



УДК 621.031: 681.332.5

Л.В. Лабунец<sup>1</sup>  
Н.Л. Лебедева<sup>2</sup>  
М.Ю. Чижов<sup>3</sup>

L.V. Labunets  
N.L. Lebedeva  
M.Yu. Chizhov

## НЕЧЕТКО-МНОЖЕСТВЕННАЯ КЛАСТЕРИЗАЦИЯ ПОСТУПЛЕНИЙ В МОСКОВСКИЙ БЮДЖЕТ ОТ РЫНКА НАРУЖНОЙ РЕКЛАМЫ

## FUZZY CLUSTERING OF THE REVENUE IN MOSCOW'S BUDGET FROM THE OUTDOOR ADVERTISING

*В статье на основе методов интеллектуального анализа данных продемонстрирована взаимосвязь вероятностных и нечетко-множественных подходов к анализу и моделированию поступлений в московский бюджет от объектов наружной рекламы (ОНР). Представлена методика лингвистического анализа распределения поступлений в бюджет в виде аппроксимации гистограммы, сглаженной сдвигом полигауссовой модели. Рассмотрены основные этапы нечеткого логического вывода и кластеризации ОНР по критерию поступлений с помощью байесовского классификатора и метода главных компонент.*

**Ключевые слова:** Data Mining, преобразование Бокса – Кокса, диаграмма рассеяния, квазистатистики, экспоненциально взвешенные оценки Мешалкина, лингвистический анализ гистограммы, EM-алгоритм, байесовские оценки, нечеткий логический вывод, метод главных компонент.

*On the basis of Data Mining methods demonstrated the interrelation of probabilistic and fuzzy-set approaches to the analysis and modeling of revenue in the Moscow's budget from the market of outdoor advertising objects. The technique of linguistic analysis for distribution of the revenue in Moscow's budget in the form of poly-Gaussian approximation of the average shift histogram is presented. The content of the main stages of fuzzy inference and clustering of outdoor advertising objects by the revenue criterion using Bayesian classifier and principal components analysis are examined.*

**Keywords:** Data Mining, Box – Cox transformations, scatter plot, quasi-statistics, Meshalkin's exponentially weighted estimates, linguistic analysis of histogram, EM-algorithm, Bayesian estimators, fuzzy inference, principal component analysis.

<sup>1</sup> Доктор технических наук, старший научный сотрудник, заведующий кафедрой информационных систем в экономике и управлении НОУ ВПО «Российский новый университет», профессор кафедры «Системы обработки информации и управления» МГТУ им. Н.Э. Баумана.

<sup>2</sup> Главный специалист Департамента информационных технологий ОАО «Банк ВТБ».

<sup>3</sup> Аспирант НОУ ВПО «Российский новый университет», заместитель руководителя Департамента средств массовой информации и рекламы города Москвы.

### Введение

Описательные и прогнозные модели, основанные на применении современных методов и алгоритмов Data Mining, позволяют выявить закономерности, содержащие знания, необходимые для оценки финансовых рисков и прогнозирования рыночных тенденций. В докладе проанализирована проблема формирования описательных моделей эффективности объектов наружной рекламы (ОНР). Надежной методической основой такого рода моделей являются

алгоритмы нечеткой кластеризации различных типов ОНР по уровню поступлений в бюджет. Представлен экспертный подход к анализу исходной информации, который позволяет учесть ее неопределенность. Плодотворным в этом смысле является применение систем обработки знаний экспертов с помощью алгоритмов нечеткого логического вывода [1; 2].

### Модель многомерных данных

Трехмерная модель исходных данных представляет собой информационный «куб» [3] со следующими измерениями:

1. **«время»** в виде последовательности месяцев  $n=1,2,\dots,N$  ( $N=48$ ) в период с 2008 по 2011 год включительно;

2. **«округ»** в виде упорядоченной последовательности административных округов г. Москвы  $m=1,2,\dots,M$  ( $M=11$ ), представленных в таблице 1 по состоянию на 2011 год;

3. **«тип ОНР»** в виде упорядоченной последовательности типов наружной рекламы  $k=1,2,\dots,K$  ( $K=30$ ), представленных в таблице 2;

Таблица 1

#### Административные округа

<i>m</i>	Наименование
1.	Общегородской заказ (ОГЗ)
2.	Восточный административный округ (ВАО)
3.	Западный административный округ (ЗАО)
4.	Зеленоградский административный округ (ЗелАО)
5.	Северный административный округ (САО)
6.	Северо-восточный административный округ (СВАО)
7.	Северо-западный административный округ (СЗАО)
8.	Центральный административный округ (ЦАО)
9.	Южный административный округ (ЮАО)
10.	Юго-восточный административный округ (ЮВАО)
11.	Юго-западный административный округ (ЮЗАО)

Кластерная структура данных исследовалась в пространстве Евклида одиннадцати административных округов г. Москвы. Исходными признаками в этом случае являются компоненты вектора столбца

$$\vec{x} = (x_1, \dots, x_M)^T$$

(1)

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

Таблица 2

#### Типы наружной рекламы

<i>k</i>	Наименование
1.	Витраж
2.	Выносная щитовая конструкция (штендер)
3.	Информационная конструкция предприятия и организации по обслуживанию населения на здании
4.	Кронштейн на здании (площадью одной стороны более 1 кв. м и высотой более 1,5 м)
5.	Кронштейн на здании (площадью одной стороны менее 1 кв. м и высотой менее 1,5 м)
6.	Кронштейн на несветовой опоре
7.	Кронштейн на световой опоре
8.	Крышная установка
9.	Маркиза
10.	Настенное панно (площадью более 2 кв. м)
11.	Настенное панно (площадью менее 2 кв. м)
12.	Носимая реклама
13.	Объемно-пространственная установка
14.	Объемно-пространственный объект (отдельно стоящий)
15.	Отдельно стоящая реклама на флагах и других мягких полотнищах
16.	Прочие нетиповые объекты (мусорные контейнеры, телефонные будки и пр.)
17.	Реклама на остановках транспорта
18.	Реклама на строительной сетке
19.	Реклама на флагах и др. мягких полотнищах
20.	Реклама на флагах и других мягких полотнищах на здании, маркизы (площадью более 2 кв. м)
21.	Реклама на флагах и других мягких полотнищах на здании, маркизы (площадью менее 2 кв. м)
22.	Стационарная установка над проезжей частью

23.	Транспарант-перетяжка
24.	Тумба отдельно стоящая
25.	Щит на временном ограждении
26.	Щит на ограждении
27.	Щит отдельно стоящий
28.	Электронный экран на здании
29.	Электронный экран на крыше
30.	Электронный экран отдельно стоящий

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

Разведочный анализ данных выполнялся после предварительного преобразования Бокса – Кокса (БКП) [4] обучающей выборки для каждого административного округа. Величина смещения БКП  $\varepsilon$  выбиралась из условий  $X_{mk} + \varepsilon > 0$ ,  $m=1,2,\dots,M$ ,  $k=1,2,\dots,K$ . Оптимальное значение параметра  $g_m$  БКП удовлетворяли критерию максимума логарифма правдоподобия  $L(g_m)$ . Зависимости логарифма правдоподобия  $L(g_m)$  от параметра  $g_m$  для поступлений в московский бюджет от ОНР за 2008 г. в одиннадцати административных округах и значения смещения  $\varepsilon = 10^{-8}$  иллюстрирует рис. 1.

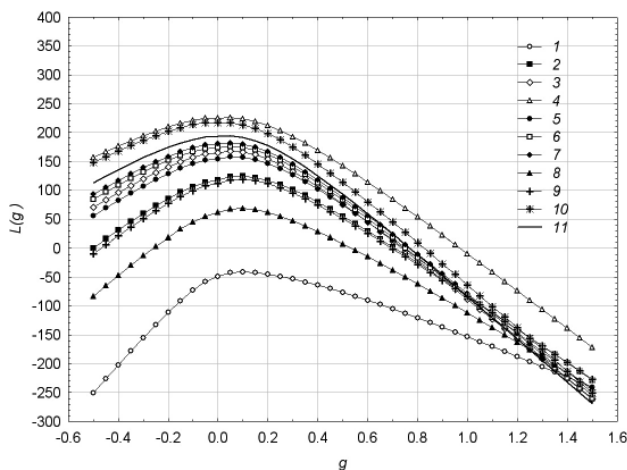


Рис. 1. Зависимость логарифма правдоподобия от параметра преобразования Бокса – Кокса

Результаты расчета показали, что оптимальные оценки параметра БКП лежат в интервале  $-0,15 \leq g_m \leq 0,1$ ,  $m=1,2,\dots,M$ . Графики БКП для указанных выше значений параметра  $g_m$  демонстрирует рис. 2.

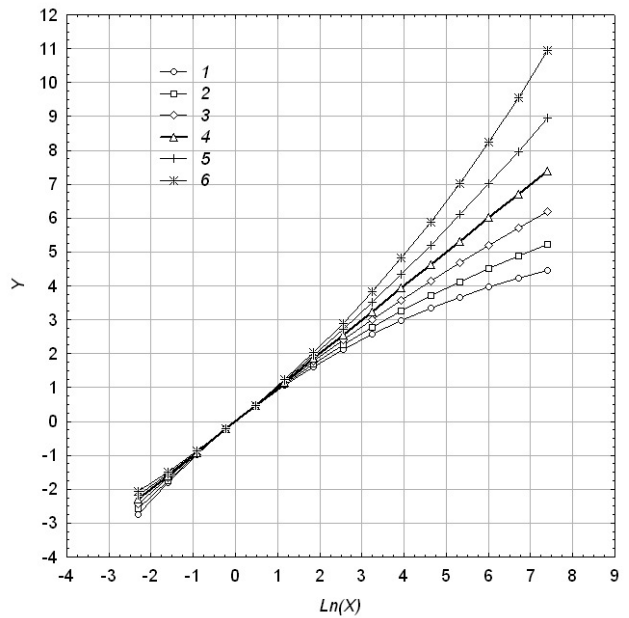


Рис. 2. БКП для различных значений параметра  $g$ :  
 1 –  $g = -0,15$ ; 2 –  $g = -0,1$ ; 3 –  $g = -0,05$ ;  
 4 –  $g = 0$ ; 5 –  $g = 0,05$ ; 6 –  $g = 0,1$

Удобным инструментом разведочного анализа данных является диаграмма рассеяния (ДР), содержащая  $K$  точек в  $M$ -мерном пространстве преобразованных признаков  $\bar{Y}_k = (Y_{1k}, \dots, Y_{Mk})^T$ ,  $k=1,2,\dots,K$ . Здесь и в дальнейшем зависимость от параметра  $g_m$  БКП опускается. Топологию ДР удобно исследовать методом “Grand Tour” динамической визуализации многомерных данных [5]. Ортогональная проекция ДР на одну из гиперплоскостей в пространстве одиннадцати административных округов г. Москвы для поступлений в 2008 г. представлена на рис. 3.

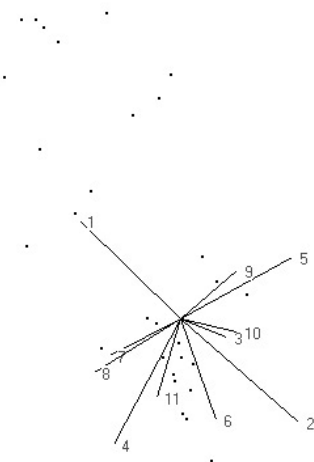


Рис. 3. Ортогональная проекция многомерной ДР на плоскость визуализации

Анализ формы ДР с различных ракурсов свидетельствует о наличии достаточно выраженных генеральных направлений в пространстве преобразованных признаков. Иными словами, имеет место значимая корреляция признаков  $\vec{y} = (y_1, \dots, y_M)^T$ .

**Нечетко-множественный и вероятностный подходы к анализу данных**

Существенной проблемой формирования адекватных описательных статистик является наличие жесткого ограничения малого объема данных. В нашем случае гипотетический объем выборки  $K = 30$  ассоциирован с количеством типов ОНР. Однако значения  $X_{mk} = 0$ , свидетельствующие об отсутствии поступлений в округа от некоторых типов рекламы, приводят к необходимости игнорировать аномально низкие величины признаков. Это снижает фактический объем данных. Известно, что в соответствии с правилом Старгеса (Sturges's rule) [6] объем выборки, необходимый для формирования гистограммы нормального распределения, содержащей как минимум 5 разрядных интервалов, рассчитывают по формуле  $K_{\min} = 2^{s-1} = 2^4 = 16$ .

Рациональное решение проблемы малого объема выборки основано на нечетко-множественном подходе к анализу данных. В рамках такого подхода конструктивным является понятие квазистатистики, введенное в работе [2, с. 28]. В частности, консолидированные по годам выборки поступлений в бюджет от 30 типов ОНР в каждый из одиннадцати округов г. Москвы будем считать достаточными, чтобы с приемлемой степенью достоверности сформировать оценки законов распределения наблюдений.

С другой стороны, нет оснований кардинально разграничивать нечетко-множественный и вероятностный подходы к интеллектуальному анализу данных. В указанном смысле следует отметить результаты теоретического анализа, представленного в разделе 1.4.6 работы [7], интерпретирующего нечеткие множества как «проекции» случайных множеств. Последующее содержание статьи демонстрирует плодотворное сочетание нечетко-множественной и вероятностной методологий.

**Исключение аномальных значений из данных**

Аномальные значения в данных существенно искажают результаты стандартного оценивания основных статистик. Одним из рациональных методов исключения влияния загрязнений выборки являются экспоненциально взвешенные

оценки (ЭВО) характеристик положения и масштаба Л. Д. Мешалкина. Одномерные робастные ЭВО математического ожидания (МО)  $a_m(\lambda)$  и среднего квадрата отклонения (СКО)  $s_m(\lambda)$ ,  $m = 1, 2, \dots, M$  представляют собой решение соответствующей системы нелинейных уравнений [8].

Структура ЭВО обеспечивает автоматическое подавление выбросов в данных, если параметр эффективности статистик  $\lambda > 0$ . Аномально большие значения  $Y_{mk}$  формируют большие расстояния Махаланобиса, поэтому взвешиваются экспоненциальными весами, достаточно малы, чтобы не вносить значимого вклада в результирующую оценку. В работах А.М. Шурьгина [8] показано, что ЭВО являются оценками минимума контраста. Однако снижение эффективности ЭВО повышает их устойчивость к нарушению гипотезы нормальности плотности распределения вероятности (ПРВ).

Процесс сходимости алгоритма расчета ЭВО с параметром  $\lambda = 1$  для характеристик положения и масштаба поступлений от ОНР за 2008 г. в случае общегородского заказа ( $m = 1$ ) иллюстрирует рис. 4

ЭВО удобны для формирования границ, отделяющих кластер «типичных» значений  $Y_{mk}$ ,  $k = 1, 2, \dots, K$  от выбросов. Результаты моделирования показали, что в качестве границ рационально выбирать правило «двух сигм», т.е. величины  $(a_m \pm 2s_m)$ . В соответствии с указанным правилом выполнялось масштабирование данных:

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

$$k = 1, 2, \dots, \tilde{K}_m,$$

очищенных от загрязнений, с целью преобразования их к стандартному интервалу от 0 до 1. Здесь  $Y_m^{(\min)}$  и  $Y_m^{(\max)}$  – робастные оценки точной нижней и верхней границ кластера «типичных» значений;  $\tilde{K}_m$  – количество наблюдений в кластере.

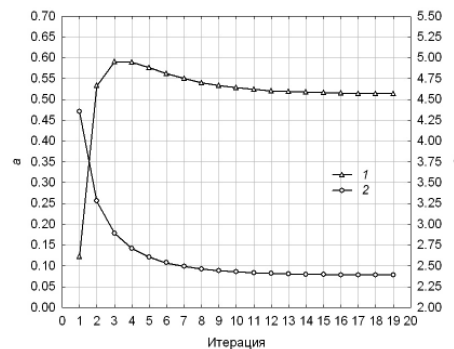


Рис. 4. ЭВО характеристик положения и масштаба: 1 – МО; 2 – СКО

**Гистограмма, сглаженная сдвигом**

В рамках понятия квазистатистики адекватной моделью закона распределения для выборки данных, очищенных от загрязнений, является ASH-оценка. Процедуру сглаживания классической гистограммы окном данных  $W_m(n)$  описывает уравнение дискретной свертки [6]. Ширина разрядных интервалов выбиралась в соответствии с робастной оценкой Фридмана – Дьякониса  $2IQ_m / \sqrt[3]{\tilde{K}_m}$ , где  $IQ_m$  – интерквартильный диапазон нормированных поступлений в бюджет в  $m$ -ом административном округе за фиксированный год. Кроме того, ширину разрядных интервалов классической гистограммы выбирали так, чтобы их количество было не менее пяти.

Окно данных  $W_m(n)$  выбирали из условия  $\sum_{n=1}^{i_m-1} W_m(n) = i_m$ . В этом случае ASH-оценка распределения интегрируема с единицей. Такой нормировке удовлетворяет обобщенное окно вида:

$$W_m(n) = i_m \text{Ker}(n/i_m) / \sum_{l=1}^{i_m-1} \text{Ker}(l/i_m),$$

где  $\text{Ker}(u)$  – положительная четная функция ядра, заданная на стандартном интервале  $[-1; 1]$  и интегрируемая с единицей. Популярные модели ядерных функций приведены в [6, с. 140]. На рис. 5 представлена ASH-оценка распределения нормированных поступлений в бюджет за 2008 г. для общегородского заказа.

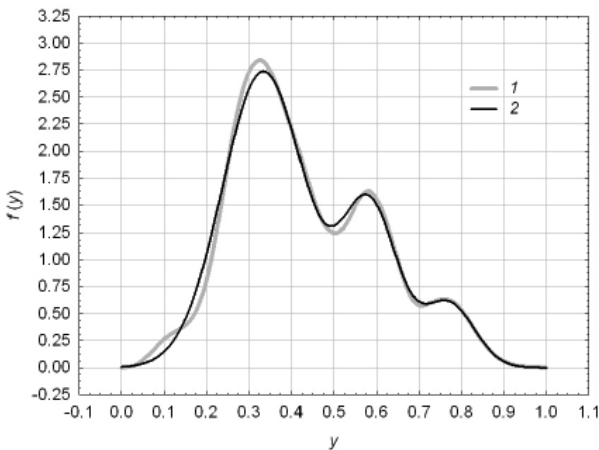


Рис. 5. Оценки распределения нормированных поступлений в бюджет: 1 – ASH; 2 – полигауссовская аппроксимация

В качестве модели функции ядра применялось трижды взвешенное ядро Епанечникова. Объем очищенной выборки данных составил  $\tilde{K}_1 = 26$  наблюдений. Количество разрядных и суженных интервалов выбиралось 6 и 180 соответственно.

**Конечная смесь стандартных распределений**

Теоретически обоснованной методической основой процедуры лингвистического анализа гистограммы [9] является модель конечной смеси стандартных распределений:

$$\hat{f}_m(y|\bar{\theta}_m) = \sum_{n=1}^{N_m} P_{mn} \varphi_{mn}(y|c_{mn}, s_{mn}), \quad (3)$$

$$\sum_{n=1}^{N_m} P_{mn} = 1,$$

где  $P_{mn}$ ,  $n=1, 2, \dots, N_m$  и  $N_m$  – априорные веса лингвистических классов уровней «доходности» ОНР и их количество для  $m$ -го административного округа;  $\varphi_{mn}(y|c_{mn}, s_{mn})$  – парциальные распределения с характеристиками положения  $c_{mn}$  и масштаба  $s_{mn}$  лингвистических классов. В дальнейшем там, где это не вызвано необходимостью, зависимость параметров модели (3) от номера  $m$  административного округа опускаем для сокращения записи.

В рамках представления (3) байесовский подход к формированию параметрической модели лингвистической переменной сводился к описанию функций принадлежности классов апостериорными весами:

$$w_n(y|\bar{\theta}) = P_n \varphi_n(y|c_n, s_n) / \hat{f}(y|\bar{\theta}), \quad n=1, 2, \dots, N \quad (4)$$

Оптимальную оценку вектора параметров  $\bar{\theta} = (P_1, \dots, P_N, c_1, \dots, c_N, s_1, \dots, s_N)$  модели (3) получают с помощью эффективного в вычислительном отношении EM-алгоритма. Стандартным критерием оптимальности параметров модели является функционал правдоподобия Фишера:

$$\bar{\theta}|_{opt} = \arg \max_{\bar{\theta}} \{L(\bar{\theta})\},$$

$$L(\bar{\theta}) = \int_0^1 \ln \{ \hat{f}(y|\bar{\theta}) \} f(y) dy.$$

Рациональным критерием оптимальности является также функционал расстояния Бхаттачария [10]:

$$\bar{\theta}|_{opt} = \arg \min_{\bar{\theta}} \{D(\bar{\theta})\},$$

$$D(\bar{\theta}) = -\ln \left\{ \int_0^1 \sqrt{\hat{f}(y|\bar{\theta}) f(y)} dy \right\}.$$

Решением указанных выше задач условной оптимизации является система нелинейных уравнений, представленная в [10]. Итерационная процедура решения такой системы уравнений [11, с. 1299] позволила аппроксимировать ASH-

оценку распределения нормированных поступлений в бюджет за фиксированный год в текущем административном округе параметрической моделью (3).

Выбор количества классов  $N$  терм множества лингвистической переменной уровней «доходности» ОНР и начальных оценок параметров модели (3) в значительной степени является субъективным и основан на мнении эксперта.

Популярной на практике является полигауссовская модель конечной смеси распределений (3), в которой в качестве парциальных применяются нормальные распределения:

$$\varphi_n(y | c_n, s_n) = \frac{1}{s_n \sqrt{2\pi}} \exp\left\{-\frac{(y - c_n)^2}{2s_n^2}\right\}.$$

Полигауссовская аппроксимация АSH-оценки распределения нормированных поступлений в бюджет за 2008 г. для общегородского заказа представлена на рис. 5. Явно выраженная модальная структура распределения свидетельствует о наличии трех классов уровней «доходности» ОНР – низкой, средней и высокой. Байесовскую модель (4) соответствующей трехуровневой лингвистической переменной демонстрирует рис. 6.

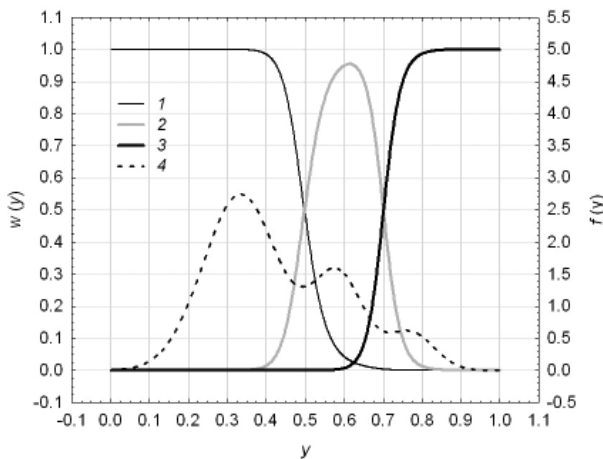


Рис. 6. Байесовская модель лингвистической переменной: 1, 2, 3 – функции принадлежности классов низкой, средней, высокой «доходности» ОНР; 4 – полигауссовская аппроксимация АSH-оценки распределения

Процесс сходимости  $EM$ -алгоритма обучения модели (3) по критерию правдоподобия иллюстрирует рис. 7. Из графиков видно, что сходимость параметров модели к оптимальным значениям достигается практически за 10 итераций как по критерию правдоподобия Фишера  $L(i)$ , так и по критерию расстояния Бхатачария  $D(i)$ .

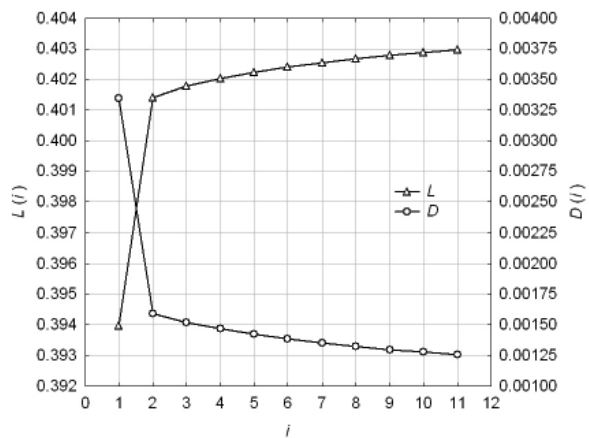


Рис. 7. Сходимость  $EM$ -алгоритма по итерациям  $i$  обучения модели

### Нечеткий логический вывод

Ранжирование ОНР по уровню поступлений в бюджет выполнялось с помощью алгоритмов нечеткого логического вывода. Одним из основных этапов формирования моделей приближенных рассуждений является агрегирование нечеткой информации об объекте анализа. В нашем случае – это композиция средних уровней «доходности» лингвистических классов. Иными словами, комплексный показатель эффективности различных типов ОНР представлен в виде байесовской модели учета неопределенности:

$$y_{mk}^{(AC)} = \sum_{n=1}^{N_m} c_{mn} w_{mn} \left( y_{mk} | \bar{\theta}_m \right), \quad m=1, 2, \dots, M,$$

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

где  $c_{mn}$  и  $y_{mk}$  – средняя «доходность» (узловая точка [2, с. 157; 9])  $n$ -го лингвистического класса и нормированные поступления в городской бюджет от  $k$ -го типа ОНР в  $m$ -ом административном округе за фиксированный год.

Последовательность значений  $y_{mk}^{(AC)}$   $k=1, 2, \dots, K$  упорядоченная по убыванию, ранжирует различные типы ОНР по их «доходности» в  $m$ -ом административном округе за фиксированный год. Кроме того, байесовский классификатор, зачисляющий  $k$ -й тип ОНР в  $n$ -й класс «доходности» по правилу победителя, т.е.

$$w_{mn} \left( y_{mk} | \bar{\theta}_m \right) > w_{mj} \left( y_{mk} | \bar{\theta}_m \right),$$

$$j=1, 2, \dots, N_m, \quad j \neq n$$

реализует лингвистическую кластеризацию ОНР. В качестве примера в таблицах 4–6 представлены результаты такого рода кластеризации по уровням поступлений в бюджет за 2008 г. для общегородского заказа.

## ОНР с высоким уровнем поступлений в городской бюджет

Тип конструкции	$k$	$y_{1k}$	$w_{11}$	$w_{12}$	$w_{13}$	$y_{1k}^{(AC)}$
Щит отдельно стоящий	27	1.23395241	0.000000	0.000000	<b>1.000000</b>	0.770188
Транспарант-перетяжка	23	1.05954391	0.000000	0.000000	<b>1.000000</b>	0.770188
Настенное панно: (площадью более 2 кв. м)	10	1.00000002	0.000000	0.000000	<b>1.000000</b>	0.770188
Крышная установка	8	0.96001369	0.000001	0.000032	<b>0.999967</b>	0.770182
Кронштейн на световой опоре	7	0.89895473	0.000002	0.000278	<b>0.999720</b>	0.770135
Электронный экран отдельно стоящий	30	0.78391064	0.000099	0.025169	<b>0.974732</b>	0.765435
Кронштейн на здании (площадью одной стороны более 1 кв.м и высотой более 1,5 м)	4	0.72123955	0.001604	0.282399	<b>0.715998</b>	0.716648
Тумба отдельно стоящая	24	0.71903489	0.001757	0.302869	<b>0.695374</b>	0.712751
Реклама на остановках транспорта	17	0.70634273	0.002893	0.435171	<b>0.561936</b>	0.687500

## ОНР со средним уровнем поступлений в городской бюджет

Тип конструкции	$k$	$y_{1k}$	$w_{11}$	$w_{12}$	$w_{13}$	$y_{1k}^{(AC)}$
Стационарная установка над проезжей частью	22	0.697657	0.003936	<b>0.533458</b>	0.462607	0.668655
Информационная конструкция предприятия и организации по обслуживанию населения на здании	3	0.607643	0.035579	<b>0.953546</b>	0.010875	0.576263
Реклама на флагах и др. мягких полотнищах	19	0.533415	0.229593	<b>0.770221</b>	0.000186	0.526039
Настенное панно (площадью менее 2 кв. м)	11	0.531127	0.243079	<b>0.756759</b>	0.000162	0.522683
Кронштейн на здании (площадью одной стороны менее 1 кв. м и высотой менее 1,5 м)	5	0.518575	0.328578	<b>0.671349</b>	0.000073	0.501415
Электронный экран на крыше	29	0.515978	0.348624	<b>0.651315</b>	0.000061	0.496430

ОНР с низким уровнем поступлений в городской бюджет

Тип конструкции	$k$	$Y_{ik}$	$W_{11}$	$W_{12}$	$W_{13}$	$Y_{ik}^{(AC)}$
Отдельно стоящая реклама на флагах и других мягких полотнищах	15	0.46345953	<b>0.796002</b>	0.203997	0.000001	0.385219
Въносная щитовая конструкция (штендер)	2	0.43031309	<b>0.942021</b>	0.057979	0.000000	0.348924
Объемно-пространственная установка	13	0.42467768	<b>0.954512</b>	0.045488	0.000000	0.345819
Щит на временном ограждении	25	0.40706334	<b>0.979532</b>	0.020468	0.000000	0.339600
Маркиза	9	0.39303412	<b>0.989563</b>	0.010437	0.000000	0.337107
Кронштейн на несветовой опоре	6	0.38571665	<b>0.992739</b>	0.007261	0.000000	0.336318
Объемно-пространственный объект (отдельно стоящий)	14	0.37392815	<b>0.996016</b>	0.003984	0.000000	0.335503
Реклама на флагах и других мягких полотнищах на здании, маркизы (площадью более 2 кв. м)	20	0.34779001	<b>0.999013</b>	0.000987	0.000000	0.334758
Электронный экран на здании	28	0.33877843	<b>0.999402</b>	0.000598	0.000000	0.334662
Прочие нетиповые объекты (мусорные контейнеры, телефонные будки и пр.)	16	0.20824337	<b>1.000000</b>	0.000000	0.000000	0.334513
Щит на ограждении	26	0.12049077	<b>1.000000</b>	0.000000	0.000000	0.334513
Витраж	1	0.11459128	<b>1.000000</b>	0.000000	0.000000	0.334513
Реклама на флагах и других мягких полотнищах на здании, маркизы (площадью менее 2 кв. м)	21	0.00000006	<b>1.000000</b>	0.000000	0.000000	0.334513
Носимая реклама	12	-0.15619906	<b>1.000000</b>	0.000000	0.000000	0.334513
Реклама на строительной сетке	18	-1.70538556	<b>1.000000</b>	0.000000	0.000000	0.334513



## Заключение

В работе проанализирована тесная взаимосвязь вероятностного и нечетко-множественного подходов к анализу и моделированию эффективности различных типов ОНР по критерию поступлений в московский бюджет. Представлена методика учета фактора неопределенности, обусловленного малым объемом выборки на основе применения фундаментальной процедуры сглаживания экспериментальных данных. В частности, проиллюстрирована эффективность лингвистического анализа распределений исходных признаков в результате комбинации непараметрической модели в виде гистограммы, сглаженной сдвигом, и параметрического описания в виде конечной смеси стандартных плотностей.

## Литература

1. Ведерников В.В. Нечетко-множественное моделирование в анализе и прогнозировании экономических явлений и процессов: исторический аспект [Электронный ресурс] // Проблемы современной экономики. – 2006. – № 1 (17). – URL: [www.m-economy.ru](http://www.m-economy.ru)
2. Недосекин А.О. Нечетко-множественный анализ рисков фондовых инвестиций. – СПб. : Сезам, 2002. – 181 с.
3. Барсегян А.А. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. – 2-е изд., перераб. и доп. – СПб. : БХВ – Петербург, 2007. – 384 с.
4. Box, G.E.P., Cox, D. R. An analysis of transformation // Journal of the Royal Statistical Society. Series B (Methodological). – 1964. – Vol. 26. – № 2. – Pp. 211–252.
5. Buja, A., Cook, D., Asimov, D., Hurley, C. Theory and computational methods for dynamic projections in high-dimensional data visualizations // Journal of Computational and Graphical Statistics. 1999. – Vol. 8. – № 3. – Pp. 1 – 24.
6. Scott D.W. Multivariate Density Estimation: Theory, Practice, and Visualization. – N.Y.: John Wiley & Sons, Inc, 1992. – 317 с.
7. Орлов А. И. Прикладная статистика. – М.: Экзамен, 2006. – 671 с.
8. Шурьгин А.М. Прикладная стохастика: робастность, оценивание, прогноз. – М. : Финансы и статистика, 2000. – 224 с.
9. Недосекин А. О., Фролова С. Н. Лингвистический анализ гистограмм экономических факторов // Вестник ВГУ. – Серия: «Экономика и управление». – 2008. – Выпуск 2. – С. 48–55.
10. Лабунец Л.В. Рандомизация многомерных распределений в метрике Махаланобиса // Радиотехника и электроника. – 2000. – Т. 45. – № 10. – С. 1214–1225.
11. Лабунец Л.В., Лукин Д.С., Червяков А.А. Реконструкция отражательных характеристик 3D-объектов в однопозиционной системе оптической локации // Радиотехника и электроника. – 2012. – Т. 57. – № 12. – С. 1289–1300.