, p.256-316.
- [16] Goguen J. L-Fuzzy Sets, *Journal of Mathematical Analysis and Applications*, 1967, vol.18, p.175-174.
- [17] Kaufmann A. Introduction a la theorie de sous-ensembles flous. Vol.3. Masson, Paris, 1974.
- [18] Novak V., Perfilieva I., Mockor J. Mathematical Principles of Fuzzy Logic. Kluwer Academic Publishers, Dordrecht, 2000.
- [19] Ovchinnikov S.V. General Negations in Fuzzy Sets Theory, *Journal of Mathematical Analysis and Applications*, 1983, vol.92, p.234-239.
- [20] Pospelov D.A. The Gray and/or the-Black-and-White, *Applied Ergonomics (Special Issue on Reflexive Processes)*, AEA, Moscow, 1994, p.25-29.
- [21] Ren P. Generalized Fuzzy Sets and Representation of Incomplete Knowledge, *Fuzzy Sets and Systems*, 1990, vol.36, p.91-96.
- [22] Schweizer B., Sklar A. Probabilistic Metric Spaces. North-Holland, Amsterdam, 1983.
- [23] Suarez Garcia F., Gil Alvarez P. Two Families of Fuzzy Integrals, *Fuzzy Sets and Systems*, 1986, vol.18, p.67-81.
- [24] Takeuti G., Titani S. Intuitionistic Fuzzy Logic and Intuitionistic Fuzzy Set Theory, *Journal of Symbolic Logic*, 1984, vol.49, 13, p.851-866.
- [25] Tarassov V.B. Nontraditional and Hybrid Logics in the Modelling of Intelligent Agents. I. Artificial Agents, Intentional Characteristics and Methods of Their Modeling, *Journal of Computers and Systems Sciences*, 2000, vol.39, 15, p.673-684.
- [26] Tarassov V.B. General Approaches to the Modelling of Soft Estimates and Beliefs in Strategic Decision Engineering, *Proceedings of 2002 IEEE International Conference on Artificial Intelligence Systems (ICAIS 2002, Divnomorskoe, Russia, 5-10 September)*, IEEE, Los Alamitos CA, 2002, p.45-49.
- [27] Weber S. A General Concept of Fuzzy Connectives, Negations and Implications Based on T-norms and T-conorms, *Fuzzy Sets and Systems*, 1983, vol.11, p.115-134.
- [28] Yager R. On the Measure of Fuzziness and Negation. II. Lattices, *Information and Control*, 1980, vol.44, p.236-260.
- [29] Zadeh L. Fuzzy Sets, *Information and Control*, 1965, vol.8, p.338-353.
- [30] Zadeh L. Fuzzy Sets as a Basis for a Theory of Possibility, *Fuzzy Sets and Systems*, 1978, vol.1, p.3-28.
- [31] Zadeh L. From Computing with Numbers to Computing with Words – From Manipulation of Measurements to Manipulation of Perceptions. In P.P.Wang (ed.). *Computing with Words*. Wiley and Sons, New York, 2001, p.35-68.

Definition of Optimum Allocation of the Service Centres

Leonid Bershtein¹, Alexander Bozhenyuk¹, Igor Rozenberg¹

¹ Scientific-technical center "Information technologies" of
Taganrog State University of Radioengineering
Oktyabrskaya ploschad, 4, 347922 Taganrog, Russia
e-mail: avb@pbox.infotecstt.ru

Abstract. In this paper the questions of definition optimum allocation of the service centres of some territory are observed. It is supposed that territory is described by fuzzy graph. In this case a task of definition optimum allocation of the service centres transforms into the task of definition of dominating vertex fuzzy set. The example of founding optimum allocation of the service centres as definition of dominating vertex fuzzy set is considered.

1 Introduction

There are many tasks of optimum allocation of the service centres. They are an allocation of radio and TV station in some region; an allocation of military bases, which control some territory; an allocation of shops, which serve some region and so on [4]. We consider that some territory is divided into n areas. There are k service centres, which may be placed into these areas. It is supposed that each centre may be placed into some stationary place of each area. From this place the centre serves all area, and also some neighbor areas with the given degree of service. The service centres can fail during the exploitation (for example, for planned or extraordinary repair). It is necessary for the given number of the service centres to define the places of their best allocation. In other words, it is necessary to define the places of k service centres into n areas such that the control of all territory is carried out with the greatest possible degree of service.

2 Service centres allocation on the base of minimal dominating vertex sets

In this paper we suppose that the service degree of region is defined as the minimal meaning from service degrees of each area. Taking into account, that the service degree can not always have symmetry property (for example, by specific character and relief of the region) the model of such task is a fuzzy oriented graph $\tilde{G} = (X, \tilde{U})$ [2]. Here, set $X = \{x_i\}$, $i \in I = \{1, 2, \dots, n\}$ is a set of vertices and $\tilde{U} = \{\langle \mu_U \langle x_i, x_j \rangle / \langle x_i, x_j \rangle \rangle, \langle x_i, x_j \rangle \in X^2\}$ is a fuzzy set of directed edges with a membership function $\mu_U: X^2 \rightarrow [0, 1]$.

Example 1. Let the region with $n=7$ areas is shown on Fig.1. Let the number of service centers is $k=4$. The fuzzy graph $\tilde{G} = (X, \tilde{U})$, corresponding this region is shown on Fig.2.

The membership function $\{\mu_U \langle x_i, x_j \rangle\}$ of graph $\tilde{G} = (X, \tilde{U})$ defines a service degree of area i in the case when a service center is placed into area j .

The notion of fuzzy dominating vertex set (external stable set) of fuzzy graph \tilde{G} was introduced in work [1].

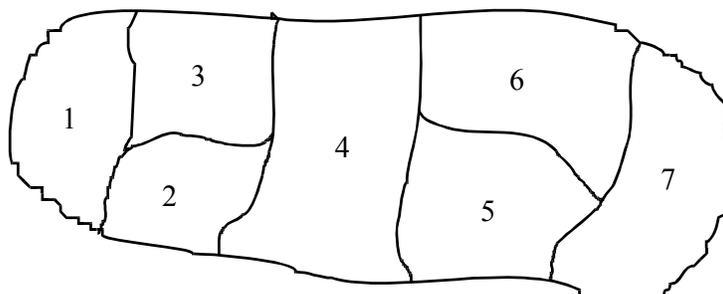


Fig. 1. A region with 7 areas

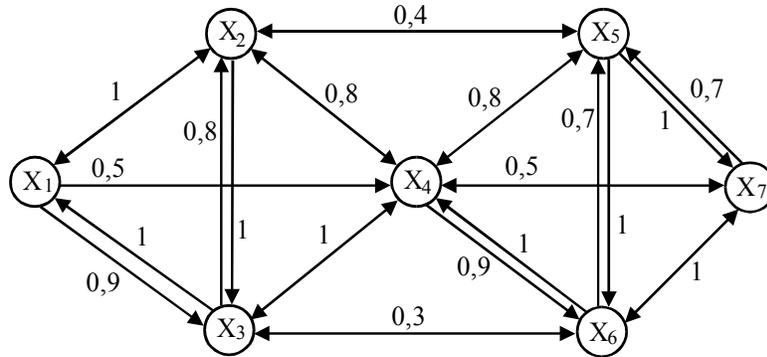


Fig. 2. Fuzzy graph of the region

Let X' be an arbitrary subset of the vertex set X . For each vertex $y \in X \setminus X'$, we define the value:

$$\gamma(y) = \max_{x \in X} \{ \mu_U \langle y, x \rangle \}. \tag{1}$$

Definition 1. The set X' is called a fuzzy dominating vertex set (external stable set) for vertex y with the degree of domination $\gamma(y)$.

Definition 2. The set X' is called a fuzzy dominating vertex set for the graph \tilde{G} with the degree of domination $\beta(X') = \min_{y \in X \setminus X'} \gamma(y)$. Using (1),

we obtain:

$$\beta(X') = \min_{y \in X \setminus X'} \max_{x \in X'} \{ \mu_U \langle y, x \rangle \}. \tag{2}$$

Note 3. In the degenerate case where the graph \tilde{G} is a nonfuzzy graph G , the value β has the value 0 or 1 for any subset X' . If $\beta(X') = 1$, then X' coincides with one of the dominating vertex sets of the nonfuzzy graph G . If $\beta(X') = 0$, then the subset X' is not a dominating vertex set of the graph G .

Now we define a minimal fuzzy dominating vertex set.

Definition 3. A subset $X' \subseteq X$ of graph \tilde{G} is called a minimal fuzzy dominating vertex set with the degree $\beta(X')$ if the condition $\beta(X'') < \beta(X')$ is true for any subset $X'' \subset X'$.

Let a set $\tau_k = \{X_{k1}, X_{k2}, \dots, X_{ki}\}$ be given where X_{ki} is a fuzzy dominating k -vertex set with the degree of domination β_{ki} . We define as $\beta_k^{\min} = \max\{\beta_{X_{k1}}, \beta_{X_{k2}}, \dots, \beta_{X_{ki}}\}$. In the case $\tau_k = \emptyset$ we define $\beta_k^{\min} = \beta_{X_{k+1}}^{\min}$.

Volume β_k^{\min} means that fuzzy graph \tilde{G} includes k -vertex subgraph with the

degree of domination β_k^{\min} and doesn't include k-vertex subgraph with the degree of domination more than β_k^{\min} .

Definition 4. A fuzzy set $\tilde{B}_X = \{ \langle \beta_1^{\min}/1 \rangle, \langle \beta_2^{\min.x}/2 \rangle, \dots, \langle \beta_n^{\min}/n \rangle \}$ is called a domination fuzzy set of fuzzy graph \tilde{G} .

Property 1. The following proposition is true:
 $0 \leq \beta_1^{\min} \leq \beta_2^{\min} \leq \dots \leq \beta_n^{\min} = 1$.

We shall consider a task of finding of domination fuzzy set with the given degree of domination $\beta \in [0, 1]$. The following algorithm may be proposed:

1⁰. For given fuzzy graph \tilde{G} and given degree of domination β we construct crisp subgraph $G' = (X, U')$, in which edge $\langle x, y \rangle \in U'$, if volume $\mu_U \langle x, y \rangle \geq \beta$;

2⁰. For construct suggraph G' we define a family of all dominating vertex sets.

The finding family is a family of dominating vertex sets of fuzzy graph \tilde{G} which degree of domination more or equal β (for example, by Maghout's method for nonfuzzy graphs [3]).

We will consider the problem of finding a family of all minimal fuzzy dominating vertex sets with the highest degree. The given method is a generalization of the Maghout's method for crisp graphs.

Assume that a set T is a fuzzy dominating vertex set of the fuzzy graph \tilde{G} with the degree $\beta(T)$. Then for an arbitrary vertex $x_i \in X$, at least one of the following conditions must be true (or both of them).

- a) A vertex x_i belongs to the considered set T ;
- b) There is a vertex x_j such that it belongs to the set T with the membership function $\mu_U \langle x_i, x_j \rangle \geq \beta(T)$.

In other words, the following statement is true:

$$(\forall x_i \in X)[x_i \in T \vee (\exists x_j)(x_j \in T \ \& \ \mu_U \langle x_i, x_j \rangle \geq \beta(T))]. \quad (3)$$

To each vertex $x_i \in X$ we assign a Boolean variable p_i that takes the value 1, if $x_i \in T$ and 0 otherwise. We assign a fuzzy variable $\xi_{ij} = \mu_U \langle x_i, x_j \rangle$ for the proposition $\mu_U \langle x_i, x_j \rangle \geq \beta(T)$. Passing from the quantifier form of proposition (3) to the form in terms of logical operations, we obtain a true logical proposition:

$$\Phi_T = \&_i (p_i \vee \vee_j (p_j \ \& \ \xi_{ij})) = 1.$$

Supposing $\xi_{ii} = 1$ and considering that the equality $p_i \vee \bigvee_j p_i \& \xi_{ij} = \bigvee_j p_j \xi_{ij}$ is true for any x_i , we finally obtain:

$$\Phi_T = \bigwedge_i \bigvee_j (\xi_{ij} p_j) = 1. \quad (4)$$

We open the parentheses in the expression (4) and reduce the similar terms using the following rule:

$$\xi' \& x \vee \xi'' \& x \& y = \xi' \& x, \text{ for } \xi' \geq \xi'' . \quad (5)$$

Here, $x, y \in \{0,1\}$ and $\xi', \xi'' \in [0,1]$.

Then the expression (4) may be write as:

$$\Phi_T = \bigvee_{i=1, \overline{j}} (p_{i_1} \& p_{i_2} \& \dots \& p_{i_k} \& a_i). \quad (6)$$

Property 2. Each disjunctive member in the expression (4) gives a minimal dominating vertex set with the largest degree of domination a_i .

Proof. We shall suppose that a following simplification is impossible in the expression (6). Let the disjunctive member

$$(p_1 \& p_2 \& \dots \& p_k \& a), k < n, a \in (0,1]. \quad (7)$$

is included into the expression (6). We suppose a subset $X' = \{x_1, x_2, \dots, x_k\}$ isn't domination set with the degree of domination a . Hence, there is a vertex $x_{k+1} \in X/X'$ for which the statement $(\forall i = \overline{1, k})(\mu_U < x_{k+1}, x_i \gg a)$ is true.

We rewrite the expression (4) as:

$$\begin{aligned} \Phi_T = & (1p_1 \vee \xi_{12}p_2 \vee \dots \vee \xi_{1n}p_n) \& (\xi_{21}p_1 \vee 1p_2 \vee \dots \vee \xi_{2n}p_n) \& \\ & \& \dots \& (\xi_{k+1,1}p_1 \vee \xi_{k+1,2}p_2 \vee \dots \vee \xi_{k+1,k}p_k \vee 1p_{k+1} \vee \dots \vee \xi_{k+1,n}p_n) \& \\ & \dots \& (\xi_{n1}p_1 \vee \xi_{n2}p_2 \vee \dots \vee 1p_n). \end{aligned} \quad (8)$$

We have $\xi_{k+1,i} < a, \forall i = \overline{1, k}$ in the expression (8). Hence, if disjunctive members in the expression (6) don't include variables $p_{k+1}, p_{k+2}, \dots, p_n$ then they must include coefficients less than a . Then the disjunctive member (7) can not be included in the expression (6). We obtain the contradiction. Then the subset $X' = \{x_1, x_2, \dots, x_k\}$ is domination set with the domination degree a .

Now we shall prove that the subset X' is a minimal dominating vertex set.

Let us assume that it isn't true. Then one of the following conditions must be true.

a) The subset X' is a minimal dominating vertex set with the degree of domination $b > a$.

b) There is a dominating vertex subset $X'' \subset X'$ with the degree of domination a .

Let the condition a) is true. Then the following proposition is true:

$$(\forall x_j, j = \overline{k+1, n})(\exists x_i, i \in \overline{1, k} \mid \mu_U < x_j, x_i \geq b).$$

Let the expression Φ_T is presented by (8). If we multiply parentheses without rule (5) then we obtain n^2 disjunctive members. Each member contains n elements, one of all parentheses of presentation (8). We select one of n^2 disjunctive members by next rule:

- We select the element Ip_1 from the first parentheses, the element Ip_2 from the second parentheses, ..., the element Ip_k from the parentheses k . Then we select the element $\xi_{k+1, i_1} p_{i_1}$ from the parentheses $(k+1)$ such that the index $i_1 \in [1, k]$ and volume $\xi_{k+1, i_1} > b$, the element $\xi_{k+2, i_2} p_{i_2}$ from the parentheses $(k+2)$ such that the index $i_2 \in [1, k]$ and volume $\xi_{k+2, i_2} > b$, ..., the element $\xi_{n, i_{n-k}} p_{i_{n-k}}$ from the parentheses (n) such that the index $i_{n-k} \in [1, k]$ and volume $\xi_{n, i_{n-k}} > b$.

Using the rule (5), the obtained disjunctive member may be present as:

$$(p_1 \ \& \ p_2 \ \& \ \dots \ \& \ p_k \ \& \ b'). \tag{9}$$

In this presentation the volume b' is defined as $b' = \min\{\xi_{k+1, i_2}, \xi_{k+2, i_2}, \dots, \xi_{n, i_{n-k}}\} \geq b > a$. What is why the member (9) reduces the disjunctive member (7). We obtain the contradiction since the expression (6) contains the disjunctive member (7). This contradiction proves that condition a) is impossible.

Let the condition b) is true. Let the subset X'' is defined as $X'' = \{x_1, x_2, \dots, x_{k-1}\}$. Let the expression Φ_T is presented by (8). We select one of n^2 disjunctive members by next rule:

- We select the element Ip_1 from the first parentheses, the element Ip_2 from the second parentheses, ..., the element Ip_{k-1} from the parentheses $(k-1)$. Then we select the element $\xi_{k, i_1} p_{i_1}$ from the parentheses (k) such that the index $i_1 \in [1, k-1]$ and volume $\xi_{k, i_1} \geq a$, the element $\xi_{k+1, i_2} p_{i_2}$ from the parentheses $(k+1)$ such that the index $i_2 \in [1, k-1]$ and volume $\xi_{k+1, i_2} \geq a$, ..., the element $\xi_{n, i_{n-k+1}} p_{i_{n-k+1}}$ from the parentheses (n) such that the index $i_{n-k+1} \in [1, k-1]$ and volume $\xi_{n, i_{n-k+1}} \geq a$.

Using the rule (5), the obtained disjunctive member may be present as:

$$(p_1 \ \& \ p_2 \ \& \ \dots \ \& \ p_{k-1} \ \& \ b). \tag{10}$$

The volume b is defined as $b = \min\{\xi_{k,i_2}, \xi_{k+1,i_2}, \dots, \xi_{n,i_{n-k+1}}\} \geq a$ in this presentation. What is why the member (10) reduces the disjunctive member (7). We obtain the contradiction too. This contradiction proves that condition b) is impossible.

The proof of Property 2 is completed.

The following method of foundation of a minimal dominating vertex sets may be propose on the base of Property 2:

- We write proposition (4) for given fuzzy graph \tilde{G} ;
- We simplify proposition (4) by proposition (5) and present it as proposition (6);
- We define all minimal dominating vertex sets, which correspond the disjunctive members of proposition (6).

Example 2. Find all minimal fuzzy dominating vertex sets for the graph presented in Fig.2. The corresponding expression (4) for this graph has the following form:

$$\begin{aligned} \Phi_T = & (p_1 \vee p_2 \vee 0,9p_3 \vee 0,5p_4) \& (p_1 \vee p_2 \vee p_3 \vee 0,8p_4 \vee 0,4p_5) \& (p_1 \vee 0,8p_2 \vee p_3 \vee p_4 \vee 0,3p_6) \\ & \& \\ & \& (0,8p_2 \vee p_3 \vee p_4 \vee 0,8p_5 \vee 0,9p_6 \vee 0,5p_7) \& (0,4p_2 \vee 0,8p_4 \vee p_5 \vee p_6 \vee p_7) \& \\ & \& (0,3p_3 \vee p_4 \vee 0,7p_5 \vee p_6 \vee p_7) \& (0,5p_4 \vee 0,7p_5 \vee p_6 \vee p_7). \end{aligned}$$

Using the rule (5), we finally obtain:

$$\begin{aligned} \Phi_T = & p_1 p_3 p_6 \vee p_1 p_4 p_6 \vee p_2 p_3 p_6 \vee p_2 p_4 p_6 \vee p_1 p_3 p_7 \vee p_1 p_4 p_7 \vee p_2 p_3 p_7 \vee p_2 p_4 p_7 \vee 0,9p_1 p_6 \vee 0,8p_2 p_6 \\ & \vee 0,9p_3 p_6 \vee 0,8p_1 p_5 p_7 \vee 0,5p_1 p_7 \vee 0,8p_2 p_7 \vee 0,9p_3 p_7 \vee 0,7p_1 p_5 \vee 0,7p_2 p_5 \vee 0,7p_3 p_5 \vee 0,5p_4 \end{aligned}$$

It follows from the last equality that the graph \tilde{G} has 19 fuzzy minimal dominating vertex sets: $T_1 = \{x_1, x_3, x_6\}$ with the domination degree 1; $T_2 = \{x_1, x_6\}$ with the domination degree 0,9; $T_3 = \{x_4\}$ with the domination degree 0,5 and others.

A domination fuzzy set of fuzzy graph \tilde{G} is defined as $\tilde{B}_X = \{<0,5/1>, <0,9/2>, <1/3>\}$.

The domination fuzzy set defines the next optimum allocation of the service centres: If we have 3 or 4 service centres then we must place these centres into areas 1, 3 and 6. The degree of service equals 1 in this case. If we have 2 service centres then we must place both centres into areas 3 and 6. The degree of service equals 0,9 in this case. If we have only one service centre then we must place it into area 4. The degree of service equals 0,5 in last case.

3 Conclusion

The task of definition of optimal allocation of the service centres as the task of definition of domination fuzzy set of fuzzy graph was considered. The method of definition of domination fuzzy set is the generalization of Maghout's method for nonfuzzy graphs. It is necessary to mark that the suggested method is the method of ordered full selection, because these tasks are reduced to the task of covering, i.e. these tasks are NP-complete tasks. However, this method is effective for the graphs which have not homogeneous structure and not large dimensionality. It is necessary to mark too that considered task may be solved by multi-objective performance.

4 Acknowledgements

This work has been supported by the Russian Fond of Fundamental Research project № 03-07-90202.

References

- 1 Bershtein LS, Bozhenyuk AV (2001) Maghout Method for Determination of Fuzzy Independent, Dominating Vertex Sets and Fuzzy Graph Kernels. *J. General Systems* 30: 45-52
- 2 Kaufmann A (1977) *Introduction a la theorie des sous-ensembles flous*. Masson, Paris
- 3 Kaufmann A (1968) *Introduction a la combinatorique en vue des applications*, Dunod, Paris
- 4 Christofides N (1975) *Graph Theory. An Algorithmic Approach*. Academic press, London

Evolutionary Algorithm of Minimization of Intersections and Flat Piling of the Graph Mathematic Models*

L.A. Gladkov, V.M. Kureichik
Taganrog State University of Radio Engineering
leo@tsure.ru; kur@tsure.ru

The problem of minimization of intersections of the graph mathematic models is considered. For its decision evolutionary algorithms are offered. It allows receiving flat piling with local optimum for polynomial time. Complexity of algorithms has a straight-line characteristic.

Introduction

At a solution of different practical problems both of engineering, and economical nature the special place is occupied with optimization problems grounded on classic graph algorithms, such, for example, as minimization of intersections and constructing of flat piling of the graph, constructing the effective coloring problem of the graph etc.

The mining of the new approaches to a solution of all set forth above problems introduces large concern, both from practical, and scientific point of view. It is conditioned by the fact that the problems of the class NP require large bulk of calculus, and in many cases and an exhaustive search. And the complication even of most simple of them is exponentially grows in accordance with magnification of the number of arguments.

1. Formulation

The relevant aspect of constructing of an effective routine of looking up is the problem of synthesizing the effective architecture.

The genetic algorithms (GA), realizing incidentally - directional policy represent a powerful search tool and determination of optimal solution on set of alternatives. As practice demonstrates, the applying of genetic algorithms allows largely to increase performance of accepted solutions at the

* Work is executed at financial support of the Russian Federal Property Fund (the project №02-01-01275)

expense of usage of the probability rules, necessity of set-up of a search pattern under conditions of a particular problem and possibilities of a choice on each pitch of operation of algorithm from set of invariant alternatives [1 - 3].

Thus, GA usage is a quality reductant of the architecture of a retrieval system, troubleshoots the realization of discrete reductant.

Usage of methods of evolutionary adapting for realization of the behavioral patterns on the basis of principles of changes of aircraft attitude (on Darwin and on Lamarck), natural culling, principle «of survival of the strongest», adapting to variation of conditions of the environment and other, allows to troubleshoot of a continuous component retrieval system [4 - 8].

Each block of the given all-level system relates to a process of decision making, that allows to section a solution of a composite problem into the set of more simple subproblems decided sequentially, pursuant to their hierarchy.

The block of analysis and choice of alternatives ensures a choice of the best solution on set of possible alternatives. The choice is carried out depending on values of a certain objective function (fitness). The given block realizes functions of discrete component.

The basic problem of the block of teaching and adapting is the concretization of set of equivocations. During operation of the block a teaching of system, takes place owing to what the size of set of equivocations is moderated by operation of the block of choice. As the instrument for realization of trial functions of the block the methods and rules of developmental adapting permitting to model process of progressing of a certain in-trabreeding population of the individuals and adapting them to variation of the environmental conditions with the purpose of survival can be utilised.

The block of teaching and adapting realizes functions of continuous reductant, constructing and permanently inflecting a dynamic mock-up of a system behavior at the expense of continuous variation of existing and generation of the new rules of behavior.

2. Minimization of intersections and flat piling of the graph

One of the most laborious graph problems. Many algorithms of minimization of intersections are known. The genetic algorithm of minimization of intersections, tendered in the present paper, allows not only to execute a

primal problem, but also to construct in passing alternative of maximal flat piling of the graph on a plain.

The McLain's theorem is trusted to in a basis of algorithm that the graph can be laid on a plain without intersections, if on set of its cycles it is possible to construct some base \mathbf{B}_p , containing \mathbf{p} - cycles such, that each edge of the graph strictly belongs to two cycles, where $\mathbf{p} = \mathbf{m} - \mathbf{n} + 2$ (\mathbf{m} - a number of edges, and \mathbf{n} - a number of tops of the graph). The basic problem of algorithm is the construction of an optimal solution or sets of solutions maximum close to it.

As input data's the modified agendas of a co occurrence of the graph are used, on a foundation of which one of the generation of a cycle is executed. The experimental researches held earlier, have shown, that for an one valued solution of a problem it is enough to investigate a set of the most short cycle of the graph, for example, cycles of length 3 - 4, that allows to reduce a burn-time of algorithm.

After a termination of process of generation of cycles the routine of an estimation of an amount and a quality of obtained set of cycles is started up. As a result of implementation of the given routine the answer to a problem should be obtained: whether there are enough of available cycles for a solution of a delivered problem or it is necessary to return to the previous routine and to prolong process of generation of cycles. After closing-up of the block of an estimation the generated set of cycles is transmitted to the mainframe of algorithm - block of the genetic operatives (BGO), where a study and generation of a solution takes place.

For a solution of a delivered problem the all-level genetic algorithm conditionally called "building" was designed. The following concept was trusted to in its basis.

The problem of determination of planarity and constructing of flat piling of the graph is esteemed as a problem of constructing of some base fitting to established yardsticks. The building of this base runs in some stages:

1 stage - forming of an original baseline solution.

2 stages - complement of an available baseline solution by new speed keys of cycles before reaching the greatest possible extent of infill, in an ideal before infill of \mathbf{p} - discharges. Depending on reached result the process is either completed, or the algorithm passes to a following stage.

3 stages - algorithm attempts by singlestep variations to improve an obtained earlier solution. In a case if it is possible a process is completed, otherwise algorithm passes to a following stage.

4 stages - obtained solution in an aspect of impossibility of further improving is stored, then some modifications are made to the given solution, and the building is iterated from the second stage.

As building methods on the maiden three stages the genetic operatives of selection (operating system), crossover (OK) and culling (OC), adapted to specificity a delivered problem are utilized. The problem of adjusting of the design is executed with the help of the operatives of mutation (OM) and inverse (OI), also modified pursuant to presented requests.

Host material for production of solutions is the earlier formed set of a cycles.

The overall performance of genetic feedback algorithms in many respects depends on a method of encoding of datas. With allowance for specificities of a decided problem a special chromosome was proposed, with a number of discharges equal to $\mathbf{p} + \mathbf{3}$. First \mathbf{p} of chromosome discharges represent numbers of cycles involved in a current solution. The additional three discharges are utilized for estimating the quality of a receivable solution. The discharge $\mathbf{p} + \mathbf{1}$ contains the information about the edges which have been not involved in the given solution, discharge $\mathbf{p} + \mathbf{2}$ - edges involved once and discharge $\mathbf{p} + \mathbf{3}$ - edges, utilized doubly. The values, which can be accepted with a function of estimation (\mathbf{m}) are in a spacing $[0; 1]$, and quality of a solution is more, than more the number of edges involved in a valued solution.

In the left-hand half chromosome there can be signifying and insignificant parts. The discharges of a signifying part are spacefilled with numbers of cycles, and insignificant - with zero points, thus, the higher is the quality of a solution, the less is the length of an insignificant part ($\mathbf{L}_0 \rightarrow \mathbf{0}$). An optimal solution \mathbf{L}_0 will be equal to zero point. I.e. the length of an insignificant part is an additional yardstick of quality estimation of the obtained solution.

Thus, the chromosome objective function of an obtained solution (\mathbf{H}) is an additive function of two variables: $(\mathbf{H}) = (\delta(\mathbf{m}); \mathbf{L}_0)$, and $f(\mathbf{H}) \rightarrow \mathbf{max}$, when $\delta(\mathbf{m}) \rightarrow 2\mathbf{m}$ & $\mathbf{L}_0 \rightarrow \mathbf{0}$.

Different alternatives of forming policy of set of original solutions are foreseen in algorithm.

1. From set of cycles we select edges repeating not more than two times. The cycles, to which one the edges, retrieved as a result of such looking up, belong form a speed key representing an original solution. If there are edges in it repeating more than 2 times, the minimum possible number of cycles containing "superfluous" edges is removed from a speed key and a speed key will be converted into chromosome.

2. The original baseline solution is reshaped at the expense of join of cycles, incidentally selected pair of cycles having as the minimum one common edge.

3. The original baseline solution is reshaped at the expense of series attachment as long as there is such a possibility, or the clandestine solution will not be derivated yet.

If there are the applicable technical facilities rather effective is a macro-evolution principle usage, when two or more populations are in bridge reshaped and developed with periodic interpopulation information exchange.

The set of the interlocutory orders is reshaped by cycle joining having identical edges. And numbers of thus receivable pair edges necessarily should coincide with numbers of edges written in discharges $\mathbf{p} + 1$ or $\mathbf{p} + 2$ baseline original solution. Besides, cycle numbers from set of the interlocutory orders should not coincide with cycle numbers of a participating in a baseline solution. The size of a population of the interlocutory orders depends on quality of an original solution or is set by the user.

After creation of set of the interlocutory orders, the selection for a choice of solutions participating in operation of a crossing-over (OK) is executed. For this purpose the following rules should be executed:

1. Minimum number of concurrences of numbers of edges in discharges $\mathbf{p} + 2$ and $\mathbf{p} + 3$ of challenger and discharge $\mathbf{p} + 3$ of parents.

2. Maximal number of identical edges in discharge $\mathbf{p} + 1$ of challenger and discharge $\mathbf{p} + 3$ of first parents.

3. Maximal number of concurrences of edges from $\mathbf{p} + 2$ of challenger with edges from discharge $\mathbf{p} + 2$ of the first parent chromosome.

On the basis of datas of the rules special all-up argument - function of operability of the interlocutory orders concerning a baseline solution $f^*(\mathbf{H})$ was designed. The quantitative value of function of operability is counted on the basis of counting number concurrence of numbers of edges in different chromosome discharges.

The improving of an original baseline solution is made with the help of the modified operative of a crossover (OK), for the purpose of that from set of the interlocutory orders the member having the best function of operability concerning the maiden parent is selected and is assigned for implementation OK. The dots of a continuous twist in such OK are selected as follows:

1. In the 1st parent 1st the dot of a continuous twist \mathbf{t}_1 is always behind the latter signifying (nonzero) chromosome element \mathbf{h}_i , 2nd the dot \mathbf{t}_2 is selected after a member $\mathbf{h}_i + \mathbf{k}$, where \mathbf{k} – is number of signifying element in 2 mother chromosome.

2. In the 2nd parent the 1st dot is always before the 1st chromosome element, and the 2nd dot is defined as well as in a case with the 1st parent.

As a result of implementation of OK there is a join of signifying parts of parents chromosomes. If a function of operability of the second parent $\delta_m^*(\mathbf{H}) \neq 1$, them from chromosomes of the descendant an edge (or edges),

repeating in the given speed key more than two time are removed. In bridge with removal of edges N_c - cycle are removed to which the given edge belongs. However again appeared in chromosome a cycle can not be deleted.

After OK implementation the count of functions of operability of all interlocutory orders concerning a new baseline solution is made and the solutions having negative estimations are removed from a population.

At deriving a solution not producing improving during considerable number of breeds, the operations of mutation and inverse are applied to it. For this purpose the casual image selects two discharges (signifying and insignificant). After interchanging the signifying member is zeroed, and all edges, inhering to the given cycle, are removed from genes $p + 2$ and $p + 3$. The schema of a fulfilling of an operation of inverse is similar to operation of mutation with the only variance, that for interchanging not one gene chromosome, is selected but a few. The number of interchanged genes is set by a casual image, however it should not be more then the numbers of insignificant discharges of inverted chromosome.

After chromosome deciphering with the help of the block of flat piling it is possible to scan alternatives of piling of the graph on a plain without intersections.

3. Results of experimental research

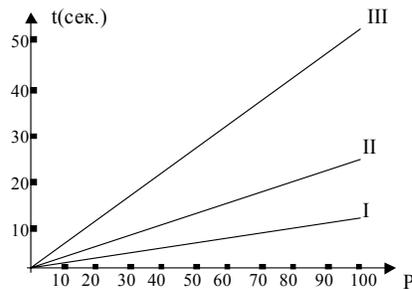
The developed algorithm is effective for the decision optimization problems of definition of planar of graphs models, revealing of the maximal planar part the column and constructions of its flat stacking. The algorithm has polinomial time complexity that proves to be true the received experimental results. The algorithm is realized as the program appendix in programming language C ++ with use of environments of programming Borland C ++ Builder 6.0 and Visual C ++ 7.0. Testing of the developed program was made on personal computers such as IBM PC c by processors Athlon XP 1700 - 2100 and Pentium IV. On the basis of the received data schedules of dependences of time and quality of received decisions from value of dimension of entrance given and basic adjustments {options} of algorithm are constructed.

At carrying out of experimental researches of the developed algorithm three were used graph models: G_1 (25 nodes, 60 edges), G_2 (50 nodes, 120 edges) и G_3 (65 nodes, 130 edges). (Tab. 1).

Table 1. Results of experimental researches

K	N_p	Br	Lo	Bc	$T(G_1)$	$T(G_2)$	$T(G_3)$
25	50	10	5	5	6.99	13.46	15,11
25	100	10	5	5	11.98	19.95	23,24
25	50	20	2	2	7.87	9.85	12,3
25	100	20	2	2	13.9	10.11	21,27
50	50	10	5	5	14.45	9,62	26,98
50	100	10	5	5	25.94	10,11	44,11
50	50	20	2	2	14.95	27,42	24,4
50	100	20	2	2	27.58	37,25	42,46
100	30	10	5	5	19.46	62,1	33,95
100	50	10	5	5	29.62	54,83	48,44
100	100	10	5	5	55.92	65,65	91,63
100	30	20	4	4	18.4	140,29	30,38
100	50	20	4	4	32.2	142,75	46,59
100	100	20	4	4	56.86	137,88	88,66

K - quantity of iterations; N_p - the size of a population; Br - number of base decisions; Lo - number of steps for definition of a local optimum; Bc - number of steps for allocation of normal basis of cycles; $T(G_1 - G_3)$ - operating time of algorithm.

**Fig. 1.** Dependence diagram of an operating time of algorithm from the size of a population

On fig.1 the schedule reflecting dependence of an operating time of algorithm on the size of a population of intermediate decisions is shown.

The conclusion

Designed algorithms has a straight-line characteristic of temporary complication. For study GA the system of the test problems (benchmarks) was designed, intended for matching GA with other techniques and the uniqueness of capacities GA for problem solving of optimization is demonstrated.

However is apparent, that at an exterior simplicity GA require considerable efforts on set-up under a particular problem. Probably, it is meaningful to speak not about complication of applying GA in general, but about adequacy of levels of complication of algorithm and decided problem. The easier problem, the more senseless are different shifts with the coding of genotypes, set-up of odds etc. In case if the fitness has an alone extremum in investigated field, the applying of GA loses any sense, since any local method will discover a solution faster and easier. On the other hand, it is impossible to tell, that there is no such a problem, which could not be decided with the help of GA.

References

1. Kureichik V.M. Genetic algorithms: the Monography. Taganrog: Publishing house TRTU, 1998.
2. Emeljanov V.V., Kureichik V.V., Kureichik V.M. Theory and practice of evolution simulation. – Moscow: “Phismatlit”, 2003.
3. Batishchev D.A. Genetic algorithms of the decision of extreme problems. Voronezh: Publishing house VGTU, 1995.
4. Holland, John H., *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Application to Biology, Control, and Artificial Intelligence*. University of Michigan, 1975.
5. *Handbook of Genetic Algorithms*. Edited by Lawrence Davis, Van Nostrand Reinhold, New York, 1991.
6. *Practical Handbook of Genetic Algorithms. Applications. Volume 1*. Edited by Lawrence D. Chambers. - CRC Press. Boca Raton, London, New York, Washington, D.C., 1995 – 1999
7. Tokinaga S., Yakabe M. An Evolutional Approach to the Control of Chaotic Economic Dynamics. // *Proceeding of 2000 International Symposium of Nonlinear Theory and Applications*. - Dresden, Germany, 2000. - p. 181 - 185.
8. Tokinaga S., Yakabe M. An Evolutional Approach to the Control of Chaotic Economic Dynamics. // *Proceeding of 2000 International Symposium of Nonlinear Theory and Applications*. - Dresden, Germany, 2000. - p. 181 - 185.

Use of a fuzzy strategy to create the shortest Hamiltonian cycle

Henryk Piech , Aleksandra Ptak

Technical University of Czestochowa, Institute of Mathematics and Informatics, tel. 604559300, e-mail:hpiech.adm.pcz.czest.pl,
ul. Dabrowskiego 73

Summary

Hamiltonian cycles, which are a basis of many algebraic analysis methods, as a rule, are subjected to procedures of minimization of their length. Such an approach is characteristic in many problems like: transport (the travelling salesman's problem), network, flow and structural problems. Most often non-directed graphs are met with in which the shortest distance between two nodes is even. The use of fuzzy strategy inference problems can be for example reduced to the selection of the successive added node to a Hamiltonian cycle using its adaptive abilities to the existing structure of connections. Thus, characteristics of the proposed node will be considered together with parameters of the created structure. There is also a possibility to use non-deterministic methods (e.g. genetic algorithms) which are controlled by means of fuzzy characteristics (belonging function, predisposition function or clearance function). Such possibilities from algorithmization and the approach of the optimal solutions perspective will be considered in this paper.

Key words: Hamiltonian cycle, logic and fuzzy modelling, genetic algorithms.

Assumptions and conceptions of the algorithmization of the approximative methods of Hamiltonian cycle length minimization

The basic assumption concerning the cycle (a shortened definition which will be used with reference to the Hamiltonian cycle) is obviously the participation of all graph nodes in it. Another assumption concerns the distances between nodes, and can be formulated as follows:

1. the distance between nodes A and B is the same as between nodes B and A (thus, a non-directed graph will probably have to be dealt with),
2. the distances between A and B nodes are different in both directions.

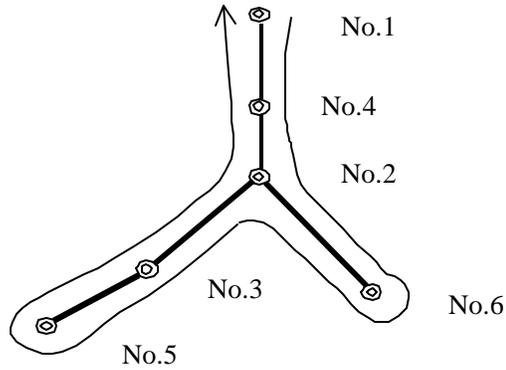
The existing methods of creating a cycle with the minimum length are mainly based on a construction of the shortest spanning tree algorithm. These methods can be rated among approximative groups which do not give a guarantee of the optimal solution obtainment. Figure 1 presents an example of the obtained solution thanks to the use of such a method. Table 1 contains the data for this example that represent distances between all nodes for the non-directed graph construction.

Table 1. Distances between nodes in the non-directed graph

	1	2	3	4	5	6
1	0	33	26	2	34	22
2	33	0	15	21	41	5
3	26	15	0	27	7	19
4	2	21	27	0	53	26
5	34	41	7	53	0	25
6	22	5	19	26	25	0

1. initial node: 1

No.	1	4	2	6	3	5	S↓
From	2	21	5	19	7	34	88



2. initial node: 3

No.	3	2	6	1	4	5		S^-
From		15	5	22	2	53	7	104

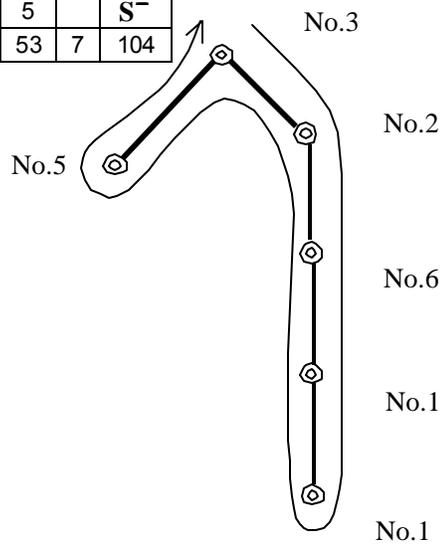


Fig. 1. Creating the two “minimized” Hamiltonian cycles for the initial points: No.1 and No.3 with use of the shortest subspanning tree algorithm (No. – number of associated node, From – metric added to the cycle, S^- denotes the total cycle length)

Adding the shortest arcs to the already existing graphic structure creates the shortest subspanning tree, on condition that they do not create a cycle [10]. In other words, it will be adding a new node which is at the shortest distance from the already built tree.

It is common knowledge that using the shortest subspanning tree method for the shortest Hamiltonian cycle search does not give the optimal solution [10]. The best result obtained in example (see Table 1), using all nodes as start points, amounts to 88, whereas the optimal solution gives 86.

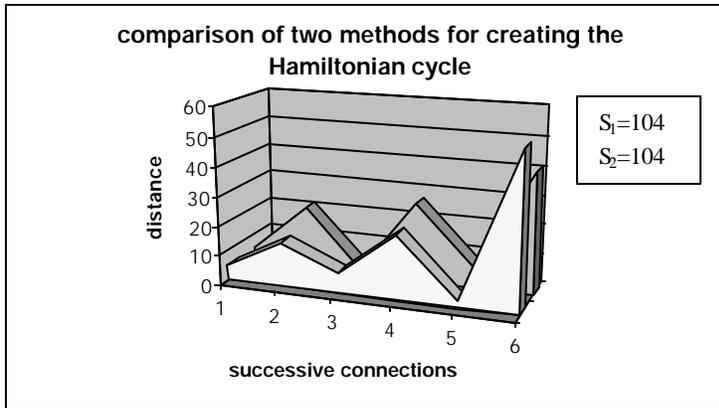
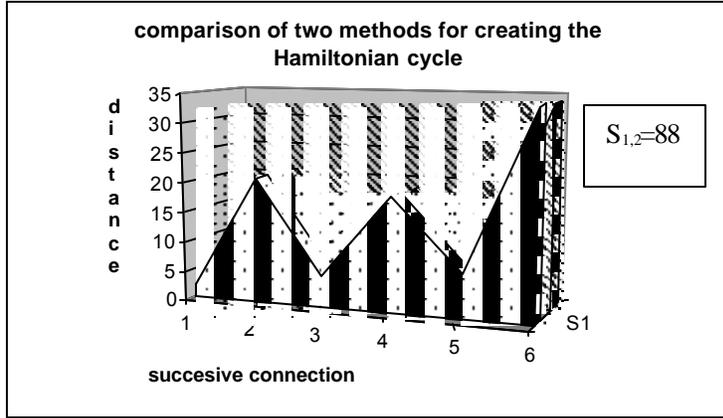
An inclusion of the above described conceptions based on modifications of the approximative method in algorithmic comparative analysis is proposed in this paper. These modifications consist of:

- replacement of the minimal subspanning tree by the shortest single linked list
- or
- replacement of the minimal subspanning tree by the shortest double linked list.

The shortest single linked list is created by adding to the last node another which is in the shortest distance. In the double linked list, however, the successive nodes are added so that they are at the shortest distance from both list ends.

Considering the fact that a by-pass of the subspanning tree or lockup of the list requires the addition of the last segment which is ended at the start point (what can to a crucial extent influence the quality of solution, that is, the cycle length) one is forced to use all possible start points for the analysis. The results obtained for both conceptions of the shortest Hamiltonian cycles searching are quoted below.

The series of investigations have demonstrated 67% effectiveness in finding the optimal solution with the shortest single linked list method and 64% effectiveness by means of the double linked list. The use of the single linked list gave the optimal result (86), however, the use of the double linked list gave the result close to but not equal to the optimal (88). Both methods can be counted among approximate methods similar in the algorithm operation and also in the obtained effects of the method based on the shortest subspanning tree.



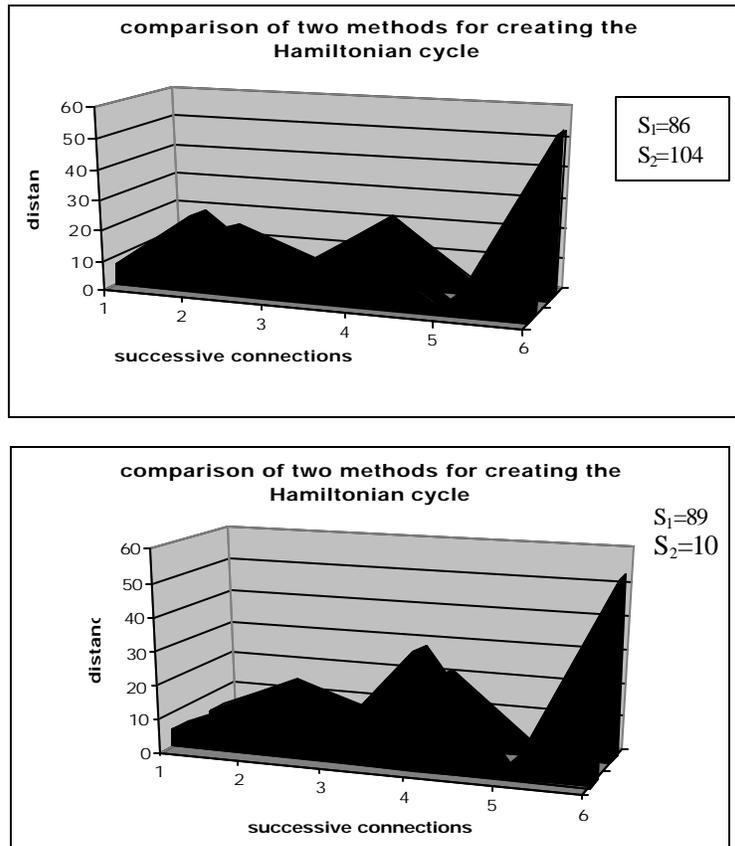
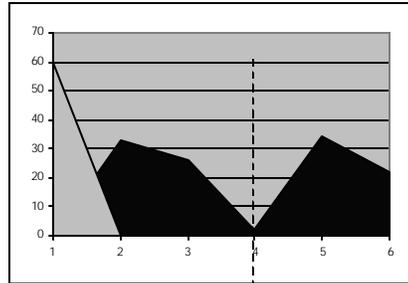


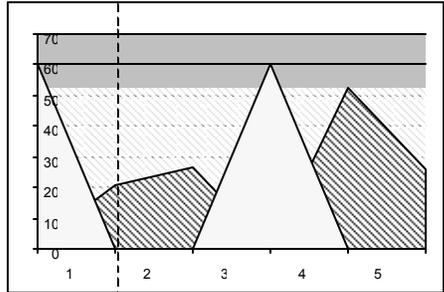
Fig. 2. Comparison between the method realization course based on the shortest single linked list (dotted surface) and the shortest double linked list (line surface).

An observation of the realization course of both methods for the creation of the minimal Hamiltonian cycle forms circumstances creating a conception of “the largest clearance” method. The idea of this method resolves itself into searching places with “the largest clearance” in the characteristic function of the currently analysed node (see Figure 2). This simultaneously is the place where the metric function shows the minimum value. The method essence is shown in Figure 3.

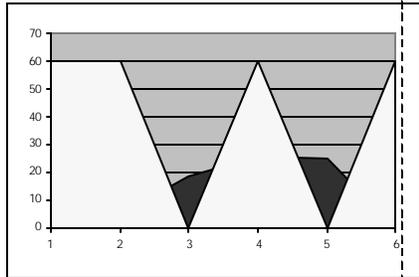
The largest clearance method



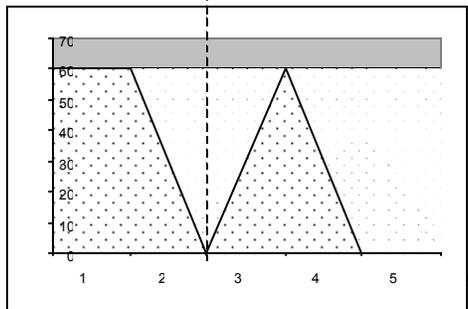
Point 1



Point 4



Point 2



Point 6

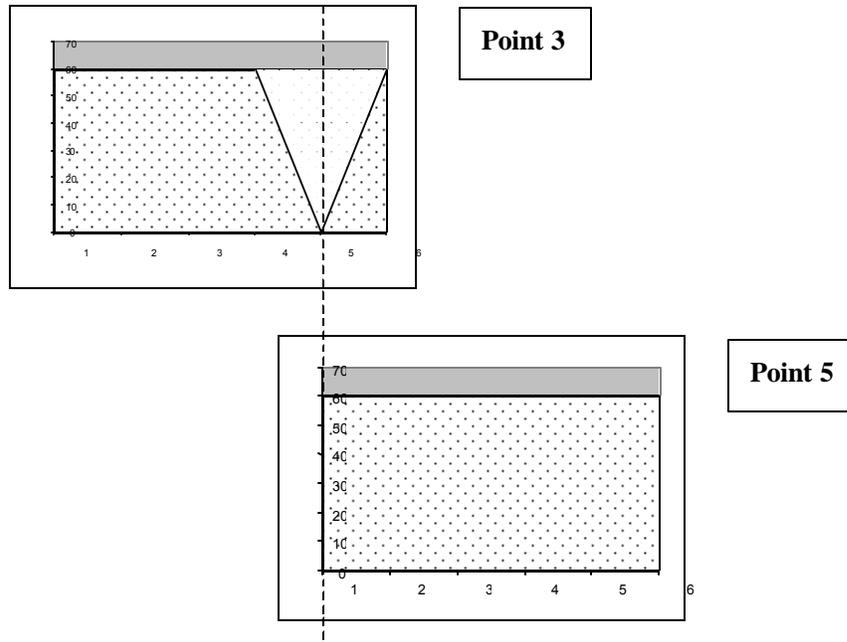


fig. 3. Graphic convention of “the largest clearance” method applied for the searching of the shortest Hamiltonian cycle

Description of “the largest clearance” method:

1. Assignment of the selection schedule of successive nodes as start points.
2. Description of the metric function from remaining nodes for the successive point (start or successive added to the list) - see Figure 6 for the metric function denotation (line surface)
3. Description of the blocking function for added nodes to the structure of the single- and double linked list - see Figure 6 for the blocking function denotation (dotted surface).
4. Search of the “clearance” quantity in the structure of the superimposed metric and blocking functions.
5. Selection of the new node in the place of “the largest clearance” of the superimposed metric and blocking functions.
6. Return to point 2 or end of the list creating the procedure in case of the all added nodes.

7. Start of search for the next start point when the list creation for a given point is completed.

References

- [1] Czogala E., Perdycz W.: Elements and methods of fuzzy sets theory. PWN, Warsaw 1985 (in Polish)
- [2] Davis L.: Genetic Algorithms and Simulated. A. Morgan Kaufmann Pub. Inc., 1988
- [3] Davis L.: Handbook of Genetic Algorithms. Nostrand Reinhold, 1991
- [4] De Groot M. H.: Optimal statistic decisions. McGraw – Hill, New York 1970
- [5] Kacprzyk J.: Fuzzy sets in system analysis. PWN, Warsaw 1986 (in Polish)
- [6] Lachwa A.: Fuzzy world of files, numbers, relations, facts, rules and decisions. Akademicka Oficyna Wydawnicza Exit, Warsaw 2001 (in Polish)
- [7] Lorayne H.: Secrets of supermemory. RAVI, Warsaw 2002 (in Polish)
- [8] Piegat A.: Fuzzy modelling and controlling. Akademicka Oficyna Wydawnicza Exit, Warsaw 2003 (in Polish)
- [9] Rutkowska D., Piliński M., Rutkowski L.: Neural networks, genetic algorithms and fuzzy systems, WN PWN, Warsaw 1997 (in Polish)
- [10] Sys³o M.M.,Deo N.,Kowalik J.S.: Discrete Optimalization Algorithms , WNT,1995.
- [11] Zadeh L. A.: Fuzzy limitations calculus; design and systems; methodological problems. Ossolineum 1980 (in Polish)

Author Index

- Abaoub Ezzeddine 378
Ahmed Uzair 89
Ahrameiko Alexey A. 360
Akbarov R.M. 332
Ali Arshad 89
Aliiev R.A. 368
Alpatsky V.V. 403
Anikin I.V. 202
Augustin Thomas 186
- Barbouch Rym 378
Batyrsin I.I. 170
Batyrsin I.Z. 39, 62, 73, 122, 170
Belluce Peter 265
Berbasova Natalya Y. 360
Bershtein Leonid 283
Bikesheva Gulnara 210
Borisov Arkady 33, 210, 466
Bozhenyuk Alexander 283
- Canfora Gerardo 457
Charfeddine Souhir 350
Cornelis Chris 54
Correa da Rosa Joel 230
- D'Alessandro Vincenzo 457
Davnis V.V. 578
De Baets Bernard 3, 162
De Cock Martine 54
De Coligny Marc 350
De Meyer Hans 3, 162
De Schuymery Bart 162
Di Nola Antonio 265
Dounias Georgios 103
Dymova L. 549
- Emelyanov Viktor V. 311
- Gilmutdinov R. 538
Gladkov L.A. 291
Glova V.I. 202
Gonera M. 549
Grabisch Michel 154
Grishina E. N. 493
- Herrera-Avelar R. 39, 62, 73
Hirota Kaouru 35
- Imanov Q.C. 332
Ionin V.K. 257
Ismagilov I. 394
- Jain V. K. 194
Jończyk M. 484
- Kacprzyk Janusz 501
Kaino Toshihiro 35
Karpov Yury 557
Katasev A.S. 202
Kerre Etienne 54
Kitchens Fred L. 386
Klimova Angelica 122, 130
Kokosh A. 423
Korchunov Valentin 474
Korolkov Mikhail 521
Kuchta Dorota 438
Kureichik V.M. 249, 291
Kureichik V.V. 249

Lajbcygier Paul 97
Leks Dariusz 570
Lettieri Ada 265
Lyubimov Boris 557

Mamedov F.C. 332
Mammadli S. F. 368
Marček Dušan 81
Mareš Milan 143
Medeiros Marcelo 230
Mora-Camino Félix 350
Muzzioli Silvia 222

Nedosekin A.O. 403, 423, 474, 521, 557
Neocleous Costas 340

Pankevich Olga 103
Phektistov O.N 202
Piech Henryk 299, 570
Pinto Dias Alves Antonio 446
Plesniewicz G.S. 257
Ptak Aleksandra 299

Reynaerts Huguette 222
Rocha-Mier Luis 321
Rozenberg Igor 283
Rozenberg P. 415
Rubtsov A.V. 394
Rudas T. 122
Ruta Dymitr 111
Ryjov Alexander 511

Salakhutdinov R.Z. 394
Schizas Christos 340
Segeda Anna 521
Sergei V. Ivliev 565
Sevastianov P. 415, 484, 549

Sheremetov L. 39, 62, 73, 321
Shtovba Serhiy 103
Siniavskaya Olga A. 360
Sitarz Sebastian 238
Suarez R. 39, 62, 73

Tarassov Valery B. 272
Tinyakova V.I. 578
Troiano Luigi 457
Trzaskalik Tadeusz 238

Utkin Lev V. 178, 186

Valishevsky Alexander 466
Veiga Alvaro 230
Velisava T. Sevrouk 565
Villareal Francisco 321

Wyrzykowski R. 549

Yahya Asjad 97
Yarushkina Nadezhda 530
Yazenin A.V. 16

Zadrożny Sławomir 501
Zaheeruddin 194
Zaib Gul 89
Zhelezko Boris A. 360

