
ФИНАНСОВЫЕ РЫНКИ

FINANCIAL MARKETS

УДК 336.7

ББК 65.262.101-09

Н 62

В.В. Никаненкова

Аспирант кафедры финансов и кредита Северо-Кавказского государственного технического университета, г. Ставрополь. Тел.: (962) 442 67 63, e-mail: vika_solnce@inbox.ru.

Кредитный скоринг как инструмент оценки кредитоспособности заемщиков

(Рецензирована)

Аннотация. В последнее время темпы роста рынка кредитования физических лиц опережают темпы роста рынка корпоративного кредитования. Это приводит к увеличению кредитных рисков банковской системы, поэтому качество управления кредитными рисками в розничном кредитовании приобретает особую актуальность. Одним из инструментов минимизации уровня кредитного риска заемщикам являются скоринговые системы. В данной статье рассмотрены основные принципы построения, задачи и виды кредитного скоринга; отражены его положительные стороны.

Ключевые слова: кредитный риск, скоринг, методы кредитного скоринга, статистические методы, линейное программирование, дерево классификаций, нейронные сети, генетический алгоритм, метод ближайшего соседа, преимущества, недостатки.

V.V. Nikanenkova

Post-graduate student of Finance and Credit Department of the North Caucasian State Technical University, Stavropol. Ph.: (962) 442 67 63, e-mail: vika_solnce@inbox.ru.

Credit scoring as a tool of an estimation of borrowers' creditworthiness

Abstract. Recently rates of increasing the crediting market for physical persons pass ahead of the rates of increasing the corporate crediting market. This leads to increase in credit risks of the bank system. Therefore quality of management of credit risks in retail crediting gets a special urgency. One of the tools of minimization of level of credit risk for the borrowers is a scoring system. This paper discusses the basic principles of construction, objections and types of a credit scoring, as well as reflects its positive sides.

Keywords: credit risk, scoring, methods of credit scoring, statistical methods, linear programming, a tree of classifications, neural networks, genetic algorithm, a method of the nearest neighbor, advantages, shortcomings.

В последнее время в России наблюдается интенсивный рост рынка кредитования и, в частности, розничного сектора. Конечно, объемные показатели по предоставлению кредитных средств физическим лицам ниже, чем по корпоративному бизнесу, однако темпы роста

являются опережающими, о чем свидетельствуют данные Центробанка (объем кредитов юридическим лицам в августе 2011 года составил 15 949 551 млн. руб., а физическим лицам — 4 721 835 млн. руб.; с января по май 2011 года объем потребительского кредитования

вырос на 15,67% , а кредитования юридических лиц — на 9,77% [1]), то есть потребительское кредитование становится одним из наиболее динамичных направлений развития банковского сектора. Возрастающие объемы кредитования приводят к увеличению кредитных рисков, которые принимают на себя как отдельные кредитно-финансовые институты, так и банковская система страны в целом. В этой ситуации качество управления кредитными рисками в розничном кредитовании приобретает особую актуальность.

На этапе рассмотрения кредитной заявки банк, прежде всего, интересуется кредитоспособность потенциального заемщика, то есть возможность полностью и в срок рассчитаться по своим долговым обязательствам. Одним из инструментов оценки уровня кредитоспособности заемщиков в основном и служат скоринговые системы.

Скоринг представляет собой математическую или статистическую модель, с помощью которой на основе кредитной истории «бывших» заемщиков банк пытается определить, насколько высока вероятность, что потенциальный заемщик вернет (либо не вернет) кредит в срок. Таким образом, скоринг является методом разбиения всей интересующей нас совокупности клиентов на различные рискованные группы, когда нам неизвестна характеристика, которая разделяет эти группы (вероятность возврата ссуды), но зато известны другие характеристики, связанные с интересующей нас переменной.

Первое упоминание о скоринге уходит в 1936 год, когда Хансом Фишером (1881–1945) была предложена классификация популяции растений на группы. Дэвид Дюран в своем исследовании *Risk Elements in Consumer Installment Financing* переложил данную методику при классификации кредитов на «плохие» и «хорошие», что было связано с недостатком квалифицированных кредитных аналитиков в период Второй мировой войны. Методика Дюрана состояла в следующем: он выделил группы факторов и их весовые значения, позволяющие определить степень кре-

дитного риска, и установил границу выдачи ссуды как 1,25 и более. Факторы, выделенные Дюраном, и баллы, присваиваемые заемщикам в зависимости от конкретных значений этих факторов, были следующими [2]):

1. Возраст: 0,1 балл за каждый год свыше 20 лет (максимум — 0,30).

2. Пол: женский (0,40), мужской (0).

3. Срок проживания в регионе: 0,042 за каждый год (максимально — 0,42).

4. Профессия: 0,55 за профессию с низким риском, 0 за профессию с высоким риском, 0,16 — другие профессии.

5. Работа: 0,21 на предприятиях общественной отрасли, 0 — другие.

6. Срок занятости: 0,059 за каждый год работы на данном предприятии.

7. Финансовые показатели: 0,45 за наличие банковского счета, 0,35 за наличие недвижимости, 0,19 — за наличие полиса по страхованию.

Широкое применение скоринга началось в период внедрения кредитных карточек. Растущие потоки людей, которые ежедневно обращались за кредитными карточками, привели к автоматизации процесса принятия решений по выдаче кредита в банках. По данным статьи Churchill G.A., Nevin J.R., Watson R.R. «The role of credit scoring in the loan decision» [3] («Роль кредитного скоринга при принятии решения в выдаче ссуды»), после внедрения скоринг-систем уровень безнадежного долга сокращался до 50%. В Россию скоринг пришел в 2005–2006 годах и изначально тоже был призван облегчить работу сотрудникам банка. Таким образом, внедрение скоринга позволило не только сократить время рассмотрения кредитной заявки, но также минимизировать уровень риска принимаемых решений.

Рассмотрим подробнее задачи, решаемые с помощью кредитного скоринга:

1. Выявление кредитоспособных клиентов; определение вероятности «утраты» клиентов и формирование стратегии по их сохранению.

2. Предсказание будущего поведения существующих должников, что

позволяет выделить недобросовестных клиентов, уменьшить вероятность возникновения проблемных ссуд.

3. Выбор оптимальных схем поведения для минимизации числа должников.

Таким образом, задача кредитного скоринга состоит не только в выявлении привлекательности заемщика, а также в привлечении добросовестных клиентов, которые формируют доходный кредитный портфель.

Выделим базовые принципы формирования скоринговой модели. Первоначальным этапом является выделение характеристик клиентов (переменные) и их признаков — значений, которые принимают данные переменные. На примере анкеты, заполняемой клиентами, характеристиками являются вопросы анкеты (возраст, семейное положение, профессия и так далее), а признаками — ответы на эти вопросы. В результате получается интегральный показатель (score) — сумма произведений переменных и их признаков; чем он выше, тем

выше надежность клиента. Интегральный показатель потенциального клиента сравнивается с неким пороговым числом, которое рассчитывается как количество добросовестных клиентов для того, чтобы компенсировать убытки от одного должника. В результате кредитные средства предоставляются только клиентам с интегральным показателем выше этого числа. Сложность заключается в выборе характеристик модели и их весовых коэффициентов.

Существует множество методов кредитного скоринга: статистические методы, основанные на дискриминантном анализе (линейная регрессия, логистическая регрессия); различные варианты линейного программирования; дерево классификации или рекурсионно-партиционный алгоритм (РПА); нейронные сети; генетический алгоритм; метод ближайших соседей.

Ниже приводится сравнительная таблица точности классификации для различных методов, составленная профессором Л. Томасом (табл. 1) [2].

Таблица 1

Точность различных методов классификации

Авторы	Линейная регрессия	Логистическая регрессия	РПА	Линейное программирование	Нейронные сети	Генетический алгоритм
Хенли (1995)	43,4%	43,3%	43,8%	—	—	—
Бойл (1992)	77,5%	—	75%	74,7%	—	—
Шринивисан (1987)	87,5%	89,3%	93,2%	86,1%	—	—
Йобас (1997)	68,4%	—	62,3%	—	62,0%	64,5%
Десаи (1997)	66,5%	67,3%	67,3%	—	64,0%	—

Наиболее распространенными являются регрессионные методы, прежде всего — линейная многофакторная регрессия (формула 1):

$$p = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n, \quad (1)$$

где p — вероятность дефолта, w — весовые коэффициенты, x — характеристики клиента.

Общее назначение множественной линейной регрессии состоит в анализе связи между несколькими независимыми переменными и зависимой переменной. Недостаток данной модели заключается в том, что в левой ча-

сти уравнения находится вероятность, которая принимает значения от 0 до 1 («плохой» / «хороший» клиент), а переменные в правой части могут принимать любые значения от $-\infty$ до $+\infty$.

Преодолеть данный недостаток позволяет логистическая регрессия (формула 2):

$$\log(p/(1-p)) = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n, \quad (2)$$

Применение логистической регрессии базируется на более сложных расчетах, целью которых является получение весовых коэффициентов (w) и,

следовательно, требует более усовершенствованного компьютерного обеспечения. В настоящее время логистическая регрессия является лидером скоринговых систем. Преимущество логистической регрессии состоит в том, что она может подразделять клиентов не только на плохих или хороших, но и на группы риска (с 1 по 4).

Основная цель регрессионных методов — построить модель с набором факторов, определив индивидуальное и совокупное их воздействие на моделируемый показатель. Включение в уравнения регрессии той или иной совокупности факторов связано, прежде всего, с представлением кредитного работника о природе взаимосвязи моделируемого показателя с другими экономическими явлениями.

Линейное программирование можно представить следующим образом. Пусть решение, которое должен принять банк, — это разделение множества (A) всех возможных вариантов ответов на вопросы анкеты $X=(X_1, X_2, \dots, X_p)$ на два подмножества: A_G и A_B , соответствующих ответам «хорошего» и «плохого» заемщика. Пусть также имеется выборка n уже проверенных заемщиков, где n_G «хороших» заемщиков соответствуют началу выборки, а n_B «плохих», $i=n_G+1, \dots, n_G+n_B$, следуют за ними. Пусть i -му заемщику соответствуют ($x_{i1}, x_{i2}, \dots, x_{ip}$) ответы на вопросы кредитной заявки. Перед нами стоит задача нахождения таких коэффициентов или весов (w_1, w_2, \dots, w_p), чтобы взвешенная сумма $w_1X_1+w_2X_2+\dots+w_pX_p$ характеристик для «хороших» заемщиков была выше выбранного значения c , а «плохих» — ниже. Если характеристики x_i трансформировать в бинарные переменные, то коэффициенты w_i будут соответствовать баллам, присваиваемым каждому из ответов.

На практике строгое разделение выборки на «хороших» и «плохих» заемщиков обычно неосуществимо, поэтому мы вводим переменную a_i , принимающую значения более или равное 0 и характеризующую ошибку классификации. Тогда, если заемщик i — «хороший», то будет верным следу-

ющее неравенство $w_1x_{i1}+w_2x_{i2}+\dots+w_px_{ip} \geq c-a_i$, для «плохого» заемщика j имеем: $w_1x_{j1}+w_2x_{j2}+\dots+w_px_{jp} \leq c+a_j$. Для нахождения весов (w_1, w_2, \dots, w_p), которые дадут минимальную сумму абсолютных значений этих отклонений, надо решить задачу линейного программирования:

$$\begin{aligned} & a_1+a_2+\dots+a_{n_G+n_B} \rightarrow \min \\ & w_1x_{i1}+w_2x_{i2}+\dots+w_px_{ip} \geq c-a_i; \quad 1 \leq i \leq n_G, \\ & w_1x_{i1}+w_2x_{i2}+\dots+w_px_{ip} \leq \\ & c+a_i; \quad n_G+1 \leq i \leq n_G+n_B, \quad (3) \\ & a_i \geq 0; \quad 1 \leq i \leq n_G+n_B \end{aligned}$$

Или для упрощения минимизировать наибольшее значение отклонения:

$$\begin{aligned} & a \rightarrow \min \\ & w_1x_{i1}+w_2x_{i2}+\dots+w_px_{ip} \geq c-a; \quad 1 \leq i \leq n_G, \\ & w_1x_{i1}+w_2x_{i2}+\dots+w_px_{ip} \leq c+a; \\ & n_G+1 \leq i \leq n_G+n_B, \quad (4) \\ & a_i \geq 0 \end{aligned}$$

Для построения скоринговой модели преимущество линейного программирования по сравнению со статистическими методами — это возможность внесения необходимых ограничений в модель. Неудобство линейного программирования состоит в том, что нельзя оценить насколько статистически верны оцененные параметры (в отличие от регрессионных методов).

Дерево классификации — один из методов автоматического анализа данных, позволяющий отнести потенциального заемщика к одному из заранее известных классов. Сущность этого метода заключается в следующем:

1. На основе данных за прошлые периоды строится дерево. При построении дерева все известные ситуации обучающей выборки сначала попадают на первый уровень, а потом, в зависимости от полученных данных о заемщике, распределяются по следующим уровням, которые, в свою очередь, также могут быть разбиты по уровням. В целях определения поля, по которому будет происходить разбиение, используется метод устранения неопределенности. Неопределенность тем выше, чем больше объектов, относящихся к различным классам, находятся на одном уровне.

2. При существенном изменении текущей ситуации на рынке дерево можно адаптировать к существующей обстановке.

Нейронные сети представляют собой метод моделирования, позволяющий воспроизводить чрезвычайно сложные зависимости. Алгоритм построения систем оценки риска на основе нейронных сетей следующий:

1. Составление базы данных на основе имеющихся данных по клиентам, разбиение совокупности данных на два множества: обучающее и тестовое (возможно разбиение на три множества: обучающее, тестовое и подтверждающее).

2. Выбор системы характеристик, выделяющих заемщиков по принципу «плохой, хороший», и преобразование данных для подачи на вход сети; выбор системы кодирования выходных значений (классическое кодирование, 2 на 2 кодирование и прочее).

3. Выбор количества слоев сети, число нейронов в слоях, функции активации нейронов, алгоритма обучения сети и прочее; оценка качества работы сети на основе подтверждающего множества, оптимизация архитектуры (уменьшение весов, прореживание пространства признаков); выбор варианта сети, который обеспечивает наилучшую способность к обобщению, и оценка качества работы по тестовому множеству.

4. Определение степени влияния различных факторов на принимаемое решение; вычисление точности классификации; при необходимости возврат на этап 2, смена способа представления образцов и базы.

Генетический алгоритм построен по аналогии с процессом естественного отбора. В сфере кредитования это выглядит следующим образом: имеется набор классификационных моделей, которые подвергаются «мутации», «скрещиваются», и в результате отбирается «сильнейший», то есть модель, дающая наиболее точную классификацию.

Предположим, для классификации кредитной заявки нам надо оценить параметры a_1, a_2, \dots, a_p , b_1, b_2, \dots, b_p и c в уравнении кредитного скоринга следующего вида:

$$f(x_i) = a_1 x_{i1}^{b1} + a_2 x_{i2}^{b2} + \dots + a_p x_{ip}^{bp} + c, \quad (5)$$

где x_{i1}, \dots, x_{ip} являются значениями ха-

рактеристик i -ого заемщика.

После того, как мы найдем значения параметров, заемщик будет классифицирован как «хороший» или «плохой» в соответствии с тем, примет ли функция положительное или отрицательное значение.

Генетический алгоритм начинает свою работу с формирования начальной популяции $I_0 = \{i_1, i_2, \dots, i_s\}$ — конечного набора допустимых решений задачи. Эти решения могут быть выбраны случайным образом. Для того чтобы использовать схему генетических алгоритмов, нам необходимо преобразовать переменные a, b, c в бинарные, принимающие значения $\{0, 1\}$.

На каждом шаге эволюции с помощью вероятностного оператора селекции выбираются два решения i_1 и i_2 . Оператор скрещивания по решениям i_1, i_2 строит новое решение j , которое затем подвергается небольшим случайным модификациям, которые принято называть мутациями. Затем решение добавляется в популяцию, а решение с наименьшим значением целевой функции удаляется из популяции.

Остановимся подробнее на основных операторах этого алгоритма: селекции, скрещивании и мутации. Среди операторов селекции наиболее распространенными являются метод рулетки и метод турнирной селекции. Метод рулетки отбирает особей с помощью n «запусков» рулетки. При этом вероятность на n -м шаге выбрать решение j в качестве одного из родителей задается формулой:

$$p_j = \left(\frac{f_j}{\sum_{j=1}^n f_j} \right), \quad \text{где } f_j > 0 \text{ для всех } i. \quad (6)$$

При таком отборе члены популяции с более высокой приспособленностью с большей вероятностью будут чаще выбираться, чем особи с низкой приспособленностью.

Турнирный отбор реализует n турниров, чтобы выбрать n особей. При этом формируется случайное подмножество из элементов популяции и среди них выбирается один элемент с наибольшим значением целевой функции.

Как только два решения выбраны, к ним применяется вероятностный оператор скрещивания — кроссовер. Существует много различных версий этого оператора, среди которых простейшим является однородный оператор. По решениям i_1, i_2 он строит решение j , присваивая каждой координате этого вектора с вероятностью 0,5 соответствующее значение одного из родителей. Если вектора i_1, i_2 совпадали, скажем, по первой координате, то вектор j «унаследует» это значение.

Оператор мутации, применяемый к решению j генетического алгоритма, с заданной вероятностью p_m (0, 1) меняет значение каждой координаты на противоположное. Таким образом, с ненулевой вероятностью решение j может перейти в любое другое решение. Отметим, что модификация решения j может состоять не только в случайной мутации, но и в частичной перестройке решения алгоритмами локального поиска. Применение локального спуска позволяет генетическому алгоритму сосредоточиться только на локальных оптимумах.

Работа генетического алгоритма представляет собой итерационный процесс, который продолжается до тех пор, пока не выполняются заданный критерий остановки, например, приспособленность индивидуумов перестает заметно увеличиваться или осуществляется заданное число поколений. На каждом поколении генетический алгоритм реализует отбор пропорционально приспособленности (значении целевой функции на этом индивидууме). Каждое следующее поколение будет в среднем лучше предыдущего.

При использовании метода ближайших соседей выбирается единица измерения для определения расстояния между клиентами. Все клиенты в выборке получают определенное пространственное положение. Каждый новый клиент классифицируется исходя из того, каких клиентов — плохих или хороших — больше вокруг него.

В целом, учитывая разнообразие методов, можно выделить следующие преимущества кредитного скоринга:

- возможность снижения издержек и минимизация операционного риска за счет автоматизации принятия решения о выдаче кредита;

- сокращение времени обработки заявлений и предоставления ответа о выдаче или отказе в кредите;

- выявление и предотвращение попыток мошенничества;

- помощь в отслеживании ситуации по счетам, управлении портфелем, большое значение для систем раннего предупреждения.

Однако кредитный скоринг имеет ряд недостатков:

- классификация выборки производится только на клиентах, которым дали кредит; при этом отсутствует возможность узнать поведение клиентов, которым в кредите было отказано;

- скоринговые модели необходимо разрабатывать на выборке из наиболее «свежих» клиентов с учетом изменения социально-культурного уровня, периодически проверять качество работы системы, и когда качество ухудшается, разрабатывать новую модель;

- программа оценивает не реального человека, а информацию, которую он о себе сообщает, и хорошо подготовленный клиент может представить данные о себе так, что практически гарантированно получит кредит;

- сотрудники банка не имеют достаточной подготовки, чтобы разрабатывать стратегии на основе скоринга и управлять ими, а также осуществлять эффективный мониторинг результатов работы модели;

- системы управленческой информации неадекватны для эффективной оценки, мониторинга и валидации модели;

- ненадлежащее применение моделей кредитного скоринга к продуктам, подгруппам заявителей или географическим регионам, не для которых они были разработаны, без проверки эффективности.

Примечания:

1. Центральный банк российской федерации. URL: www.cbr.ru.
2. Эйтшгтон В.Н., Анохин С.А. Прогнозирование банкротства: основные методики и проблемы. М.: ИНФРА-М, 2007.

3. Churchill G.A., Nevin J.R., Watson R.R. The role of credit scoring in the loan decision // Credit World. 1977. March.

References:

1. Central bank of the Russian Federation. URL: www.cbr.ru.
2. Eitshgton V.N., Anokhin S.A. Bankruptcy forecasting: the main techniques and problems. M: INFRA-M, 2007.
3. Churchill G.A., Nevin J.R., Watson R.R. The role of credit scoring in the loan decision//Credit World. 1977. March.