

## **Формулировка модели идентификации налогоплательщиков на основе системы нечеткого вывода**

Гонохова В.А., аспирант ДВГТУ

Задача отбора налогоплательщиков, требующих дополнительных мер налогового контроля – это задача построения математической модели по результатам наблюдений, или другими словами задача идентификации объектов. Идентификация с помощью нечеткого логического вывода является эффективным методом построения моделей, базирующихся на лингвистических высказываниях типа «Если расхождение между выручкой от реализации и стоимостью отгруженной продукции высоко, то предприятие требует дополнительных мер налогового контроля» [4, 5]. Поэтому в качестве инструмента разработки экономико-математической модели идентификации состояний налогоплательщиков выбрана система нечеткого вывода.

*Нечетким логическим выводом* называется получение заключения в виде нечеткого множества, соответствующего текущим значениям входов, с использованием нечеткой базы знаний и нечетких операций [5]. Для этого системы нечеткого вывода должны содержать базу правил нечетких продукций и реализовывать нечеткий вывод на основе посылок и условий, представленных в форме нечетких лингвистических высказываний. Информацией, которая поступает на вход системы нечеткого вывода, являются измеренные некоторым образом входные переменные. Эти переменные соответствуют реальным переменным процесса управления. Информация, которая формируется на выходе системы нечеткого вывода, соответствует выходным переменным, которыми являются управляющие переменные процесса управления. Типовая модель на основе нечеткого логического вывода (Рис. 1) содержит следующие блоки [6]:

- фаззификатор, преобразующий фиксированный вектор влияющих факторов в вектор нечетких множеств, необходимых для выполнения нечеткого логического вывода;

- нечеткая база знаний, содержащая информацию о зависимости выходной переменной от входных переменных в виде лингвистических правил нечетких продукций;
- машина нечеткого логического вывода, которая на основе правил базы знаний определяет значение выходной переменной в виде нечеткого множества, соответствующего нечетким значениям входных переменных;
- дефаззификатор, преобразующий выходное нечеткое множество в четкое число.

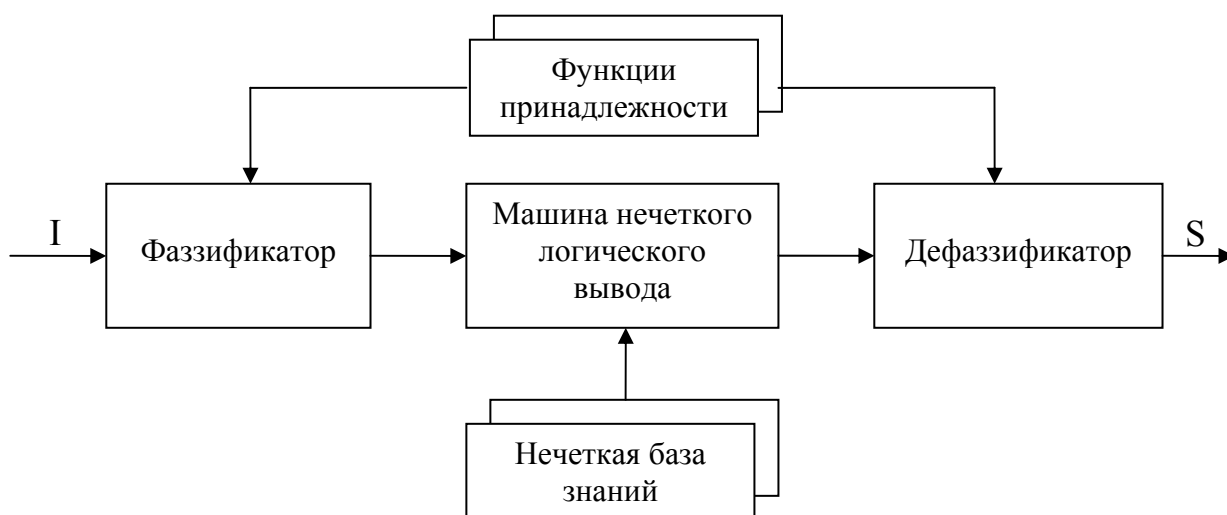


Рис. 1. Типовая структура модели нечеткого логического вывода

Разработка и применение систем нечеткого вывода включают в себя ряд этапов, которые могут быть реализованы неоднозначным образом, поскольку включают в себя отдельные параметры, которые должны быть фиксированы и специфицированы. Наиболее используемыми являются алгоритмы Мамдани [3, 5] и Сугэно [3, 5].

В результате компьютерных экспериментов Штовбой С.Д. [6] установлено, что использование лингвистической информации в виде экспертных правил «ЕСЛИ–ТО» позволяет значительно снизить необходимый объем обучающей выборки для нечеткой идентификации. При больших

объемах выборки экспериментальных данных идентификация с помощью модели типа Сугэно обеспечивает, как правило, большую точность. Однако при этом возникают трудности с содержательной интерпретацией параметров нечеткой модели и с объяснением логического вывода. С моделью типа Мамдани таких трудностей не возникает, ее параметры и после обучения легко интерпретируются содержательно. Процедура нечеткого логического вывода в модели типа Мамдани интуитивно понятна пользователям модели (в нашем случае – налоговым инспекторам). Так как для решения задачи отбора налогоплательщиков, требующих дополнительных мер налогового контроля более важным является объяснение, обоснование принятого решения (см. главу 1), то при построении модели идентификации состояний налогоплательщиков использовался алгоритм Мамдани [5]. Диаграмма основных этапов алгоритма представлена на Рис. 2., рассмотрим каждый из них подробнее.

## 1. Формирование базы правил

База правил систем нечеткого вывода предназначена для формального представления эмпирических знаний или знаний экспертов в той или иной проблемной области. В системах нечеткого вывода используются *правила нечетких продукций*, в которых условия и заключения сформулированы в терминах нечетких лингвистических высказываний [3]. Совокупность таких правил называется базой нечетких продукций [3].

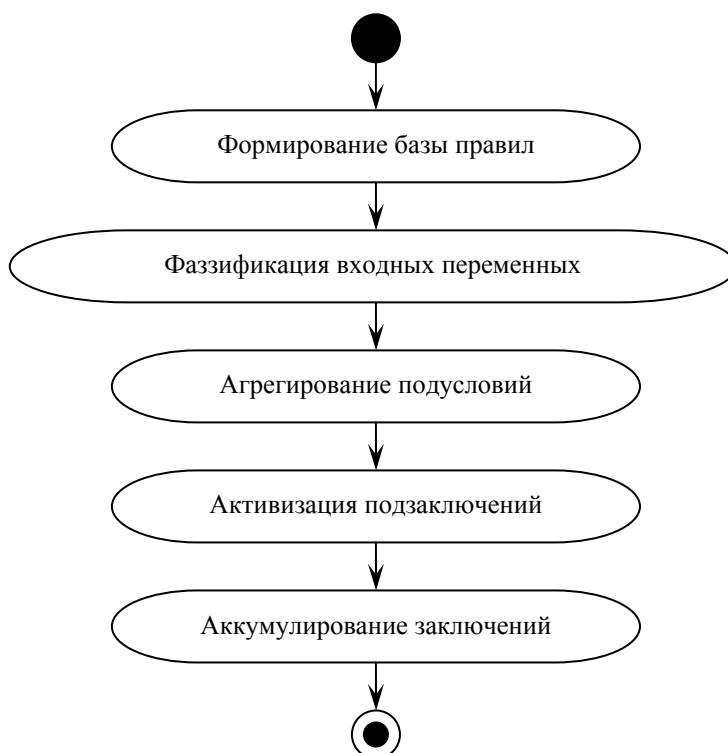


Рис. 2. Процесса нечеткого вывода в форме диаграммы деятельности языка UML

При задании базы нечетких продукций необходимо, во-первых, определить множество входных лингвистических переменных. В разработанной диссертантом модели в их роли выступают показатели, характеризующие состояние налогоплательщика  $I = \{i_1, i_2, \dots, i_m\}$ . Такими переменными являются, к примеру, отношение темпов роста объема лова к темпам роста выручки от реализации, темпы роста объема добытого полезного ископаемого, темпы роста цены реализации добытого полезного ископаемого. Или отношение стоимости отгруженной продукции, согласно грузовым таможенным декларациям (ГТД) к выручке от реализации продукции, согласно налоговой и бухгалтерской отчетности налогоплательщика (показатель  $i_1$ ). Достаточно высокая величина указанного показателя сигнализирует о возможных нарушениях налогового законодательства в части формирования налогооблагаемой базы по налогу на прибыль.

Во-вторых, следует определить выходную лингвистическую переменную, которой в нашем случае является уровень подозрительности налогоплательщика  $S$ .

В-третьих, необходимо сформулировать множество правил нечетких продукций  $R=\{r_1, r_2, \dots, r_n\}$ . Данные правила определяют взаимосвязи между входными и выходной переменными и представляются в форме структурированного текста [3]:

**ЕСЛИ «Условие 1» ТО «Заключение 1»**

Например, описанную выше зависимость можно представить в виде следующего правила нечетких продукций:

**ЕСЛИ « $i_1$  высокий», ТО « $S$  высокий»**

Все лингвистические переменные в базе нечетких продукций представляются как нечеткие множества, заданные соответствующими функциями принадлежности.

База правил систем нечеткого вывода может трактоваться как некоторое разбиение пространства влияющих факторов на подобласти с размытыми границами, в каждой из которых функция отклика принимает значение, заданное соответствующим нечетким множеством. Правило в базе знаний представляет собой «информационный сгусток», отражающий одну из особенностей зависимости «входы–выход». Такие «сгустки насыщенной информации» или «гранулы знаний» могут рассматриваться как аналог «вербального кодирования, которое, как установили психологи, происходит в человеческом мозге при обучении» [6]. Поэтому формирование нечеткой базы знаний в конкретной предметной области, обычно не составляет особых трудностей.

## 2. Фаззификация

В контексте нечеткой логики под *фаззификацией* [4] понимается не только отдельный этап выполнения нечеткого вывода, но и собственно процесс

или процедура нахождения значений функций принадлежности нечетких множеств (термов) на основе обычных (не нечетких) исходных данных. Фаззификацию еще называют «введением нечеткости» [3].

Целью этапа фаззификации является установление соответствия между конкретным значением отдельной входной переменной системы нечеткого вывода и значением функции принадлежности соответствующего ей термина входной лингвистической переменной. После завершения этого этапа для всех входных переменных должны быть определены конкретные значения функций принадлежности по каждому из лингвистических термов, которые используются в подусловиях базы правил системы нечеткого вывода.

*Функции принадлежности* для входных и выходной переменных представляют собой термы трапециевидных функций принадлежности (более подробно см. в Приложении 1), конкретные формы которых определялись методом статистической обработки экспертной информации [1, 5].

Нечеткие описания в структуре модели появляются в связи с неуверенностью эксперта, которая возникает в ходе классификации уровня факторов. Например, эксперт не может четко разграничить понятия «высокой» и «максимальной» цены реализации. Или провести границу между средним и низким уровнем значения отношения выручки от реализации к статистической стоимости отгруженной продукции. Тогда применение нечетких описаний производится следующим образом. Эксперт фиксирует показатель и его количественное значение. Затем эксперт строит лингвистическую переменную со своим терм-множеством значений. Например, переменная « $i_1$ » может обладать терм-множеством значений, «Низкий», «Средний», «Высокий». Далее эксперт каждому значению лингвистической переменной (которое, по своему построению, является нечетким подмножеством значений интервала  $(0,1)$  – области значений показателя) сопоставляет функцию принадлежности тому или иному нечеткому подмножеству.

В экономико-математической модели идентификации состояний налогоплательщиков используются трапециевидные функции принадлежности (Рис. 3), как наиболее удобные для расчетов [4]. Верхнее основание трапеции соответствует полной уверенности эксперта в правильности своей классификации, а нижнее – уверенности в том, что никакие другие значения интервала  $(0,1)$  не попадают в выбранное нечеткое подмножество.

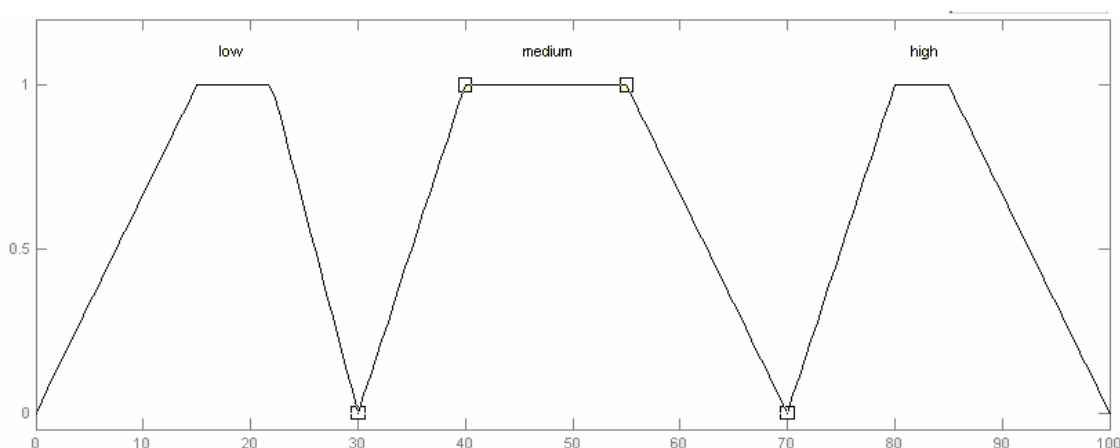


Рис. 3. Функции принадлежности значений лингвистической переменной  $i_1$

Продemonстрируем метод статистической обработки экспертной информации на простейшем примере: построим функции принадлежности термов «Низкий», «Средний», «Высокий», используемых для лингвистической оценки переменной « $i_1$ ». Результаты опроса пяти экспертов приведены в таблице 1.

Таблица 1.

Результаты опроса экспертов

	термы	[0,15)	[15,25)	[25,30)	[30,40)	[40,55)	[55,70)	[70,80)	[80,85)	[85,100)
Эксперт 1	низкий	1	1	1	0	0	0	0	0	0
	средний	0	0	0	1	1	1	0	0	0
	высокий	0	0	0	0	0	0	1	1	1
Эксперт 2	низкий	1	1	1	0	0	0	0	0	0

	термы	[0,15)	[15,25)	[25,30)	[30,40)	[40,55)	[55,70)	[70,80)	[80,85)	[85,100)
Эксперт 3	средний	0	0	0	1	1	1	0	0	0
	высокий	0	0	0	0	0	1	1	1	1
	низкий	1	0	0	0	0	0	0	0	0
Эксперт 4	средний	0	1	1	1	1	1	1	0	0
	высокий	0	0	0	0	0	0	1	1	1
	низкий	1	1	1	0	0	0	0	0	0
Эксперт 5	средний	0	0	1	1	1	0	0	0	0
	высокий	0	0	0	0	0	1	1	1	1
	низкий	1	1	1	0	0	0	0	0	0

Результаты обработки экспертных мнений представлены в таблице 2. Числа над пунктирной линией – это количество голосов, отданных экспертами за принадлежность нечеткому множеству соответствующего элемента универсального множества. Числа под пунктирной линией – степени принадлежности.

Таблица 2.

#### Результаты обработки мнений экспертов

термы	[0,15)	[15,25)	[25,30)	[30,40)	[40,55)	[55,70)	[70,80)	[80,85)	[85,100)
низкий	5	4	4	0	0	0	0	0	0
	1	<b>0.8</b>	<b>0.8</b>	0	0	0	0	0	0
средний	0	1	2	5	5	4	1	0	0
	0	0.2	0.4	<b>1</b>	<b>1</b>	<b>0.8</b>	0.2	0	0
высокий	0	0	0	0	0	3	5	5	5
	0	0	0	0	0	0.6	<b>1</b>	<b>1</b>	<b>1</b>



Таким образом, результатом опроса экспертов для терма «Низкий» являются 3 интервала вещественной оси:  $[0,15)$ ,  $[15,25)$ ,  $[25,30)$ . Тогда множество вершин трапецевидной функции принадлежности:  $(0, 0)$ ,  $(15, 1)$ ,  $(25, 1)$ ,  $(30,0)$ . Таким же образом определяются остальные функции принадлежности. Графики функций принадлежностей термов лингвистической оценки переменной « $i_1$ » показаны на Рис. 3.

### 3. Агрегирование

Агрегирование представляет собой процедуру определения степени истинности условий  $T' = \{t'_k\}$  по каждому из правил системы нечеткого вывода [3].

Формально процедура агрегирования выглядит следующим образом. До начала этого этапа предполагаются известными значения истинности всех подусловий системы нечеткого вывода, то есть множество значений  $T = \{t_i\}$ . Далее рассматривается каждое из условий правил системы нечеткого вывода ( $R_k$ ).

Если условие правила представляет собой нечеткое высказывание вида:

**ЕСЛИ «Условие 1» ТО «Заключение 1»,**

то степень его истинности равна соответствующему значению  $t_i$ .

Если же условие представляет собой нечеткое высказывание вида:

**ЕСЛИ «Условие 1» И «Условие 2» ТО «Заключение 1»,**

или

**ЕСЛИ «Условие 1» ИЛИ «Условие 2» ТО «Заключение 1»,**

то определяется степень истинности сложного высказывания на основе известных значений истинности подусловий. При этом для определения нечеткой конъюнкции (связки «И») используется основная формула *логической конъюнкции нечетких высказываний*.

А для определения результата нечеткой дизъюнкции (связки «ИЛИ») используется основная формула *логической дизъюнкции нечетких высказываний*.

#### 4. Активизация

Активизация в системах нечеткого вывода представляет собой процесс нахождения степени истинности каждого из подзаключений правил нечетких продукций [3]. При формировании базы правил системы нечеткого вывода могут быть заданы весовые коэффициенты для каждого правила, в описываемой модели предполагается, что коэффициенты для всех правил равны 1. Соответственно, значения степеней истинности заключений в нашем случае равны соответствующим значениям  $t'_k$ .

После нахождения степеней истинности определяются функции принадлежности каждого из подзаключений для рассматриваемых входных лингвистических переменных. Для этой цели используется *метод min-активизации*:

$$\mu'(y) = \min \{ t'_k, \mu(y) \},$$

где  $\mu(y)$  – функция принадлежности терма, который является значением выходной переменной  $S$ .

Для иллюстрации выполнения этого этапа рассмотрим пример процесса активизации заключения в следующем правиле нечеткой продукции:

**ЕСЛИ** «величина  $i_1$  высокая», **ТО** «предприятие подозрительное»

Входной лингвистической переменной в этом правиле является  $i_1$ , а выходной переменной является  $S$  – уровень подозрительности налогоплательщика. Предположим, что текущее значение показателя  $i_1$  равно 75%. Поскольку агрегирование условия этого правила  $t'_i = 0.5$ , а весовой коэффициент равен 1 (по умолчанию), то значение 0.5 будет использоваться для получения результата активизации (см. Рис. 4, результат изображен более темным цветом).

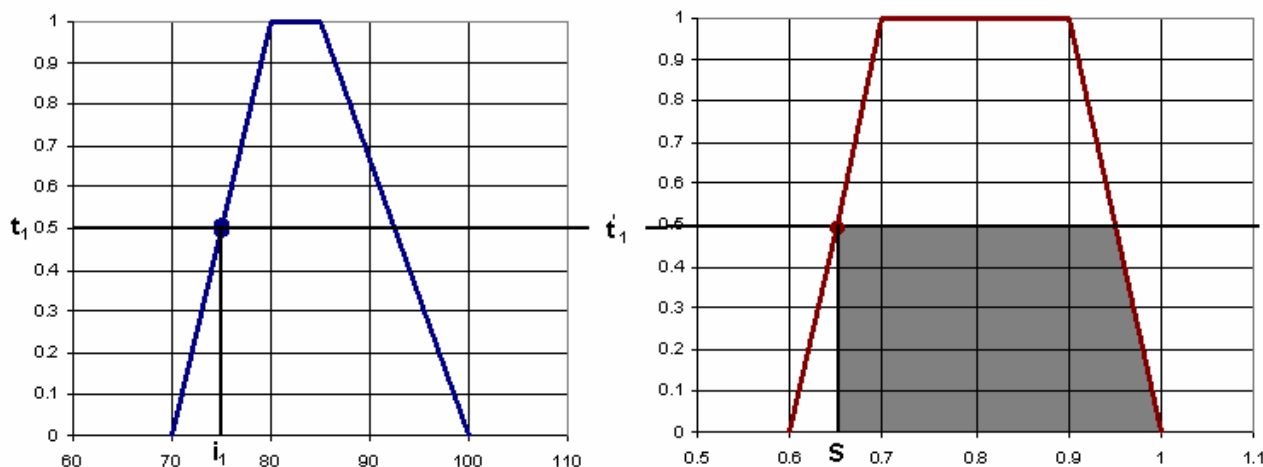


Рис. 4. Пример активизации заключения для правила нечеткой продукции

## 5. Аккумуляция

Аккумуляция в системах нечеткого вывода – это процедура нахождения функции принадлежности выходной переменной  $S$  [3].

Цель аккумуляции заключается в том, чтобы объединить все степени истинности заключений для получения функции принадлежности выходной переменной. Причина необходимости выполнения этого этапа состоит в том, что заключения, относящиеся к одной и той же выходной переменной, принадлежат различным правилам.

## 6. Дефаззификация

Дефаззификация в системах нечеткого вывода представляет собой процедуру или процесс нахождения обычного (не нечеткого) значения выходной переменной  $S$  [4]. Цель дефаззификации заключается в том, чтобы, используя результаты аккумуляции выходной переменной, получить обычное количественное значение выходной переменной.

В модели идентификации состояний налогоплательщиков результат дефаззификации определяется по *методу центра тяжести*. При этом методе обычное значение функции равно абсциссе центра тяжести площади,

ограниченной графиком кривой функции принадлежности соответствующей переменной [3].

Пример дефаззификации методом центра тяжести функции принадлежности выходной лингвистической переменной «уровень подозрительности предприятий» (S) изображен на Рис. 5. В этом случае приближенное значение результата дефаззификации равно 0.45.

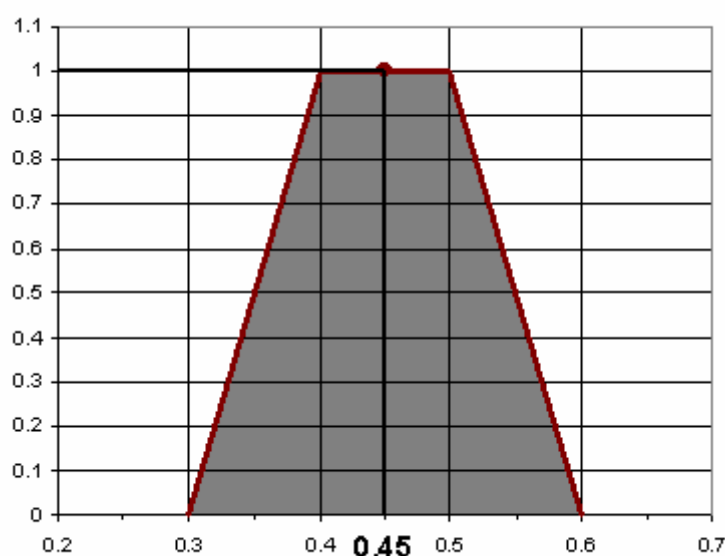


Рис. 5. Пример дефаззификации выходной лингвистической переменной «уровень подозрительности предприятий» методом центра тяжести

При разработке программной реализации вышеописанной модели был проведен сравнительный анализ некоторых инструментов создания систем нечеткого вывода, а также рассмотрена возможность написания собственного программного продукта.

Начиная с 1995 года, на российском рынке стали появляться программные продукты для персональных компьютеров, рассчитанные на их массовое использование [2]. Математический аппарат, предоставляющий такие возможности, детально описанный в специальной литературе и в полной мере реализованный в программных пакетах, спрятан «за кадром», что делает

процесс освоения этих инструментов более доступным и интуитивно понятным для любого пользователя. Поэтому большинство повседневных задач, в которых возникает необходимость приближенного задания условий и, соответственно, получения столь же приближенных результатов, в том числе и задачу идентификации налогоплательщиков, возможно быстро и с приемлемой точностью решать, не прибегая к созданию собственного программного продукта.

Разработанная диссертантом программная реализация экономико-математической модели идентификации состояний налогоплательщиков состоит из нескольких компонент, диаграмма деятельности процесса анализа с ее помощью приведен на Рис. 6.

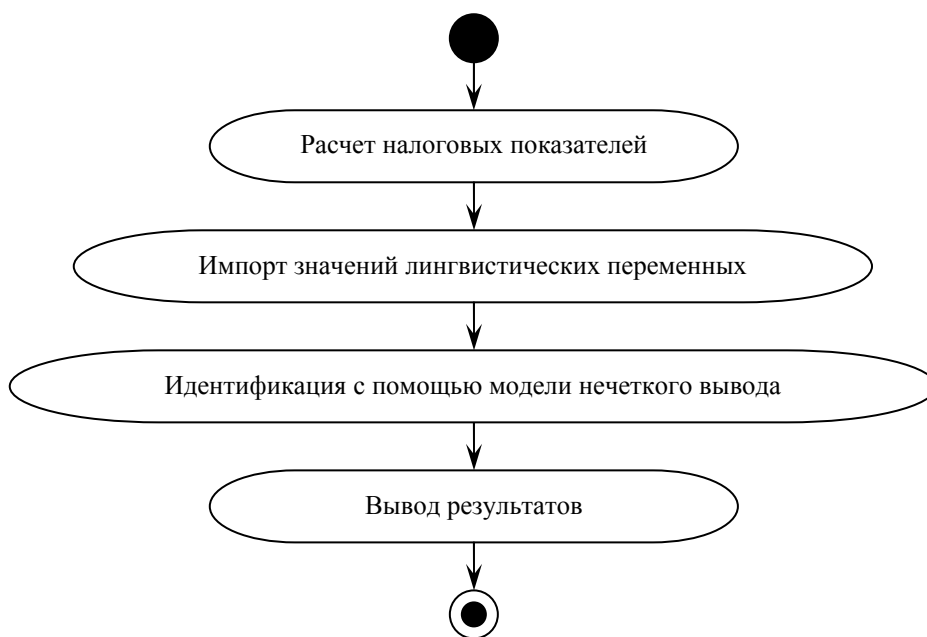


Рис. 6. Процесс идентификации налогоплательщиков с помощью модели нечеткого вывода в форме диаграммы деятельности языка UML

Рассмотрим этапы функционирования указанного программного средства. На первом этапе происходит расчет значений налоговых показателей с помощью скриптов на языке T-SQL на основе данных, содержащихся в

автоматизированной информационной системе налогового органа. Далее с помощью компонента Database Toolbox (версия 3), производится импорт полученных значений из базы данных в модель нечеткого вывода, реализованную с помощью компонента Fuzzy Logic Toolbox (версия 2) системы MatLab (версия 7.0.1). Затем производится анализ значений и выдача на экран результатов.

Разработанная модель удовлетворяет всем требованиям, предъявляемыми к ней сотрудниками налоговых органов.

**Обработка большого объема сведений, содержащихся в базах данных налоговых органов, и выделения из него полезной информации.** В налоговых органах Приморского края уже накоплен огромный объем разнообразных данных о налогоплательщиках, данные находятся в базах данных под MS SQL Server 2000 и недоступны для открытого использования. Набор SQL-скриптов для расчета налоговых показателей с открытым кодом позволяет рассчитывать показатели на сервере налогового органа на основе данных, содержащихся в автоматизированной информационной системе налогового органа. А использование компонента Database Toolbox, позволяющего, находясь в среде MATLAB, использовать команды языка запросов SQL для чтения данных из ODBC-совместимых баз данных (каким и является MS SQL Server 2000), обеспечивает предоставление в модель уже только значений входных переменных.

**Многокритериальность и многофакторность.** Разработчики системы MatLab не ограничивают предельно допустимое количество лингвистических переменных или правил нечетких продукций.

**Учет субъективного характера правил оценки альтернативных вариантов, то есть принятие решений в условиях информационной неопределенности.** Использование теории нечетких множеств и нечеткой логики для создания описанной модели позволяет оперировать с неопределенностями слов и понятий и учитывать в модели качественные

аспекты, не имеющие точной числовой оценки. Оказывается возможным совмещать в оценке учет количественных и качественных признаков, что резко повышает уровень адекватности применяемой методики.

**Прозрачность объяснения и обоснования принятого решения.** При использовании модели нечеткого вывода типа Мамдани не возникает трудностей с содержательной интерпретацией ее параметров и с объяснением логического вывода. Процедура нечеткого логического вывода в модели типа Мамдани интуитивно понятна ее пользователям.

При апробации экономико-математической модели идентификации налогоплательщиков, требующих дополнительных мер налогового контроля, в налоговых органах Приморского края был проведен сравнительный анализ результатов, получаемых с помощью модели, с выводами системы МДК и с данными реальных выездных налоговых проверок по одним и тем же организациям. По результатам анализа установлено, что, во-первых, значение уровня подозрительности в целом соответствует порядку доначислений по результатам ВМП (0.5 – до 50 тыс. руб., 0.8 – от 50 тыс. руб. до 600 тыс. руб., 1 – свыше 600 тыс. руб.). Во-вторых, по нескольким предприятиям, по которым в результате проведения МДК не были выявлены нарушения, а в результате проведения ВМП были произведены доначисления, моделью был рассчитан довольно высокий уровень подозрительности (от 0.8. до 1). В-третьих, по некоторым предприятиям в результате МДК были обнаружены существенные нарушения, но доначисления по результатам ВМП были сравнительно не велики (около 40 тыс. руб.), рассчитанный же уровень подозрительности по указанному предприятию оказался невысоким (0.5).

Таким образом, разработанная модель не только выявляет «подозрительные» предприятия, но и обеспечивает большую точность результатов по сравнению с используемыми в данный момент техниками анализа.



**International Fuzzy Economics Lab (Россия), [www.ifel.ru](http://www.ifel.ru), [info@ifel.ru](mailto:info@ifel.ru)**  
Международная научная лаборатория по внедрению нечетко-множественных  
подходов в экономических исследованиях



## Список литературы

1. Борисов А.Н., Алексеев А.В., Крумберг О.А. и др. Модели принятия решений на основе лингвистической переменной. - Рига: Зинатне, 1982. - 256 с.
2. Круглов В.В., Дли М.И., Голунов Р.Ю. Нечеткая логика и искусственные нейронные сети. Физматлит, 2001. - 224 с.
3. Леоненков А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH. - СПб.: БХВ-петербург, 2003. - 736 с.
4. Ротштейн А.П. Интеллектуальные технологии идентификации: нечеткая логика, генетические алгоритмы, нейронные сети. - Винница: УНИВЕРСУМ-Винница, 1999. - 320 с.
5. Штовба С.Д. Введение в теорию нечетких множеств и нечеткую логику. — На сайте: <http://matlab.exponenta.ru/fuzzylogic/book1/index.php>
6. Штовба С.Д. Идентификация нелинейных зависимостей с помощью нечеткого логического вывода в системе MATLAB. // Exponenta Pro. Математика в приложениях, №2, 2003. — с. 9-15