

УДК 519.862.6

**КРИТЕРИИ ИДЕНТИФИКАЦИИ ЛОГИКО-ВЕРОЯТНОСТНЫХ  
МОДЕЛЕЙ КРЕДИТНОГО РИСКА ПО СТАТИСТИЧЕСКИМ  
ДАНЫМ****Д.С. Строков, Е.Д. Соложенцев**

Выполнен анализ приложений логико-вероятностных (ЛВ) моделей риска, показана важность процедуры идентификации ЛВ моделей риска по статистическим данным. Приведено краткое математическое описание ЛВ моделей риска. Предложены и исследованы разные критерии идентификации и даны рекомендации по их применению.

**Ключевые слова:** модель, статистика, система, состояния, логика, идентификация, критерий, вероятность, градиенты, Монте-Карло, алгоритм, оптимизация, база знаний.

**Введение**

Для оценки кредитных рисков физических и юридических лиц применяются методики классификации на «хорошие» и «плохие» кредиты на основе линейного (LDA) и квадратичного (QDA) дискриминантного анализа, кластерного анализа (CARD) и нейронных сетей (NN) [1]. Эти методики имеют в два раза меньшую точность классификации, чем логико-вероятностные (ЛВ) модели риска [2]. Однако процесс идентификации (обучения, оптимизации) ЛВ моделей риска по статистическим данным о ранее выданных кредитах банка отличается исключительно высокой, до нескольких часов, вычислительной сложностью. Это связано с целочисленным критерием оптимизации (число корректно классифицируемых кредитов, у которых классификация совпала по модели и по статистике) и большим числом оцениваемых коэффициентов-вероятностей (до 100), которые, к тому же, нужно вычислять до 6–7 знака после запятой. Поэтому выбор и исследование других критериев оптимизации с меньшей вычислительной сложностью, например, дискретно-непрерывных, является актуальной задачей.

**Приложения ЛВ моделей риска**

ЛВ модели широко используются в технике для решения задач надежности и безопасности, в которых инициирующие события и итоговое событие принимают только два значения (0 и 1) [3]. ЛВ модели риска и эффективности имеют также многочисленные приложения в экономике, где инициирующие события имеют много значений-градаций (до 50 и более) и необходимо решать задачи идентификации ЛВ моделей риска (оценки вероятностей инициирующих событий) по статистическим данным.

ЛВ модели риска неуспеха применяются в следующих приложениях [2–5]:

- оценка и анализ кредитного риска физических и юридических лиц,
- анализ риска и эффективности экономических и социальных процессов.

В ряде приложений ЛВ моделей риска задача классификации является основной. Состояния (объекты) системы классифицируются на хорошие и плохие (возможно большее число классов). Задача идентификации формулируется так: максимизировать число корректно распознанных хороших и плохих состояний системы, которые рассматриваются как случайные, имеющие вероятности; используется статистическая база данных (БД), и решается обратная оптимизационная задача. В проблемах эффективно-

сти задача идентификации решается для детального анализа риска и эффективности системы по вкладам процессов, влияющих на итоговый процесс.

Новизна ЛВ подхода для управления риском и эффективностью по статистическим данным мониторинга в экономических и социальных системах и процессах состоит в следующем:

- Представление экономических и социальных систем как структурно-сложных с использованием для их описания Л-переменных и случайных событий. На статистических данных состояний сложной системы рассматриваются два типа событий: появление состояний (объектов, кредитов) и неуспех состояний (объектов, кредитов);
- Введение в статистическую табличную базу данных (БД) конечных множеств (групп несовместных событий) для значений параметров, что позволяет получить систему Л-уравнений или базу знаний (БЗ), использовать ЛВ исчисление Рябинина и формулу Байеса для связи вероятностей и решать задачи риска, эффективности и управления.

Технология ЛВ управления риском в приложениях включает в себя процедуры:

1. формулировка сценария риска и запись Л- и В-функций риска для всех состояний;
2. идентификация ЛВ модели риска системы по статистическим данным;
3. анализ риска состояния и всех состояний по вкладам параметров и градаций параметров, описывающих состояния;
4. управление риском и эффективностью системы.

Оценка и анализ кредитных рисков является ярким примером задач классификации. Поэтому реальные исследования проблемы выбора критериев идентификации ЛВ моделей риска по статистическим данным проведены на примере кредитных рисков физических и юридических лиц.

Преимущества ЛВ модели риска на примере кредитных рисков подробно изложены в [2, 6]. ЛВ модели кредитного риска разительно отличаются от распространенных скоринговых методик. Идентификация ЛВ модели риска по статистическим данным позволяет решать следующие задачи:

- строить В-модель кредитного риска, определяя вероятности событий-градаций;
- выполнять анализ точности ЛВ модели риска;
- определять вклады событий-параметров, описывающих кредит, и их градаций в точность, робастность и прозрачность оценки кредитного риска;
- управлять кредитным риском банка, изменяя число параметров в описании кредита и градаций в параметре, асимметрию распознавания хороших и плохих кредитов.

Наряду с большими достоинствами ЛВ моделей риска, их идентификация по статистическим данным отличается большой вычислительной сложностью из-за большого числа оцениваемых вероятностей (для кредитного риска их число доходит до 100), наличия локальных экстремумов из-за ступенчатости целевой функции, учета связей вероятностей в группах несовместных событий (ГНС) и вычисления логических разностей. Поэтому необходимы тщательные исследования методик идентификации ЛВ моделей по статистическим данным, что приведет к их более широкому распространению.

### **Математическое описание ЛВ моделей риска**

Общим для ЛВ моделей является одинаковое табличное представление статистических данных. Табличная БД содержит информацию об однородных объектах (кредитах) или состояниях системы в разные моменты времени (портфель ценных бумаг). В таблице количество столбцов может достигать нескольких десятков, а количество строк – нескольких сотен. В ячейках таблицы находятся значения параметров (качественные или количественные, целые или дробные), характеризующие объекты или состояния



Системы (1) и (2) будем называть *базой знаний* (БЗ) и использовать для получения новых знаний. В ЛВ теории риска с ГНС события-параметры связаны Л-операциями *AND*, *OR*, *NOT*, и могут иметься циклы. Событиям-параметрам соответствуют Л-переменные, которые могут быть зависимыми, но не изначально, а только потому, что они содержатся в Л-формуле, которая и определяет зависимость между ними. События-градации для каждого параметра являются зависимыми и образуют ГНС.

### Идентификация ЛВ модели риска неуспеха

Задача идентификации решается алгоритмическими итеративными методами и подробно описана в [2]. Предложена следующая схема решения задачи. Пусть известны в первом приближении оценки вероятностей для градаций  $P_{jr}$ ,  $r = 1, 2, \dots, N_j$ ;  $j = 1, 2, \dots, n$ , и вычислены риски  $P_i$ ,  $i=1, 2, \dots, N$  кредитов статистических данных. Определим допустимый риск  $P_{ad}$  так, чтобы принятое нами расчетное число хороших кредитов  $N_{gg}$  имело риск меньше допустимого и соответственно число плохих кредитов  $N_{bb}=N-N_{gg}$  имело риск больше допустимого. В индексах «*bb*» и «*gg*» первая буква означает классификацию по статистике, а вторая – по модели. На шаге оптимизации нужно так изменить вероятности  $P_{jr}$ ,  $r = 1, 2, \dots, N_j$ ;  $j = 1, 2, \dots, n$ , чтобы число распознаваемых кредитов увеличилось.

Разработаны следующие программные средства (ПС) для идентификации ЛВ модели риска и анализа риска:

1. демоверсия для оценки и анализа кредитных рисков;
2. ПС для дистанционного обслуживания кредитной деятельности банков;
3. ПС для оценки кредитов и управления кредитной деятельностью банка.

ПС работают в среде MS Windows и используют статистику по объектам в виде файла, который загружается в программу.

### Критерии идентификации ЛВ модели риска

Для идентификации можно использовать следующие целевые функции:

1. Число корректно классифицируемых состояний

$$F = N_{bb} + N_{gg} \rightarrow \max_{P_{jr}}, \quad (4)$$

где  $N_{gg}$ ,  $N_{bb}$  – соответственно числа состояний, классифицируемых как хорошие и плохие и статистикой, и В-моделью (корректные оценки);

2. Энтропия вероятностей корректно классифицируемых состояний

$$H = -\sum_{k=1}^{N_k} P_k \cdot \ln P_k \rightarrow \max_{P_{jr}}, \quad (5)$$

где  $P_k$  – вероятности корректно классифицируемых состояний;

3. Сумма вероятностей корректно классифицируемых состояний

$$S = \sum_{k=1}^{N_k} P_k \rightarrow \max_{P_{jr}}. \quad (6)$$

Критерий  $F$  является целочисленным, изменяется дискретно и равен числу корректно распознанных хороших и плохих состояний (4). Критерии  $H$  и  $S$  являются непрерывно-дискретными, так как их значения зависят от числа корректно распознанных состояний и от их вероятностей (риска). Дискретно-непрерывные критерии оптимизации  $H$  и  $S$  позволяют применить большой арсенал методов нелинейной оптимизации [7].

Итеративная алгоритмическая идентификация ЛВ модели риска выполняется по формуле

$$\Delta P1_{jr} = K_1 \frac{N_{opt} - N_v}{N_{opt}} K_3 P1_{jr}, j = 1, 2, \dots, n; r = 1, 2, \dots, N_j, \quad (7)$$

где  $K_1$  – коэффициент, равный  $\sim 0,05$ ;  $N_{opt}$ ,  $N_v$  – число оптимизаций и номер текущей оптимизации,  $K_3$  – случайное число в интервале  $[-1,+1]$ . В процессе итеративной алгоритмической оптимизации  $\Delta P1_{jr}$  стремится к нулю. Формула (7) обеспечивает простое задание максимального приращения вероятностей и определение точности оценки вероятностей по величине приращений на шаге последней оптимизации.

Различные критерии оптимизации сведены в табл. 1. Критерии по некорректно распознанным состояниям  $gb$  и  $bg$  являются дополнениями критериев корректно распознанных состояний  $gg$  и  $bb$ .

Критерии по порядку	Энтропия состояний (H-критерии)	Число состояний (F-критерии)	Сумма вероятностей (S-критерии)	Примечание
1	$H_{gg}$	$F_{gg}$	$S_{gg}$	Хороших корректно распознанных
2	$H_{bb}$	$F_{bb}$	$S_{bb}$	Плохих корректно распознанных
3	$H=H_{gg}+H_{bb}$	$F=F_{gg}+F_{bb}$	$S=S_{gg}+S_{bb}$	Хороших и плохих корректно распознанных

Таблица 1. Критерии оптимизации

### Исследование критериев идентификации

Исследования выполнялись с использованием статистических данных о 1000 кредитах, приведенных в работе [1]. Проводилась оценка 96 вероятностей событий-градаций. Исследовалось влияние шага отступлений на конечное значение целевой функции. Результаты исследований при оптимизации по  $H$  и  $S$  приведены в табл. 2 и 3 соответственно.

№	$\Delta H$	$H$	$F$	$S$
1	0,075	224,73	840	190,89
2	0,09	224,43	842	190,95
3	0,1	224,78	842	191,33
4	0,105	225,65	845	190,95
5	0,11	225,65	845	190,95
6	0,12	225,21	844	190,34
7	1,376	221,89	826	190,75

Таблица 2. Зависимость критериев от величины «отступления» при оптимизации по  $H$

№	$\Delta S$	$H$	$F$	$S$
1	0,06	221,91	828	197,0
2	0,1	222,44	828	195,4
3	0,12	223,98	830	197,13
4	0,18	222,25	826	195,52
5	0,972	221,74	826	193,52

Таблица 3. Зависимость критериев от величины «отступления» при оптимизации по  $S$

Оптимальные величины «отступлений» для критериев оптимизации составили  $\Delta F=4$  и  $\Delta H = \Delta S = 0,1125$ . Результаты исследований для разных критериев оптимизации приведены в табл. 4.

Различие процессов оптимизации по разным критериям характеризуют величины «отступлений» в случае неуспеха попыток оптимизаций на шаге. Отступление для кри-

терия  $F$  равно  $\Delta F=4$  и составляет примерно  $4/800=1/200$  от оптимального значения целевой функции. «Отступлению»  $\Delta F=4$  соответствуют «отступления»  $\Delta H=1,376$  (строка 7 в табл. 2) и  $\Delta S=0,972$  (строка 5 в табл. 3). Оптимальные «отступления» для критериев  $H$  и  $S$  равны  $\Delta H = \Delta S \approx 0,1125$  и составляют примерно  $0,1125/225=1/2000$  от оптимального значения целевых функций, т.е. для непрерывно-дискретных критериев  $H$  и  $S$  высота ступенек при оптимизации уменьшается в десять раз.

Оптимизация по критериям	Значения		
	$F$	$H$	$S$
$F$	844	223,35	182,84
$H$	842	225,21	190,34
$S$	830	223,98	197,13

Таблица 4. Результаты оптимизации по разным критериям

Наибольшее значение числа корректно распознанных кредитов или, что то же самое, наибольшая точность ЛВ модели риска достигается при оптимизации с использованием прямого целочисленного критерия  $F$  (табл. 4). Косвенные непрерывно-дискретные критерии  $H$  и  $S$ , хотя сами принимают наибольшие значения при оптимизации по ним, распознают меньшее число кредитов, их следует использовать для определения начальных значений вероятностей  $P_{1j}$  и  $P_{jr}$  при оптимизации по критерию  $F$ , а также контроля оптимизации по этому критерию. Эти критерии дают неоптимальные и смещенные оценки для числа корректно распознанных кредитов  $F$ . Косвенный критерий оптимизации  $H$  предпочтительнее косвенного критерия  $S$ , так как дает более высокое значение прямого критерия  $F$  и меньшее смещение его от оптимального значения.

### Выводы

1. Анализ приложений ЛВ моделей риска показал, что в технологии ЛВ управления риском и эффективностью систем важной и самой сложной в вычислительном отношении является процедура идентификация ЛВ модели риска по статистическим данным.
2. Для идентификации ЛВ модели риска по статистическим данным методами Монте-Карло и градиентов предложены формулы одинаковой структуры, отличающиеся простотой и прозрачностью и обеспечивающие приемлемое время вычислений, сходимость процесса обучения, простое задание начальных значений.
3. Метод Монте-Карло и метод градиентов дают одинаковые результаты при оптимизации и позволяют взаимно контролировать результаты идентификации.
4. Идентификация методом градиентов требует меньшего времени вычислений, и ей следует отдать предпочтение для определения достаточно хороших начальных значений вероятностей. Окончательную оптимизацию следует выполнять методом Монте-Карло.
5. Оптимизация по дискретно-непрерывным критериям  $H$  и  $S$  имеет меньшую вычислительную сложность и позволяет использовать весь арсенал нелинейных методов оптимизации. Но число корректно распознаваемых состояний меньше, поэтому эти критерии следует использовать для оценки начальных приближений вероятностей.
6. Изложенные результаты и методика нашли применение при разработке ПС для разных типов и классов ЛВ моделей риска (кредитные риски, риск портфеля ценных бумаг, риск коррупции и взятки и др.).

## Литература

1. Seitz J., Stickel E. Consumer Loan Analysis Using Neural Network // Proc. of the Bankai Workshop: Adaptive Intelligent Systems. – Brussels, 14–19 October 1996.
2. Solojentsev E.D. Scenario Logic and Probabilistic Management of Risk in Business and Engineering. – Second edition. – Springer, 2008. – 500 p.
3. Рябинин И.А. Надежность и безопасность структурно-сложных систем. – 2-е изд. – СПб: Изд-во СПбГУ, 2007. – 276 с.
4. Соложенцев Е.Д., Степанова Н. В., Карасев В.В. Прозрачность методик оценки кредитных рисков и рейтингов. – СПб: Изд-во СПбГУ, 2005. – 200 с.
5. Исследование рисков. Методические указания к проведению лабораторных работ «Логико-вероятностная теория кредитных рисков» / Н.С. Медведева, Е.Д. Соложенцев, Д.С. Строков. – СПб: СПбГУАП, 2007. – 23 с.
6. Соложенцев Е.Д. Управление риском и эффективностью в экономике: Логико-вероятностный подход. – СПб: Изд-во СПбГУ, 2009. – 259 с.
7. Аттетков А.В., Галкин С.В., Зарубин В.С. Методы оптимизации: Учеб. для вузов / Под ред. В.С. Зарубина, А.П. Крищенко. – 2-е изд., стереотип. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2003. – 440 с.

*Строков Дмитрий Сергеевич* – Санкт-Петербургский государственный университет аэрокосмического приборостроения, аспирант, [dima.src@gmail.com](mailto:dima.src@gmail.com)  
*Соложенцев Евгений Дмитриевич* – Институт проблем машиноведения РАН, доктор технических наук, профессор, [esokar@gmail.com](mailto:esokar@gmail.com)