

# ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ МОДЕЛИ

А.Н. Бирюков, канд. экон. наук, доцент,<sup>1</sup>  
г. Стерлитамак

## БАЙЕСОВСКИЙ ПОДХОД К РЕГУЛЯРИЗАЦИИ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ КЛАСТЕРИЗАЦИИ ЭКОНОМИЧЕСКИХ ОБЪЕКТОВ (НА ПРИМЕРЕ МОДЕЛИ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПО НАЛОГОВОМУ РЕГУЛИРОВАНИЮ)

В статье построена нейросетевая модель кластеризации предприятий – налогоплательщиков. Используются для сравнения два метода – самоорганизующихся карт Кохонена и  $k$ -средних. Построение модели изложено на основе теории байесовской кластеризации, разработанной А.С. Шумским. Расчеты показали близость полученных результатов кластеризации по обеим моделям, которые будут полезны в разработке управленческих решений по налоговому регулированию.

**Ключевые слова:** нейросетевая модель (НСМ); байесовский подход; лицо, принимающее решение (ЛПР); метод SOM (сеть Кохонена); метод  $k$ -средних; регуляризация; кластеризация; налоговое администрирование.

### 1. Особенности постановки задач кластеризации в аспекте байесовского подхода к обучению нейросети

Байесовы сети – это весьма революционная технология последних двух десятилетий. Основное их достоинство – универсальное и интуитивно понятное представление моделей, основанных на данных. В противоположность нейронным сетям графические байесовы модели допускают прямую интерпретацию результатов моделирования. Байесова сеть представляет собой оргграф, состоящий из узлов и ребер, отражающих зависимости между переменными. Отсутствие ребра между узлами означает независимость соответствующих переменных.

<sup>1</sup> Бирюков Александр Николаевич – кандидат экономических наук, доцент кафедры экономики и управления Стерлитамакского института физической культуры (филиал) Уральского государственного университета физической культуры; e-mail: guzsa@ufamts.ru.

Для общей постановки задачи кластеризации многомерных данных акцентируем внимание на двух ее особенностях:

1. В данных  $D_{ij}$  в вектор-строках наблюдений  $\{X_i\}$ ,  $i = 1, N$  переменные (кластеризующие признаки) не разделяются на объясняющие и результативные. Соответственно не выделяются шумовая составляющая  $\eta$  в записи:

$$y_i = [y(\bar{x})]_i + \eta_i(\beta), \quad (1)$$

где  $y_i$  – измеренное значение результативного признака в  $i$ -ом наблюдении;  $[y(\bar{x})]_i$  – регулярная часть случайной величины  $Y$  в  $i$ -ом измерении;  $\eta_i(\beta)$  – функция шума;  $\beta$  – амплитуда шума.

2. Наиболее полным статистическим описанием наблюдаемых многомерных данных  $D_{ij}$  является совместная плотность распределения вероятности  $P(x_1, x_2, \dots, x_n)$  в  $n$ -мерном пространстве признаков.

Эти две особенности значительно расширяют возможности различных аналитических построений в рамках байесовского подхода к решению задачи кластеризации. Например, не надо делать никаких предположений о законе распределения шумовой составляющей в измерениях. Если дополнительно ввести допущения о гауссовом распределении векторов  $\{\bar{X}_i\}$ , то можно воспользоваться результатами теории байесовской кластеризации А.С. Шумского [11].

## 2. Байесова кластеризация (модель гауссовой смеси)

Пусть набор эмпирических данных состоит из  $N$   $n$ -мерных векторов:  $D = \{x^{(1)}, \dots, x^{(N)}\}$ . Требуется на основании этой выборки найти наилучшую кластерную модель порождения этих данных  $P(x|h)$ . В решение ищется в виде смеси  $M$  независимых источников данных, каждый из которых относительно прост, а именно вероятность порождения данных источником зависит лишь от расстояния до его центра  $\bar{r}_m$ :

$$P(\bar{x}|h) = \sum_{m=1}^M P(\bar{x}|m)P(m);$$

$$P(\bar{x}|m) \propto \exp(-\beta E(|\bar{x} - \bar{w}_m|)); \quad (2)$$

$$\bar{r} = \bar{x} - \bar{w}_m,$$

где  $E(\bar{r})$  – функция расстояния;  $m$  – номер источника;  $P(m)$  – интенсивность  $m$ -го источника;  $\propto$  – символ, заменяющий слова «изменяется по закону...»;  $\bar{w}_m$  – центр источника в  $n$ -мерном пространстве.

В [11] считается, что все источники, аппроксимирующие крупномасштабные флуктуации плотности распределения векторов  $\{\bar{o}_i\}, i = 1, N$ , т.е. кластеры-гауссовы. Такая модель кластеризации называется гауссовой смесью:

$$E(|\bar{w}_m - \bar{x}|) = \frac{1}{2}(\bar{w}_m - \bar{x})^2. \quad (3)$$

Априорной гипотезе  $h_q$  в байесовском ансамбле нейросетей о порождении данных соответствует набор этих источников и их относительная интенсивность:

$$h = \{w_1, \dots, w_M, P(1), \dots, P(M)\}. \quad (4)$$

Мета-гипотеза  $H$  определяет число  $M$  и дисперсию  $\beta^{-1}$  источников, а в общем случае и конкретный вид функции ошибки  $E$ .

Как обычно, максимизация апостериорной вероятности  $P(h|D, H)$  дает наилучшую гипотезу, а максимизация  $P(D|H)$  – наилучшую мета-модель данных.

### Оптимальная гипотеза

Пусть отсутствуют сведения априорных предпочтений относительно различных гипотез ( $P(h|H) = \text{const}$ ). В этом случае оптимальная гипотеза максимизирует правдоподобие данных:

$$h_{ML} = \arg \max_h \ln P(D|h, H) =$$

$$= \sum_i \ln \sum_m P(\bar{x}_i|m)P(m). \quad (5)$$

Здесь мы встречаемся с ситуацией, когда под логарифмом производится суммирование по неким альтернативам. Такая задача сводится к минимизации «свободной энергии» [11]:

$$F(P, h) = \sum_{m,i} P(m|i) \ln P(m|i) -$$

$$- \sum_{m,i} P(m|i) \ln P(\bar{x}_i, m) =$$

$$= \sum_i \ln \sum_m P(x_i, m) + \sum_{m,i} P(m|i) \ln \frac{P(m|i)}{P(m|x_i)}. \quad (6)$$

Решение для  $P(m|i)$  дается формулой Байеса:

$$P(m|i) = P(m|\bar{x}) = \frac{P(x_i|m)P(m)}{\sum_m P(x_i|m)P(m)} \quad (7)$$

а наилучшая гипотеза максимизацией усредненного по этому распределению логарифма совместной вероятности:

$$\begin{aligned}
 w_m &= \arg \max_{w_m} \sum_{m,n} P(m|i) \ln P(\bar{x}_i, m) = \\
 &= \arg \max_{w_m} \sum_{m,n} P(m|i) \ln P(x_i | m), \\
 P(m) &= \arg \max_{P(m)} \sum_{m,i} P(m|i) \ln P(\bar{x}_i, m) = \\
 &= \arg \max_{P(m)} \sum_{m,i} P(m|i) \ln P(m).
 \end{aligned} \tag{8}$$

Отсюда легко получить:

$$\begin{aligned}
 \sum_n P(m|n) \frac{\partial E(|w_m - x^{(n)}|)}{\partial w_m} &= \\
 &= \sum_n P(m|x^{(n)}) x^{(n)}, \\
 = 0 \Rightarrow w_m &= \frac{\sum_n P(m|x^{(n)}) x^{(n)}}{\sum_n P(m|x^{(n)})}, \\
 P(m) &= \frac{1}{N} \sum_n P(m|n).
 \end{aligned} \tag{9}$$

Иными словами, оптимальная гипотеза находится путем последовательных итераций следующего **EM**-алгоритма:

- **Е шаг:** Фиксируем источники  $\{w_m, P(m)\}$  и находим вероятности принадлежности к ним точек данных:

$$P(m|\bar{x}_i) = \frac{P(m) \exp(-\beta E(|\bar{x}_i - \bar{w}_m|))}{\sum_m P(m) \exp(-\beta E(|\bar{x}_i - \bar{w}_m|))} \tag{10}$$

- **М шаг:** Фиксируем распределение данных по источникам  $P(\overline{\min}_i)$  и находим новые характеристики источников:

$$\bar{w}_m = \frac{\sum_i P(m|\bar{x}_i) x_i}{\sum_i P(m|\bar{x}_i)}. \tag{11}$$

$$P(m) = \frac{1}{N} \sum_n P(m|\bar{x}_i). \tag{12}$$

Повторяем эти итерации до гарантированной сходимости.

Заметим, что в пределе  $\beta \rightarrow 0$  приведенный выше алгоритм совпадает с хорошо известной кластеризацией методом  $k$ -средних [1]. Во-первых, на каждом шаге определяется жесткая привязка точек к своим кластерам:

$$\begin{aligned}
 P(m|\bar{x}_i) &= \delta_{m,m^{(i)}}, \\
 m^{(i)} &= \arg \min_m |\bar{x}_i - \bar{w}_m|
 \end{aligned} \tag{13}$$

и, во-вторых, новые центры кластеров помещаются в центры тяжести принадлежащих им точек:

$$\bar{w}_m = \sum_n \delta_{m,m_i} \bar{x}_i. \tag{14}$$

Гауссовы смеси с конечным  $\beta$  осуществляют мягкую или нечеткую кластеризацию.

### Сколько кластеров в данных?

Хотя при изложении считалось, что число источников в модели известным (и равным  $M$ ), на самом деле этот параметр явным образом не определен. Действительно, описанный выше **EM**-алгоритм допускает слияние источников.

Если в какой-то момент положения двух источников совпадут:  $\bar{w}_m^t = \bar{w}_k^t$  согласно (10)–(12), эти источники сольются, т.е. и на всех последующих итерациях мы получим  $\bar{w}_m^{t+T} = \bar{w}_k^{t+T}$ . Таким образом, реальное число кластеров в модели может существенно отличаться от начального.

В зависимости от значения  $\beta$  и конкретной конфигурации данных некоторые источники будут притягиваться друг к другу и сливаться. Численные эксперименты показывают, что флуктуация плотности данных приводит к слиянию источников, оказавшихся в ее окрестности, если их радиус  $\beta^{-1/2}$  превышает масштаб этой флуктуации. Иными словами, такой масштаб неод-

нородностей в данных модель уже не различает.

Таким образом, чем больше  $\beta$ , т.е. меньше радиус взаимодействия источников, тем больше возможное число кластеров в модели. Напротив, достаточно малые  $\beta$ , такие, что характерный радиус источников превышает масштаб разброса всех данных, приводят к слиянию источников в один большой кластер.

Можно взглянуть на эту ситуацию следующим образом. Допустим, что мы смотрим на имеющееся распределение данных с расстояния, пропорционального  $\beta^{-1/2}$ . На большом расстоянии все данные сливаются в одно пятно. По мере нашего приближения становятся различимы все новые детали неоднородностей в кластерной структуре данных, и число различимых кластеров возрастает. В пределе  $\beta \rightarrow 0$  становятся различимы отдельные точки данных, т.е. «равновесное» число кластеров стремится к числу примеров, хотя их реальное количество будет, естественно, ограничено начальным значением.

Оптимальная модель данных соответствует в этой аналогии оптимальному масштабу, с которого структура данных видна наилучшим образом, т.е. когда мелкомасштабные флуктуации не мешают разглядеть общую картину.

В двумерном или трехмерном случае человек легко находит наиболее информативный масштаб округления данных. Однако для многомерных данных определение «оптимального разрешения» модели уже не столь тривиально. Сколько кластеров в данных на самом деле? Байесовский подход позволяет дать ответ на этот непростой вопрос.

### Оптимальная модель

Оптимальный параметр  $\beta$  и соответствующее ему число кластеров  $M$  определяется максимизацией *Evidence*, которую можно вычислить приближенно, используя метод перевала [11]:

$$\begin{aligned} P(D|\beta) &= \int dh P(D|h, \beta) = \\ &= \int dh \exp(\ln P(D|h, \beta)) \approx \\ &\approx \int \approx (2\hbar)^{|h|/2} |A|^{-1/2} P(D|h_{ML}, \beta). \end{aligned} \quad (15)$$

Здесь  $|h| = Mn + M$  – размерность пространства гипотез, а  $|A|$  – детерминант матрицы вторых производных в точке максимума *Likelihood*:  $A = -\nabla\nabla \ln P(D|h, \beta)|_{h_{ML}}$ . Можно убедиться, что эта матрица в нашем случае диагональна и ее детерминант равен

$$|A| = (\beta N)^{Md} N^M \left( \prod_m P(m) \right)^{d-1}. \quad (16)$$

С учетом этого факта, приравнивая нулю производную

$$\begin{aligned} 0 &= \frac{\partial}{\partial \beta} \ln P(D|\beta) \approx \frac{\partial}{\partial \beta} \ln P(D|h_{ML}, \beta) - \\ &- \frac{1}{2} \frac{\partial}{\partial \beta} \ln |A|, \end{aligned} \quad (17)$$

получим следующее выражение для оптимальной  $\beta$ :

$$\beta_{ML}^{-1} = -\frac{1}{(N-M)d} \sum_{m,i} P(m|\bar{x}_i) (\bar{w}_m - \bar{x}_i)^2. \quad (18)$$

Заметим, что если бы мы определяли  $\beta$  из условия максимума *Likelihood*, то получили бы аналогичный результат только без уменьшения числа данных на число источников  $M$ . Что касается самого значения *Evidence* для оптимальной модели, то его главные члены, растущие с числом данных, даются следующим выражением:

$$\begin{aligned} \ln P(D|\beta_{ML}) &\approx \frac{Nd}{2} \ln \beta_{ML} - \\ &- \sum_{m,n} P(m|\bar{x}_i) \ln \frac{P(m|\bar{x}_i)}{P(m)} - \frac{M(d+1)}{2} \ln N. \end{aligned} \quad (19)$$

Как видим, качество модели возрастает с ростом  $\beta_{ML}$ , чему соответствует уменьшение масштаба ошибки. Таким образом, первый член способствует увеличению числа кластеров. Однако вто-

рой и третий члены, напротив, «штрафуют» излишне сложные модели с большим числом кластеров. *Оптимальная модель представляет собой баланс сложности и точности воспроизведения ее структуры данных.* При наличии нескольких вариантов кластеризации предпочтение следует отдавать той модели (мета-гипотезе  $H$ ), которой соответствует наибольшая *Evidence*.

Заметим, что при выводе (19) мы считали слившиеся источники единым кластером, т.е. как число  $M$  во всех формулах, так и все распределения вероятностей по кластерам соответствуют различающимся между собой источникам/

### 3. Кластеризация налогоплательщиков при выработке управленческих решений по налоговому регулированию

Задача оперативного и гибкого налогового администрирования является частью государственного налогового менеджмента. От эффективности решения этой задачи зависит, с одной стороны, увеличение наполнения бюджетов всех уровней, а с другой стороны, сохранение экономической стабильности предприятия-налогоплательщиков как крупных корпоративных, так и средних.

Выработка правильных управленческих решений в указанных задачах требует знания, достаточно достоверного (т.е. не искаженного сознательно) финансово-экономического состояния налогоплательщиков. Источником информации здесь может служить массив налоговых деклараций за некоторый предшествующий временной период по интересующему аналитика множеству объектов налогообложения.

В принципе, объективную для налогового регулирования информацию могли бы дать выездные налоговые проверки. Однако на практике такой вид получения информации для этих целей неэффекти-

вен, что связано с рядом его недостатков и ограничений, к которым относят:

- законодательные ограничения периодичности проведения выездных проверок (не чаще одного раза в три года);
- трудоемкость (1–2 месяца работы выездной бригады налоговых инспекторов);
- возможные искажения истинного финансово-экономического положения налогоплательщика во всем объеме первичной документации.

Примером тому служат многочисленные разбирательства дел в арбитражных судах.

Отмеченные недостатки не позволяют оперативно и с малыми затратами получать достаточно достоверную информацию о финансово-экономическом состоянии налогоплательщиков и тем более об их группах (кластерах).

В данной статье предлагается мало-затратная иерархическая схема получения необходимой информации для налогового администрирования, в частности, информации о субъектах налогообложения на основе математического моделирования, использующего только один источник – исторические данные по налоговым декларациям (рис. 1), разумеется, с пересчетом показателей по известным детерминированным формулам в области финансового менеджмента предприятий.

#### 3.1. Спектр моделей налогового администрирования регионально-го уровня

В проблеме совершенствования системы налогового администрирования в России важной задачей является использование новейших информационных технологий и телекоммуникаций. На мезо- и микроуровне налоговой системы эта задача предусматривает, в частности, построения математических мо-

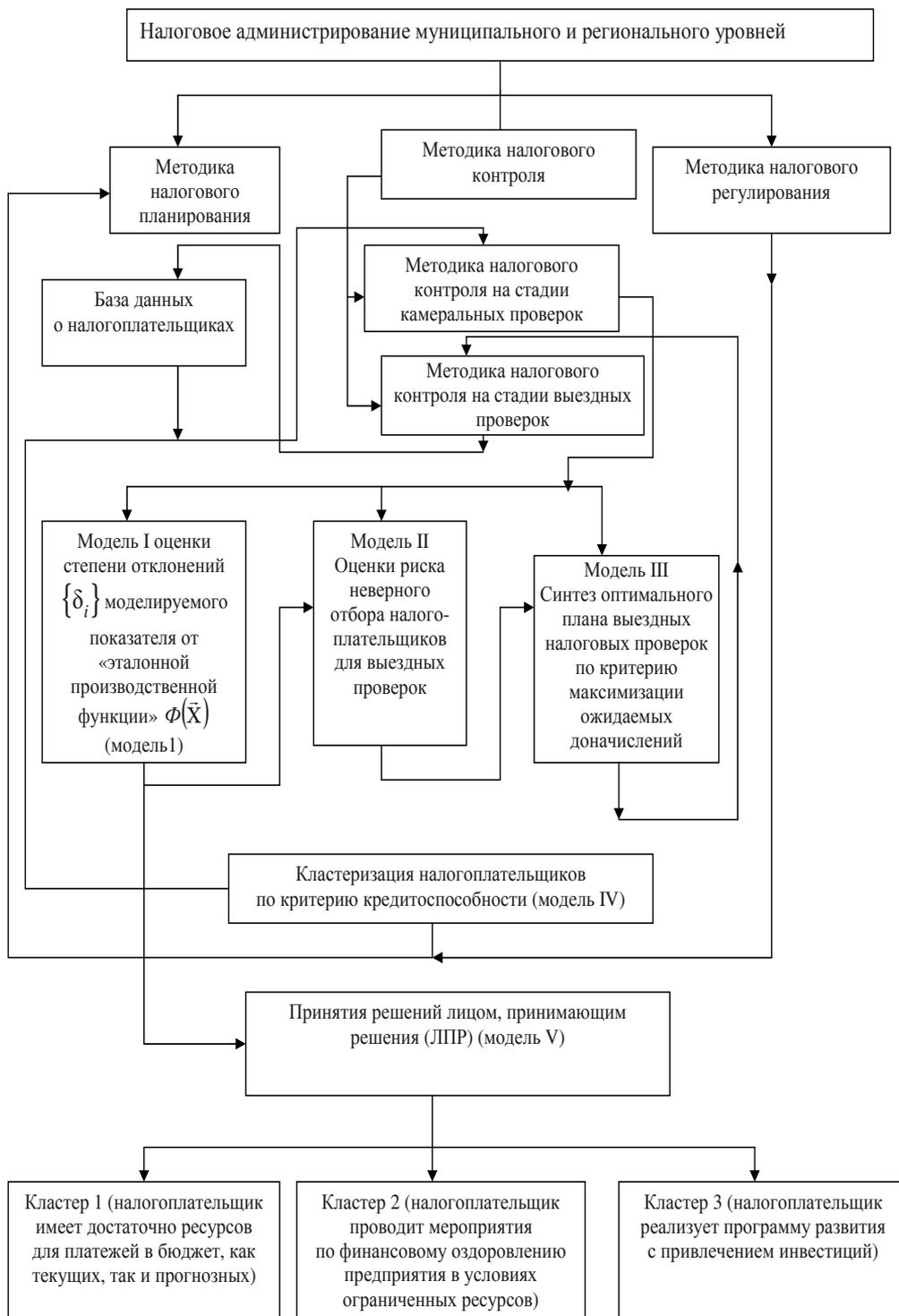


Рис. 1. Спектр подзадач и математических моделей налогового администрирования

делей поддержки принятия управленческих решений [12, 13]. Спектр этих задач показан на схеме рис. 1. Изложим коротко взаимодействие всех пяти математических моделей и иерархической схемы рис. 1. В модели I используется нейросетевой базис [5, 9, 10]. На вход модели подаются вектор-строки примеров из налоговых деклараций, то есть кортежи

$\langle \Phi_i, \vec{x}_i \rangle, i = \overline{1, N}$ , где  $\Phi(\vec{x})$  – моделируемый показатель, имеющий смысл виртуального «объема продукции» или обобщенной «производственной функции» для кластера примерно однородных налогоплательщиков;  $\vec{x}$  – вектор входных факторов;  $i$  – текущий номер наблюдений в статистической выборке;  $N$  – количество примеров. Модель I представляет собой адаптивную аппроксимационную модель, восстанавливающую многомерную нелинейную «производственную функцию»  $\Phi(\vec{x})$ , «зашитую» в данных. Такая задача восстановления (или интерпретации) относится к классу обратных задач и требует специальных мер регуляризации.

Выходом модели I является множество относительных отклонений

$\delta_j = |(\Phi_j - \hat{\Phi}_j) / \Phi_j, j = \overline{1, N}$ , где  $\Phi_j, \hat{\Phi}_j$  – экспериментальные и рассчитанные нейросетью значения  $\Phi$  соответственно. Модель I изложена достаточно подробно в [2] и в данной статье не рассматривается.

Модель II оценки риска ошибок неверного отбора налогоплательщиков в замкнутой форме в виде вероятностной модели (коэффициента риска) получена в [7] в предложении о нормальном законе распределения плотности вероятности отклонений  $\{\delta_j\}$ . Поскольку нормальный закон распределения для отклонений  $\{\delta_j\}$ , не всегда выполняется на практике, исследования по оценке риска отбора продолжены в [2] с использовани-

ем фрактальных методов и методов нечеткой логики, что является предметом отдельной публикации.

Модель III синтеза оптимального плана отбора налогоплательщиков для выездных проверок представляет собой модель задачи линейного (либо нелинейного) программирования, подробно описанную в [7], поэтому останавливаться на ней не будем.

Модель V поддержки принятия решений лицом, принимающим решения (ЛПР), в настоящее время проработана только на концептуальном уровне. На рис. 1 показан пример использования результатов кластеризации налогоплательщиков для поддержки принятия решения ЛПР. Пусть в зависимости от финансового состояния налогоплательщиков и имеющихся у них ресурсов выделено три кластера по уровню кредитоспособности налогоплательщика. Тогда возможными решениями по налоговому регулированию могут быть:

- для кластера 1 – уплата текущих налогов и постепенное увеличение сумм погашения задолженности;
- для кластера 2 – уплата текущих налоговых платежей и «замораживание» выплат недоимок;
- для кластера 3 – полная или частичная отсрочка по текущим платежам и «замораживание» недоимок.

Предметом подробного исследования в данной статье является модель IV кластеризации налогоплательщиков по критерию кредитоспособности. Следует отметить, что в теоретическом плане рассматриваемая задача для модели IV тесно примыкает к более широкому классу задач оценки кредитного риска [4].

Подобная задача встречается в экономике довольно часто. Приведем некоторые примеры. В настоящее время в России для ряда регионов выде-

лены антикризисные фонды (помимо целевых субсидий для градообразующих и оборонных предприятий). При распределении антикризисных фондов возникает задача объективного и справедливого распределения суммы кредитов с учетом эффективности работы предприятия-заемщика и риска невозврата кредита в полном объеме. С данной задачей тесно связана подзадача расчета допустимого размера кредита для данного предприятия. Другой пример – это задача распределения фонда материального развития (ФМР), то есть части субвенций, передаваемой из бюджета РФ ее субъектам, между муниципальными учреждениями (МУ) [6]. Третий пример – задача входного финансового контроля для корпоративных заемщиков, которая решалась Голландским национальным банком при предоставлении кредитов польским банкам и финансовым корпорациям [4]. Здесь с помощью нейросетевой модели решалась задача кластеризации корпоративных заемщиков с извлечением знаний экспертов в области кредитного дела из сделанных ими оценок по группе из 44 качественных (переменных Арженти) и финансовых количественных переменных. Заемщики разбивались на 3 группы:

- «группа 0» – предприятия выживут, по крайней мере, в течение года;
- «группа 1» – пограничные случаи;
- «группа 2» – назревает банкротство в ближайшее время.

Четвертый пример – прогноз финансовой устойчивости и конкурентоспособности предприятия при выработке управленческих решений на перспективу руководителем предприятия. Примеры рассматриваемого класса можно было бы продолжить. Предметом исследования в статье является общая часть (ядро) данного класса задач оценки кредитного риска – модель кластеризации пред-

приятий в том или ином аспекте. Подчеркнем также специфику источника информации: в нашем случае – это исторические данные по налоговым декларациям совокупности налогоплательщиков.

### 3.2. Моделирования на основе самоорганизующихся карт Кохонена

Для построения обобщенной производственной функции  $\Phi(\vec{X}(t), t)$  используем морфологический метод. Его идея состоит в том, что в качестве «осей» морфологического ящика экспертно выделяются аддитивные свертки показателей (признаков), которые затем сворачиваются мультипликативно. Конкретно в нашей задаче кластеризации налогоплательщиков главная полезная функция (обобщенный критерий эффективности)  $\Phi(\vec{X}(t), t)$  имеет вид:

$$\Phi(\vec{X}, t) = \Phi_1(\vec{X}, t) \cdot \Phi_2(\vec{X}, t) \cdot \Phi_3(\vec{X}, t) \cdot \Phi_4(\vec{X}, t), \quad (20)$$

где  $\vec{X}$  – вектор входных факторов;  $t$  – время;  $\Phi_1(\vec{X}, t)$  – показатель, характеризующий эффективность производственного процесса;  $\Phi_2(\vec{X}, t)$  – показатель финансового состояния предприятия;  $\Phi_3(\vec{X}, t)$  – организационный структурный показатель;  $\Phi_4(\vec{X}, t)$  – показатель состояния внешней экономической среды.

Перемножение частных критериев  $\{\Phi_k\}$  в (20) отражает взаимосвязь всех показателей в общей оценке  $\Phi$  эффективности работы предприятия: этот показатель будет максимальным, если одновременно будут максимальными все агрегаты  $\Phi_1, \dots, \Phi_p$ .

Основное достоинство морфологического подхода – это усиление корреляционной связи выходной переменной моделирования  $\Phi$  с вектором экзогенных переменных  $\vec{X}(t)$  по сравнению с частными показателями и аддитивными свертками:

$$\{\Phi_k(\vec{X}, t)\}, k = \overline{1, 4}.$$

При конструировании главной полезной функции  $\Phi(\bar{X}(t))$  и частных критериев  $\Phi_1(\bar{X}(t)), \Phi_2(\bar{X}(t)), \Phi_3(\bar{X}(t)), \Phi_4(\bar{X}(t))$  будем придерживаться следующих правил.

1. В аддитивных свертках  $\Phi_1, \Phi_2, \Phi_3, \Phi_4$ , если суммируемые члены разнородны (имеют разные единицы измерения), эти члены должны быть предварительно приведены к безразмерной форме, причем после такого приведения (нормировки) все члены весовых сумм должны стать однонаправленными, т.е. улучшению кредитоспособности моделируемого объекта должно соответствовать увеличение (либо уменьшение) одновременно всех членов сумм.

2. Аналогичные требования безразмерности и однонаправленности должны соблюдаться и для частных критериев  $\Phi_1, \Phi_2, \Phi_3, \Phi_4$ , перемножаемых в мультипликативной свертке  $\Phi$  в (20).

3. В аддитивных свертках все суммируемые члены должны быть примерно одного порядка. Соответственно чувствительности функций (сверток  $\Phi_k(\Phi_{kr})$ ) по переменным  $\Phi_{kr}$ , определяемые как частные производные ( $\partial\Phi_k / \partial\Phi_{kr}$ ), будут одного порядка. Если этого правила не соблюдать, то некоторые свертки  $\{\Phi_k(\Phi_{kr})\}$  могут оказаться нечувствительными к вариации некоторых своих аргументов  $\{\Phi_{kr}\}$  членов свертки. Данное требование не распространяется на сами свертки  $\Phi_1, \Phi_2, \Phi_3, \Phi_4$  в выражении обобщенного критерия  $\Phi$  (20), то есть сомножители могут иметь разный порядок в произведении. Для удобства вычислений всегда есть возможность умножить  $\Phi$  на некоторое масштабирующее число, одинаковое для всех строк таблицы базы данных – это не изменит результатов кластеризации и других статистических оценок.

Отметим, что «платой» за компрессию данных путем образования агрега-

тов  $\Phi_1, \Phi_2, \Phi_3, \Phi_4$  и затем главной полезной функции  $\Phi(\bar{X}(t))$  в (20) является неизбежная потеря исходной интерпретируемости, которой обладают суммируемые члены аддитивных свертки. Вместо этой интерпретируемости возникает новая, обобщенная интерпретируемость:  $\Phi_1, \Phi_2, \Phi_3, \Phi_4$  и  $\Phi$  является числовыми мерами одного главного свойства кластеризуемых налогоплательщиков – их кредитоспособности в аспектах соответственно эффективности производства, финансового состояния, организационной структуры, внешней среды и в целом по перечисленным аспектам.

Теперь детализируем показатели (характеристики), к примеру, торговых предприятий-налогоплательщиков. Агрегат  $\Phi_1$  определим как линейную свертку из показателей эффективности производства и ресурсных переменных:

$$\Phi_1 = \sum_{r=1}^{q_i} C_r \Phi_{1r}, C_r \geq 0, \sum_{r=1}^{q_i} C_r = 1. \quad (21)$$

Здесь рассмотрим:

$\Phi_1$  – рентабельность производства (чистая прибыль, приходящаяся на единицу затрат).

Поскольку рассматриваемый подход к кластеризации на основе предлагаемого морфологического принципа требует описания обоих указанных методов во всей полноте их алгоритмов, нетривиальных для понимания, указанные методы далее изложены подробно: метод карт Кохонена и метод  $k$ -средних.

### 3.3. Метод кластеризации $k$ -средних

Название метода связано с тем, что здесь число образуемых кластеров задается априори и равно  $k$ . Термин «средних» означает, что алгоритм разбиения точек в многомерном пространстве на кластеры определяется через центры тяжести кластеров, т.е. по средним координатам элементов. При этом в

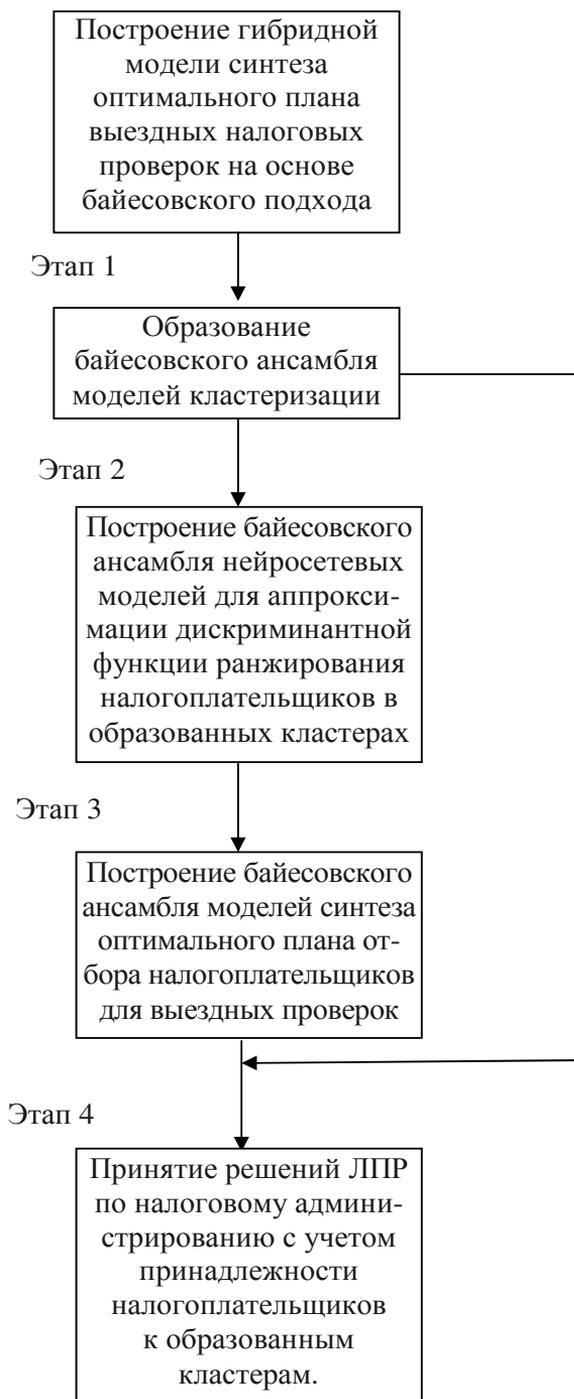


Рис. 2. Иерархия моделей поддержки принятия решений по налоговому администрированию на основе байесовского подхода

качестве меры удаления точек (векторов многомерных наблюдений)  $\bar{X}_q, \bar{X}_p$  друг от друга и от центра кластера используется евклидово расстояние:

$$d(\bar{X}_p, \bar{X}_q) = \sqrt{\sum_{j=1}^n (x_{pj} - x_{qj})^2};$$

$$p, q = 1, 2, \dots, N;$$

$$j = 1, 2, \dots, n, \quad (22)$$

где  $j$  – номер компоненты вектора  $\bar{X} = (x_1, x_2, x_3, \dots, x_n)$  в признаковом пространстве.

Смысл алгоритма  $k$ -средних состоит в последовательном уточнении эталонных точек  $\{\bar{E}_q^{(v)}\}$ , где  $v$  – номер итерации,  $v = 0, 1, 2, \dots, k$  – число эталонов, равное числу образуемых кластеров. Эталоны играют роль зародышей (центров) будущих кластеров. На начальном шаге итераций ( $v = 0$ ) в качестве эталонов  $\{\bar{E}_q^{(0)}\}, q = 1, k$  можно принять случайно выбранные  $k$  точек (векторов) в  $n$ -мерном признаковом пространстве  $\bar{X}$ . Смысл уточнения состоит в том, чтобы на каждом очередном шаге итераций  $v = 0, 1, 2, \dots$  расстояния между точками и центром внутри кластера по выбранной числовой мере, например евклидовой (22), сокращались, а расстояния между образуемыми кластерами по одной из числовых мер междуклассового расстояния – увеличивались [1].

Обычно в программных продуктах пользователю предлагается несколько типов междуклассовых расстояний: «ближнего соседа», по «центрам тяжести», по принципу «средней связи», степенного среднего, предложенного А.Н. Колмогоровым и др.

В алгоритме  $k$ -средних на каждом шаге производится пересчет приписываемых эталонам  $\{\bar{E}_q^{(v)}\}, q = 1, k$  весов:

$$\Omega^{(v)} = \{\omega_1^{(v)}, \omega_2^{(v)}, \dots, \omega_q^{(v)}, \omega_k^{(v)}\}.$$

При этом на нулевом шаге веса всех выбранных эталонов равны 1, т.е.  $\omega_q^{(0)} = 1, q = 1, k$ .

Затем на 1-ом шаге «извлекается» из базы данных очередная точка  $\bar{X}_{k+1}$  и выясняется, к какому из эталонов  $\{\bar{E}_q^{(0)}\}$  она оказалась ближе всего по мере (22). Именно этот, самый близкий к  $\bar{X}_{k+1}$  эталон заменяется эталоном, определяемым как центр тяжести старого эталона и присоединенной к нему точки  $\bar{X}_{k+1}$  (с увеличением на единицу соответствующего ему веса), а все другие эталоны остаются неизменными (с прежними весами) и т.д. Таким образом, пересчет эталонов и весов на  $v$ -ом шаге итераций, т.е. при извлечении из данных очередной точки  $\bar{X}_i = \bar{X}_{k+v}$ , происходит по следующему правилу:

$$\bar{E}_q^{(v)} = \begin{cases} \omega_q^{(v-1)} \cdot \bar{E}_q^{(v-1)} + X_{k+v}, & \text{если} \\ (X_{k+v}, \bar{E}_q^{(v-1)}) = \min_{1 \leq q \leq k} (\bar{X}_{k+v}, \bar{E}_q^{(v-1)}) \\ \bar{E}_q^{(v-1)} & \text{– в противоположном случае.} \end{cases} \quad (23)$$

$$\omega_q^{(v)} = \begin{cases} \omega_q^{(v-1)} + 1 & \text{если } (X_{k+v}, \bar{E}_q^{(v-1)}) = \\ \min_{1 \leq q \leq k} (\bar{X}_{k+v}, \bar{E}_q^{(v-1)}) \\ \omega_q^{(v-1)} & \text{– в противоположном случае.} \end{cases}$$

Если при этом обнаруживается несколько по  $q$  одинаковых минимальных расстояний  $\rho(\bar{X}_i = \bar{X}_{k+v}, \bar{E}_q^{(v-1)})$ , то можно условиться относить предъявляемую точку  $\bar{X}_i = \bar{X}_{k+v}$  к эталону с наименьшим порядковым номером.

Недостатком данного алгоритма является зависимость результата кластеризации от начального положения эталонов  $\{E_q^{(0)}, q = 1, k$  в  $n$ -мерном признаковом пространстве  $\{X_j\}, j = 1, n$ . Однако этот недостаток слабо проявляется даже при весьма широких ограничениях на природу исследуемых наблюдений, если соблюдены два условия:

а) при достаточно большом объеме классифицируемых наблюдений ( $i = 1, N$ );

б) при большом числе итераций  $v$ .

Другими словами, условия а) и б) обеспечивают сходимость эталонов  $\{\bar{E}_q^{(v)}\}$  в определенном смысле (по выбранной числовой мере качества кластеризации) к некоторому пределу  $\{\bar{E}_q^*\}$  при  $v \rightarrow \infty$ . В этом случае пересчет эталонных точек (центров образованных кластеров) не требуется.

Если же в какой-то конкретной задаче алгоритм «не успел добраться» до практически устойчивых (по  $v$ ) значений эталонов  $\{\bar{E}_q^{(v)}\}$ , то можно воспользоваться одним из приемов:

1) либо «зациклить» алгоритм, прогоняя его после рассмотрения последней точки  $\bar{X}_{k+(N-k)}$  и снова предъявляя уже полученным эталонам все точки данных  $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_N$  и т.д.;

2) либо произвести многократное повторение алгоритма, используя в качестве начальных эталонов  $\{\bar{E}_q^{(0)}\}$  различные комбинации из  $k$ -точек исследуемой совокупности и выбрать наиболее повторяющийся финальный эталон  $\{\bar{E}_i^{(N-k)}\}$ .

Теперь остановимся на оценке качества проведенной кластеризации при заданном числе кластеров. Пусть исследователем уже выбрана метрика  $d$  в пространстве признаков  $\{X_j\}$ ,  $j = 1, n$  и пусть  $S = (S_1, S_2, \dots, S_\ell, \dots, S_k)$  – некоторое фиксированное разбиение наблюдений  $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_N$  на заданное число  $k$  кластеров. За функционалы качества кластеризации часто берутся следующие характеристики:

- Взвешенная сумма внутрикластерных дисперсий, где центром

рассеяния служит «центр тяжести» кластера  $\bar{X}(1)$ :

$$Q_1(S) = \sum_{l=1}^k \sum_{\bar{X}_q \in S_l} d^2(\bar{X}_q, \bar{X}(l)). \quad (24)$$

- Сумма попарных внутрикластерных расстояний между элементами  $\bar{X}_p, \bar{X}_q$ :

$$Q_2(S) = \sum_{l=1}^k \sum_{\bar{X}_p, \bar{X}_q \in S_l} d^2(\bar{X}_p, \bar{X}_q) \quad (25)$$

либо:

$$Q_2(S) = \sum_{l=1}^k \frac{1}{N_l} \sum_{\bar{X}_p, \bar{X}_q \in S_l} d^2(\bar{X}_p, \bar{X}_q). \quad (26)$$

- Обобщенная внутрикластерная дисперсия

$$Q_3(S) = \det \left( \sum_{l=1}^k N_l \hat{\sum}_l \right), \quad (27)$$

где  $\det(\bullet)$  – символ определителя матрицы.

$\hat{\sum}_l$  – выборочная ковариационная матрица по наблюдениям, попавшим в какой-то один кластер  $S_\ell$ ; элементы  $\hat{\sigma}_{qp}(l)$  выборочной ковариационной матрицы  $\hat{\sum}_l$  кластера  $S_\ell$  подсчитываются по формуле

$$\hat{\sigma}_q(l) = \frac{1}{N_X} \sum_{\bar{X} \in S} (x_{pj} - \bar{x}_j(l))(x_{qj} - \bar{x}_j(l)), \quad (28)$$

$$j = 1, 2, \dots, n,$$

где  $X_{pj}$  –  $j$ -ая компонента многомерного наблюдения (вектора)  $\bar{X}_p$ ;  $\bar{x}_j(l)$  – среднее значение  $j$ -ой компоненты по всем точкам в  $\ell$ -ом кластере.

- Мультипликативная обобщенная внутриклассовая дисперсия:

$$Q_4(S) = \prod_{l=1}^k \left( \det \hat{\sum}_l \right)^{N_l}. \quad (29)$$

Как видно из формулы (27), обобщенная внутриклассовая дисперсия

$Q_3(S)$  является одной из характеристик рассеивания многомерных наблюдений одного фиксированного  $\ell$ -го кластера около своего «центра тяжести». При этом функционал  $Q_3(S)$  является среднеарифметической (по всем классам) характеристикой обобщенной внутриклассовой дисперсии, в то время как функционал  $Q_4(S)$  в (29) пропорционален средней геометрической характеристике тех же величин.

### 3.4. Пример решения задачи кластеризации налогоплательщиков

В качестве примера рассмотрена задача кластеризации для группы из 24 торговых предприятий – налогоплательщиков.

Данные являются реальными и опубликованы в монографиях [3, 7] (выборка  $Z^{\text{II}}$ , номера торговых предприятий закодированы). Прикладная цель решения задачи кластеризации – это группировка совокупности налогоплательщиков по степени их кредитоспособности, т.е. получения модели поддержки принятия решений ЛПР по налоговому регулированию (см. модель V на рис.1).

Теоретический аспект примера задачи кластеризации – это количественная апробация предложенной на основе морфологического принципа концепции агрегирования переменных с использованием реальных данных деклараций налогоплательщиков. Важно, что в реальных, сильно зашумленных данных зашумление (вплоть до сознательного искажения) носит естественный характер, свойственный для моделируемой группы объектов (см. табл.1). Введены обозначения:  $X_1$  – первоначальная стоимость основных средств, тыс. руб.;  $X_2$  – износ (амортизационные отчисления) за квартал, тыс. руб.;  $X_3$  – оборотные активы, тыс.руб.;  $X_4$  – запасы, тыс.руб.;  $X_5$  – среднесписочная численность работающих, человек;  $X_6$  – дебиторская задолжен-

ность, тыс. руб.;  $X_8$  – себестоимость реализации товаров за квартал, тыс. руб.;  $Y$  – выручка предприятия, тыс.руб.

Кластеризация проведена по пяти переменным: агрегатам (сверткам)  $\Phi_1, \Phi_2, \Phi_3, \Phi_4$  и главной полезной функции  $\Phi$  согласно (20). Последняя позволяет наглядно и просто интерпретировать результаты расчета (см. ниже).

В качестве показателей  $\{\Phi_{kz}\}$ , входящих в свертки  $\Phi_1, \Phi_2, \Phi_3, \Phi_4$ , использовались:

$\Phi_{11}$  – рентабельность производства:

$$\Phi_{11} = \frac{Y - X_8}{X_8}, \quad (30)$$

где в числителе стоит чистая прибыль.

$\Phi_{13}$  – удельные основные средства:

$$\Phi_{13} = X_1 / X_8; \quad (31)$$

$\Phi_{21}$  – чистая маржа:

$$\Phi_{21} = (Y - X_8) / Y. \quad (32)$$

$\Phi_{24}$  – индекс роста объема продаж:

$$\Phi_{24} = y_{i+1} / \left( \frac{1}{n_i} \sum_{i=1}^i y_i \right), \quad (33)$$

$$i = 1, 2, \dots, 12.$$

Здесь  $i$  – текущий номер наблюдений, упорядоченных для каждого предприятия в порядке возрастания времени (квартала) в данных панельного типа. В знаменателе (33) стоит среднее значение  $Y$  за предшествующий период наблюдений. Показатель  $\Phi_{24}$  характеризует динамику объема продаж и показывает: во сколько раз возрос этот объем по отношению к среднему значению за предшествующий период.

Все показатели  $\Phi_{11}, \Phi_{13}, \Phi_{21}, \Phi_{24}, \Phi$  определялись в выборке  $Z^{\text{II}}$  как среднее значение по всем кварталам  $i = 1, 12$ .

В последней колонке  $\bar{\Phi}$  – это усредненное по кварталам и по элементам кластера значение обобщенного критерия  $\Phi$ .

Из табл. 2 и 3 видно, что метод  $k$ -средних и метод SOM (сеть Кохонена)

Таблица 1

Исходные данные для кластеризации группы торговых предприятий (выборка  $Z^1$ )

од предприятия	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_8$	$Y$
1	43,66282	1,178545	223,0073	161,6862	12,45455	0,571909	302,1623	335,2882
2	136,0155	2,612455	270,0109	218,2636	32,90909	33,37182	1236,573	1437,018
3	46,11588	0,994125	540,9274	48,13813	16,15	512,0149	977,1925	1145,555
4	45,6035	1,3108	36,2	26,92	9,75	0	265,35	307,54
5	425,8971	11,11571	3649,1433	3171,286	22,14286	5,428571	4952,714	5527
6	724,78	23,7491	3358,86	2173,7	126,8	76,8	7748,02	8808,9
7	540,72	13,9863	2641,088	884,9097	42,6	1348,242	410,087	1768,083
8	175,5245	2,316273	737,8555	417,4818	39,09091	161,6636	3525,171	2761,041
9	427,3075	0,5523	1378,019	649,431	42,8	0	2462,274	3573,793
10	268,361	2,7577	335,195	344,335	19,7	169,333	823,284	964,154
11	302,09	6,933	686,7	407,7	27,3	136,26	1594,03	1543
12	65,33609	1,853091	168,1309	124,9445	6,572727	8,246545	701,6318	821,3109
13	131,356	7,3933	569,17	211,106	22,9	16,94	536,359	899,46
14	0	0	647,1543	351,9729	28,71429	24,47	2606,986	2798,929
15	168,258	3,5222	231,1248	193,5293	27,3	0	599,955	1434,368
16	14,065	0,440364	820,0227	235,9627	10,81818	62,54545	667,11	733,0709
17	0	0	2202,914	1612,743	15	0	1725,132	973,0457
18	8,986	0,256	1217,545	924,43	6,5	182,28	320,8075	392,54
19	25,90543	2,348857	261,3986	122,2291	13,28571	61,62214	791,3343	893,7971
20	0	0	53,66667	30,33333	16	0	589,3333	356,6667
21	141,622	2,8098	161,96	106,34	18,95	21,77	883,03	989,85
22	227,08	3,656	221,9	174,8	18	0	744,1	850,4
23	11,918	0,248667	114,7125	89,3085	13,83333	11,80017	474,3422	481,013
24	71,39833	1,885333	257,565	31,8995	3,766667	149,1967	181,0593	335,3692

Таблица 2

Результаты кластеризации по SOM (сеть Кохонена)

Код предприятия	$\Phi_{11}$	$\Phi_{13}$	$\Phi_{21}$	$\Phi_{24}$	$\Phi$	Номер кластера	Расстояние до центра кластера	$\bar{\Phi}$
1	1,0856083922	0,3205368465	0,1473559092	0,7348109896	0,3101136961	1	0,186279484	
2	0,1635909278	0,11033834	0,1403682376	0,909368968	0,071888436	1	0,08959292	
3	0,1752658824	0,0555322268	0,1490068817	1,1249584919	0,0735071998	1	0,1089983605	
4	0,1586680854	0,2159803302	0,1367753467	0,8111802878	0,088789889	1	0,0987253236	
5	0,1418138848	0,101626026	0,1214875853	1,1514365959	0,0774701373	1	0,1015272117	
6	0,1370976988	0,1571028141	0,1103852474	1,7973531057	0,1403144005	1	0,2418434309	
9	1,1978081342	0,4566134779	0,219897646	1,2537676595	0,6095159326	1	0,2269582094	
10	0,2359581397	0,4418339503	0,1549966648	0,7660423182	0,1560682343	1	0,186279484	
11	0,2398989855	0,2514795746	-0,1072624308	0,7410069138	0,0778521129	1	0,2355140896	
12	0,1865261399	0,0774207635	0,1561640682	1,466179391	0,1070531331	1	0,1142827463	
13	1,7800406731	0,3570444674	0,3650155937	0,7508128722	0,5961551084	1	0,3508268835	
14	0,0759819347	0,0	0,0703190687	1,0854164841	0,0219537558	1	0,1504765472	
15	1,2263618711	0,289923746	0,1867303407	1,8166078002	0,7594082023	1	0,259109935	
16	0,1036988808	0,0284237717	0,0915536107	1,3025588809	0,460484601	1	0,1533634089	0,188
18	0,2197773601	0,023942236	0,1789277469	3,3867716005	0,2172577012	1	0,5631647739	
19	0,1301099448	0,0343453804	0,1147978263	0,7329765942	0,0348552545	1	0,1735227699	
21	0,1219197753	0,1604537466	0,1062448157	1,0034199019	0,0783349836	1	0,08959292	
22	0,1428571429	0,3051740357	0,125	0,0	0,0140009743	1	0,3601476049	
23	0,0712464906	0,063861628	0,0585244261	2,7552121056	0,9500396622	1	0,4105281994	
8	0,6625125808	0,1131906807	-0,5450982655	1,4576351345	0,1769644564	2	0,2471376323	
17	6,3895760611	0,0	-0,8278283689	0,7507074093	-0,1231925594	2	0,7709907822	0,01355
20	-0,2786585643	0,0	-0,5945950598	0,7833742151	-0,0131512321	2	0,2480176218	
7	3,0880400586	1,3753851211	0,210485508	2,8224377367	3,384314077	3	0,7740355297	
24	1,8348203434	0,453822268	0,277556912	2,573147596	1,63113222	3	0,1618886589	2,5

Таблица 3

Результаты кластеризации по методу  $k$ -средних

Код предприятия	Номер кластера	Расстояние от центра кластера	Обобщенный критерий $\Phi$	$\bar{\Phi}$	
1	1	1,00	0,31010	0,2457	
2	1	0,623	0,07189		
3	1	0,594	0,07351		
4	1	0,682	0,08879		
5	1	0,433	0,07747		
6	1	0,689	0,14030		
8	1	2,243	0,17700		
9	1	1,226	0,60950		
10	1	1,174	0,15607		
11	1	0,990	0,7785		
12	1	0,552	0,10705		
13	1	1,681	0,59610		
14	1	0,746	0,02195		
15	1	1,024	0,79408		
16	1	0,603	0,04605		
18	1	2,712	0,21725		
19	1	0,914	0,03485		
20	1	2,610	-0,01315		
21	1	0,449	0,07833		
22	1	1,696	0,01400		
23	1	1,900	0,09504		
24	1	2,239	1,3611		
17	2	0,000	-0,1232		-0,1232
7	3	0,000	3,3843		3,3843

дают близкие результаты, т.е. подтверждают друг друга. Наилучшим в аспекте кредитоспособности оказался кластер с номером 3, куда попали предприятия № 7 и № 24 по методу SOM и  $k$ -средних соответственно. Усредненный обобщенный критерий здесь имеет большие значения ( $\bar{\Phi}=2,5$  и  $3,3843$  соответственно для метода SOM и  $k$ -средних).

Во второй кластер для обеих методов попало предприятие № 17 со значением  $\Phi=-0,1232$ . Это наихудший по финансовому положению кластер. В методе SOM сюда же попали предприятия

№ 8 с низким значением  $\Phi=0,1769$  и № 20 со значением  $\Phi=-0,01315$ , т.е. предприятие № 8 занимает пограничное положение между кластерами. Первый кластер включает основную часть предприятий для обоих методов SOM и  $k$ -средних и соответственно  $\bar{\Phi}=0,188$  и  $0,2457$  (промежуточное финансовое состояние между кластерами 3 (хорошее состояние) и кластера 2 (назревает банкротство)). Таким образом, полученные данные могут быть использованы при разработке ЛПР управляющих воздействий по налоговому администрированию.

**Выводы.** Показано, что в задачах кластеризации многомерных данных в отсутствие каких-либо допущений о характере распределения шумов в наблюдениях и в рамках моделей кластеризации типа гауссовых смесей можно использовать эффективный байесовский **EM**-алгоритм, разработанный А.С. Шумским [11].

В примере, основанном на реальных данных налоговых деклараций, показана эффективность оригинального метода

агрегирования переменных задач путем образования комплексной аддитивно-мультипликативной свертки в качестве обобщенного критерия  $\Phi$  с использованием морфологического принципа.

Данные кластеризации предприятий – налогоплательщиков по обобщенному критерию их кредитоспособности представляют собой удобный инструмент поддержки принятия решений в системе налогового администрирования.

### Список использованных источников

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. М.: ЮНИТИ, 1998. 1023 с.
2. Бирюков А.Н. Теоретические основы разработки нейросетевых моделей в системе налогового администрирования. Уфа: ГИЛЕМ, 2011. 380 с.
3. Букаев Г.И., Бублик Н.Д., Горбатков С.А., Саттаров Р.Ф. Модернизация системы налогового контроля на основе нейросетевых информационных технологий. М.: Наука, 2001. 344 с.
4. Бэстенс Д.-Э., Ван дер Берг В.-М., Вуд Д. Нейронные сети и финансовые рынки: принятие решений в торговых операциях. М.: ТВП, 1997. 236 с.
5. Галушкин А.И. Теория нейронных сетей. : учеб. пособие / Под общ. ред. А.И. Галушкина. Кн. 1. М.: ИПРЖ, 2000. 416 с.
6. Гатауллин Р.Ф., Горбатков С.А., Бирюков А.Н., Глущенко О.И. Моделирование бюджетных процессов на муниципальном уровне на основе нейросетей. Уфа: Изд-во Восточного университета, 2008. 216 с.
7. Горбатков С.А., Полупанов Д.В. Методы нейроматематики в налоговом контроле. Уфа: Изд. Башгосуниверситета, 208. 134 с.
8. Дебок Г., Кохонен Т. Анализ финансовых данных с помощью самоорганизующихся карт / Пер. с англ. М.: АЛПИНА, 2001. 317с.
9. Ежов А.А., Шумский С.А. Нейрокомпьютинг и его применение в экономике и бизнесе: учебник / Под ред. проф. В.В. Харитонов. М.: МИФИ, 1998. 224 с.
10. Хайкин С. Нейронные сети: полный курс, 2-е изд. М.: Вильямс, 2006. 1104 с.
11. Шумский А.С. Байесова регуляризация обучения // Лекции для школы – семинар «Современные проблемы нейроинформатики» (23–25 января 2002 г.). М.: МИФИ, 2002. (file://НейроОК Интелсофт.htm ). 33 с.
12. Вылкова Е.С., Тарасевич А.Л. Концептуальные основы реформирования налога на доходы физических лиц в России. СПб.: Изд-во СПбГУЭФ, 2010. 212 с.
13. Зайковская Л.Г., Нижегородцев Р.М., Хакимов З.Р. Нелинейные регрессивные методы параметрического ценообразования информационных продуктов // Вестник УГТУ–УПИ. Серия экономика и управление. 2010. № 5.