АЛГОРИТМЫ КРЕДИТНОГО СКОРИНГА НА ОСНОВЕ ЭКОНОМЕТРИЧЕСКИХ МОДЕЛЕЙ

В. С. Дежемесова, Б. А. Барановский

На практике управление кредитным риском коммерческим банком предполагает решение задач анализа и классификации заемщиков по степени кредитоспособности. Важным инструментом анализа и принятия решений является *кредитный скоринг* (*credit scoring*), осуществляемый с помощью систем автоматической классификации потенциальных заемщиков коммерческого банка на основе доступной информации.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ И ЗАДАЧИ ИССЛЕДОВАНИЯ

Будем предполагать, что при оценке (прогнозировании) состояния заемщика на момент окончания срока кредитного договора наряду с вектором контролируемых признаков $x=(x_1,...x_N)'\in\Re^N$ (балансовых показателей и показателей качества обслуживания долга) используется вектор экзогенных переменных (факторов) $z=(z_1,...,z_M)'\in Z\subset\Re^M$, описывающих влияние на состояние заемщиков со стороны общих внешних экономических факторов.

Предположим, что существующая между векторами $x\in\Re^N$ и $z\in Z$ статистическая зависимость для t-го заемщика описывается эконометрической моделью многомерной линейной регрессии:

$$x_t = B_l z_t + u_t, \quad l\in S(L)=\{1,...,L\},\ t=1,...,T, \tag{1}$$

где для класса заемщиков Ω_l : $B_l=(b_{lje})$ – $(N\times M)$ -матрица коэффициентов регрессии, $Z\subset\Re^M$ – ограниченная замкнутая область, такая, что $(B_i-B_j)z\neq 0, (i\neq j)$ при $z\in Z$; $u_t\in\Re^N$ – гауссовский случайный вектор с нулевым средним значением и невырожденной ковариационной $(N\times N)$ -матрицей $\Sigma_l=(\sigma_{ljk})$.

Пусть истинные значения параметров $\{\pi_l\},\{B_l\},\{\Sigma_l\}(l\in S(L),L\geq 2)$ модели (1) не известны. Имеется неклассифицированная обучающая выборка значений признаков $X=\{x_1,...,x_T\}$ объема T из классов $\{\Omega_l\}(l\in S(L))$, соответствующая последовательности значений факторов $Z=\{z_1,...,z_T\}$, где значение вектора признаков $x_t\in\Re^N$ определяется на основании (1) для заданного номера класса l и значения вектора факторов $z_t\in Z$ $(t=1,...,T)$.

В рамках рассматриваемой модели наблюдений выборку регрессионных наблюдений $X = \{x_1,...,x_T\}$, соответствующую последовательности значений факторов $Z = \{z_1,...,z_T\}$, можно рассматривать как случайную выборку из смеси распределений регрессионных наблюдений, плотность распределения вероятностей для которой имеет вид:

$$p_\pi(x,z,\theta) = \sum_{l=1}^{L} \pi_i n_N(x \mid B_l z, \Sigma_l) = \sum_{l=1}^{L} \pi_l p(x,z,\theta_l), \quad x \in \Re^N, \quad z \in Z, \tag{2}$$

где $\theta \in \Theta \subset \Re^m$ – составной вектор всех параметров смеси, образованный из независимых элементов матриц $\{B_l\}$, $\{\Sigma_l\}$ и априорных вероятностей $\{\pi_l\}$ $(l \in S(L))$; $p(x,z,\theta_l) \equiv n_N(x \mid B_l z, \Sigma_l)$ $(x \in \Re^N)$ – условная плотность нормального распределения для заданного значения $z \in Z$; θ_l – составной вектор параметров модели для класса Ω_l.

Имеют место следующие задачи кластерного анализа смеси (2) по неклассифицированной обучающей выборке $\{X, Z\}$:

1) статистическое оценивание вектора параметров $\theta \in \Theta \subset \Re^m$ (вычисление оценок максимального правдоподобия $\{\pi_l\},\{B_l\},\{\Sigma_l\}(l \in S(L))$;

2) классификации, обучающей выборки $\{X, Z\}$, т.е. оценивание вектора классификации выборки $d = (d_t) \in S^T(L)$.

3) классификация вновь поступающих наблюдений (x_τ, z_τ) $(\tau = T+1,...,T+n,\ n \ge 1)$.

АЛГОРИТМ КРДИТНОГО СКОРИНГА

Для решения задач 1, 2 предлагается итерационный алгоритм расщепления смеси распределений регрессионных наблюдений (2) из класса EM-алгоритмов, позволяющий одновременно вычислять оценки максимального правдоподобия параметров $\{\hat{\pi}_l\},\{\hat{B}_l\},\{\hat{\Sigma}_l\}$, а также апостериорные вероятности классов. Решающее правило классификации на основе данного алгоритма эквивалентно решающему правилу по максимуму апостериорной вероятности классов вида:

$$d_t^{(k)} = \operatorname{argmax}\left\{p^{(k)}(l \mid x_t, z_t, \theta)\right\} \text{ по } l \in S, \tag{3}$$

где $\left\{p^{(k)}(l \mid x_t, z_t, \theta)\right\}$ $\left(l \in S(L)\right)$ – апостериорная вероятность отнесения регрессионного наблюдения (x_t, z_t) к классу Ω_l на итерации k $(1 \le k \le K)$, определяемая по формуле:

$$p^{(k)}(l \mid x_t, z_t, \theta^{(k-1)}) = \frac{\pi_l^{(k-1)} p_l(x_t, z_l \mid \theta_l^{(k-1)})}{p_\pi(x_t, z_t, \theta^{(k-1)})},$$

где $\theta^{(0)}$, $\{\theta_l^{(0)}\}$ – заданные начальные значения векторов параметров.

При использовании в (3) известных значений параметров получаем решающее правило, эквивалентное байесовскому решающему правилу (БРП).

Получим представления для оценок максимального правдоподобия параметров смеси распределений регрессионных наблюдений (2), зависящие от апостериорных вероятностей. $\left\{p^{(k)}(l \mid x_t, z_t, \theta)\right\}$ $\left(l \in S(L)\right)$ отнесения регрессионного наблюдения (x_t, z_t) к классу Ω_l .

Теорема. В условиях описываемой выше модели наблюдения оценки максимального правдоподобия (МП-оценки) $\{\hat{\pi}_l\}, \{\hat{B}_l\}, \{\hat{\Sigma}_l\}$ $l \in S(L)$, параметров смеси (2) по неклассифицированной выборке регрессионных наблюдений $X = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$, $Z = \{\mathbf{z}_1, ..., \mathbf{z}_N\}$ из классов $\{\Omega_l\}$, допускают представления:

$$\hat{B}_l = \sum_{t=1}^{T} x_t z_t' p(l \mid x_t, z_t, \theta) \left(\sum_{i=1}^{T} z_i z_i' p(l \mid x_t, z_t, \theta)) \right)^{-1},$$

$$\hat{\Sigma}_l = \frac{\sum_{t=1}^{T} p(l \mid x_t, z_t, \theta)(x_t - \hat{B}_l z_t)(x_t - \hat{B}_l z_t)'}{\sum_{i=1}^{T} p(l \mid x_t, z_t, \theta)}, \; \hat{\pi}_l = \frac{1}{T} \sum_{t=1}^{T} p(l \mid x_t, z_t, \theta).$$

ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ АЛГОРИТМОВ

В таблице 1 и таблице 2 для алгоритма кластерного анализа приводятся оценки следующих характеристик: P^L и P^E – безусловная вероятность ошибки при классификации обучающей (learning) и экзаменационной (examine) выборки; $P^{БРП}$ – риск байесовского решающего правила в условиях полной априорной информации.

Тестовый пример 1. В данном эксперименте для моделирования данных использовались значения параметров: $L = 2$, $N = 2$, $M = 3$,

$$n_1 = n_2 = 1000, \quad \Sigma = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}, \quad B_0 = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 0 & 3 \end{pmatrix}.$$

Полагалось, что $B_1 = B_2 + H$, где H – фиксированная матрица, определяющая расстояние Махаланобиса $\Delta(z)$ между классами, характеризующее степень разделимости классов

$$\Delta^2(z_0) = z_0{}^T H^T \Sigma^{-1} H z_0 \text{ , } z_0 = (0.5;0.5;0.5).$$

Результаты экспериментов представлены в таблице 1. Как и следовало ожидать, при уменьшении межклассового расстояния вероятности ошибочной классификации увеличиваются, при этом точность классификации обучающей и новой экзаменационной выборки практически идентична и заметно уступают при большом межклассовом расстоянии точности БРП.

Таблица 1

Результаты анализа точности классификации

Δ	P^L	P^E	$P^{БРП}$
6	6,15%	6,8%	0,1%
4	11,75%	12,45%	2,2%
2,5	14,5%	14,15%	10,6%
1	47,65%	47,95%	30,85%
0	42,8%	42,6%	50%

Тестовый пример 2. Целью данного эксперимента было выяснения влияние точности задания начальной классификации обучающей выборки на итоговую классификацию. В качестве начальных значений вектора классификации $d = (d_t) \in S(L)^T$ использовались векторы, полученные из истинного вектора классификации внесением 5%, 10%, 30% ошибок. Актуальность данной проблемы отмечалась в [3].

На основе приведенных в таблице 2 результатов можно сделать вывод, что увеличение доли ошибок во входных данных приводит к существенному увеличению безусловной вероятности ошибки.

Таблица 2

Влияние входного вектора классификации на итоговую классификацию

Δ	Доля ошибок 5%		Доля ошибок 10%		Доля ошибок 30%		$P^{БРП}$
	P^L	P^E	P^L	P^E	P^L	P^E	
6	7,25%	7,9%	7,5%	7,75%	8,9%	8,6%	0,1%
4	13,05%	13,35%	13,05%	14,3%	13,2%	13,45%	2,2%
2,5	16,1%	16,7%	16,5%	16,95%	17,55%	18%	10,6%
1	48,1%	48%	47,65%	47,95%	48,2%	48,15%	30,85%
0	43,75%	43,15%	46,8%	46,6%	47,8%	48,55%	50%

Литература

1. *Гринь Н.В., Малюгин В.И.* Исследование точности методов классификации многомерных данных в задачах кредитного скоринга / Н.В. Гринь, В.И. Малюгин // Вестник ГрГУ. Сер.2. 2008. №1. С. 77–85.

2. *Thomas, L.C.* A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers / L. C. Thomas // International Journal of Forecasting. 2000. V. 6. P. 49–172.
3. *Малюгин В.И.* Исследование эффективности алгоритмов классификации заемщиков банков на основе балансовых коэффициентов / В.И. Малюгин, О.И. Корчагин, Н.В. Гринь // Банковский Вестник № 4 2009. С. 27–32.

РАЗРАБОТКА ТЕСТОВ КОНТРОЛЯ ЦИФРОВЫХ СИСТЕМ НА ВЕРХНИХ УРОВНЯХ ПРОЕКТИРОВАНИЯ

А. В. Ильинкова

Анализируется состояние проблемы контроля сверхбольших интегральных схем (СБИС). Рассматриваются задачи направленного построения тестов контроля по поведенческому описанию объекта на уровне межрегистровых передач (RTL) на языке VHDL. Класс функциональных неисправностей, рассматриваемых при направленном построении теста, соответствует неисправностям константного типа реализаций СБИС на элементах соответствующих библиотек проектирования. Предлагается метод направленного построения теста, который позволяет на ранних этапах проектирования анализировать контролепригодность проекта и ее зависимость от применяемых технологических библиотек проектирования.

Требования к надежности цифровых систем постоянно возрастают не только в областях, в которых отказ может привести к катастрофическим событиям, но и во всех других применениях. Одной из самых важных компонентов систем обеспечения высокой надежности электронных систем является способность определить наличие или отсутствие ошибки функционирования системы. Проблема построения тестов на всем интервале развития интегральной схемотехники является одной из наукоемких проблем, которые до настоящего времени не получили эффективного теоретического и практического решения. Задача построения тестов принадлежит к классу NP- трудных проблем. В связи с высокой сложностью построения тестов контроля и диагностирования функционально-сложных цифровых систем применяются методы контролепригодного проектирования, которые направлены на снижение сложности задачи. В то же время проблема построения тестов остается высоко актуальной, так как все используемые подходы к проектированию не решают по разным причинам задачу контроля цифровых систем.

Тесты необходимы на всех этапах жизненного цикла цифровой системы. Актуальность тестов обусловлена потребностью в тестах на этапе проектирования для анализа корректности проектов на всех этапах про-