

# БАНКОВСКОЕ ДЕЛО

Ю. В. Сажин

заведующий кафедрой статистики Мордовского государственного университета им. Н. П. Огарева

## СИСТЕМА СКОРИНГА ДЛЯ ОЦЕНКИ КРЕДИТОСПОСОБНОСТИ ЗАЕМЩИКА

Важным критерием предоставления кредитов банками выступает кредитоспособность заемщика, т. е. способность погашать ссудную задолженность. Ее оценка представляет собой оценку банком заемщика с точки зрения возможности и целесообразности предоставления ему кредита. Она определяет вероятность его своевременного возврата и выплаты процентов.

Исследование кредитоспособности осуществляется с целью качественной оценки заемщика до решения вопроса о выдаче кредита и его условиях, определения способности и готовности клиента вернуть взятые им в долг средства в соответствии с кредитным договором.

Основными задачами определения кредитоспособности заемщика являются изучение его финансового положения, предупреждение потерь кредитных ресурсов вследствие неэффективной деятельности заемщика, а также возможностей стимулирования заемщика в направлении повышения эффективности его деятельности.

Изучение банками разнообразных факторов, которые могут повлечь за собой непогашение кредитов, или, напротив, обеспечивают их своевременный возврат, составляет содержание банковского анализа кредитоспособности.

Для решения такого рода задач можно предложить использовать систему *скоринга*, заключающуюся не только в том, чтобы выяснить, в состоянии клиент выплатить кредит или нет, но и степень надежности и обязательности клиента. Иными словами, скоринг оценивает, насколько клиент «creditworthy», т. е. насколько он «достоин» кредита. Скоринг представляет собой статистическую модель, с помощью которой на основе кредитной истории «прошлых» клиентов банк определяет, насколько велика вероятность, что конкретный потенциальный заемщик вернет кредит в срок.

При оценке и прогнозировании банковских кредитных рисков могут быть использованы различные подходы, в частности, методы статистической классификации (Гаскаров, Голинкевич, Мозгалевский, 1974; Загоруйко, 1972; Затолокин, 1985). Данные методы позволяют на основе информативных признаков клиентов банка классифицировать их на различные группы, одной из которых может быть группа, имеющая высокую степень вероятности возврата кредита (платежеспособная), а другой – группа, имеющая очень низкую степень вероятности возврата кредита (неплатежеспособная).

Применение статистической классификации включает следующие основные этапы: постановка задачи, выбор информативных социально-экономических признаков, формирование обучающей выборки, построение решающего правила, оценка достоверности проведенной классификации, клиентов.

На этапе постановки задачи для оценки и прогнозирования был определен банк и подготовлена соответствующая информационная база по информативным признакам.

Для нахождения совокупности признаков, обеспечивающих наиболее высокую достоверность оценки рисков, предварительно составляют первоначальный набор их, исходя из социально-экономических соображений, с учетом данных ранее проводимых исследований, анализа рисков и опыта работы. После этого оцениваются различные совокупности из данного набора и выбирают наиболее информативный комплекс согласно соответствующим критериям. Окончательный выбор наиболее информативной совокупности признаков проводится на основе результатов апробирования с учетом соображений экономического и социального характера.

В итоге имеется множество объектов, каждый из которых характеризуется  $n$ -мерным вектором признаков. Задача распознавания включает выполнение двух этапов: обучения и распознавания.

На первом этапе задается критерий разбиения единиц на классы. За векторы признаков принимаются некоторые социально-экономические параметры клиентов. По результатам испытаний обучающей выборки известно, к какому из классов принадлежит каждый из них. На основании этого строится решающее правило. В качестве решающего правила в данном случае выбран логарифм отношения правдоподобия. Необходимо стремиться получить в достаточной степени представительные выборки объектов каждого класса.

Число клиентов в обучающих выборках каждого класса должно быть не меньше размерности пространства признаков. Это необходимо для того, чтобы ковариационная матрица признаков не оказалась особой и, следовательно, имела бы обратную матрицу, требующуюся для вычисления логарифма отношения правдоподобия.

На втором этапе на основании использования обучающей выборки осуществляется классификация клиентов: тот или иной клиент относится с помощью построенного правила классификации к одному из классов в соответствии со значениями его признаков.

При построении решающего правила на основании обучающей выборки первоначально оцениваются параметры многомерного нормального распределения — векторы средних значений и ковариационные матрицы для различных классов. Оценки компонентов векторов средних значений ( $m_j^{(h)}$ ) и элементов ковариационных матриц ( $\sigma_k^{(h)}$ ) для каждого из классов определяются по формулам

$$m_j^{(1)} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_{ij}; \quad (1)$$

$$m_j^{(2)} = \frac{1}{n_2} \sum_{i=1}^{n_2} x_{ij}; \quad (2)$$

$$\sigma_{kl}^{(1)} = \frac{n_1}{n_1 - 1} \left( \frac{1}{n_1} \sum_{i=1}^{n_1} x_{ik} x_{il} - m_k^{(1)} m_l^{(1)} \right); \quad (3)$$

$$\sigma_{kl}^{(2)} = \frac{n_2}{n_2 - 1} \left( \frac{1}{n_2} \sum_{i=1}^{n_2} x_{ik} x_{il} - m_k^{(2)} m_l^{(2)} \right); \quad j, k, l = 1, 2, \dots, P, \quad (4)$$

где  $n_1, n_2$  — число объектов, принадлежащих к первому и второму классам. Логарифм отношения правдоподобия при гауссовской аппроксимации плотностей  $f_1(x)$  и  $f_2(x)$  равен

$$L(x) = \frac{1}{2} (x - m_1)^* S_1^{-1} (x - m_1) + \frac{1}{2} (x - m_2)^* S_2^{-1} (x - m_2) + \frac{1}{2} \ln \frac{|S_2|}{|S_1|}. \quad (5)$$

В случае равных ковариационных матриц решающее правило получается линейным и имеет вид

$$L(x) - X^* S_1^{-1} (m_1 - m_2) - \frac{1}{2} (m_1 + m_2)^* S_2^{-1} (m_1 - m_2), \quad (6)$$

где  $S_1^{-1}$ ,  $S_2^{-1}$  — обратные матрицы  $S_1$  и  $S_2$ ,  $|S_1|$ ,  $|S_2|$  — определители матриц;  $X$  — текущий вектор-столбец;  $*$  — знак транспонирования.

При классификации  $i$ -го клиента значения его признаков подставляются в выражение (6). Полученное значение  $L(X_i)$  сравнивается с пороговым значением  $\Lambda$ , при  $L(X) \geq \Lambda$  клиент относится к первому классу, при  $L(X) < \Lambda$  — ко второму классу. Величина порога  $\Lambda$  определяется априорными вероятностями классов  $q_1$  и  $q_2$  и задается матрицей потерь  $W = (w_{ij})$ . Порог рекомендуется выбирать из условия равенства средних потерь, вызванных ошибками первого и второго рода  $P_{12}w_{12} = p_{21}w_{21}$ , при этом порог определяется по формуле

$$\Lambda = \ln \frac{q_2 w_{12}}{q_1 w_{21}}. \quad (7)$$

В качестве характеристик достоверности отнесения клиента к различным группам можно использовать вероятности ошибок при разделении классов, вероятность правильной классификации, средний риск. В данном случае качество решающего правила оценивается ожидаемым риском и рассчитывается следующим образом:

$$M = w_{12} q_{12} \alpha_{12} + w_{21} q_{21} \alpha_{21}, \quad (8)$$

где  $\alpha_{12} = \int_{\lambda_1}^{\infty} \varphi_2(L) dL$  — вероятность ошибки первого рода (когда объект из второго

класса относится решающим правилом к первому классу);  $\alpha_{21} = \int_{-\infty}^{\lambda_2} \varphi_1(L) dL$  —

вероятность ошибки второго рода (когда объект из первого класса относится решающим правилом ко второму классу);  $\varphi_1(L)$ ,  $\varphi_2(L)$  — плотность распределения логарифма отношения правдоподобия при условии, что объекты принадлежат соответственно классам первому и второму.

Блок-схемы расчета правила классификации и классификации клиентов представлены на рис. 1–2.

Реализация описанного метода осуществлена нами на примере клиентов одного из банков г. Саранска, которые обслуживались данным банком более трех лет. Была отобрана одна выборка, в которую входили платежеспособные клиенты и вторая — в которую входили неплатежеспособные. В результате этого каждый клиент описывался совокупностью  $n$  социально-экономических признаков и было известно, к какому из классов (платежеспособных или неплатежеспособных) он принадлежит.

Из 120 клиентов банка к группе платежеспособных отнесен 101 человек, к группе неплатежеспособных — 19 человек. Следует отметить, что отдельные информативные признаки носят качественный характер, поэтому предлагается провести градацию внутри них на основе экспертной оценки и присваивать каждой из них соответствующий рейтинг. Предпочтение отдается наиболее высокому значению рейтинга. Так, например,  $X_1$  — возраст (лица до 21 года имеют рейтинг — 1, лица 55 лет и старше имеют рейтинг 2, лица от 26 до 55 лет имеют рейтинг 3). Аналогично проведена балльная оценка остальных переменных:

$X_2$  — пол (мужчины — 2, женщины — 1);

$X_3$  — адрес прописки (г. Саранск — 3, др. город — 2, поселок и село — 1);

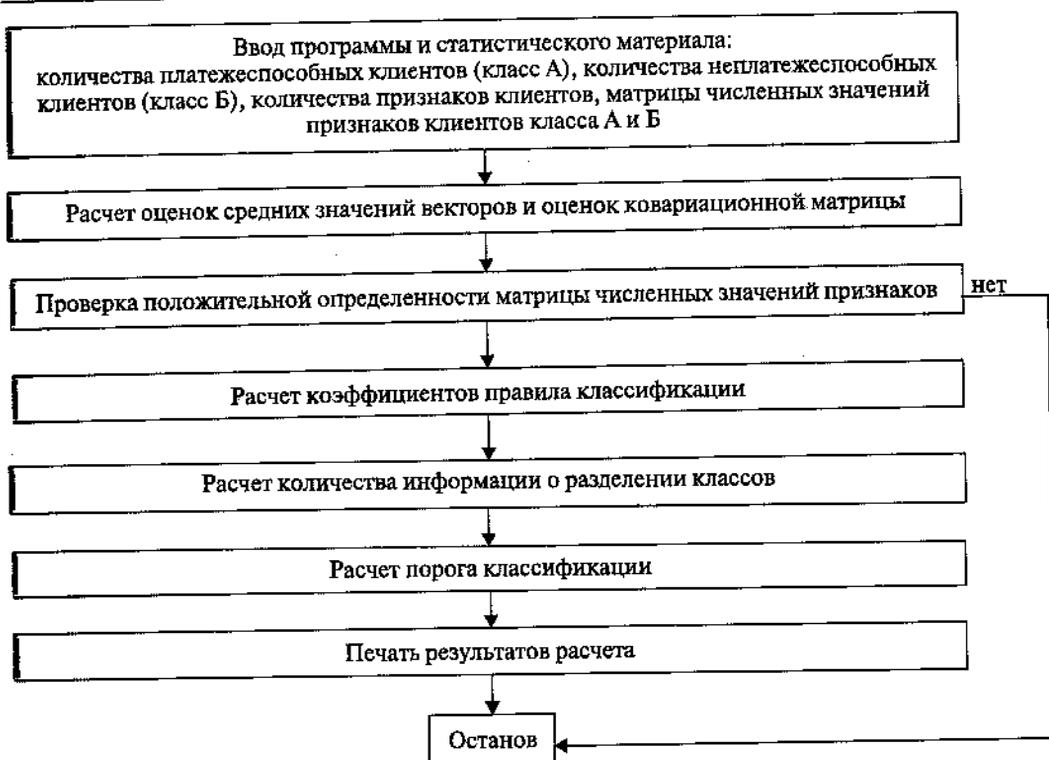


Рис. 1. Блок-схема расчета правила классификации

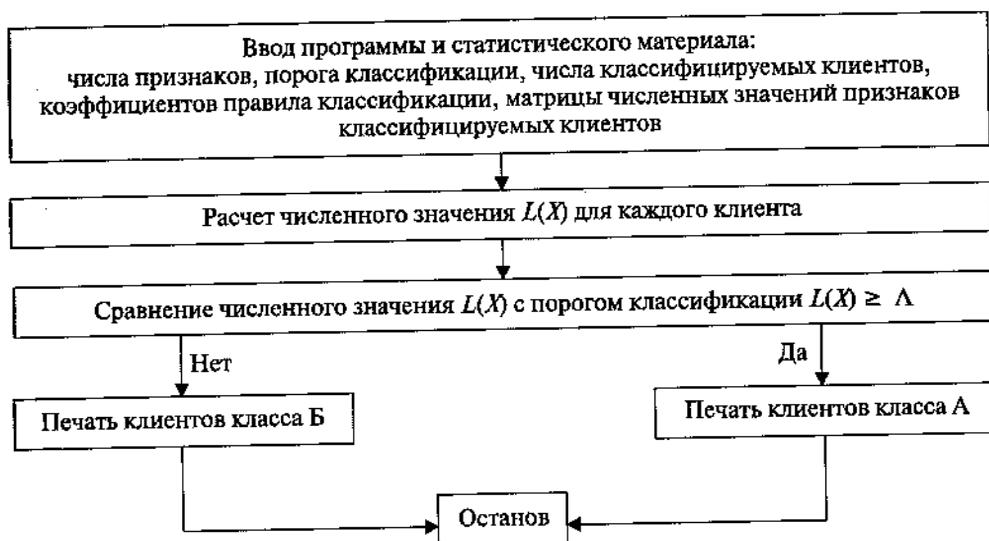


Рис. 2. Блок-схема классификации клиентов

- $X_4$  – место работы (бюджетные учреждения – 4, крупные ОАО и т.д. – 3, пенсионеры – 2, индивидуальные частные предприятия – 1);
- $X_5$  – продолжительность работы, в годах;
- $X_6$  – семейное положение (зам/жен – 1, нет – 0);
- $X_7$  – количество иждивенцев, чел.;
- $X_8$  – место работы созаемщика (бюджетное учреждение – 4, крупное ОАО и т.д. – 3, пенсионер – 2, индивидуальное частное предприятие – 1);
- $X_9$  – доход, руб.;

$X_{10}$  — наличие движимого и недвижимого имущества (и то и другое — 2, только что-то одно — 1, ничего — 0).

В результате сбора и обработки статистической информации получены числовые значения признаков, на основании которых разработано линейное решающее правило, которое построено с использованием пакета прикладных программ Statistica, раздел «Дискриминантный анализ» (см. Загоруйко, 1972). Порог классификации равен  $\Lambda = 0$ . Правило линейной классификации имеет вид:

$$(-0,974)(x_1 - 2,865) + 0,5(x_2 - 1,354) + 0,556(x_3 - 2,491) + (-0,214)(x_4 - 3,021) + 0,109(x_5 - 6,495) + 2,469(x_6 - 0,551) + (-0,597)(x_7 - 1,066) + 0,009(x_8 - 1,196) + 0,001(x_9 - 3898,9) + 0,212(x_{10} - 0,315) \geq 0.$$

Далее была проведена классификация клиентов полученной выборки по разработанному правилу и порогу классификации. Клиенты принадлежат к классу А (платежеспособным), если выполняется неравенство:

$$\sum_{i=1}^n a_i (X_i - d_i) \geq \Lambda, \quad (9)$$

где  $a_i$  — коэффициент решающего правила;  $n$  — число признаков;  $X_i$  — значения соответствующих признаков, вновь предъявляемых для классификации клиентов;  $\Lambda = 0$  — порог классификации;  $d_i$  — среднее значение средних векторов показателей клиентов для классов А и Б.

Если неравенство (9) не выполняется, то клиенты принадлежат к классу Б. При классификации клиентов определяются ошибки первого и второго рода. В табл. 1 приведены значения коэффициентов решающего правила, рисков первого и второго рода.

Таблица 1

### Расчетные значения коэффициентов правила классификации

Коэффициент правила классификации		$a_i$	$b_i$
Информативные параметры			
$X_1$		-0,974	2,865
$X_2$		0,500	1,354
$X_3$		0,556	2,491
$X_4$		-0,214	3,021
$X_5$		0,109	6,495
$X_6$		2,469	0,551
$X_7$		-0,597	1,066
$X_8$		0,009	1,1996
$X_9$		0,001	3898,9
$X_{10}$		0,212	0,315
$\alpha_{12}$		0,033	
$\alpha_{21}$		0,410	

Число клиентов, правильно классифицированных в классе платежеспособных, составило 79 человек из 101, значит, процент правильной классификации равен 78,22%. Число клиентов, правильно расклассифицированных во втором классе — неплатежеспособных, равен 16 из 19, т.е. составит 84,21%. Общий процент правильной классификации составляет 79,12%, что обусловлено доста-

точно высокой степенью точности и качества построенного правила классификации.

Таким образом, подготовив статистический информационный материал по ранее определенным информативным параметрам по платежеспособной и неплатежеспособной клиентуре банка, можно, используя дискриминантный анализ, построить правило классификации на основе собранного массива информации и оценить качество правила по соответствующему критерию — проценту правильной классификации объектов.

На основании построенного решающего правила можно классифицировать вновь поступающих клиентов на основе их индивидуальных информативных признаков, подставляя данные значения в решающее правило, и определить к какому классу относится данный клиент. И если клиент по набору информативных признаков относится правилом классификации к классу неплатежеспособных, то необходимо более осторожно подходить к оформлению ему соответствующего кредита, поскольку вероятность невозврата кредита в этом случае достаточно высокая. Это позволит значительно повысить достоверность оценки платежеспособности клиентов банка и вероятность возврата выданных кредитов.

Обратим внимание на то, что эффективность применения метода статистической классификации во многом зависит от выбора информативных признаков. Их количество может колебаться от нескольких единиц до десятков и сотен. Естественно стремление свести к минимуму их число как с точки зрения простоты построения правила классификации, так и с точки зрения сокращения затрат на сбор исходной информации.

С помощью соответствующих математических методов можно оценить информативность каждого признака и отобрать систему наиболее информативных из них. При этом достигается оптимальная вероятность ошибки распознавания, что позволяет при сравнительно небольшом объеме статистических данных построить наиболее точное решающее правило, сократив затраты, необходимые для измерения дополнительных параметров, и существенно уменьшив время получения оценки.

Экспериментальные расчеты значений вероятностей ошибок при различных сочетаниях признаков (при условии, что число признаков не превосходит числа объектов) показали, что для некоторых из них значение вероятностей ошибок получается в два и более раз меньше, чем для всей совокупности.

Для выделения оптимального комплекса необходимо определить количественную информативность каждого признака, а также любого произвольного комплекса. Как показано (Розин, 1973; Перроте, Карташев, Цветаев, 1968) все величины, входящие в формулы для количественных оценок информативности, могут быть оценены на основе статистического материала.

На основании статистической информации, полученной в результате ее сбора и первичной обработки, определяются оценки  $H(A)$  — энтропии множества  $A$  классов  $A_r$  ( $r = 1; 2$ , так как рассматриваются два класса — класс платежеспособных и неплатежеспособных) и  $J(x_j)$  — информативность каждого признака согласно (Henley, 1995; Boyle and others, 1992; Srinivasan, Kim, 1987; Yobas, Stook, Ross, 1997; Desai and others, 1997). Величина  $J(x_j)$  может принимать любые значения в интервале от нуля до  $H(A)$ . Равенство  $J(x_j) = H(A)$  означает, что признак  $x_j$  содержит полную информацию о принадлежности объекта к какому-либо из классов. При  $J(x_j) = H(A)$  наиболее информативная система параметров будет состоять только из одного признака. Если это условие не выполняется, то необходимо найти комплекс, состоящий из минимального числа признаков, информативность которого не ниже заданного порога.

Процедура выделения оптимального комплекса признаков будет следующей. Из всех рассматриваемых признаков выбирается один, имеющий наибольшую информативность, к нему добавляется еще один из оставшихся ( $n - 1$ ) и прово-

дится перебор информативностей комплексов ( $n - 1$ ), состоящих из двух признаков. Если среди этих комплексов не окажется информативности больше пороговой, то пересматриваются аналогично комбинации ( $n - 2$ ), состоящие из трех признаков и т. д. Такой процесс перебора комплексов продолжается до тех пор, пока не найдется комплекс, информативность которого не ниже заданной пороговой величины. Он считается оптимальным. Информативность произвольного комплекса, состоящего из  $m$  признаков, оценивается согласно (Henley, 1995).

В соответствии с изложенной методикой был разработан алгоритм, блок-схема и программа расчетов для ПЭВМ по определению информативности каждого признака, комплекса признаков и выделению оптимального из них.

Оптимальный комплекс признаков определялся для клиентов банка, для которых ранее было построено правило классификации по 10 признакам. В табл. 2 и 3 представлены результаты расчетов по определению информативности каждого признака, комплекса признаков и выделение оптимального из них.

Таблица 2

### Результаты определения информативности признаков

Информативный признак	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$	
Информативность	0,001	0,001	0,051	0,002	0,032	0,221	0,083	0,003	0,325	0,088	0,807

В качестве порога информативности, согласно вышеизложенному, принималась сумма информативности статистически независимых признаков, составляющих комплекс из десяти признаков. Он равен 0,807 (табл. 2). Данные табл. 3 показывают, что оптимальным является комплекс, состоящий из двух признаков:  $X_9$  — доход,  $X_{10}$  — семейное положение.

Таблица 3

### Результаты расчетов по определению информативности комплекса признаков

Информативные признаки, входящие в комплекс	Информативность комплекса
$X_9$	0,325
$X_9, X_6$	0,546
$X_9, X_6, X_{10}$	0,546
$X_9, X_6, X_{10}, X_3$	0,546
$X_9, X_6, X_{10}, X_3, X_7$	0,546

Проведенная оценка решающего правила, построенного по двум информативным признакам, показала, что процент правильной классификации платежеспособных клиентов составил 73,5%, для неплатежеспособных клиентов 80,3%, а общий процент правильной классификации составляет 75,3%. Это достаточно близко к правилу классификации, построенному по десяти признакам. Следует отметить, что в среднем процент правильного распознавания находится в пределах 70–80%.

Таким образом, нами показана целесообразность применения метода статистической классификации к решению задач классификации клиентов банка на два класса — платежеспособных и неплатежеспособных. При этом эффективность метода значительно может быть повышена за счет использования процедуры выбора оптимального комплекса признаков, что позволяет значительно сократить трудоемкость сбора и обработки информации и повысить достоверность результатов.

## Источники

- Гаскаров Д.В., Голикевич Т.А., Мозгалецкий А.В. Прогнозирование технического состояния и надежности радиоэлектронной аппаратуры. М., 1974.
- Загоруйко Н.Г. Методы распознавания и их применение. М., 1972.
- Затолокин В.М. Методы анализа качества продукции. М., 1985.
- Лерроте А.И., Карташев Г.Д., Цветаев К.Н. Основы ускоренных испытаний радиоэлементов на надежность. М., 1968.
- Розин Б.Б. Теория распознаваний образов в экономических исследованиях. М., 1973.
- Boyle M., Crook J.N., Hamilton R., Thomas L.C. Methods for credit scoring applied to slow payers in Credit Scoring and Credit Control. Oxford University Press, 1992.
- Desai V.S., Conway D.G., Crook J.N., Overstreet G.A. Credit scoring models in the credit union environment using neural networks and genetic algorithms // TMA.T Mathematics applied in business and industry. 1997. N 8.
- Henley W.E. Statistical aspects of credit scoring. Ph. D. thesis. Open university, 1995.
- Srinivasan V., Kim Y.H. Credit granting: a comparative analysis of classification procedures // Journal of Finance. 1987. Vol. 42.
- Yobas M.B., Crook J.N., Ross P. Credit scoring using neural and evolutionary techniques // Credit research Centre, University of Edinburgh. Working Paper 1997. N 2.