9. ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И ИНФОРМАТИКА

9.1. НЕЧЕТКО-МНОЖЕСТВЕННАЯ КЛАССИФИКАЦИЯ ОБЪЕКТОВ НА ОСНОВЕ ПРИБЛИЖЕНИЯ МНОГОМЕРНОГО ИНТЕГРАЛА ВЕРОЯТНОСТЕЙ

Лабунец Л.В., д.т.н., профессор, кафедра систем обработки информации и управления, Московский государственный технический университет им. Н.Э. Баумана, г. Москва, профессор, кафедра информационных систем в экономике и управлении, Российский новый университет, г. Москва;
Березовская Е.Л., старший администратор систем ДБО, АКБ «РосЕвроБанк», г. Москва;
Лебедева Н.Л., главный специалист, Департамент информационных технологий, ОАО «Банк ВТБ», г. Москва

Перейти на ГЛАВНОЕ МЕНЮ

Предложена нелинейная модель приближенных рассуждений в задаче нечеткой классификации объектов. Модель основана на процедуре лингвистического анализа ковариационного приближения для многомерного распределения вероятностей информативных признаков. Рассмотрены основные этапы нечеткого логического вывода, предусматривающего формирование байесовского классификатора с помощью EM-подобного алгоритма. Инвариантность методики к содержанию предметной области иллюстрируют примеры скоринга объектов наружной рекламы и биржевых активов.

ВВЕДЕНИЕ

Современный этап развития инновационных технологий поддержки управленческих решений характеризуется интенсивным применением методов и алгоритмов Data Mining. Методология этого раздела теории искусственного интеллекта позволяет совершенствовать существующие структуры и механизмы управления сложными социально-экономическими системами. Примерами подобного рода задач являются планирование бюджетов крупных мегаполисов, а также оценка инвестиционной привлекательности биржевых активов.

Представленный в работе [2] эконометрический анализ и моделирование нестационарных временных рядов помесячных поступлений в московский бюджет от объектов наружной рекламы (ОНР) позволяют сформировать различные сценарии прогноза. Учет неопределенности в моделях, подобного рода, опирается на альтернативный экспертный подход к исследованию исходной информации. Рациональным в этом смысле является применение систем обработки знаний экспертов, основанных на алгоритмах нечеткого логического вывода [1, 7].

Предложенные в работах [4, 5] методики агрегирования нечеткой информации об объекте анализа основаны на популярной процедуре двукратной свертки [7] исходных одномерных признаков по их лингвистическим классам и формировании комплексного показателя одним из методов экспертного оценивания [8, 9]. Такого рода алгоритм приближенных рассуждений реализует нейро- нечеткая сеть Такаги – Сугено – Канга и, по сути, воспроизводит линейный метод главных компонент.

Учет нелинейных связей возможен, на наш взгляд, на основе лингвистического анализа многомерного распределения информативных признаков. Данная работа посвящена рассмотрению такого подход на примерах нечеткой классификации ОНР по критерию поступлений в городской бюджет, а также инвестиционной привлекательности биржевых активов с помощью байесовской многомерной модели лингвистической переменной «уровень показателя».

1. Модель многомерных данных

Многомерная модель исходных данных представляет собой информационный куб. В задаче анализа поступлений в городской бюджет от ОНР данные имеют следующие измерения [2, 4]:

- время в виде последовательности месяцев $n = 1, 2, \ldots, N$ $(N = 48)$ в период с 2008 по 2011 гг. включительно;
- округ в виде упорядоченной последовательности административных округов г. Москвы $m = 1, 2, \ldots, M$ $(M = 11)$ по состоянию на 2011 г.;
- тип ОНР в виде упорядоченной последовательности типов наружной рекламы $k = 1, 2, \ldots, K$ $(K = 30)$.

Кластерную структуру данных исследуют в пространстве административных округов г. Москвы. Исходными признаками в этом случае являются компоненты вектора столбца

$$\vec{x} = (x_1, \ldots, x_M)^T, \qquad (1)$$

в виде поступлений в московский бюджет от ОНР, консолидированных по месяцам в периоды 2008-го $(1 \le n \le 12)$, 2009-го $(13 \le n \le 24)$, 2010-го $(25 \le n \le 36)$ и 2011-го $(37 \le n \le 48)$ гг. Здесь и далее зависимость векторов от индекса времени n опускаем для упрощения записи.

Для каждого года выборка обучающих примеров представляет собой блочную матрицу $X = (\vec{X}_1 \vdots \cdots \vdots \vec{X}_K)$ размером $M \times K$. Текущий столбец $\vec{X}_k = (X_{1k}, \ldots, X_{Mk})^T$ матрицы – это поступления в бюджет от k-го типа рекламы в административные округа за фиксированный год.

В задаче скоринга биржевых активов данные имеют следующие измерения [3, 5, 6]:

- время в виде последовательности кварталов $n = 1, 2, \ldots, N$ $(N = 20)$ в период с 2007 по 2011 гг. включительно;
- мультипликатор в виде упорядоченной последовательности фундаментальных финансовых показателей деятельности компаний $m = 1, 2, \ldots, M$ $(M = 4)$;
- ценная бумага в виде упорядоченной последовательности акций компаний $k = 1, 2, \ldots, K$ $(K = 108)$.

Вектор признаков (1) в этом случае представляет собой набор из M фундаментальных финансовых показателей эмитента. Компонентами текущего столбца $\vec{X}_k = \left(X_{1k}, \ldots, X_{Mk} \right)^T$ блочной матрицы $X = \left(\vec{X}_1 \vdots \cdots \vdots \vec{X}_K \right)$ являются значения мультипликаторов для k-го эмитента по состоянию на 4-й квартал фиксированного года.

Разведочный анализ данных выполняют после предварительного преобразования Бокса – Кокса (БКП) обучающей выборки для каждого показателя. Это преобразование сжимает динамический диапазон признаков (1) и в определенной мере приближает их к нормальному распределению. Для m-го показателя БКП имеет вид

$$Y_{mk}\left(g_m \right) = \begin{cases} \dfrac{1}{g_m}\left\{ \left(X_{mk} + \varepsilon \right)^{g_m} - 1 \right\}, & g_m \neq 0 \\ \ln\left(X_{mk} + \varepsilon \right), & g_m = 0 \end{cases},$$

$$k = 1,2,\ldots,K$$

где величину смещения ε выбирают из условий $X_{mk} + \varepsilon > 0$, $m = 1,2,\ldots,M$, $k = 1,2,\ldots,K$. Оптимальное значение параметра g_m БКП выбирают по критерию максимума логарифма правдоподобия. Результаты расчета показывают [4], что для ОНР оптимальные оценки параметра БКП лежат в интервале $-0{,}15 \le g_m \le 0{,}1$.

Удобным инструментом разведочного анализа данных является диаграмма рассеяния, содержащая K точек в M-мерном пространстве преобразованных признаков $\vec{Y}_k = \left(Y_{1k}, \ldots, Y_{Mk} \right)^T$, $k = 1,2,\ldots,K$. Анализ формы диаграммы рассеяния методом «Grand Tur» [13] динамической визуализации многомерных данных свидетельствует о наличии значимой корреляции признаков $\vec{y} = \left(y_1, \ldots, y_M \right)^T$ [3, 4].

2. Исключение аномальных значений из данных

Аномальные значения в данных существенно искажают результаты оценивания основных статистик. Одним из методов исключения влияния загрязнений выборки являются экспоненциально взвешенные оценки (ЭВО) характеристик положения и масштаба Л. Д. Мешалкина [10]. Робастные ЭВО вектора математического ожидания $\vec{A}(\lambda)$ и ковариационной матрицы $B(\lambda)$ выборки векторов

$$\vec{Y}_k = \left\{ Y_{1k}, \cdots, Y_{Mk} \right\}^T, \; k = 1,2,\ldots,K$$

получают в результате решения системы нелинейных уравнений [12, с. 75]

$$\begin{cases} \vec{A}(\lambda) = \sum\limits_{k=1}^{K} q^{\lambda}\left\{ D_k(\lambda) \right\} \vec{Y}_k \Big/ \\ \Big/ \sum\limits_{k=1}^{K} q^{\lambda}\left\{ D_k(\lambda) \right\} \\ B(\lambda) = (1+\lambda) \sum\limits_{k=1}^{K} q^{\lambda}\left\{ D_k(\lambda) \right\} \left\{ \vec{Y}_k - \vec{A}(\lambda) \right\} \\ \left\{ \vec{Y}_k - \vec{A}(\lambda) \right\}^T \Big/ \\ \qquad \Big/ \sum\limits_{k=1}^{K} q^{\lambda}\left\{ D_k(\lambda) \right\} \end{cases}$$

например, методом последовательных приближений. Здесь

$$D_k(\lambda) = \left\{ \vec{Y}_k - \vec{A}(\lambda) \right\}^T B^{-1}(\lambda) \left\{ \vec{Y}_k - \vec{A}(\lambda) \right\},$$

$$k = 1,2,\ldots,K$$

M-мерная метрика Махаланобиса в пространстве преобразованных признаков, $\lambda > 0$ – параметр устойчивости статистик к загрязнениям.

ЭВО удобны для формирования границы, отделяющей кластер типичных значений $\vec{Y}_k$, $k = 1,2,\ldots,K$ от выбросов. Результаты моделирования показали [3-5], что дискриминантную границу рационально формировать по правилу «двух сигм» [12, с. 82]:

$$\left| \frac{1}{KM} \sum_{k=1}^{K} D_k(\lambda) - 1 \right| \le 2\sigma_{\lambda},$$

$$\sigma_{\lambda} = \frac{1}{\sqrt{KM}} \sqrt{\frac{\left(2 + 4\lambda + 3\lambda^2 \right)(1+\lambda)^{\frac{M}{2}}}{(1+2\lambda)^{\frac{M}{2}+2}} - 2},$$

ограниченной поверхностью гиперэллипсоида в пространстве признаков. В соответствии с указанным правилом выполнялось масштабирование данных:

$$y_{mk} = \left\{ Y_{mk} - Y_m^{(min)} \right\} \Big/ \left\{ Y_m^{(max)} - Y_m^{(min)} \right\},$$

$$k = 1,2,\ldots,\tilde{K},$$

очищенных от загрязнений, с целью преобразования их к стандартному интервалу от нуля до единицы. Здесь $Y_m^{(min)}$ и $Y_m^{(max)}$ – робастные оценки нижней и верхней границ m-го признака для кластера типичных значений; $\tilde{K}$ – количество наблюдений в кластере.

3. Нечеткая универсальная шкала

На первом этапе синтеза правил приближенных рассуждений формируют экспертную модель в виде универсальной шкалы субъективных вероятностей p для лингвистических уровней показателей. С этой целью в результате опроса экспертов выбирают функции принадлежности $W_l(p)$, $l = 1,2,\ldots,L$ нечеткого классификатора объектов, например, по методике, представленной в разделе 2.4.3 работы [9]. В наших вычислительных экспериментах для уровней поступлений от ОНР применялась стандартная трехуровневая (L = 3) трапезоидная модель нечеткого разбиения

$$W_1(p)=\begin{cases}1 & ,0\le p\le \Phi_1\\ (\Phi_2-p)/\Phi_0 & ,\Phi_1<p\le\Phi_2;\\ 0 & ,\Phi_2<p\le 1\end{cases}$$

$$W_2(p)=\begin{cases}0 & ,0\le p\le \Phi_1\\ (p-\Phi_1)/\Phi_0 & ,\Phi_1<p\le\Phi_2\\ 1 & ,\Phi_2<p\le\Phi_3;\\ (\Phi_4-p)/\Phi_0 & ,\Phi_3<p\le\Phi_4\\ 0 & ,\Phi_4<p\le 1\end{cases}$$

$$W_3(p)=\begin{cases}0 & ,0\le p\le \Phi_3\\ (p-\Phi_3)/\Phi_0 & ,\Phi_3<p\le\Phi_4\\ 1 & ,\Phi_4<p\le 1\end{cases}$$

с параметрами $\Phi_0=0{,}2$, $\Phi_1=0{,}25$, $\Phi_2=0{,}45$, $\Phi_3=0{,}55$, $\Phi_4=0{,}75$ (рис. 1).

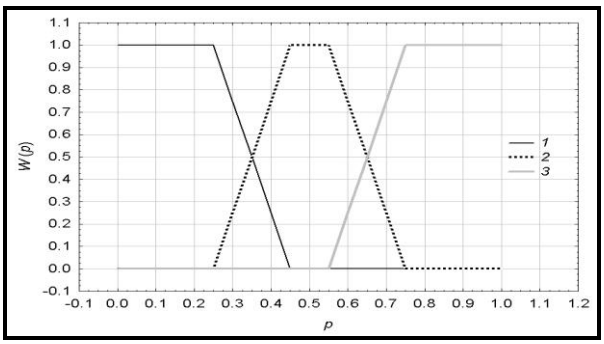


Рис. 1. Лингвистическая шкала субъективных вероятностей уровней поступлений от ОНР: 1 – низкий; 2 – средний; 3 – высокий

Границы ядер и носителей нечетких подмножеств определяются диапазонами значений вероятностей низких, средних и высоких уровней показателей. Соответствующие интервалы представлены в табл. 1.

Таблица 1

ПРАВИЛА ПРИБЛИЖЕННЫХ РАССУЖДЕНИЙ

Уровень показателя	Низкий	Средний	Высокий
Достоверно (ядро)	$p\le\Phi_1$	$\Phi_2<p\le\Phi_3$	$p>\Phi_4$
Возможно (носитель)	$p\le\Phi_2$	$\Phi_1<p\le\Phi_4$	$p>\Phi_3$

4. Ковариационное приближение многомерного распределения

В рамках интервального подхода к формированию правил принятия решений гиперповерхности для дискриминантных границ рассмотренных выше нечетких подмножеств рационально рассчитывать с помощью квантилей интегральной функции распределения $\Phi(\vec{y})$ в многомерном пространстве нормированных показателей $\vec{y}=(y_1,\ldots,y_M)^T$. Удобным описанием в этом случае является ковариационное приближение интегрального распределения [3]:

$$\Phi(\vec{y})=\sum_{n_{1,2}}\cdots\sum_{n_{(M-1),M}}\frac{b_{1,2}^{n_{1,2}}\cdot\ldots\cdot b_{(M-1),M}^{n_{(M-1),M}}}{n_{1,2}!\cdot\ldots\cdot n_{(M-1),M}!}\prod_{m=1}^{M}\frac{d^{n_m}}{dy_m^{n_m}}\left\{\Phi_{Y_m}(y_m)\right\},\qquad(2)$$

$$0\le\sum_{i=1}^{M-1}\sum_{j=i+1}^{M}n_{i,j}\le N_{max};\ n_m=\sum_{i=1}^{m-1}n_{i,m}+\sum_{i=m+1}^{M}n_{m,i}$$

в виде усеченного степенного ряда по элементам верхней треугольной матрицы $\left\{b_{i,j}\right\}_{i=1,(M-1)}^{j=(i+1),M}$ выборочных оценок ковариаций признаков.

Наш опыт интеллектуального анализа данных показал, что в качестве оценок интегральных распределений $\Phi_{Y_m}(y_m)$, $m=1,2,\ldots,M$ признаков рационально применять гистограмму сглаженную сдвигом [14] с ее последующей аппроксимацией полигауссовской моделью [2-5]. В этом случае соответствующие производные распределения достаточно просто вычисляют с помощью ортонормированных полиномов Чебышева – Эрмита.

5. Нечеткая классификация объектов по лингвистическим классам

Наблюдения $\vec{y}_k$, $k=1,2,\ldots,\tilde{K}$ из выборки типичных данных заносят в перекрывающиеся кластеры Ω_1 – низких, Ω_2 – средних и Ω_3 – высоких уровней показателей в соответствии с правилами базы знаний:

Если $\Phi(\vec{y}_k)\le\Phi_2$ то $\vec{y}_k\in\Omega_1$;

Если $\Phi_1<\Phi(\vec{y}_k)\le\Phi_4$ то $\vec{y}_k\in\Omega_2$; $\qquad(3)$

Если $\Phi(\vec{y}_k)>\Phi_3$ то $\vec{y}_k\in\Omega_3$.

В результате формируют три пересекающиеся подмножества типичных данных $\vec{y}_{k_l}$, $k_l=1,2,\ldots,K_l$, где l = 1, 2, 3 и $K_1+K_2+K_3\ge\tilde{K}$.

6. Байесовская модель скоринга

На втором этапе лингвистического анализа распределения (2) выполняют байесовскую нечеткую классификацию объектов. Рациональной методической основой скоринга является, на наш взгляд, описание генеральной совокупности наблюдений моделью конечной смеси стандартных многомерных распределений:

$$f(\vec{y}\,|\,\vec{\theta})=\sum_{l=1}^{L}P_l f_l(\vec{y}\,|\,\vec{c}_l,B_l),\ \sum_{l=1}^{L}P_l=1,\qquad(4)$$

где $\vec{\theta}=(P_1,\ldots,P_L,\vec{c}_1,\ldots,\vec{c}_L,B_1,\ldots,B_L)$ – вектор параметров смеси, P_l, $l=1,2,\ldots,L$ – априорные веса лингвистических классов уровней показателей; $f_l(\vec{y}\,|\,\vec{c}_l,B_l)$ – парциальные распределения с характеристиками положения $\vec{c}_l=(c_{1l},\ldots,c_{Ml})^T$ и масштаба $B_l=\left\{b_{i,j}^{(l)}\right\}_{i=\overline{1,M}}^{j=\overline{1,M}}$ классов.

В наших вычислительных экспериментах в качестве парциальных распределений применялась гауссовская модель

$$f_l\left(\vec{y}\,|\,\vec{c}_l,B_l\right)=$$
$$=\frac{1}{\sqrt{(2\pi)^M\,det(B_l)}}exp\left\{-\frac{1}{2}D\left(\vec{y}\,|\,\vec{c}_l,B_l\right)\right\},$$

где $D\left(\vec{y}\,|\,\vec{c}_l,B_l\right)=\left(\vec{y}-\vec{c}_l\right)^T B_l^{-1}\left(\vec{y}-\vec{c}_l\right)$ – многомерная метрика Махаланобиса. Априорные веса P_l, а также характеристики положения $\vec{c}_l=\left(c_{1l},\ldots,c_{Ml}\right)^T$ и масштабы B_l классов оценивают с помощью процедуры, подобной EM-алгоритму [11], а именно

$$P_l=\frac{w_l}{\tilde{K}},\quad w_l=\sum_{k_l=1}^{K_l}W_l\left\{\Phi\left(\vec{y}_{k_l}\right)\right\},$$

$$\vec{c}_l=\frac{1}{w_l}\sum_{k_l=1}^{K_l}W_l\left\{\Phi\left(\vec{y}_{k_l}\right)\right\}\vec{y}_{k_l} \qquad (5)$$

$$B_l=\frac{1}{w_l}\sum_{k_l=1}^{K_l}W_l\left\{\Phi\left(\vec{y}_{k_l}\right)\right\}\vec{y}_{k_l}\vec{y}_{k_l}^T-\vec{c}_l\vec{c}_l^T,$$

$$l=1,2,3.$$

Отметим, что в силу свойства нечеткого разбиения для моделей функций принадлежности лингвистических классов имеет место тождество

$$\sum_{l=1}^{3}W_l\left\{\Phi\left(\vec{y}_k\right)\right\}\equiv 1,$$

откуда несложно получить

$$\tilde{K}=\sum_{k=1}^{\tilde{K}}\sum_{l=1}^{3}W_l\left\{\Phi\left(\vec{y}_k\right)\right\}=$$
$$=\sum_{l=1}^{3}\sum_{k=1}^{K_l}W_l\left\{\Phi\left(\vec{y}_k\right)\right\}=\sum_{l=1}^{3}w_l \quad\Rightarrow\quad \sum_{l=1}^{3}\frac{w_l}{\tilde{K}}=\sum_{l=1}^{3}P_l=1.$$

Этот результат согласуется со структурой модели конечной смеси (4). Более того, оценки (5) параметров лингвистических классов имеют наглядный практический смысл. Веса P_l, $l=1,2,3$ естественно трактовать как субъективные вероятности низких, средних и высоких значений показателей. Компоненты вектора $\vec{c}_l$ и элементы матрицы B_l рационально интерпретировать как средние значения для низких, средних и высоких уровней показателей и меры их волатильности соответственно.

7. Мягкая и жесткая классификация объектов

В рамках представления (4) байесовский подход к нечеткой классификации объектов сводится к описанию многомерных функций принадлежности для классов апостериорными весами

$$Wr_l\left(\vec{y}\,|\,\vec{\theta}\right)=P_l f_l\left(\vec{y}\,|\,\vec{c}_l,B_l\right)\Big/f\left(\vec{y}\,|\,\vec{\theta}\right),$$
$$l=1,2,\ldots,L.$$

Вектор $\vec{y}_k$ признаков для k-го объекта ассоциируют с l-м лингвистическим классом уровней показателей с надежностью $Wr_l\left(\vec{y}_k\,|\,\vec{\theta}\right)$. Такого рода мягкая классификация позволяет модели функционировать в нестационарной среде за счет адаптации статистик интегрального распределения $\Phi\left(\vec{y}\right)$ к изменяющимся внешним условиям. Иными словами, обеспечивается возможность формирования динамической модели лингвистической переменной.

Жесткая классификация предполагает применение правила победителя, т.е. k-й объект зачисляется в l-й класс уровней показателей, если выполняется условие $Wr_l\left(\vec{y}_k\,|\,\vec{\theta}\right)>Wr_j\left(\vec{y}_k\,|\,\vec{\theta}\right)$ для всех $j=1,2,\ldots,L$ и $j\neq l$.

Важно отметить, что ковариационное приближение (2) интегрального распределения позволяет предельно просто получить маргинальное распределение любого подмножества компонент вектора признаков $\vec{y}$ и в соответствии с изложенной выше методикой выполнить скоринг объектов для интересующей аналитика группы показателей.

8. Вычислительные эксперименты

На рис. 2 представлено маргинальное ковариационное приближение интегрального распределения нормированных поступлений от ОНР консолидированных по месяцам за 2008 г. для общегородского заказа y_{1k} (m = 1) и центрального административного округа y_{8k} (m = 8), $k=1,2,\ldots,\tilde{K}$ [4].

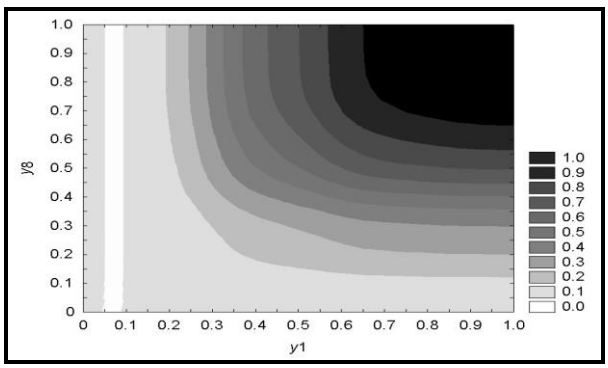


Рис. 2. Ковариационное приближение двумерного интегрального распределения поступлений от ОНР

Предварительное преобразование исходных данных сводилось к операции логарифмирования $Y_{mk}=ln\left(X_{mk}+10^{-7}\right)$ для $k=1,2,\ldots,30$. Объем выборки очищенной от аномальных значений составил $\tilde{K}=23$, а количество членов разложения степенного ряда (2) $N_{max}=3$. В качестве одномерных распределений $\Phi_{Y_m}\left(y_m\right)$ применялись их *ASH*-оценки, аппроксимированные полигауссовскими моделями [4]. Ковариация типичных нормированных поступлений y_1 и y_8 составила величину $b_{18}=0.0784$.

Нечеткая кластеризация ОНР по лингвистическим классам в соответствии с правилами (3) и последующая идентификация параметров

$$P_l,\ \vec{c}_l = \left(c_{l1}, c_{l8}\right)^T,$$

$$B_l = \begin{pmatrix} \sigma_{l1}^2 & \rho_{l18}\sigma_{l1}\sigma_{l8} \\ \rho_{l18}\sigma_{l1}\sigma_{l8} & \sigma_{l8}^2 \end{pmatrix},\ l = 1,2,3$$

трехуровневого байесовского нечеткого классификатора по формулам (5) позволила получить значения, сведенные в табл. 2.

Таблица 2

ПАРАМЕТРЫ БАЙЕСОВСКОГО НЕЧЕТКОГО КЛАССИФИКАТОРА ОНР

l	Параметры					
	P_l	c_{l1}	c_{l8}	σ_{l1}	σ_{l8}	ρ_{l18}
1	0,5136	0,3629	0,1757	0,1309	0,1336	0,6482
2	0,2157	0,6126	0,4781	0,0817	0,0343	0,1596
3	0,2707	0,8057	0,6297	0,0995	0,0761	0,8119

Рис. 3 демонстрирует полигауссовскую модель (4) плотности распределения вероятностей (ПРВ), соответствующую указанным выше оценкам. Важно отметить, что распределение имеет явно выраженную кластерную структуру, содержащую три класса, а классификация ОНР хорошо согласуются с данными, представленными в работе [4]. Результаты жесткой кластеризации ОНР по правилу победителя иллюстрируют табл. 3-5.

Скоринг акций выполнялся в рамках фундаментального анализа ценных бумаг по четырем финансовым мультипликаторам [6].

$X_1 = PlE$; $X_2 = EV/EBITDA$; $X_3 = P/S$; $X_4 = ROE$.

Значения этих показателей подвергались предварительному преобразованию $y_n = \left\{ ln\left(X_n + \varepsilon \right) - x_n^{(min)} \right\} / R_n$, n = 1, 2, 3, 4 к стандартному диапазону от нуля до единицы с параметрами, представленными в работе [5].

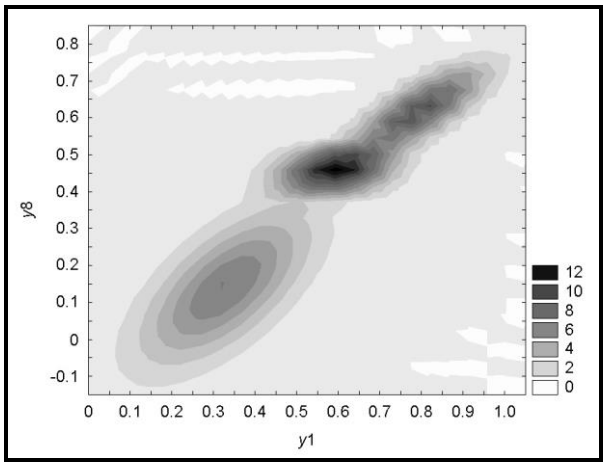


Рис. 3. Полигауссовская модель ПРВ поступлений для трехуровневого байесовского нечеткого классификатора ОНР

Таблица 3

ОНР С НИЗКИМ УРОВНЕМ ПОСТУПЛЕНИЙ В ГОРОДСКОЙ БЮДЖЕТ

Тип конструкции	k	y_{1k}	y_{8k}	$\Phi(\vec{y})$	$Wr_1(\vec{y})$	$Wr_2(\vec{y})$	$Wr_3(\vec{y})$
Витраж	1	0,114591	0,075756	0,0000	1,0000	0,0000	0,0000
Электронный экран на крыше	29	0,515978	-1,570484	0,0000	1,0000	0,0000	0,0000
Электронный экран на здании	28	0,338778	-1,570484	0,0000	1,0000	0,0000	0,0000
Щит на ограждении	26	0,120491	-1,570484	0,0000	1,0000	0,0000	0,0000
Щит на временном ограждении	25	0,407063	-1,570484	0,0000	1,0000	0,0000	0,0000
Реклама на флагах и других мягких полотнищах на здании, маркизы (площадью менее 2 кв. м)	21	0,000000	-1,570484	0,0000	1,0000	0,0000	0,0000
Реклама на строительной сетке	18	-1,705386	-1,570484	0,0000	1,0000	0,0000	0,0000
Прочие нетиповые объекты (мусорные контейнеры, телефонные будки и пр.)	16	0,208243	-1,570484	0,0000	1,0000	0,0000	0,0000
Объемно-пространственный объект (отдельно стоящий)	14	0,373928	0,000001	0,0000	1,0000	0,0000	0,0000
Носимая реклама	12	-0,156199	-1,570484	0,0000	1,0000	0,0000	0,0000
Маркиза	9	0,393034	0,064762	0,0031	1,0000	0,0000	0,0000
Кронштейн на несветовой опоре	6	0,385711	0,154384	0,0398	1,0000	0,0000	0,0000
Отдельно стоящая реклама на флагах и других мягких полотнищах	15	0,463460	0,154399	0,0518	1,0000	0,0000	0,0000
Реклама на флагах и других мягких полотнищах на здании, маркизы (площадью более 2 кв. м)	20	0,347790	0,214808	0,0634	1,0000	0,0000	0,0000
Объемно-пространственная установка	13	0,424678	0,257558	0,1203	1,0000	0,0000	0,0000
Выносная щитовая конструкция (штендер)	2	0,430313	0,370771	0,2021	0,9854	0,0109	0,0036
Настенное панно (площадью менее 2 кв. м)	11	0,531127	0,362197	0,2445	0,9728	0,0177	0,0095

Таблица 4

ОНР СО СРЕДНИМ УРОВНЕМ ПОСТУПЛЕНИЙ В ГОРОДСКОЙ БЮДЖЕТ

Тип конструкции	k	y_{1k}	y_{8k}	$\Phi(\vec{y})$	$Wr_1(\vec{y})$	$Wr_2(\vec{y})$	$Wr_3(\vec{y})$
Кронштейн на здании (площадью одной стороны	5	0,518575	0,472426	0,4118	0,0703	0,9105	0,0192

Тип конструкции	k	y_{1k}	y_{8k}	$\Phi(\vec{y})$	$Wr_1(\vec{y})$	$Wr_2(\vec{y})$	$Wr_3(\vec{y})$
менее 1 кв. м и высотой менее 1,5 м)							
Реклама на флагах и др. мягких полотнищах	19	0,533415	0,478440	0,4337	0,0553	0,9192	0,0255
Стационарная установка над проезжей частью	22	0,697657	0,436171	0,4717	0,0723	0,8960	0,0317
Информационная конструкция предприятия и организации по обслуживанию населения на здании	3	0,607643	0,480504	0,5104	0,0287	0,8898	0,0815

Таблица 5

ОНР С ВЫСОКИМ УРОВНЕМ ПОСТУПЛЕНИЙ В ГОРОДСКОЙ БЮДЖЕТ

Тип конструкции	k	y_{1k}	y_{8k}	$\Phi(\vec{y})$	$Wr_1(\vec{y})$	$Wr_2(\vec{y})$	$Wr_3(\vec{y})$
Реклама на остановках транспорта	17	0,706343	0,554729	0,7036	0,0100	0,1152	0,8748
Тумба отдельно стоящая	24	0,719035	0,553886	0,7075	0,0085	0,1028	0,8886
Электронный экран отдельно стоящий	30	0,783911	0,548766	0,7215	0,0068	0,0899	0,9034
Кронштейн на здании (площадью одной стороны более 1 кв. м и высотой более 1,5 м)	4	0,721240	0,651465	0,8253	0,0055	0,0000	0,9945
Кронштейн на световой опоре	7	0,898955	0,674252	0,9017	0,0001	0,0000	0,9999
Крышная установка	8	0,960014	0,749888	0,9697	0,0000	0,0000	1,0000
Настенное панно (площадью более 2 кв. м)	10	1,000000	0,815800	0,9939	0,0000	0,0000	1,0000
Щит отдельно стоящий	27	1,233952	1,000000	1,0000	0,0000	0,0000	1,0000
Транспарант-перетяжка	23	1,059544	0,896578	1,0000	0,0000	0,0000	1,0000

Правила приближенных рассуждений (3) относительно инвестиционного качества эмитентов формулировались на основе экспертной модели универсальной шкалы субъективных вероятностей p для низких, средних и высоких уровней мультипликаторов (L = 3). Соответствующие функции принадлежности $W_l(p)$, $l = 1,2,3$ нечеткого классификатора биржевых активов иллюстрирует рис. 4.

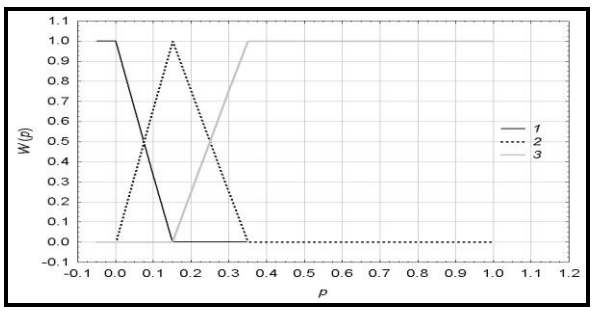


Рис. 4. Лингвистическая шкала субъективных вероятностей качества акций: 1 – низкое; 2 – среднее; 3 – высокое

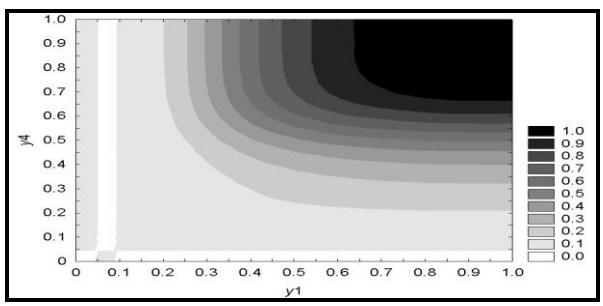


Рис. 5. Маргинальное двумерное ковариационное приближение интегрального распределения мультипликаторов акций

В качестве одномерных распределений $\Phi_{Ym}(y_m)$, $m = 1,2,3,4$ в четырехмерном (M = 4) ковариационном приближении (2) применялись их ASH-оценки, аппроксимированные полигауссовскими моделями с параметрами, представленными в табл. 2-5 работы [5]. Корреляция типичных нормированных значений факторов y_1, …, y_4 оценивалась по данным портала (подробнее см. http://www.-2stocks.ru/main/invest/stocks/info/finance) за IV квартал 2010 г. Элементы ковариационной матрицы составили величины: $b_{12} = 0.0137$; $b_{13} = 0.0230$; $b_{14} = -0.0159$; $b_{23} = 0.0039$; $b_{24} = 0.0058$; $b_{34} = -0.0142$ [3]. Рис. 5 демонстрирует маргинальное двумерное ковариационное приближение для интегрального распределения мультипликаторов y_1 и y_4.

Параметры трехуровневого байесовского нечеткого классификатора эмитентов, рассчитанные по формулам (5), содержат табл. 6 и 7.

Таблица 6

ПАРАМЕТРЫ БАЙЕСОВСКОГО НЕЧЕТКОГО КЛАССИФИКАТОРА ЭМИТЕНТОВ (ОДНОМЕРНЫЕ РАСПРЕДЕЛЕНИЯ)

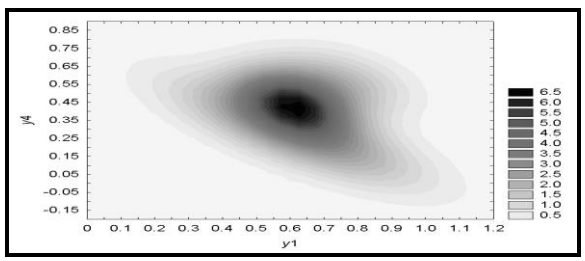
Параметры	l		
	1	2	3
P_l	0,4894	0,3366	0,1740
c_{l1}	0,4903	0,4867	0,5969
c_{l2}	0,3631	0,4965	0,6336
c_{l3}	0,4877	0,4680	0,5431
c_{l4}	0,3380	0,4936	0,6060
σ_{l1}	0,2172	0,1077	0,1340
σ_{l2}	0,1712	0,1417	0,1715
σ_{l3}	0,2006	0,1381	0,3100
σ_{l4}	0,1867	0,1196	0,1234

Таблица 7

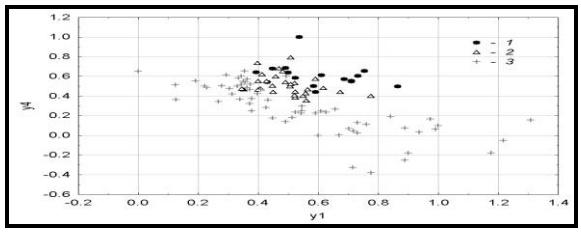
ПАРАМЕТРЫ БАЙЕСОВСКОГО НЕЧЕТКОГО КЛАССИФИКАТОРА ЭМИТЕНТОВ (КОВАРИАЦИОННОЯ МАТРИЦА)

Параметры	*l*		
	1	2	3
ρ_{l12}	0,4247	0,1082	0,5626
ρ_{l13}	0,7993	0,3789	0,1233
ρ_{l14}	-0,7505	-0,4339	-0,3233
ρ_{l23}	0,1367	-0,0634	0,0969
ρ_{l24}	-0,1612	-0,1740	-0,2940
ρ_{l34}	-0,6423	-0,3677	-0,0660

Анализ маргинальных ПРВ полигауссовской модели (4) классификатора показал, что распределение с полученными параметрами не имеет явно выраженной кластерной структуры. В качестве примера на рис. 6а изображена маргинальная двумерная ПРВ мультипликаторов y_1 и y_4. Рис. 6б демонстрирует результаты жесткой классификации акций. Маркерами 1, 2 и 3 отмечены активы с высоким, средним и низким инвестиционным качеством соответственно. Аналогичные результаты в пространстве трех мультипликаторов y_1, y_2 и y_4 иллюстрирует рис. 7.



а



б

Рис. 6. Полигауссовская маргинальная двумерная ПРВ мультипликаторов для байесовского нечеткого классификатора акций

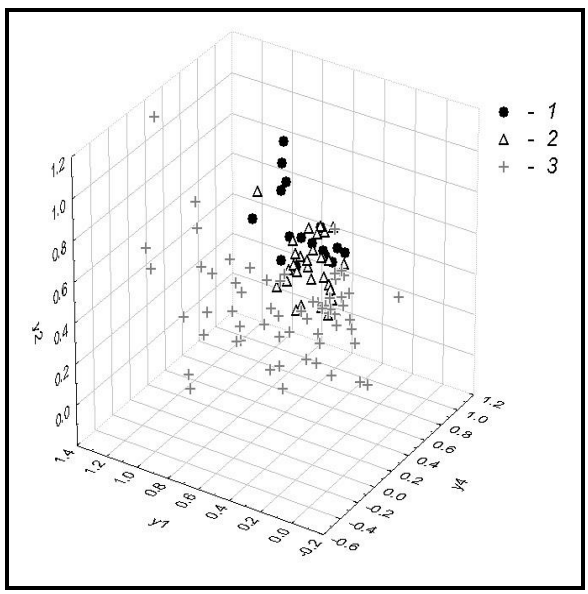


Рис. 7. Байесовская нечеткая классификация акций в пространстве трех мультипликаторов

Результаты жесткой классификации эмитентов по правилу победителя сведены в табл. 8-10.

Таблица 8

ЭМИТЕНТЫ С ВЫСОКИМ ИНВЕСТИЦИОННЫМ КАЧЕСТВОМ

Эмитенты	*k*	y_{1k}	y_{2k}	y_{3k}	y_{4k}	$\Phi(\vec{y}_k)$	$Wr_1(\vec{y}_k)$	$Wr_2(\vec{y}_k)$	$Wr_3(\vec{y}_k)$
Уралкалий	96	0,753915	0,659735	0,999999	0,656542	0,7936	0,0000	0,0000	1,0000
НОВАТЭК	59	0,733734	0,570432	0,918376	0,604600	0,6320	0,0000	0,0000	1,0000
КАЛИНА	37	0,611905	0,547526	0,582907	0,611912	0,5398	0,0000	0,0000	1,0000
Полиметалл	70	0,686000	0,558117	0,849024	0,571993	0,5288	0,0000	0,0000	1,0000
Сильвинит	86	0,711362	0,521055	0,809239	0,552928	0,4665	0,0000	0,0000	1,0000
Челябинский трубопрокатный завод	104	0,537546	0,572895	0,388645	0,999999	0,4317	0,0000	0,0000	1,0000
Вимм-Билль-Данн	17	0,864688	0,668855	0,649491	0,499053	0,4091	0,0000	0,0000	1,0000
Магнит	53	0,709900	0,555565	0,469077	0,548403	0,3773	0,0000	0,0000	1,0000
Армада	5	0,492058	0,436305	0,495784	0,684972	0,3599	0,0000	0,0000	1,0000
ТрансКредитБанк	95	0,393865	0,668909	0,572833	0,641397	0,3491	0,0000	0,0046	0,9954
Центр Телеком	102	0,522949	0,421204	0,583996	0,585195	0,3091	0,0000	0,2047	0,7953
Сбербанк	81	0,430511	0,654822	0,587230	0,541101	0,2818	0,0000	0,3411	0,6589
Фармстандарт	99	0,499980	0,368025	0,648700	0,638502	0,2815	0,0000	0,3425	0,6575
МТС	52	0,447257	0,406616	0,480180	0,677787	0,2657	0,0000	0,4213	0,5787
ЮТК	108	0,584371	0,545516	0,487262	0,500310	0,2646	0,0000	0,4268	0,5732
Банк ВТБ Северо-Запад	8	0,590320	0,732175	0,654947	0,439158	0,2522	0,0000	0,4893	0,5107

Таблица 9

ЭМИТЕНТЫ СО СРЕДНИМ ИНВЕСТИЦИОННЫМ КАЧЕСТВОМ

Эмитенты	k	y_{1k}	y_{2k}	y_{3k}	y_{4k}	$\Phi(\vec{y}_k)$	$Wr_1(\vec{y}_k)$	$Wr_2(\vec{y}_k)$	$Wr_3(\vec{y}_k)$
Балтика	7	0,482400	0,341794	0,626975	0,650328	0,2365	0,0000	0,5674	0,4326
Сибирьтелеком	84	0,521483	0,451449	0,490713	0,532476	0,2343	0,0000	0,5786	0,4214
Кузбасская ТК	41	0,617801	0,511094	0,470970	0,479981	0,2260	0,0000	0,6167	0,3833
Силовые машины	85	0,398063	0,413586	0,461327	0,736808	0,2250	0,0000	0,6248	0,3752
Полюс Золото	71	0,776328	0,575830	0,842341	0,398867	0,2219	0,0000	0,6403	0,3597
Уралсвязьинформ	97	0,459447	0,375728	0,532388	0,598736	0,2110	0,0000	0,6948	0,3052
РНТ	73	0,492295	0,382949	0,576246	0,539921	0,1926	0,0000	0,7869	0,2131
Номос - банк	63	0,407253	0,736362	0,586158	0,475856	0,1804	0,0000	0,8482	0,1518
ИНТЕР РАО ЕЭС	33	0,409581	0,574409	0,735371	0,475393	0,1753	0,0000	0,8733	0,1267
Мостотрест	55	0,674760	0,622449	0,385813	0,439695	0,1682	0,0000	0,9092	0,0908
ВТБ	15	0,525029	0,752284	0,605633	0,384019	0,1626	0,0000	0,9370	0,0630
ВолгаТелеком	18	0,559562	0,460153	0,526723	0,426080	0,1529	0,0000	0,9853	0,0147
М- видео	44	0,591746	0,503476	0,265108	0,574363	0,1466	0,0224	0,9776	0,0000
Фосагро	100	0,507655	0,372681	0,539257	0,496095	0,1433	0,0448	0,9552	0,0000
ТД Копейка	89	0,508321	0,461880	0,237687	0,791064	0,1288	0,1410	0,8590	0,0000
НЛМК	57	0,523158	0,394524	0,570526	0,444828	0,1283	0,1449	0,8552	0,0000
Распадская	76	0,448107	0,356507	0,677504	0,503080	0,1249	0,1675	0,8325	0,0000
Дальсвязь	30	0,428066	0,363951	0,457339	0,550337	0,1219	0,1875	0,8125	0,0000
ЭОН-Россия (бывш. ОГК-4)	106	0,550767	0,410862	0,657241	0,400134	0,1210	0,1941	0,8059	0,0000
НМТП	58	0,412542	0,286948	0,678192	0,621047	0,1201	0,1992	0,8008	0,0000
Северо-Западный телеком	82	0,522740	0,434026	0,513374	0,406168	0,1149	0,2340	0,7660	0,0000
РусгрэйнХолдинг	80	0,564552	0,686871	0,289296	0,466590	0,1126	0,2494	0,7506	0,0000
Банк Санкт-Петербург	10	0,351635	0,719803	0,469077	0,469138	0,1019	0,3206	0,6794	0,0000
Акрон	4	0,449800	0,421898	0,481973	0,440588	0,1014	0,3241	0,6759	0,0000
ГМК НорНикель	20	0,509647	0,277156	0,687589	0,520349	0,0975	0,3499	0,6501	0,0000
Дальневосточный Банк	29	0,346130	0,673768	0,444890	0,475239	0,0933	0,3779	0,6221	0,0000
Верофарм	16	0,399101	0,306465	0,548502	0,550563	0,0891	0,4060	0,5940	0,0000
Волжская ТГК (ТГК-7)	19	0,560461	0,442840	0,474703	0,352050	0,0867	0,4218	0,5782	0,0000
РОСИНТЕР	75	0,470287	0,479335	0,212765	0,681241	0,0790	0,4734	0,5266	0,0000
Мечел	54	0,399791	0,494469	0,344921	0,463205	0,0757	0,4956	0,5044	0,0000

Таблица 10

ЭМИТЕНТЫ С НИЗКИМ ИНВЕСТИЦИОННЫМ КАЧЕСТВОМ

Эмитенты	k	y_{1k}	y_{2k}	y_{3k}	y_{4k}	$\Phi(\vec{y}_k)$	$Wr_1(\vec{y}_k)$	$Wr_2(\vec{y}_k)$	$Wr_3(\vec{y}_k)$
Группа Черкизово	25	0,335184	0,386263	0,371033	0,583126	0,0687	0,5418	0,4582	0,0000
МГТС	45	0,546628	0,425849	0,530986	0,295906	0,0678	0,5478	0,4522	0,0000
Аэрофлот	6	0,363490	0,436629	0,279960	0,572717	0,0632	0,5788	0,4212	0,0000
Роснефть	77	0,350376	0,326244	0,465238	0,547488	0,0622	0,5855	0,4145	0,0000
Иркутскэнерго	36	0,342680	0,290870	0,490713	0,600124	0,0620	0,5865	0,4135	0,0000
ТНК-ВР	91	0,355376	0,270840	0,444890	0,655224	0,0613	0,5912	0,4088	0,0000
ДИОД	28	0,545487	0,545516	0,516416	0,250799	0,0605	0,5966	0,4034	0,0000
Башнефть	13	0,374800	0,356743	0,377073	0,519578	0,0587	0,6085	0,3915	0,0000
Энел ОГК-5	107	0,627142	0,549612	0,503966	0,237146	0,0580	0,6131	0,3869	0,0000
Дорогобуж	31	0,356199	0,295671	0,495784	0,542626	0,0556	0,6293	0,3707	0,0000
Синергия	87	0,397020	0,408103	0,385813	0,423798	0,0525	0,6501	0,3499	0,0000
ИКРУСС-ИНВЕСТ	32	0,343547	0,289896	0,721945	0,518289	0,0498	0,6680	0,3320	0,0000
ТрансКонтейнер	94	0,841525	0,670028	0,568192	0,194331	0,0440	0,7065	0,2935	0,0000
Новосибирскэнерго	61	0,432931	0,476413	0,318986	0,361376	0,0421	0,7194	0,2806	0,0000
МРСК Центра	50	0,363490	0,369811	0,318986	0,486261	0,0385	0,7432	0,2568	0,0000
Татнефть	92	0,372503	0,336390	0,361626	0,468663	0,0365	0,7565	0,2435	0,0000
ТМК	90	0,655903	0,673976	0,270179	0,265565	0,0347	0,7684	0,2316	0,0000
Башкирэнерго	12	0,591746	0,426868	0,377073	0,227884	0,0266	0,8224	0,1776	0,0000
Красноярская ГЭС	40	1,309453	1,024336	1,152952	0,154788	0,0242	0,8387	0,1613	0,0000
МОЭСК	47	0,278550	0,365998	0,371033	0,502942	0,0220	0,8536	0,1464	0,0000
ОГК-2	65	0,544149	0,417872	0,355106	0,227454	0,0218	0,8544	0,1456	0,0000
Газпромнефть	22	0,329751	0,302521	0,318986	0,500588	0,0213	0,8583	0,1417	0,0000
Таттелеком	93	0,313187	0,360959	0,377073	0,420140	0,0207	0,8619	0,1381	0,0000
ТГК-1	88	0,338298	0,352924	0,394196	0,366029	0,0191	0,8726	0,1274	0,0000
АвтоВАЗ	3	0,493710	0,732649	0,142340	0,600676	0,0180	0,8801	0,1199	0,0000
ОГК-1	64	0,426526	0,309746	0,385813	0,283268	0,0166	0,8890	0,1110	0,0000
Квадра (бывш. ТГК-4)	39	0,371732	0,269410	0,433906	0,322864	0,0132	0,9117	0,0883	0,0000
ДИКСИГРУПП	27	0,975063	0,721473	0,284681	0,166128	0,0114	0,9241	0,0759	0,0000

Эмитенты	k	y_{1k}	y_{2k}	y_{3k}	y_{4k}	$\Phi(\vec{y}_k)$	$Wr_1(\vec{y}_k)$	$Wr_2(\vec{y}_k)$	$Wr_3(\vec{y}_k)$
АФК Система	1	0,408918	0,452351	0,168591	0,464823	0,0099	0,9339	0,0661	0,0000
Нижнекамскнефтехим	60	0,293313	0,241033	0,259909	0,614533	0,0090	0,9401	0,0599	0,0000
Группа ЛСР	23	0,731327	0,677749	0,455316	0,129736	0,0089	0,9406	0,0594	0,0000
ЛУКОЙЛ	42	0,304213	0,271195	0,279960	0,476010	0,0085	0,9436	0,0564	0,0000
Мосэнерго	56	0,446972	0,380640	0,318986	0,176961	0,0057	0,9618	0,0382	0,0000
Банк Возрождение	9	0,759972	0,994106	0,500732	0,113081	0,0052	0,9656	0,0344	0,0000
Газпром	21	0,229448	0,228339	0,451209	0,485966	0,0039	0,9737	0,0263	0,0000
ОГК-6	66	0,491584	0,464868	0,298225	0,139205	0,0037	0,9756	0,0244	0,0000
МРСК Урала	49	0,377831	0,357923	0,176645	0,251578	0,0019	0,9877	0,0123	0,0000
Седьмой континент	83	0,512496	0,647377	0,132787	0,183750	0,0007	0,9952	0,0048	0,0000
Институт стволовых клеток человека	34	1,000000	0,803517	0,711119	0,098147	0,0007	0,9956	0,0044	0,0000
Казаньоргсинтез	38	0,378208	0,616978	0,112234	0,374138	0,0003	0,9983	0,0017	0,0000
Челябинский цинковый завод	105	0,221052	0,117047	0,270179	0,501005	0,0002	0,9989	0,0011	0,0000
ПАВА	67	0,776178	1,076287	0,160208	-0,378826	0,0000	1,0000	0,0000	0,0000
ВСМПО АВИСМА	14	0,936765	0,789440	0,561028	0,031909	0,0000	1,0000	0,0000	0,0000
Петрокоммерц	69	0,889232	1,207940	0,433906	-0,251000	0,0000	1,0000	0,0000	0,0000
РусГидро	79	0,731148	0,581497	0,399600	0,026053	0,0000	1,0000	0,0000	0,0000
РОСБАНК	74	1,176340	1,101420	0,666898	-0,176752	0,0000	1,0000	0,0000	0,0000
Новошип	62	0,714836	0,640517	0,419899	-0,326925	0,0000	1,0000	0,0000	0,0000
ММК	46	0,668826	0,656415	0,355106	0,001097	0,0000	1,0000	0,0000	0,0000
Ленэнерго	43	0,126836	0,355081	0,279960	0,364569	0,0000	1,0000	0,0000	0,0000
Ростелеком	78	0,990551	0,999999	0,854044	0,061809	0,0000	1,0000	0,0000	0,0000
РАО ЭС Востока	72	0,191603	-37,723825	-0,000001	0,553488	0,0000	1,0000	0,0000	0,0000
ФСК ЕЭС	98	0,600131	0,530048	0,655714	-0,000001	0,0000	1,0000	0,0000	0,0000
Группа Разгуляй	24	0,900170	-6,261502	0,122767	-0,179047	0,0000	1,0000	0,0000	0,0000
МРСК Волги	48	0,716364	0,592701	0,249099	0,045048	0,0000	1,0000	0,0000	0,0000
Холдинг МРСК	101	0,266423	0,435981	0,089393	0,343152	0,0000	1,0000	0,0000	0,0000
Иркут	35	0,123685	0,423797	0,017867	0,512796	0,0000	1,0000	0,0000	0,0000
Центральный Телеграф	103	1,216824	0,922197	0,511835	-0,052644	0,0000	1,0000	0,0000	0,0000
Авиакомпания ЮТэйр	2	0,609566	0,830781	0,063691	0,246465	0,0000	1,0000	0,0000	0,0000
Банк Таврический	11	0,889939	1,036654	0,529575	0,075076	0,0000	1,0000	0,0000	0,0000
МРСК Центра и Приволжья	51	0,701245	0,648734	0,249099	0,071008	0,0000	1,0000	0,0000	0,0000
ДВМП	26	0,000000	-0,000001	0,433906	0,649953	0,0000	1,0000	0,0000	0,0000
ПРОТЕК	68	0,523367	0,586304	-0,000001	0,237146	0,0000	1,0000	0,0000	0,0000

Сравнительный анализ результатов классификации эмитентов по критериям инвестиционного качества $\Phi(\vec{y}_k)$ и a_k, исследованных в данной статье и работе [5], представлен в табл. 11-13.

Таблица 11

ЭМИТЕНТЫ, КЛАССИФИЦИРОВАННЫЕ БЕЗ РАСХОЖДЕНИЙ

Эмитенты	k	$\Phi(\vec{y}_k)$	a_k	Качество по критерию Ф	Качество по критерию *а*
Уралкалий	96	0,7936	0,5997	3	3
НОВАТЭК	59	0,6320	0,5748	3	3
КАЛИНА	37	0,5398	0,5288	3	3
Полиметалл	70	0,5288	0,5566	3	3
Сильвинит	86	0,4665	0,5644	3	3
Вимм – Билль - Данн	17	0,4091	0,6099	3	3
Магнит	53	0,3773	0,5442	3	3
Транс Кредит Банк	95	0,3491	0,4987	3	3
Сбербанк	81	0,2818	0,4999	3	3
ЮТК	108	0,2646	0,5068	3	3
Банк ВТБ Северо-Запад	8	0,2521	0,5734	3	3
Балтика	7	0,2365	0,4401	2	2
Сибирьтелеком	84	0,2343	0,4766	2	2
Силовые машины	85	0,2250	0,4143	2	2
Уралсвязьинформ	97	0,2110	0,4315	2	2
РНТ	73	0,1926	0,4497	2	2
ИНТЕР-РАО-ЕЭС	33	0,1753	0,4782	2	2
ЭОН-Россия (бывш. ОГК-4)	106	0,1209	0,4794	2	2
НМТП	58	0,1201	0,4181	2	2
Северо-Западный телеком	82	0,1149	0,4742	2	2
М- видео	44	0,1466	0,4582	2	2
Фосагро	100	0,1433	0,4486	2	2
ТД Копейка	89	0,1288	0,4243	2	2
НЛМК	57	0,1283	0,4634	2	2
Распадская	76	0,1249	0,4305	2	2
Банк Санкт-Петербург	10	0,1019	0,4940	2	2

Эмитенты	k	$\Phi(\vec{y}_k)$	a_k	Качество по критерию Ф	Качество по критерию *a*
Акрон	4	0,1014	0,4347	2	2
ГМК НорнНикель	20	0,0975	0,4472	2	2
Дальневосточный Банк	29	0,0933	0,4718	2	2
Верофарм	16	0,0891	0,4129	2	2
Волжская ТГК (ТГК-7)	19	0,0867	0,4768	2	2
РОСИНТЕР	75	0,0790	0,4141	2	2
Мечел	54	0,0757	0,4300	2	2
Группа Черкизово	25	0,0687	0,3865	1	1
Аэрофлот	6	0,0632	0,3872	1	1
Роснефть	77	0,0622	0,3907	1	1
Иркутскэнерго	36	0,0620	0,3958	1	1
ТНК-ВР	91	0,0613	0,3852	1	1
Башнефть	13	0,0587	0,3826	1	1
Дорогобуж	31	0,0556	0,3978	1	1
Синергия	87	0,0525	0,4010	1	1
МРСК Центра	50	0,0385	0,3721	1	1
Татнефть	92	0,0365	0,3782	1	1
МОЭСК	47	0,0220	0,3807	1	1
ОГК-2	65	0,0218	0,3957	1	1
Газпромнефть	22	0,0213	0,3648	1	1
Таттелеком	93	0,0207	0,3808	1	1
ТГК-1	88	0,0191	0,3813	1	1
ОГК-1	64	0,0166	0,3704	1	1
Квадра (бывш. ТГК-4)	39	0,0132	0,3801	1	1
Нижнекамскнефтехим	60	0,0090	0,3455	1	1
АФК Система	1	0,0099	0,3878	1	1
ЛУКОЙЛ	42	0,0085	0,3508	1	1
Мосэнерго	56	0,0057	0,3275	1	1
Газпром	21	0,0039	0,3845	1	1
ОГК-6	66	0,0037	0,3711	1	1
МРСК Урала	49	0,0019	0,3114	1	1
Седьмой континент	83	0,0007	0,3924	1	1
Казаньоргсинтез	38	0,0003	0,4080	1	1
Челябинский цинковый завод	105	0,0002	0,3473	1	1
ПРОТЕК	68	0,0000	0,3978	1	1
Группа Разгуляй	24	0,0000	0,3885	1	1
ДВМП	26	0,0000	0,3814	1	1
Холдинг МРСК	101	0,0000	0,3707	1	1
Иркут	35	0,0000	0,3657	1	1
Ленэнерго	43	0,0000	0,3537	1	1
РАО ЭС Востока	72	0,0000	0,3376	1	1

Таблица 12

ЭМИТЕНТЫ, КЛАССИФИЦИРОВАННЫЕ С РАСХОЖДЕНИЕМ НА ОДИН КЛАСС

Эмитенты	k	$\Phi(\vec{y}_k)$	a_k	Качество по критерию Ф	Качество по критерию *a*
Челябинский трубопрокатный завод	104	0,4317	0,4820	3	2
Армада	5	0,3599	0,4607	3	2
ЦентрТелеком	102	0,3091	0,4767	3	2
Фармстандарт	99	0,2815	0,4500	3	2
МТС	52	0,2657	0,4257	3	2
Кузбасская ТК	41	0,2267	0,5072	2	3
Полюс Золото	71	0,2219	0,5823	2	3
Номос-банк	63	0,1804	0,5238	2	3
Мостотрест	55	0,1682	0,5267	2	3
ВТБ	15	0,1626	0,5611	2	3
Волга Телеком	18	0,1529	0,4989	2	3
Русгрэйн Холдинг	80	0,1126	0,4968	2	3
Дальсвязь	30	0,1219	0,4019	2	1
МГТС	45	0,0678	0,4688	1	2
ДИОД	28	0,0605	0,4710	1	2
ЭнелОГК-5	107	0,0580	0,4781	1	2
ИК РУССИНВЕСТ	32	0,0498	0,4120	1	2
Новосибирскэнерго	61	0,0421	0,4248	1	2
АвтоВАЗ	3	0,0180	0,4762	1	2
ТМК	90	0,0347	0,4823	1	2
Новошип	62	0,0000	0,4880	1	2
РусГидро	79	0,0000	0,4784	1	2
Авиакомпания ЮТэйр	2	0,0000	0,4732	1	2
ММК	46	0,0000	0,4674	1	2
ФСК_ЕЭС	98	0,0000	0,4622	1	2
МРСК Центра и Приволжья	51	0,0000	0,4495	1	2
МРСК Волги	48	0,0000	0,4411	1	2
Башкирэнерго	12	0,0266	0,4137	1	2

Таблица 13

ЭМИТЕНТЫ, КЛАССИФИЦИРОВАННЫЕ С РАСХОЖДЕНИЕМ НА ДВА КЛАССА

Эмитенты	k	$\Phi(\vec{y}_k)$	a_k	Качество по критерию Ф	Качество по критерию *a*
Транс Контейнер	94	0,0440	0,5505	1	3
Красноярская ГЭС	40	0,0242	0,5766	1	3
ДИКСИ ГРУПП	27	0,0114	0,5096	1	3
Группа ЛСР	23	0,0089	0,5142	1	3
Банк Возрождение	9	0,0052	0,5559	1	3
Институт стволовых клеток человека	34	0,0007	0,5746	1	3
Ростелеком	78	0,0000	0,5749	1	3
РОСБАНК	74	0,0000	0,5748	1	3
ВСМПОАВИСМА	14	0,0000	0,5716	1	3
Банк Таврический	11	0,0000	0,5681	1	3
Центральный	103	0,0000	0,5641	1	3

Эмитенты	k	$\Phi(\vec{y}_k)$	a_k	Качество по критерию Ф	Качество по критерию a
Телеграф					
Петрокоммерц	69	0,0000	0,5458	1	3
ПАВА	67	0,0000	0,4988	1	3

Цифрами 1, 2 и 3 пронумерованы классы соответственно низкой, средней и высокой привлекательности акций. Классификация 67 активов из 108 совпадает по обоим показателям (табл. 11). Расхождения на один класс качества низкое $\Leftrightarrow$ среднее или среднее $\Leftrightarrow$ высокое наблюдаются для 28 активов (табл. 12).

Значимые расхождения на два класс качества низкое $\Leftrightarrow$ высокое наблюдаются для классификации 13 активов (табл. 13). Причиной таких невязок является, на наш взгляд, применение субъективного в значительной мере метода парных сравнений Т. Саати. Экспертное оценивание этим методом [6] дает явно завышенный вес важности фактора y_4 в процедуре двукратной свертки.

ЗАКЛЮЧЕНИЕ

В работе предложена методика скоринга объектов на основе процедуры лингвистического анализа многомерного вероятностного распределения информативных признаков. Рациональное сочетание экспертного и статистического подходов к учету факторов неопределенности позволило сформулировать этапы нечеткого логического вывода, инвариантного к содержанию предметной области. В частности, синтез нелинейной модели приближенных рассуждений основан на комбинации нечеткой универсальной шкалы субъективных мнений экспертов и ковариационном приближении многомерного распределения признаков. Адаптация параметров байесовской процедуры классификации объектов с помощью ***EM***-подобного алгоритма позволяет модели адекватно функционировать в нестационарной среде.

Вместе с тем, результаты классификации объектов существенным образом зависят от выбора экспертной модели универсальной шкалы субъективных вероятностей для лингвистических уровней показателей. Мера неопределенности может быть снижена, по нашему мнению, с помощью обучения этой модели на исторических данных для выбранного критерия качества объектов анализа.

Литература

1. Ведерников В.В. Нечетко-множественное моделирование в анализе и прогнозировании экономических явлений и процессов: исторический аспект [Электронный ресурс] / В.В. Ведерников // Проблемы современной экономики. – 2006. – №1. Режим доступа: http://www.m-economy.ru.
2. Лабунец Л.В. и др. Анализ динамики поступлений в бюджет от московского рынка наружной рекламы [Текст] / Л.В. Лабунец, Н.Л. Лебедева, М.Ю. Чижов // Вестник Российского нового университета. – 2013. – №4. – С. 66-81.
3. Лабунец Л.В. и др. Кумулянтное приближение многомерного вероятностного распределения [Текст] / Л.В. Лабунец, Н.Л. Лебедева, М.Ю. Чижов // Радиотехника и электроника. – 2012. – Т. 57; №6. – С. 638-648.
4. Лабунец Л.В. и др. Нечетко-множественный анализ московского рынка наружной рекламы [Текст] / Л.В. Лабунец, Н.Л. Лебедева, М.Ю. Чижов // Аудит и финансовый анализ. – 2014. – №2. – С. 139-148.
5. Лабунец Л.В. и др. Скоринг биржевых активов на основе байесовской нечеткой классификации [Текст] / Л.В. Лабунец, Е.Л. Лабунец, Н.Л. Лебедева // Аудит и финансовый анализ. – 2014. – №3. – С. 450-459.
6. Лабунец Л.В. и др. Экспертная модель скоринга российских акций [Текст] / Л.В. Лабунец, Е.Л. Лабунец, Н.Л. Лебедева // Аудит и финансовый анализ. – 2014. – №4. – С. 102-110.
7. Недосекин А.О. Нечетко-множественный анализ рисков фондовых инвестиций [Текст] / А.О. Недосекин. – СПб. : Сезам, 2002. – 181 с.
8. Сидельников Ю.В. Системный анализ экспертного прогнозирования [Текст] / Ю.В. Сидельников. – М. : МАИ, 2007. – 453 с.

9. Орлов А.И. Теория принятия решений [Текст] : учеб. пособие / А.И. Орлов. – М. : Экзамен, 2005. – 656 с.
10. Прикладная статистика: исследование зависимостей [Текст] : справ. изд. / С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин ; под ред. С.А. Айвазяна. – М. : Финансы и статистика, 1985. – 487 с.
11. Прикладная статистика: классификация и снижение размерности [Текст] : справ. изд. / С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин ; под ред. С.А. Айвазяна. – М. : Финансы и статистика, 1989. – 607 с.
12. Шурыгин А.М. Прикладная стохастика: робастность, оценивание, прогноз [Текст] / А.М. Шурыгин – М. : Финансы и статистика, 2000. – 224 с.
13. Buja A. et al. 14 – Computational methods for high-dimensional rotations in data visualization [Text] / A. Buja, D. Cook, D. Asimov, C. Hurley // Rao C.R. et al. Handbook of statistics, data mining and data visualization. – 2005. – Vol. 24. – Pp. 391-413.
14. Scott D.W. Multivariate density estimation: theory, practice, and visualization [Text] / D.W. Scott. – N.-Y. : John Wiley & Sons, Inc, 1992. – 317 p.

Ключевые слова

Data mining; экспертное оценивание; нечеткая шкала; ковариационное приближение; лингвистический анализ распределения; EM-алгоритм; байесовские оценки; нечеткий логический вывод; скоринг.

Лабунец Леонид Витальевич

Березовская Елена Леонидовна

Лебедева Наталья Леонидовна

РЕЦЕНЗИЯ

Проблема скоринга и классификации объектов является ключевой в процессе проектирования систем поддержки принятия управленческих решений в технике, экономике, социологии, экологии. Сложность подобного рода задач обусловлена в значительной мере необходимостью учета неопределенности исходной информации. Фундаментальная методология, позволяющая корректно обрабатывать данные, опирается на рациональное сочетание моделей прикладной статистики, теории нечетких множеств и экспертного оценивания.

Популярная на практике методика агрегирования вероятностных методов и экспертных суждений получила воплощение в сетях нечеткого логического вывода, реализующих модель двукратной свертки информативных признаков. Этот метод, как известно, предусматривает лингвистический анализ одномерных гистограмм распределений факторов и их взвешивание с соответствующими весами важности. Такой мультипликативный подход к формированию агрегированного критерия классификации в значительной мере не учитывает нелинейную топологию исходных данных. В этой связи, представленная в рецензируемой работе модель приближенных рассуждений является актуальной темой исследования, поскольку реализует анализ многомерного распределения вероятностей информативных признаков, что позволяет учесть нелинейный характер топологии признаков.

Научная новизна работы состоит в предложении авторов применять нестандартный подход к синтезу алгоритма нечеткого логического вывода, основанный на анализе ковариационного приближения многомерного интегрального распределения факторов. Архитектура этого приближения позволяет предельно просто вычислять вероятности лингвистических классов «низких», «средних» или «высоких» уровней показателей и оценивать параметры байесовского классификатора с помощью ***EM***-алгоритма. Такая последовательность обработки данных обеспечивает возможность исследования динамических нечетких множеств с помощью адаптивной системы приближенных рассуждений, адекватно функционирующей в нестационарной среде.

Статья содержит практически важные результаты относительно скоринга и классификации объектов наружной рекламы по критерию поступлений в бюджет мегаполиса, а также оценки инвестиционной привлекательности биржевых активов. Эти примеры наглядно демонстрируют инвариантность предложенной методики, моделей и алгоритмов нечеткого логического вывода к содержанию предметной области.

Заключение: рецензируемая работа отвечает требованиям, предъявляемым к научным статьям, и рекомендуется к публикации в журнале «Аудит и финансовый анализ».

Клименко И.С., д.ф.-м.н., профессор кафедры информационных систем в экономике и управлении факультета информационных систем и компьютерных технологий Российского нового университета, г. Москва.

Перейти на ГЛАВНОЕ МЕНЮ