

ОЦЕНКА КРЕДИТНОЙ ИСТОРИИ ФИЗИЧЕСКИХ ЛИЦ НА ОСНОВЕ НЕЧЕТКИХ МОДЕЛЕЙ

Кузнецов Л. А., Перевозчиков А. В.

(Липецкий государственный технический университет,
Липецк)

kuznetsov@stu.lipetsk.ru, perevozchikov@lipetsk.ru

Предложена методика формализации задачи оценки кредитной истории физических лиц на основе использования нечетких моделей. Рассматриваются принципы построения баз знаний с помощью основных показателей, применяемых в ряде российских банков для определения качества выполнения заемщиком обязательств по выплате основного долга и процентов. Показаны возможные методы анализа кредитной истории, на основе которых могут быть построены алгоритмы теории нечетких множеств.

Ключевые слова: нечеткая логика, функция принадлежности, лингвистическая переменная, кредитная история, база знаний.

Введение

Проблема своевременного возвращения кредитов актуальна для банков. Ее решение в значительной мере зависит от «качества» оценки потенциальных заемщиков, которая осуществляется экспертами на основании ретроспективной информации о кредитных историях физических лиц, хранящейся в базе данных банка.

Бухгалтерская отчетность дает достаточно полную информацию о финансовом состоянии юридических лиц и позволяет выработать объективные критерии принятия решения о выдаче кредита. Оценка кредитоспособности физических лиц часто осуществляется по различного рода косвенным характеристикам.

кам, содержащимся в кредитной истории и допускающим достаточно широкий спектр толкования.

В системе кредитования большого количества банков, оценка кредитной истории производится экспертом, который, в основном, опирается на свой опыт и интуицию, что может приводить к внесению в решение не имеющих достаточных оснований субъективных соображений. В реальной ситуации мнения разных аналитиков часто различаются, особенно если рассматриваются спорные ситуации, имеющие множество альтернативных решений.

Вследствие этого в оценке чрезмерный вес может приобретать субъективное мнение эксперта и следующая из него некомпетентная или преднамеренная интерпретация информации, приводящая к принятию решений, ущербных для банка.

Особенно сложным является описание характеристик, определяющих кредитную историю заемщика. Задание жестких (четких) ограничений на значения ее составляющих (количество дней на просрочке за определенный период, общее количество просрочек и т. п.), если их диапазоны узки, может привести к исключению из рассмотрения целого ряда потенциальных клиентов, и снижению прибыли финансово-кредитной организации, и наоборот, излишнее «расширение» границ сопровождается ухудшением качества кредитного портфеля и повышением рисков банка.

Снижение возможности влияния эксперта на решение и повышение в нем доли объективных факторов может быть обеспечено формализацией прогноза поведения заемщика и процедуры принятия решения о выдаче кредита.

В отношении физических лиц основой для обучения подобной системы являются ретроспективные кредитные истории, которые представляют собой в значительной степени документы вербальные и вследствие этого нечеткие. Одним из способов формализации вербальных величин и преобразования их в количественные, допускающие применение математических операций и упорядочивания, представляется теория нечетких отношений и множеств Л. А. Заде [4].

В будущем оценка кредитной истории будет осуществляться посредством запроса в Бюро кредитных историй, в котором уже содержатся определенные данные о договорах, по которым рассматриваемое физическое лицо имеет обязательства перед банком.

Предоставленная информация в основном касается параметров кредитной сделки, таких как срок, сумма, вид кредита, а данные, относящиеся к срокам погашения задолженности и фактам наличия просроченных платежей, не будут содержать рекомендаций об оценке уровня платежной дисциплины. В результате перед экспертом ставится задача по обобщению и систематизации ряда факторов, отражающих качество выполнения обязательств физического лица, по кредитному договору.

В работе приведен метод анализа кредитного портфеля банка на предмет оценки качества исполнения обязательств заемщиков по погашению задолженности, основанный на построении функций принадлежности (алгоритм Мамдани), применение которого с определенной степенью позволит разделить кредитные истории клиентов, расплатившихся по своим обязательствам, на «положительные» и «отрицательные».

В дальнейшем, имея в наличии информацию о «хороших» и «плохих» кредитах и сведения о соответствующих им клиентах (анкетные данные, справки о доходах, наличии в собственности имущества и т. п.) возможно построение функционального соответствия, представляющего зависимость качества кредитной истории от характеристик заемщика. Таким образом, на этапе рассмотрения кредитной заявки рассчитанная математическая модель позволит сделать вывод о наиболее вероятном уровне платежной дисциплины потенциального клиента.

1. Описание характеристик кредитной истории с использованием лингвистических переменных

Разработка математической модели для анализа качества выполнения обязательств заемщика требует наличия адекватно-

го формального представления, которое учитывало бы особенности кредитования физических лиц.

Для изучения систем, на поведение которых сильное влияние оказывают суждения, восприятия или эмоции человека (гуманистические системы), Л. А. Заде предложил использовать так называемые лингвистические переменные [4], т. е. переменные, значениями которых являются слова или предложения естественного языка.

Процесс оценки кредитной истории может быть описан в терминах теории нечетких множеств с использованием лингвистических переменных.

Лингвистическая переменная может быть задана в виде набора $\langle X, T, U, G, M \rangle$ [1].

Применительно к задаче анализа кредитной истории переменным может быть приписан следующий содержательный смысл:

X – лингвистическая переменная с именем «кредитная история»;

T – терм-множество переменной X , т. е. множество значений лингвистической переменной X , областью определения каждого из которых является множество U .

В банковской практике кредитную историю наиболее часто классифицируют по следующим категориям:

«положительная» – за период действия договора отсутствуют факты задержки оплаты, или нормативные документы кредитной организации позволяют отнести кредитную историю с некоторым количеством просроченных платежей к данной категории;

«приемлемая» – наиболее часто встречающаяся ситуация, когда имеет место несколько фактов задержки платежей;

«отрицательная» – выдача кредита нецелесообразна вследствие систематического нарушения условий договора по погашению задолженности.

Поэтому множество значений кредитной истории может быть, например следующим:

$T = \{ \langle \text{«положительная»}, \text{«приемлемая»}, \text{«отрицательная»} \rangle \}$.

Множество U представляет собой набор количественных характеристик, на основании которых возможно определить принадлежность кредитной истории к значениям, входящим в T . Например, оно может иметь вид: $U = \{\text{«количество просроченных платежей»}, \text{«количество дней в течении которых погашение не производилось»} \text{ и т. п.}\}$.

G – синтаксические правила, часто в виде грамматики, порождающие название термов.

Элементы множества G предназначены для формирования новых значений X , детализирующих кредитные истории. На основе комбинаций элементов $t \in T$ и $g \in G$, могут быть введены дополнительные значения множества T .

Например, при $G = \{\text{«не»}, \text{«очень»}, \text{«более-менее»}\}$, кредитной истории могут быть приданы следующие лингвистические значения: «не отрицательная», «более-менее приемлемая», «не положительная».

M – семантические правила, задающие функции принадлежности нечетких термов, порожденных синтаксическими правилами G .

Нечеткие множества, полученные с помощью семантических правил M , характеризуются простотой определения функций принадлежности, для расчета которых не требуется применения специализированных методов, а достаточно использовать стандартные математические операции.

Семантические правила M представлены в таблице 1.

Таблица 1. Правила расчета функций принадлежности

Квантификатор $t \in T$	Функция принадлежности $\mu_t(u)$ при $u \in U, t \in T$
не t	$1 - \mu_t(u)$
очень t	$(\mu_t(u))^2$
более-менее t	$\sqrt{\mu_t(u)}$

Функция принадлежности $\mu_t(u) \in [0; 1]$ ставит в соответствие значению $u \in U$ число из интервала $[0; 1]$, характеризующее

степень принадлежности u к терм-множеству $t \in T$. Фактически это позволяет формализовать процедуру выбора наиболее подходящей кредитной истории по вербальным характеристикам заемщика.

2. Построение функций принадлежности на основе метода парных сравнений

Следующим этапом после определения структуры лингвистической переменной является переход к представлению зависимости статуса кредитной истории от выбранной характеристики, т. е. требуется найти степень принадлежности рассматриваемой кредитной истории к значениям, определенным в множестве T .

Исходной информацией для построения подобной функциональной зависимости являются экспертные парные сравнения. Для каждой пары элементов множества U эксперт оценивает преимущество одного элемента над другим по отношению к значениям нечеткой переменной.

Парные сравнения удобно представлять следующей матрицей:

$$(1) \quad A = \begin{matrix} & \begin{matrix} u_1 & u_2 & \dots & u_n \end{matrix} \\ \begin{matrix} u_1 \\ u_2 \\ \dots \\ u_n \end{matrix} & \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \end{matrix}.$$

где a_{ij} – уровень преимущества элемента u_i над u_j ($i, j = \overline{1, n}$), определяемый по девятибалльной шкале (см. таблицу 2), разработанной математиком Т. Саати [8] для решение задач, ориентированных на анализ множества альтернатив, как правило, соответствующих естественному ходу человеческого мышления. Эффективность использования данной шкалы подтверждена многолетней практикой, она применяется в ряде систем поддержки принятия решений.

Таблица 2. Шкала преимуществ Саати

Степень важности	Определение	Комментарии
1	<i>Отсутствует преимущество</i> элемента u_i над элементом u_j	Два значения вносят одинаковый вклад в достижение цели.
3	Имеется <i>слабое преимущество</i> u_i над u_j	Опыт и суждение дают легкое предпочтение одному значению над другим.
5	имеется <i>существенное преимущество</i> u_i над u_j	Опыт и суждение дают сильное предпочтение одному значению над другим.
7	имеется <i>явное преимущество</i> u_i над u_j	Преимущество одного значения над другим очень сильно.
9	имеется <i>абсолютное преимущество</i> u_i над u_j	Свидетельства в пользу предпочтения одного значения в высшей степени убедительны.
2,4,6,8	Промежуточные значения между соседними значениями шкалы.	Ситуации, когда необходимы компромиссные решения.

Исследуются все показатели u_i , включаемые в U , например: количество дней на просрочке за определенный период; общее количество просрочек; отношение количества просроченных платежей к «удачным» и т. п. Конкретное наполнение элементов U должно отображать политику финансовой организации по отношению к кредитным рискам.

Построение матрицы A должно быть произведено для каждого значения нечеткой переменной $t \in T$.

Для снижения размерности матрицы A на основании априорных сведений о t целесообразно исключить все элементы u ,

степень принадлежности которых, к рассматриваемому значению равна единице.

Матрица парных сравнений является обратно симметричной: $(a_{ij} = \frac{1}{a_{ji}}, i, j = \overline{1, n})$.

Степени принадлежности принимаются равными соответствующим координатам собственного вектора $W = (w_1, w_2, \dots, w_n)$ матрицы парных сравнений:

$$(2) \quad \mu_i(u_i) = w_i, i = \overline{1, n}.$$

Собственный вектор матрицы A находится из следующей системы уравнений [10]:

$$(3) \quad \begin{cases} A * W = \lambda_{\max} * W, \\ w_1 + w_2 + \dots + w_n = 1. \end{cases}$$

где λ_{\max} – максимальное собственное значение матрицы A .

Для примера рассмотрим определение значений функции принадлежности для нечеткого термина «несколько просроченных платежей» = {3, 5, 7, 8}, в качестве элементов которого было принято количество пропущенных периодов оплаты по кредиту, взятых для четырех кредитных историй.

Матрица парных сравнений для этого случая имеет вид:

$$A = \begin{matrix} & \begin{matrix} 3 & 5 & 7 & 8 \end{matrix} \\ \begin{matrix} 3 \\ 5 \\ 7 \\ 8 \end{matrix} & \begin{bmatrix} 1 & 3 & 5 & 6 \\ 1/3 & 1 & 3 & 5 \\ 1/5 & 1/3 & 1 & 3 \\ 1/6 & 1/5 & 1/3 & 1 \end{bmatrix} \end{matrix}$$

Первым этапом является определение максимального собственного числа λ_{\max} матрицы парных сравнений A . Для этого составляется и решается характеристическое уравнение матрицы A . Максимальное собственное значение оказывается равным $\lambda_{\max} = 4,095$

Далее находится собственный вектор W , соответствующий λ_{\max} , элементы которого удовлетворяют второму равенству системы уравнений (3):

$$W = \{0,562; 0,256; 0,122; 0,06\}$$

Нечеткое множество с функцией принадлежности, определяемой вектором W , получилось субнормальным [1]. Для нормализации все степени принадлежности делятся на максимальное значение 0,562. Результаты представлены в таблице 3.

Таблица 3. Нормализация нечеткого множества

u_i	3	5	7	8
$\mu_{\text{несколько просроченных платежей}}(u_i)$ (субнормальное нечеткое множество)	0,562	0,256	0,122	0,06
$\mu_{\text{несколько просроченных платежей}}(u_i)$ (нормальное нечеткое множество)	1	0,06	0,217	0,106

Данная функция принадлежности позволяет определить для каждой кредитной истории степень, с которой ее допустимо считать относящейся к нечеткому терму «несколько просроченных платежей».

Для того чтобы полученные результаты были адекватны ситуации, в которой принимается решение, необходимо, чтобы в матрице парных сравнений достигались требуемые уровни согласованности данных.

Под согласованностью матрицы парных сравнений понимается численная (кардинальная) согласованность и транзитивная (порядковая) согласованность.

Пример кардинальной несогласованности. Пусть параметр A (четыре просроченных платежа) лучше параметра B (пять просроченных платежей) в 3 раза, а параметр B лучше параметра C (шесть просроченных платежей) в 4 раза, таким образом, A лучше C в $3 * 4 = 12$ раз. Нарушение этого равенства в рамках выбранной шкалы (шкала Саати имеет градацию от 1 до 9) считается кардинальной несогласованностью.

Пример транзитивной несогласованности. Пусть параметр A предпочтительнее параметра B (обозначим как $A > B$), а параметр B предпочтительнее параметра C ($B > C$), следовательно, A предпочтительнее C ($A > C$). Нарушение последнего неравенства называется транзитивной несогласованностью.

Таким образом, для оценки однородности суждений эксперта необходимо использовать отклонение величины максимального собственного значения λ_{max} от порядка матрицы N . Однородность суждений оценивается индексом однородности (ИО) или отношением однородности (ОО) в соответствии со следующими выражениями:

$$ИО = (\lambda_{max} - N)/(N - 1); \quad ОО = ИО/M(ИО).$$

где $M(ИО)$ – среднее значение (математическое ожидание) индекса однородности случайным образом составленной матрицы парных сравнений, основанное на экспериментальных данных.

В качестве допустимого используется значение $ОО \leq 0,10$. Если для матрицы парных сравнений отношение однородности $ОО > 0,10$ то это свидетельствует о нарушении логичности суждений, допущенных при заполнении матрицы, поэтому эксперту предлагается пересмотреть данные, использованные для построения матрицы, чтобы улучшить однородность.

При согласованных парных сравнениях процедура построения функций принадлежности значительно упрощается. В этом случае матрица A обладает следующими свойствами:

- она обратна симметрична, т. е. элементы, симметричные относительно главной диагонали, связаны зависимостью $a_{ij} = 1/a_{ji}, \quad i, j = \overline{1, n}$;
- она транзитивна, т. е. $a_{ik} * a_{kj} = a_{ij}, \quad i, j, k = \overline{1, n}$.

Наличие этих свойств позволяет определить все элементы матрицы парных сравнений, если известно $(n - 1)$ недиагональных элементов. Например, если известна k -тая строка, то произвольный элемент a_{ij} определяется так:

$$(4) \quad a_{ij} = a_{kj}/a_{ki}, \quad i, j, k = \overline{1, n}.$$

После определения всех элементов матрицы парных сравнений, степени принадлежности нечеткого множества вычисляются по формуле:

$$(5) \quad \mu_i(u_i) = \frac{1}{a_{1i} + a_{2i} + \dots + a_{ni}}, \quad i = \overline{1, n}.$$

Таким образом, возможно рассчитать значения функции принадлежности не применяя трудоемких вычислительных процедур поиска собственных чисел и векторов.

3. Аналитическое построение функций принадлежности на основе априорной информации о нечетком множестве

Другой способ построения функций принадлежности основывается на обобщении знаний и опыта эксперта по отношению к характеристикам, определяющим кредитную историю заемщика.

Анализ терм-множества $t \in T$ можно начать на основании субъективных причин и неформальных сведений о выбранном показателе просроченной задолженности и возможных ограничениях, накладываемых на область определения функции $\mu_i(u)$, которые задаются экспертом и согласуются с внутренними документами кредитной организации (при наличии таковых).

Естественным является также и требование непрерывности функции $\mu_i(u)$, которое формализует представление о том, что если два элемента из множества U отличаются друг от друга лишь незначительно, то значения соответствующих функций принадлежности также близки.

Конкретный вид функции принадлежности определяется на основе различных дополнительных предположений о свойствах этой функции (симметричность, монотонность и т. д.) с учетом специфики имеющейся неопределенности [1].

Подобные предположения позволяют упростить процедуру построения функции принадлежности, в частности, выделить значения, при которых степень принадлежности элементов

$u \in U$ к рассматриваемому терм-множеству $t \in T$ равна нулю и единице, т. е. найти область определения функции $\mu(u)$.

Например, кредитную историю можно считать отрицательной, при отсутствии платежей за период в три месяца, т. е. $\mu(u) = 1$ при $u \in [3 \text{ месяца; срок кредита}]$, $t = \text{«отрицательная»}$. Результаты позволяют сделать следующий вывод: когда заемщик задерживает с выплатой по кредиту более чем на три месяца, его кредитная история относится к «отрицательной» категории.

Аналитическое представление дает возможность не применять трудоемкие вычислительные процедуры расчета степеней принадлежности. В качестве недостатков такого подхода можно отметить более низкую точность по сравнению с математическими методами и высокие требования, предъявляемые к квалификации эксперта.

Следует отметить, что важной составляющей предлагаемых подходов является непосредственное участие эксперта как при построении матрицы парных сравнений (см. таблицу 2), так и при выборе аналитической формы функции принадлежности.

Необходимость разработки унифицированной методологии особенно актуальна в случае отсутствия в финансово-кредитной организации нормативных документов, регламентирующих процесс определения качества кредитной истории физических лиц и порядок разрешения спорных ситуаций.

Рассмотренные методы позволяют получить математическую интерпретацию опыта эксперта, отражающую общую стратегию принятия решений по оценке кредитной истории.

Применение единой методологии на основе синтеза знаний высококвалифицированной группы экспертов необходимо для устранения субъективного влияния кредитных аналитиков, не обладающих достаточным опытом работы в сфере кредитования физических лиц, а также минимизации рисков от возможного злоупотребления служебным положением со стороны инсайдеров банковской организации.

4. Построение базы знаний

На практике процесс анализа кредитной истории представляет собой рассмотрение ряда показателей, применяемых для оценки исполнения обязательств заемщика по погашению задолженности. В зависимости от подхода, применяемого в кредитной организации, наиболее распространенными являются следующие характеристики: количество платежей, количество просроченных платежей, отношение общего просроченных платежей к сроку кредита и т. п.

При анализе кредитной истории рассуждения эксперта состоят в выборе нескольких параметров, оказывающих, по его мнению, наиболее сильное влияние на качество кредитной истории заемщика, на основании которых он составляет соответствующие заключения.

Для приведения задачи классификации кредитной истории к форме, позволяющей учитывать совокупное влияние показателей на принимаемое решение, целесообразно использовать понятие составной лингвистической переменной [1]. В этом случае процесс классификации разбивается на анализ взаимодействия ряда частей, которые будут включены в состав лингвистической переменной «кредитная история», что позволит сформировать зависимость результата от комбинаций элементов данной переменной.

Для задания подобных многомерных зависимостей строится набор условных правил нечеткого логического вывода вида «Если «Набор условий», то «Вывод»» [9].

Нечеткая база знаний может быть представлена в следующем виде:

$$\begin{aligned} & \text{ЕСЛИ } (x_1 = a_{1,1,1}) \text{ И } (x_2 = a_{2,1,1}) \text{ И } \dots \text{ И } (x_n = a_{n,k,1}) \\ & \text{ИЛИ } (x_1 = a_{1,1,2}) \text{ И } (x_2 = a_{2,1,2}) \text{ И } \dots \text{ И } (x_n = a_{n,k,2}) \\ (6) \quad & \dots \\ & \text{ИЛИ } (x_1 = a_{1,k,p}) \text{ И } (x_2 = a_{2,k,p}) \text{ И } \dots \text{ И } (x_n = a_{n,k,p}), \\ & \text{ТО } y = d_j, \end{aligned}$$

где $a_{i,k,p}$ – нечеткий терм (*немного, несколько, много и т. п.*), которым оценивается лингвистическая переменная $x_i \in X =$

$= (x_1, x_2, \dots, x_n)$ (количество просроченных платежей, временной интервал задержки платежа и т. п.) в правиле под номером p , $i = \overline{1, n}$, $k = \overline{1, q}$, $p = \overline{1, l}$; $y = (d_1, d_2, \dots, d_m)$ – значения нечеткого логического вывода y_j («положительная», «приемлемая», «отрицательная» и т. п.), $j = \overline{1, t}$; t – количество значений нечеткого логического вывода; n – количество лингвистических переменных; l – количество правил в базе знаний; q – количество нечетких термов, которыми оценивается лингвистическая переменная $x_i \in X = (x_1, x_2, \dots, x_n)$.

Система условных обозначений используется для отражения смысла применяемых математических преобразований.

Приведем простейший пример базы знаний, показывающей зависимость между количеством просроченных платежей (x) и возможностью оценки качества кредитной истории (y), как «отрицательная»:

- (7) ЕСЛИ x = несколько, ТО y = низкая;
ЕСЛИ x = много, ТО y = высокая.

5. Классификация кредитной истории с использованием нечетких моделей типа Мамдани

Применительно к процессу оценки кредитной истории при кредитовании физических лиц задача классификации состоит в определении качества выполнения заемщиком обязательств по выплате основного долга и процентов по ряду критериев, сформулированных кредитным экспертом.

Рассмотрим алгоритм Мамдани [10] как один из возможных вариантов решения поставленной задачи. В качестве основных этапов данного метода можно выделить следующую последовательность операций:

1. Построение нечеткой базы знаний.
2. Фаззификация входных параметров [10].
3. Определение результирующего нечеткого множества.
4. Дефаззификация.

Построение нечеткой базы знаний

Построение нечеткой базы знаний заключается в представлении опыта эксперта в виде определенного набора правил (6), отражающих процесс принятия решения в рассматриваемой области.

С помощью операций \cup (ИЛИ) и \cap (И) перепишем (6) в более компактном виде:

$$(8) \quad \bigcup_{p=1}^l \left(\bigcap_{i=1}^n x_i = a_{i,k,p} \right) \rightarrow y = d_j, \quad j = \overline{1, m}.$$

Фаззификация входных параметров

Фаззификацией, или введением нечеткости, называется процесс построения функций принадлежности для входных переменных системы нечеткого вывода $x_i \in X = (x_1, x_2, \dots, x_n)$ на основе соответствующих им терм-множеств $a_{i,k,p}$.

Определение результирующего нечеткого множества

На данном этапе происходит определение подмножеств, построение которых производится по всем нечетким логическим выводам для соответствующих правил из базы знаний.

Для дальнейшего рассмотрения алгоритма Мамдани введем следующие обозначения: $\mu_{k,p}(x_i)$ – функция принадлежности лингвистической переменной x_i нечеткому терму $a_{i,k,p}$; $\mu_{d_j}(y)$ – функция принадлежности значения нечеткого логического вывода терму d_j .

Степень принадлежности лингвистической переменной x_i нечетким термам d_j из базы знаний определяется следующей системой нечетких логических уравнений:

$$(9) \quad \mu_{d_j}(X) = \bigvee_{p=1, l} \bigwedge_{i=1, n} [\mu_{k,p}(x_i)], \quad j = \overline{1, m}.$$

где \vee (\wedge) – операция из s -нормы (t -нормы), т. е. из множества реализаций логической операций ИЛИ (И). Наиболее часто используются следующие реализации: для операции ИЛИ – нахождение максимума и для операции И – нахождение минимума.

Далее нечеткие подмножества, назначенные для каждой выходной переменной, объединяются вместе, чтобы сформировать одно результирующее нечеткое множество.

Нечеткое множество \tilde{y} , соответствующее вектору X , определяется следующим образом:

$$(10) \tilde{y} = \text{agg} \left(\sum_{j=1, m} \text{imp} (\mu_{d_j}(X), \mu_{d_j}(y)) / y \right),$$

где *imp* – импликация, обычно реализуемая как операция нахождения минимума; *agg* – агрегирование нечетких множеств, которое наиболее часто реализуется операцией нахождения максимума.

Дефаззификация

Четкое значение y , соответствующее вектору X определяется в результате дефаззификации нечеткого множества \tilde{y} .

Наиболее часто применяется дефаззификация по методу центра тяжести:

$$(11) y = \frac{\sum_{j=1, m} d_j \mu_{d_j}(X)}{\sum_{j=1, m} \mu_{d_j}(X)}.$$

Физическим аналогом этой формулы является нахождение центра тяжести плоской фигуры, ограниченной осями координат и графиком функции принадлежности нечеткого множества \tilde{y} .

Таким образом, после выполнения нечеткого логического вывода статус кредитной истории определяется по максимальной степени принадлежности результата дефаззификации к оценкам качества выполнения обязательств по погашению задолженности, сформулированных экспертом в базе знаний.

Для иллюстрации работы метода рассмотрим пример нечеткого логического вывода по базе знаний (7) при значении входной переменной $x = 4$. В качестве зависимости, представляющей функцию принадлежности для нечеткого термина «несколько просроченных платежей» (рис. 1а), взяты рассчитанные

параметры нормализованного нечеткого множества из таблицы 3. Используя метод парных сравнений, аналогично была определена функция принадлежности для нечеткого термина «много просроченных платежей» (рис. 1в).

Треугольные функции принадлежности $\mu_d(y)$ значений нечеткого логического вывода (возможность оценки качества кредитной истории как «отрицательная») терминам «низкая» и «высокая» были выбраны на основе аналитического представления. Следует отметить, что подобная зависимость является простейшей и служит для демонстрации работы метода.

Исследование логики суждений, применяемых при формировании выводов, для каждого правила из базы знаний заключалось в определении характера поведения функции $\mu_d(y)$ на интервале $[0; 1]$. В связи с этим численное увеличение принадлежности кредитной истории к «отрицательной» категории предполагает, что возможность оценить ее как «низкая» должна уменьшаться и, соответственно, возрастет при «высокой».

Данные функции и нечеткие логические выводы по правилам из базы знаний представлены на рис. 1б и рис. 1г.

Графическая интерпретация результатов заключается в выделении множества, полученного при отсечении $\mu_d(y)$ на уровне значения функции принадлежности для входной переменной $x = 4$.

Следующим этапом решения задачи является агрегирование нечетких множеств по (10). Графически данный алгоритм заключается в выделении множества, ограниченного линиями функций, представляющих результаты нечетких логических выводов и имеющих наибольшее значение на всей области определения. Операция агрегирования методом нахождения максимума представлена на рис. 2.

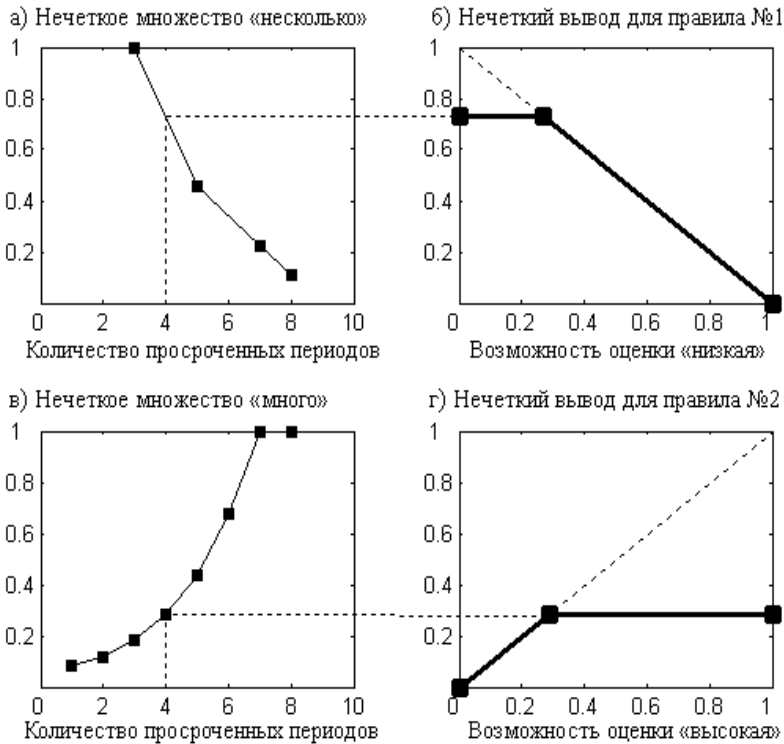


Рис.1. Нечеткие логические выводы по базе знаний

Координаты вершин полученной фигуры, определяют результирующее нечеткое множество в следующем виде:

$$\tilde{y} = \{0/0; 0,73/0; 0,73/0,267; 0,286/0,712; 0,286/1; 0/1\}$$

Далее, применяя к нечеткому множеству \tilde{y} дефазификацию методом центра тяжести (11), находится четкое значение y :

$$y = \frac{0*0 + 0*0,73 + 0,267*0,73 + 0,712*0,286 + 1*0,286 + 1*0}{0 + 0,73 + 0,73 + 0,286 + 0,286 + 0} = 0,392.$$



Рис. 2. Агрегирование нечетких множеств

Результаты позволяют сделать вывод о том, что полученная оценка характеризует кредитную историю заемщика, пропустившего четыре платежных периода, как имеющую «низкую» степень принадлежности к отрицательной категории.

Необходимо отметить, что в примере отражены общие подходы, используемые для решения поставленной задачи. Наиболее эффективное применение метода на этапе формирования базы знаний предполагает рассмотрение всех возможных комбинаций причинно-следственных связей между предпосылками модели и соответствующими им выводами.

Проведение подобного анализа для всего кредитного портфеля позволит разделить кредитные истории на категории с «положительной» и «отрицательной» оценкой качества выполнения заемщиками условий договора по погашению задолженности.

Для проверки достоверности результатов, полученных при использовании предлагаемого подхода, рассмотрим метод нечеткой классификации по сравнению с одним из распространенных способов многомерного статистического анализа (дискриминантный анализ [5]) на примере кредитного портфеля одного из коммерческих банков.

Оценка кредитных историй физических лиц была проведена по двум группам: «положительная» и «отрицательная».

В качестве элементов, входящих в первую категорию, были выбраны кредитные истории без просроченных платежей и с одной просрочкой, которая была погашена в десятидневный срок, количество – 3 289.

Вторая категория была сформирована по выборке из кредитных договоров, по которым для взыскания задолженности требовалось обращение в судебные органы, а также договора со сроком нахождения на просрочке свыше 180 дней, т. е. задолженность по которым классифицирована как «безнадежная» [7], количество – 637.

На основании традиционной практики, применяемой кредитными работниками в рассматриваемом банковском учреждении для оценки качества исполнения заемщиком обязательств по погашению задолженности, были сформулированы основные критерии, по которым осуществлялся анализ кредитной истории.

В качестве независимых переменных были выбраны следующие показатели:

- количество просрочек за весь срок пользования кредитом (x_1);
- максимальный срок недоплаты в днях (x_2).

Зависимая переменная – оценка кредитной истории (y).

В связи с особенностями алгоритма Мамдани, требующего для своей работы наличие нечеткой базы знаний, был сформулирован свод правил, представляющих собой синтез знаний и опыта эксперта в сфере оценки кредитоспособности физических лиц.

Для построения базы знаний были использованы нечеткие термы, которыми в повседневной жизни для принятия оптимальных решений оперирует каждый человек.

Применяемая база знаний

1. ЕСЛИ x_1 = «много», ТО y = «отрицательная».
2. ЕСЛИ x_1 = «небольшое количество» И x_2 = «небольшой», ТО y = «более-менее положительная».

3. ЕСЛИ $x_1 =$ «небольшое количество» И $x_2 =$ «совсем небольшой» ТО $y =$ «положительная».
4. ЕСЛИ $x_1 =$ «небольшое количество» И $x_2 =$ «большой», ТО $y =$ «отрицательная».
5. ЕСЛИ $x_2 =$ «большой» ТО $y =$ «отрицательная».

Результаты классификации представлены в таблице 4.

Таблица 4. Сравнительные оценки работы методов

Оценка кредитной истории	Качество предсказания (%)	Количество ошибок	Минимальное значение оценки принадлежности к группе	Максимальное значение оценки принадлежности к группе
<i>Дискриминантный анализ (выполнен в среде Statistica 6.0)</i>				
Положительная	100	0	≈ 1	≈ 1
Отрицательная	74,72	161	0,52	1
<i>Метод нечеткой классификации (выполнен в среде Matlab 7.0.1)</i>				
Положительная	100	0	0,91	0,94
Отрицательная	100	0	0,72	0,94

Результаты позволяют сделать вывод о недостаточной точности классификации кредитных историй, входящих в «отрицательную» категорию, методом дискриминантного анализа по сравнению с алгоритмом Мамдани.

Полученный диапазон оценки принадлежности значений к «отрицательной» группе (0,52-1) характеризует традиционный метод как позволяющий принимать заявки с высоким уровнем риска и, соответственно, оказывающий более сильное влияние на возможное снижение качества кредитного портфеля, чем предлагаемый.

Дальнейшее исследование предполагает выделение характеристик физических лиц, кредитная история которых попадает в «отрицательную» категорию, с целью создания «образа» потенциального должника и построение модели, позволяющей установить платежную дисциплину клиента, обратившегося за получением денежных средств.

Выводы

Показана возможность формального подхода к представлению задачи анализа кредитной истории физических лиц с использованием математического аппарата теории нечетких множеств. Рассмотрены методы построения функций принадлежности и баз знаний с помощью составных лингвистических переменных, а также последующее применение нечетких логических выводов на их основе. Задачи, представленные в работе, интерпретированы в терминах известных методов нечетко-математического моделирования, что открывает широкое поле для их практического использования.

Литература

1. АЛТУНИН А. Е., СЕМУХИН М. В. *Модели и алгоритмы принятия решений в нечетких условиях*: Монография. Тюмень: Издательство Тюменского государственного университета, 2000. – 352 с.
2. ГАНТМАХЕР Ф. Р. *Теория матриц*. М.: Наука, 1966. – 576 с.
3. ЕДРОНОВА В. Н., ХАСЯНОВА С. Ю. *Модели анализа кредитоспособности заемщиков*. М.: Финансы и кредит, 2002, № 6. С. 9-15.
4. ЗАДЕ Л. А. *Понятие лингвистической переменной и его применение к принятию приближенных решений*. М.: Мир, 1976. – 161 с.
5. КИМ ДЖ. О, МЬЮЛЛЕР Ч. У., КЛЕККА У. Р. И ДР. *Факторный, дискриминантный и кластерный анализ*. М.: Финансы и статистика, 1989. – 215 с.

6. КУЗНЕЦОВ Л. А. *Применение нечетких моделей для решения задач управления качеством проката*. Известия ВУЗОВ Черная металлургия, 2001, № 5. С. 61-65.
7. ПОЛОЖЕНИЕ ЦБ РФ от 26.03.2004 № 254-П. *Положение о порядке формирования кредитными организациями резервов на возможные потери по ссудам, по ссудной и приравненной к ней задолженности*. [Электронный ресурс]: <http://www.bestpravo.ru/fed2004/data06/tex20171.htm>.
8. СААТИ Т. *Принятие решений. Метод анализа иерархий*. М.: Радио и связь, 1993. – 316 с.
9. ЧЕРНОВ В. Г., ИЛЛАРИОНОВ А. В. *Методика оценки кредитоспособности предприятий сферы малого бизнеса, основанная на нечеткомножественной математической модели*. М.: Финансы и кредит, 2006, № 20. С. 72-78.
10. ШТОВБА С. Д. *Введение в теорию нечетких множеств и нечеткую логику*. [Электронный ресурс]: <http://matlab.exponenta.ru/fuzzylogic/book1>.
11. ALTMAN E. I., AVERY, R. B., EISENBEIS, R. A., SINKEY J. *Application of Classification Techniques in Business, Banking and Finance*. JAI Press, Greenwich, CT, 8 (1981). P. XX-418.
12. JONSON R. W., THOMAS L. C., CROOK J. N. AND EDELMAN D. B. *Legal, Social and Economic Issues Implementing Scoring in the US*. Credit Scoring and Credit Control, Oxford University Press, Oxford, 1992. P. 19-32.

*Статья представлена к публикации
членом редакционной коллегии Ф.Т. Алескеровым.*