

УДК 519.22:336.144.36

ПРОГНОЗИРОВАНИЕ НАДЕЖНОСТИ БАНКОВ С ПОМОЩЬЮ ДЕРЕВЬЕВ КЛАССИФИКАЦИИ

Кириллов К. В.¹, Халафян А. А.²

BANK RELIABILITY FORECASTING USING CLASSIFICATION TREES

Kirillov K. V., Halafyan A. A.

The authors use the methods of classification analysis to forecast the reliability of banks. The classification trees, which use balance sheet items with the smallest binary correlation as attributes, are made. Special aspects of modelling trees for the group of banks under consideration are described, as well as those of the formation of training and testing samples and selection of criteria for truncating of branches. The model developed was tested on real data. Financial condition of 557 Russian banks was analysed based on the data for 2008, when a significant increase in the number of bankruptcy was recorded.

Keywords: classification trees, training and testing sample, logical derivation, truncation criterium.

Введение

Банковский бизнес во всем мире выступает одной из самых важных отраслей экономики. Кредитные институты, являясь основополагающей составляющей экономического развития, всегда отображают в своём развитии состояние всей экономики, что наглядно продемонстрировали периоды резкого падения экономики в 1998 г. и 2008 г., когда количество банкротств кредитных организаций резко возросло. В связи со сложившейся ситуацией органы надзора как национального, так и международного уровня выдвигают новые, более жесткие регламентационные требования к ведению отчётности об экономической деятельности банков.

В некоторых случаях кредитные организации прибегают к фальсификации отчётности экономической деятельности, усложняя органам пруденциального надзора осуществление деятельности по выявлению проблемных субъектов и прогнозированию экономической ситуации в отрасли. Поэтому «классические» показатели надёжности дают все менее адекватную оценку состояния субъекта и ситуации в целом. Возникшую проблему можно решить с помощью перехода

от оценки одного субъекта к оценке группы субъектов, обладающих схожими признаками. Следовательно, задача достоверной оценки и прогноза надёжности кредитной организации сводится к задаче сегментации всей совокупности имеющихся банков за счет классификации банков на группы однородности.

Для анализа надёжности банков как в России, так и за рубежом часто используется рейтинговая оценка деятельности кредитных организаций. Важную группу методов анализа надёжности банков составляют подходы, основанные на применении дискриминантного анализа — статистического метода изучения различий между двумя или более группами объектов по совокупности нескольких финансовых показателей. Объекты (в данном случае банки) разбиваются на несколько попарно непересекающихся групп на основании ряда показателей, характеризующих надёжность и успешность их работы. J. Argenti [1] первым отметил важность качественных показателей в вопросе о банкротстве корпораций и выделил двенадцать переменных, из которых восемь являются причинными факторами, а остальные четыре — симптомами банкротства.

¹Кириллов Кирилл Валерьевич, аспирант кафедры прикладной математики Кубанского государственного университета; e-mail: k.kirillov@mail.ru.

²Халафян Алексан Альбертович, д-р техн. наук, профессор кафедры прикладной математики Кубанского государственного университета; e-mail: khalaphyan@kubannet.ru.

Разработке различных алгоритмов применения дискриминантного анализа для исследования финансового состояния кредитных организаций посвящены работы Э. Альтмана, Р. Таффлера, А. Кумара, Д. Чессера, Г. Холдмэна, П. Нараяна, Марэ, Д. Паттелла, И.Г. Оводова, С.Н. Шевченко, Т. Шумвейя, А.А. Пересецкого, А.М. Карминского. В последнее время появились модели оценки надежности банков, основанные на теории марковских процессов (Э. Альтман, Д. Као, работы специалистов рейтингового агентства Standard&Poor's).

Для прогнозирования банковских банкротств также применялись деревья логического вывода [2–6]. А.Ф. Атия, Дж. Баек, С. Чоу, М.Л. Насир, Р.И. Джон, С.К. Беннетт [7–9] развили в своих работах различные алгоритмы нейронно-сетевых и нечетко-множественных моделей оценки финансово-экономического состояния банков.

Банки активно применяют деревья решений для определения кредитоспособности клиентов при выдаче кредита [10, 11].

Целью исследований, которым посвящена эта статья, является построение наилучшего относительно размерности и точности прогнозирования дерева, позволяющего оценить надежность выбранного банка. Задача деревьев классификации — создать на основе некоего обучающего множества известных объектов модель, с помощью которой новые объекты могут быть разделены на классы на основе их свойств. При этом заранее известны классы и принадлежность к ним объектов обучающего множества. К преимуществам деревьев решений можно отнести возможность графического представления сгенерированных логических правил классификации и простоту интерпретации. Другим важным отличительным свойством является широкая применимость — исследуемые данные не обязательно должны подчиняться некоторому общему закону распределения и в отличие от классического дискриминантного анализа могут быть измерены в номинальной или порядковой шкале.

1. Математическая обработка статистических данных

Для создания прогнозирующей математической модели надежности кредитных организаций в условиях финансовой неустойчивости были построены деревья классифика-

ции группы российских банков — обучающей выборки. С помощью ресурса “Bankscope” была сформирована выборка банков, состоящая из 814 действующих организаций и 107 организаций, лицензии которых были отозваны после 2008 г. Каждый субъект характеризуется более чем 90 показателями из статей бухгалтерского баланса, отчётов о прибыли и убытках, отчётов о достаточности капитала на покрытие ссуд и прочих активов, а также из различных нормативов экономической деятельности. Для получения адекватных результатов анализа исследуемые объекты должны обладать схожими по размерности показателями, так как наличие нетипичных величин способствует искажению сегментации объектов. Например, сложно сравнивать Сбербанк с региональными банками, у которых объём экономической деятельности в тысячи раз меньше. Исходя из этих соображений, из анализируемой выборки были отсечены банки, сумма активов и пассивов которых превышает 1000 миллионов или меньше 50 миллионов долларов США. В ходе отсеечения в выборке осталось 557 банков, из которых у 48 были отозваны лицензии после кризиса 2008 года.

Было проведено логарифмирование статистических данных, хотя этот шаг не является обязательным для построения деревьев логического вывода. Предварительно было сокращено количество показателей, характеризующих экономическую деятельность кредитной организации. При сильной парной зависимости отдельных характеристик они могут описывать одну и ту же информацию и построенные деревья будут иметь нечеткую структуру. В связи с этим были удалены все показатели с высокой ($> 0,8$) парной корреляцией, а также нормативные характеристики, так как их информационная составляющая и так будет отражаться в полученных факторах при совпадающих исходных статьях баланса и отчётов. После проведенной структуризации данных анализировалась выборка из 557 кредитных организаций, каждая из которых характеризуется следующими 24 статьями бухгалтерской отчетности: 1 — сумма активов и пассивов, 2 — собственный капитал, 3 — чистые доходы (расходы), 4 — колебания процентной ставки, 5 — проценты по активам, 6 — проценты по собственному капиталу, 7 — прочие обязательства, 8 — прочие ценные бумаги, 9 — основные средства, 10 — резерв на возможные потери по ссудам, 11 — вклады физических

лиц, 12 — средства кредитных организаций, 13 — прочие депозиты и краткосрочные займы, 14 — долгосрочные вложения в ценные бумаги, 15 — наращенные доходы и авансовые расходы, 16 — просроченная задолженность, 17 — прочая операционная прибыль, 18 — чистый комиссионный доход, 19 — безотзывные обязательства кредитной организации, 20 — объем выплат процентов по депозитам клиентов, 21 — доходы от торговли иностранной валютой, 22 — обязательные резервы, 23 — неиспользованная прибыль за отчетный период, 24 — средства в Центральном банке.

2. Особенности построения дерева классификации для рассмотренной выборки

Дерево классификации строили на платформе Deductor. Существуют различные алгоритмы построения деревьев классификации: CART, C4.5, Newld, ITrule, CHAID, CN2 и т.д. В исследовании был применен алгоритм C4.5, в котором для оптимального выбора атрибута используется критерий максимизации величины $Info(T) - Info_x(T)$, где $Info(T)$ — энтропия исходного множества объектов,

$$Info_x(T) = \sum_{i=1}^n \frac{|T_i|}{|T|} Info(T_i),$$

а $|T|$ — мощность множества T . Множества $T_1, \dots, T_n$ являются результатом разбиения множества T по атрибуту X . Было построено дерево, использующие в качестве атрибутов выделенные 24 статьи балансовой отчетности. В табл. 1 они обозначены соответственно перечисленным номерам. Столбец «Условие» отображает правила разбиения на группы, называемые также ветвями дерева. В столбце «Следствие» расположены значения листьев («банкрот» или «действующий»). Значению «банкрот» соответствует класс обанкротившихся банков, а значению «действующий» — класс действующих банков. Столбец «Поддержка» говорит о том, сколько объектов из общего количества было классифицировано по данному правилу и какова их доля в общей выборке. Столбец «Достоверность» показывает количество и процентное соотношение правильно определённых объектов по данному правилу. Дерево состоит из 11 ветвей и 12 листьев.

Если для формирования тестирующей выборки использовать случайный выбор объектов, то в выборку, например, из 28 объектов может попасть 2 обанкротившихся и 26 действующих банков в соответствии с изначальным соотношением 48 на 509. При таком соотношении правильная классификация обанкротившихся банков гораздо важнее классификации действующих. Поэтому разбиение на обучающую и тестирующую выборки следует проводить так, чтобы количество обанкротившихся банков в тестирующей выборке было сравнимо с количеством действующих. Для этого банки были упорядочены по статье отчетности «Сумма активов и пассивов», которая, как показал факторный анализ рассмотренной выборки банков [12], в наибольшей степени влияет на построение факторов и на формирование ветвей дерева. Для тестирующей выборки были выбраны банки, имеющие различные значения вышеупомянутой статьи отчетности. Сформированная таким образом тестирующая выборка содержала 8 обанкротившихся и 20 действующих банков.

В табл. 2 приведены уровни значимости каждой статьи в классификационной модели. Как и следовало ожидать, это оказались статьи из числа наиболее сильно влияющих на каждый фактор, выделенный авторами с помощью факторного анализа [12]: сумма активов и пассивов (1), средства кредитных организаций (12), чистые доходы (расходы) (3), собственный капитал (2) и чистый комиссионный доход (18). Статьи 20, 19 и 7 в меньшей степени влияли на факторы, значения их уровней значимости невелики по сравнению с другими статьями.

3. Оценка точности построенного дерева

Целью анализа с помощью деревьев классификации является получение максимально точной классификации, которая связана с наименьшей долей неправильно классифицированных объектов. Чем больше размерность дерева, тем точнее классификация, но сложнее интерпретация результатов. Дерево должно содержать информацию, улучшающую точность прогноза и игнорирующую ветвления, не улучшающие прогноз. Пользователь статистических пакетов может сам устанавливать максимальные размеры дерева. Обычно это делается на основе информации об уже построенных деревьях, диагно-

Таблица 1. Правила логического вывода

№ правила	Условие			Следствие	Поддержка		Достоверность	
	Статья	Знак	Значение		Кол-во	%	Кол-во	%
1	1	<	10,3698	банкрот	12	2,27	11	91,67
2	1	>=	10,3698	банкрот	3	0,57	2	66,67
	18	<	6,8323					
3	1	>=	10,3698	действующий	3	0,57	3	100,00
	18	>=	6,8323					
	3	<	10,9235					
	2	<	11,1679					
4	1	>=	10,3698	банкрот	4	0,76	4	100,00
	18	>=	6,8323					
	3	<	10,9235					
	2	>=	11,1679					
5	1	>=	10,3698	действующий	5	0,95	5	100,00
	18	>=	6,8323					
	3	>=	10,9235					
	7	<	6,4057					
	20	<	-2,4959					
6	1	>=	10,3698	банкрот	7	1,32	5	71,43
	18	>=	6,8323					
	3	>=	10,9235					
	7	<	6,4057					
	20	>=	-2,4959					
	19	<	1,6621					
7	1	>=	10,3698	действующий	2	0,38	2	100,00
	18	>=	6,8323					
	3	>=	10,9235					
	7	<	6,4057					
	20	>=	-2,4959					
	19	>=	1,6621					
	18	>=	6,8323					
	3	>=	10,9235					
	7	>=	6,4057					
	1	<	10,7201					
	12	<	5,3234					
	1	>=	10,3698					
9	18	>=	6,8323	банкрот	12	2,28	9	75,00
	3	>=	10,9235					
	7	>=	6,4057					
	1	<	10,7201					
	12	>=	5,3234					
	1	>=	10,3698					
10	18	>=	6,8323	действующий	17	3,21	17	100,00
	3	>=	10,9235					
	7	>=	6,4057					
	1	>=	10,7201					
	1	<	10,9613					
	12	<	5,3234					
11	18	>=	6,8323	банкрот	2	0,38	2	100,00
	3	>=	10,9235					
	7	>=	6,4057					
	1	>=	10,7201					
	1	<	10,9613					
	12	>=	5,3234					
12	18	>=	6,8323	действующий	449	85,25	441	98,22
	3	>=	10,9235					
	7	>=	6,4057					
	1	>=	10,9613					

Таблица 2. Уровни значимость атрибутов

Атрибут (статья)	Значимость, %
1	57,821
3	13,372
12	13,077
2	8,591
18	7,139

Таблица 3. Таблица сопряженности

Обучающая выборка	Классифицировано		
Фактически	Банкрот	Действующие	Итого
Банкрот	33	7	40
Действующие	9	480	489
Итого	42	487	529
Тестирующая выборка	Классифицировано		
Фактически	Банкрот	Действующие	Итого
Банкрот	6	2	8
Действующие	2	18	20
Итого	8	20	28
Вся выборка	Классифицировано		
Фактически	Банкрот	Действующие	Итого
Банкрот	39	9	48
Действующие	11	498	509
Итого	50	507	557

стических сообщений используемого пакета программ или, в крайнем случае, интуитивно [13, 14]. Но следует помнить, что так называемая «ранняя остановка» соответствует менее точным классификационным моделям. Признанные эксперты в этой области [15] рекомендуют вместо остановки использовать отсечение. В качестве критерия отсечения эвристическим путём был выбран уровень доверия 40 %. При заданном уровне построенное дерево обладает оптимальным числом правил.

Для того чтобы оценить точность построенного дерева, была построена таблица сопряженности (табл. 3). Она содержит данные о качестве классификации с помощью построенного дерева по двум выборкам (обучающей и тестирующей), а также по всей совокупности объектов. На пересечении строки «банкрот» и столбца «действующие» указано количество ошибок классификации 1-го рода, когда обанкротившиеся банки классифицируются как действующие. Соответственно на пересечении строки «действующие» и столбца «банкрот» указано количество ошибок классификации 2-го рода, когда действующие банки классифицируются как обанкротившиеся. Шесть из восьми обанкротившихся

банков отнесены в правильный класс, из двадцати действующих восемнадцать классифицированы правильно. Как отмечалось ранее, при значительно более высоком количестве действующих банков по сравнению с числом обанкротившихся правильная классификация вторых является более важной.

Таким образом, правила логического вывода верно классифицируют 96,98 %, 85,71 % и 96,41 % объектов для обучающей, тестирующей и совокупной выборки соответственно.

Заключение

Предложен метод прогнозирования банкротств кредитных организаций в условиях нестабильности на финансовых рынках на основе деревьев классификации. Построены правила и выявлены закономерности отнесения рассматриваемых банков к классу «банкрот» или «действующий». Для тестирования построенных классификаторов проведена предварительная обработка данных более чем 500 банков РФ накануне финансового кризиса 2008 года с целью изучения структуры их финансовых показателей.

Проведенные исследования показали, что для построения дерева, обладающего опти-

мальным числом правил, в качестве критерия отсекающего достаточен уровень доверия 40%. Полученное дерево обладает высокой достоверностью построенных правил логического вывода. Разработаны практические рекомендации по формированию обучающей и тестирующей выборок, выбора критериев отсекающих ветвей. Показано, что предложенная методика может также служить для выделения статей, наиболее влияющих на надежность банка.

Литература

1. *Argenti J.* Corporate Collapse // England: McGraw-Hill, 1976. 193 p.
2. *Bryant S.M.* A case-based reasoning approach to bankruptcy prediction modeling. *Intelligent Systems in Accounting // Finance and Management.* 1997. Vol. 6. P. 195–214.
3. *Curram S.P., Mingers J.* Neural networks, decision trees induction and discriminant analysis: An empirical comparison // *Journal of the Operational Research Society.* 1994. Vol. 45. No. 4. P. 440–450.
4. *Elhadi M.T., Vamos T.* An IR-CBR approach to legal indexing and retrieval in bankruptcy law // In *Tenth proceedings in database and expert systems applications.* 1999. P. 769–774.
5. *Jo H., Han I., Lee H.* Bankruptcy prediction using case-based reasoning, neural networks, and discriminant analysis // *Expert Systems with Applications.* 1997. Vol. 13. No. 2. P. 97–108.
6. *Park C., Han I.* A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction // *Expert Systems with Applications.* 2000. Vol. 23. No. 3. P. 255–264.
7. *Atiya A.F.* Bankruptcy prediction for credit risk using neural networks: A survey and new results // *IEEE Transactions on Neural Networks.* 2001. Vol. 12. No. 4. P. 929–935.
8. *Baek J., Cho S.* Bankruptcy prediction for credit risk using an autoassociative neural network in Korean firms // In *IEEE international conference on computational intelligence for financial engineering.* 2003. P. 25–29.
9. *Nasir M.L., John R.I., Bennett S.C.* Predicting corporate bankruptcy using modular neural networks // In *IEEE international conference on computational intelligence for financial engineering.* 2000. P. 86–91.
10. *Chen M.-Y.* Predicting corporate financial distress based on integration of decision tree classification and logistic regression // *Expert Systems with Applications.* 2011. Vol. 38. P. 11261–11272.
11. *Pang Su-lin, Gong Ji-zhang* C5.0 Classification Algorithm and Application on Individual Credit Evaluation of Banks // *Systems Engineering – Theory & Practice.* 2009. Vol. 29. Iss. 12. P. 94–102.
12. *Кириллов К.В.* Анализ деятельности банка с помощью факторного анализа // Международная молодежная научно-практическая конференция «Математическое моделирование в экономике, страховании и управлении рисками». 05–08 ноября 2013, г. Саратов (принято в печать).
13. *Халафян А.А.* STATISTICA 6. Математическая статистика с элементами теории вероятностей: учебник // Москва: Издательство Бином. 2011. 496 с.
14. *Пелипенко Е.Ю., Халафян А.А.* Оценка платежеспособности предприятия на основе метода «дерева классификации» // *Современные тенденции в экономике и управлении: новый взгляд: сборник материалов III Международной научно-практической конференции: в 2-х частях.* Ч. 2. 2010. С. 34–39.
15. *Breiman L., Friedman J.H., Olshen R.A., Stone C.T.* Classification and Regression Trees // Wadsworth, Belmont, California, 1984. 368 p.

Ключевые слова: деревья классификации, обучающая и тестирующая выборка, правила логического вывода, критерий отсекающего.