

Идентификация скоринговой модели принятия решения о выдаче кредита

Недосекин Алексей Олегович, д.э.н., к.т.н., вице-президент IFEL Rus, академик МАНЭБ
Соложенцев Евгений Дмитриевич, д.т.н., проф., заслуженный деятель науки РФ

Введение

Любое банковское решение о выдаче кредита юридическому или физическому лицу опирается на две оценки:

1. Оценка кредитоспособности заемщика в принципе – насколько он является потенциально платежеспособным и ответственным на сколь-нибудь длительную перспективу. То есть, делаются предположения об ожидаемой волатильности доходов заемщика. Чем ниже кредитоспособность, тем выше ожидаемая волатильность.
2. Насколько тот кредит, который берет заемщик, является удобным для него, как он соотносится с уровнем текущих доходов заемщика с точки зрения напряженности текущих платежей.

В итоге, сравнивая кредитоспособность заемщика и напряженность платежей, кредитный комитет банка принимает решение – давать кредит или нет. При этом кредитный комитет руководствуется заключением кредитного эксперта, изучившего заемщика, а также своим собственным опытом работы, в том числе сформировавшимся под влиянием сложившейся кредитной практики в данном банке. Чем выше кредитоспособность заемщика, тем выше допустимая напряженность текущих выплат, и наоборот.

Чтобы осмыслить сложившуюся практику кредитования, нужно построить скоринговую модель кредитоспособности. Параметры такой модели должны быть **идентифицированы** (уточнены) на основе исторических данных по кредитным историям, собранным в определенном формате. В настоящей статье мы продемонстрируем пути построения такой модели и идентификации ее параметров.

1. Теоретические предпосылки

Рассмотрим довольно простую скоринговую модель для заемщиков - физических лиц (на примере подхода к оценке заемщика, сложившегося в одном из банков Германии), состоящую из 5 блоков и 11-и критериев в составе этих блоков. Модель основана на данных из базы демонстрационной версии модели [1].

Блок А. Данные по кредитной истории заемщика. Критерии:

- **A1** – Признак кредитной истории. Градации:
 - A1.1 – нет займов, все предыдущие займы возвращены
 - A1.2 – прошлые задолженности обслуживались аккуратно
 - A1.3 – есть текущий займ в этом же банке, который обслуживается нормально
 - A1.4 – есть проблемы в обслуживании взятых кредитов
 - A1.5 – счет в критическом состоянии, множество займов в других банках
- **A2** – Число кредитов, включающих заявляемый кредит. Градации:
 - A2.1 – 1
 - A2.2 – 2-3
 - A2.3 – 4-5
 - A2.4 – свыше 5

Блок В. Текущая занятость заемщика. Критерии:

- **V1** – Продолжительность работы заемщика на одном месте. Градации:
 - V1.1 – в настоящее время безработный
 - V1.2 – занят менее 1 года на данном месте
 - V1.3 – 1-4 года на данном месте
 - V1.4 – 4-7 лет на данном месте
 - V1.5 – свыше 7 лет на данном месте работы
- **V2** – Профессиональный статус заемщика. Градации:
 - V2.1- безработный, неквалифицированный рабочий или нет жилья
 - V2.2 – неквалифицированный работник, но есть постоянное место проживания
 - V2.3 - квалифицированный рабочий, клерк, служащий средней квалификации
 - V2.4 – менеджер, высококвалифицированный служащий или работающий на собственный бизнес
- **V3** – Признак заемщика-гастарбайтера. Градации:
 - V3.1. Да
 - V3.2. Нет

Блок С. Социальное положение заемщика. Критерии:

- **C1** – Продолжительность проживания по одному адресу. Градации:
 - C1.1 – до одного года
 - C1.2 – от одного года до 4-х лет
 - C1.3 - От 4 до 7 лет
 - C1.4 – свыше 7 лет
- **C2.** Число лиц на иждивении заемщика. Градации:
 - C2.1 – 0-2
 - C2.2 – 3 и более

Блок D. Финансовое положение заемщика. Критерии:

- **D1** – Состояние сберегательного счета и депозитов. Градации:

- D1.1 – меньше 100 евро
- D1.2 – 100-500 евро
- D1.3 – 500-1000 евро
- D1.4 – свыше 1000 евро
- D1.5 – нет счета или нет информации
- **D2** – Наличие солидарной ответственности или гарантии по кредитам третьих лиц. Градации:
 - D2.1 – нет
 - D2.2 – солидарная ответственность
 - D2.3 – гарант
- **D3** – Наличие собственности или имущественных прав. Градации:
 - D3.1 – недвижимость
 - D3.2 – недвижимости нет, но есть контракт на строительство или полис страхования жизни
 - D3.3 – ничего из D3.1 и D3.2, но есть машина или схожие ценности
 - D3.4 – нет ничего из перечисленного или неизвестно

Блок Е. Напряженность кредитных выплат. Блок состоит всего из одного фактора **E1** – соотношение текущих кредитных платежей и располагаемого текущего дохода заемщика, с градациями:

- E1.1 – свыше 35%
- E1.2 – 25%-35%
- E1.3 – 20% - 25%
- E1.4 – менее 20%

Произведем лингвистическую классификацию градаций всех факторов модели по трехуровневому принципу «низкий уровень градации – средний уровень градации – высокий уровень градации». При этом будем понимать «низкий» или «высокий» не в количественном плане, а с точки зрения относимости этого качественного уровня к интегральному уровню доверия к ситуации со стороны лиц, принимающих решение. Чем «выше» уровень градации по критерию, тем выше шансы заемщика на получение кредита.

Результаты лингвистической классификации распознавания градаций приведены в таблице 1.

Таблица 1. Лингвистическая классификация градаций факторов

№	Блок	Вес блока	Критерий	Вес критерия в блоке	Отнесение градаций критериев к уровням:		
					<i>низкий</i>	<i>средний</i>	<i>высокий</i>
1	А	P _A	A1	P _{A1}	A1.5	A1.4	A1.1, A1.2, A1.3
2			A2	P _{A2}	A2.3, A2.4	A2.2	A2.1
3	В	P _B	B1	P _{B1}	B1.1, B1.5	B1.2, B1.3	B1.4
4			B2	P _{B2}	B2.1	B2.2, B2.3	B2.4
5			B3	P _{B3}	B3.1	B3.2	

6	C	P _C	C1	P _{C1}	C1.1	C1.2, C1.3	C1.4
7			C2	P _{C2}	C2.2	C2.1	
8	D	P _D	D1	P _{D1}	D1.1, D1.2, D1.5	D1.3	D1.4
9			D2	P _{D2}	D2.2, D2.3		D2.1
10			D3	P _{D3}	D3.4	D3.2, D3.3	D3.1
11	E	P _E =0.5	E1	P _{E1} =1	E1.1	E1.2, E1.3	E1.4

Сопоставим каждому качественному уровню числовую оценку в интервале от 0 до 1: «низкий» = 0, «средний» = 0.5, «высокий» = 1. Тогда **уровень доверия** к ситуации выдачи кредита Y можно оценить по формуле:

$$Y = P_A(P_{A1}*\lambda_{A1} + P_{A2}*\lambda_{A2}) + P_B(P_{B1}*\lambda_{B1} + P_{B2}*\lambda_{B2} + P_{B3}*\lambda_{B3}) + P_C(P_{C1}*\lambda_{C1} + P_{C2}*\lambda_{A2}) + P_D(P_{D1}*\lambda_{D1} + P_{D2}*\lambda_{D2} + P_{D3}*\lambda_{D3}) + P_E*\lambda_{E1}, \quad (1)$$

где $\lambda = 0, 0.5, 1$ – числовая оценка сложившегося качественного уровня по каждой градации.

По построению, уровень доверия к ситуации выдачи кредита изменяется от 0 («кредит выдавать нельзя ни в коем случае») до 1 («кредит выдавать можно при любых прочих условиях»). При этом сумма весов всех блоков и всех критериев в пределах одного блока строго равна 1. Фактически, формула (1) есть не что иное как вырожденный случай метода матричного агрегирования качественных оценок факторов, предлагаемого во множестве работ д.э.н. А.О.Недосекина и его научной школы.

Теперь рассмотрим набор из N всех кредитных историй банка, в отношении состава которых мы будем идентифицировать построенную модель. Пусть каждая кредитная история представлена вектором

$$(i, \underline{Y}, A1, A2, B1, B2, B3, C1, C2, D1, D2, D3, E1), \quad (2)$$

где i – номер кредитной истории по порядку от 1 до N , \underline{Y} – **признак успешности кредита**: $\underline{Y} = 1$, если кредит беспрепятственно обслуживался и был погашен точно в срок и $\underline{Y} = 0$, если в ходе реализации кредита возникали серьезные проблемы, вплоть до дефолта.

Тогда можно построить интегральную функцию ошибки идентификации модели на основе сводной кредитной истории банка следующего вида:

$$Z = \sum_{i=1}^N (Y_i - \underline{Y}_i)^2 / N, \quad (3)$$

где уровень доверия к ситуации кредитования и признак успешности кредита определяются для каждой i -ой истории в общем сводке кредитных историй банка.

По построению, ошибка Z изменяется в пределах от 0 (когда достигается максимальное совпадение модельной оценки и признака успешности кредита) до 1 (когда есть полное несовпадение модели и факта кредитной истории).

Тогда задача идентификации модели – это оптимизационная задача поиска весов блоков и критериев в формуле (1), которые минимизируют уровень ошибки идентификации модели (3) в условиях стандартных ограничений на размер весов.

Пример идентификации

Рассмотрим демонстрационный пример свода кредитных историй [2], составленной по образцу (2). Применяя к нему последовательно процедуру лингвистической классификации по таблице 1 и процедуру оптимизации, приходим к следующим оптимальным весам для модели (таблица 2).

Таблица 2. Оптимальные веса блоков и факторов модели

Блок	Критерий	Веса блоков	Веса критериев
A	A1	0.131	0.504
	A2		0.496
B	B1	0.190	0.167
	B1		0.833
	B1		0.000
C	C1	0.119	0.953
	C2		0.047
D	D1	0.560	0.024
	D2		0.400
	D3		0.576
E	E1	0.000	1.000

Соответствующий оптимальным весам уровень интегральной ошибки Z составляет 0.214. Это довольно существенный уровень расхождения модели и реальности, который говорит о том, что в целом предложенная скоринговая модель слишком проста для того, чтобы быть инструментом для практического использования. Также ясно, что некоторые факторы, которые должны были бы участвовать в оценке, в свод кредитных историй не попали.

Из анализа вытекают 2 соображения, одно из которых является парадоксальным. Во-первых, мы не обнаружили четко выраженной связи между дефолтами по кредитам и напряженностью выплат (критерий E1). Более подробный анализ истории показал, что в рамках выделенных историй брались сравнительно небольшие суммы, и напряженность выплат не имела в структуре дефолтов решающего значения.

Зато с не меньшей убедительностью выяснилось, что наиболее существенным блоком в оценке является блок финансовой состоятельности заемщика (в частности, наличие у

него недвижимости). Напрашивается вывод, что состоятельность – это не сколько факт владения собственностью, сколько набор адекватных психологических установок, направленных на накопление и удержание богатства. Вероятно, соответствующий психологический анализ заемщика позволил бы заведомо отодвинуть от кредитов людей беспечных, инфантильных, не умеющих отвечать по взятым на себя имущественным обязательствам.

Заключение

Процедура идентификации скоринговой модели – это инструмент проверки качества модели, ее состоятельности. Действуя итеративно, можно совершенствовать модель, добавляя к ней и убавляя из нее новые блоки и факторы. Критерием истины является неуклонное снижение ошибки. Там, где результаты оптимизации носят парадоксальный (с точки зрения аналитика) характер – возникает потребность в дополнительном исследовании.

Следующий шаг в настройке модели – это более квалифицированный подход к лингвистической классификации отдельных параметров и выстраивания градаций. Из описания модели и по истории [2] видно, что такая работа в банке была выполнена без должного тщания.

Таким образом, успех моделирования опирается на три вещи:

1. Определение чувствительных факторов модели и объединение их в блоки соответствующей направленности.
2. Адекватная лингвистическая классификация уровней факторов.
3. Механизм обратной связи: анализ причин дефолтов и увязывание этих дефолтов с параметрами заемщика. Обогащение представлений кредитного эксперта и аналитика о типичном портрете «хорошего» и «плохого» заемщика.

Список литературы

1. Модель логико-вероятностной оценки и анализа кредитного риска от бюро «Ино_Тел». Демонстрационная версия // Банки и Риски, № 1(7), 2007.
2. Свод кредитных историй (демонстрационный пример). – На сайте: http://www.ifel.ru/br8/soft/Кредитная_история.xls.