

# 1. НАЛОГООБЛОЖЕНИЕ И ФИНАНСОВЫЙ КОНТРОЛЬ

## 1.1. НЕЙРОСЕТЕВАЯ МОДЕЛЬ КЛАСТЕРИЗАЦИИ НАЛОГОПЛАТЕЛЬЩИКОВ В ЗАДАЧАХ НАЛОГОВОГО РЕГУЛИРОВАНИЯ

Бирюков А.Н., к.э.н., старший преподаватель

Башкирский государственный университет

Работа посвящена вопросу объективной количественной кластеризации налогоплательщиков – юридических лиц на базе нейросетей – самоорганизующихся карт Кохонена (СОК). Необходимость разработки такой модели кластеризации часто возникает в аспекте модели поддержки принятия решений в процедурах налогового регулирования (налоговых льгот, каникул, налоговых кредитов, налоговых вычетов и др.).

Основная идея работы заключается в следующем. Из показателей налоговых деклараций целесообразно образовывать несколько обобщенных критериев – агрегатов (линейных сверток), характеризующих:

- эффективность производства и ресурсы налогоплательщика;
- финансовое состояние (аналог  $Z$  – индикатора финансовой устойчивости Альтмана);
- организационной структуры предприятия (этот критерий включает в себя, в основном, качественные переменные);
- состояние внешней экономической среды.

Образованные критерии – агрегаты (линейные свертки) затем сворачиваются мультипликативно, образуя главную полезную функцию  $\Phi$ , положенную в основу кластеризации.

Приведены данные вычислительных экспериментов на реальных данных деклараций группы закодированных торговых предприятий города Уфы. Для оценки адекватности модели кластеризации сравнивались результаты кластеризации для двух методов – СОК и  $k$ -средних.

Работа предназначена для научных работников и работников налоговых органов, аспирантов, магистрантов и студентов, занимающихся проблемами совершенствования налогового и финансового контроля, нейросетевого моделирования и автоматизации экономических объектов.

Задача оперативного и гибкого налогового регулирования является частью государственного налогового менеджмента. От эффективности решения этой задачи зависит, с одной стороны, увеличение наполнения бюджетов всех уровней, а с другой – 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 субъектов налогообложения.

В принципе исчерпывающую для налогового регулирования информацию могли бы дать выездные налоговые проверки. Однако на практике такой вид получения информации для этих целей неэффективен, что связано с рядом его недостатков и ограничений, к которым относят:

- законодательные ограничения периодичности проведения выездных проверок;
- трудоемкость (один-два месяца работы выездной бригады налоговых инспекторов);
- не исключается полностью искажение истинного финансово-экономического положения налогоплательщика во всем объеме первичной документации; примером тому служат многочисленные разбирательства дел в арбитражных судах, взаимоисключающие документы россий-

ских и зарубежных таможенных и пограничных служб, практика проведения встречных налоговых проверок и др. Отмеченные недостатки не позволяют оперативно и с малыми затратами получать достаточно достоверную информацию о финансово-экономическом состоянии и тем более об их группах (кластерах).

В данной статье предлагается малозатратная иерархическая схема получения необходимой для налогового регулирования информации о субъектах налогообложения на основе математического моделирования, использующего только один источник – исторические данные по налоговым декларациям (рис. 1).

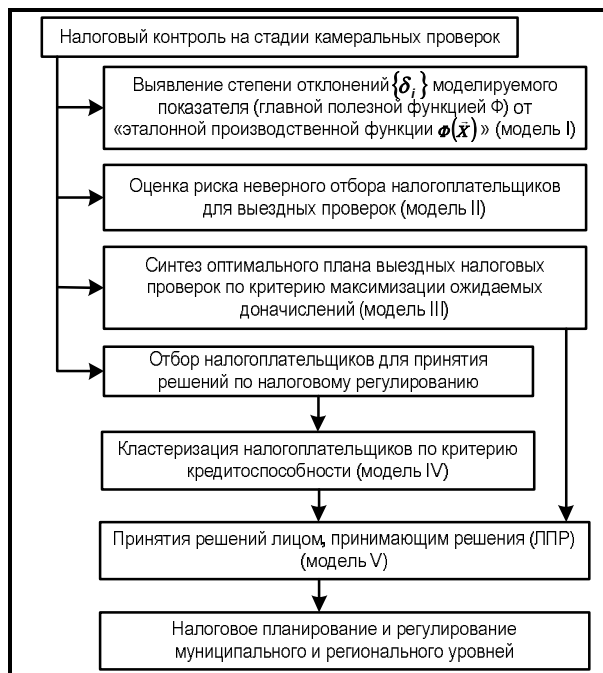


Рис. 1. Получение необходимой информации о субъектах налогообложения на основе математического моделирования

Изложим коротко взаимодействие всех пяти математических моделей этой иерархической схемы. В модели I используется нейросетевой базис [7, с. 28]. На вход модели подаются вектор – строки примеров из налоговых деклараций, т.е. кортежи:

$$\langle \Phi, \bar{x}_i \rangle, \quad i = \overline{1, N},$$

где

$\Phi$  – моделируемый показатель, имеющий смысл виртуального «объема продукции» в обобщенной «производственной функции»  $\Phi(\bar{x})$  для кластера примерно однородных налогоплательщиков;

$\bar{x}$  – вектор входных факторов;

$i$  – текущий номер наблюдений в статистической выборке;

$N$  – количество примеров. Модель I суть адаптивная аппроксимационная модель, восстанавливающая многомерную нелинейную «производственную функцию»  $\Phi(\bar{x})$ , «зашитую» в данных.

Такая задача восстановления относится к классу обратных задач и требует специальных мер регуляриза-

ции [4, с. 47]. Выходом модели I является множество отклонений:

$$\delta_i = |(\Phi_i - \hat{\Phi}_i) / \Phi_i|, \quad i = \overline{1, N},$$

где  $\Phi_i, \hat{\Phi}_i$  – экспериментальные и рассчитанные нейросетью значения  $\Phi$  соответственно.

- Модель I рассмотрена в [7, с. 28] и в данной статье подробно не рассматривается.
- Модель II оценки риска ошибок неверного отбора налогоплательщиков в замкнутой форме в виде вероятностной модели (коэффициента риска) получена в [7, с. 39] в предположении о нормальном законе распределения плотности вероятности отклонений  $\{\delta_i\}$ . Поскольку нормальный закон распределения для отклонений  $\{\delta_i\}$ , не всегда выполняется на практике, исследования по оценке риска отбора мы продолжаем с использованием 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 лицами, принимающими решения (ЛПР), в настоящее время проработана только на концептуальном уровне, поэтому в данной статье не описывается.

Предметом исследования в 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 [6, с. 135]. Подобная задача встречается в экономике довольно часто. Приведем некоторые примеры. В настоящее время в Российской Федерации для ряда регионов выделены 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 [3, с. 145]. Другой пример – это задача распределения фонда материального развития (ФМР), т.е. части субвенций, передаваемых из бюджета РФ ее субъектам, между муниципальными учреждениями (МУ) муниципальных образований (МО), на территории которых находятся МУ [3, с. 136]. Третий пример – задача входного финансового контроля для корпоративных заемщиков, которая решалась Голландским национальным банком при предоставлении кредитов мелким польским банкам и финансовым корпорациям [6, с. 129]. Здесь с помощью нейросетевой модели решалась задача классификации 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 три группы:

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

Четвертый пример – прогноз финансовой устойчивости и конкурентоспособности предприятия при выработке

управленческих решений на перспективу руководителем, т.е. ЛПР. Примеры рассматриваемого класса задач можно было бы продолжить. Предметом исследования в нашей статье является общая часть (ядро) данного класса задач оценки кредитного риска – модель ранжирования предприятий в том или ином аспекте. Подчеркнем также специфику источника информации: в нашем случае – это исторические данные по налоговым декларациям совокупности налогоплательщиков.

1. Актуальность темы работы. Многие модели предсказания банкротств убедительно демонстрируют свои способности к прогнозированию в задачах классификации «задним числом» (ex post). Подробный анализ этих задач приведен в [6, с. 155]. В нынешней кризисной ситуации, когда резервы выживания и кредитные резервы взаимосвязаны, эти модели работают плохо. Поэтому актуальна задача разработки новых моделей оценки кредитного риска, которые могли бы служить объективной научной основой принятия решений при распределении антикризисных кредитов [11].
2. Классические модели предсказания банкротства. При использовании количественных моделей исходят из того что, что шансы компании на выживание можно оценить, отслеживая, как меняются со временем соотношения между различными финансовыми показателями. Наиболее известными являются два метода [6, с. 121]:
  - метод множественного дискриминантного анализа (multiple discriminant analysis, *MDA*);
  - модель множественной линейной регрессии (linear probability models, *LPM*).

**MDA** – это статистический метод изучения различий между двумя и более группами объектов по совокупности нескольких финансовых показателей. Объекты (предприятия, компании, обратившиеся с просьбами о предоставлении займа) разбиваются на несколько попарно непересекающихся групп на основании ряда показателей их работы. Основные допущения здесь состоят в том, что имеется два и более возможных исходов (групп) и что описывающие модель переменные распределены в своем интервале изменения в соответствии с многомерным нормальным законом распределения. Остается открытым вопрос о том, можно ли пользоваться многомерным нормальным распределением применительно к финансовым показателям, которые распределены ненормально, что часто наблюдается на практике [11].

Другое затруднение применения **MDA**-моделей – это отсутствие оптимальной процедуры выбора совокупности переменных. Здесь наблюдается чрезмерное многообразие описательных переменных, и процедура отбора теоретически не обоснована. Однако можно сделать общую важную рекомендацию, которую первым отметил Арженти [14]: помимо количественных финансовых показателей в модель **MDA** следует вводить качественные переменные (их набирается до 20 и более). По мнению Арженти, качественные переменные описывают причины банкротства, а количественные финансовые показатели – его симптомы.

Отметим, что в статье автора [3, с. 123] предпринята попытка разработки морфологического принципа спецификации переменных, вытекающего из общесистемного закона энтропийного равновесия в открытых информационных системах [9, с. 111]. Полученные здесь количественные оценки вселяют уверенность в перспективности этого подхода.

В регрессионных **LPM**-моделях наиболее подходят для рассматриваемой задачи модели probit и logit (соответственно со стандартной и логистической функцией преоб-

разования), поскольку преобразование является монотонным, его выходные значения ограничены нулем и единицей и стремятся к нулю на хвостах распределения.

Анализ известных моделей банкротств, подробная библиография которых приведена в [6, с. 105; 11], позволяет сделать вывод, что современным эффективным инструментарием для условий сильной волатильности показателей внешней экономической среды являются методы нейроматематики, фрактальные методы и методы нечетной логики [11]. Автор статьи имел возможность практически убедиться в этом применительно к моделям налогового контроля и муниципального бюджетирования [9, с. 144].

**Экономико-математическая концепция моделирования**

Будем предполагать, что показатель надежности и успешности деятельности предприятия, можно оценить как мультипликативную свертку нескольких обобщенных (агрегированных) показателей:

$$\Phi(\bar{X}, t) = \Phi_1(\bar{X}, t) * \Phi_2(\bar{X}, t) * \Phi_3(\bar{X}, t) * \Phi_4(\bar{X}, t), \quad (1)$$

где  $\bar{X}$  – вектор входных факторов;  
 $t$  – время;

$\Phi_1(\bar{X}, t)$  – показатель, характеризующий эффективность производственного процесса;

$\Phi_2(\bar{X}, t)$  – показатель финансового состояния предприятия;

$\Phi_3(\bar{X}, t)$  – организационный показатель;

$\Phi_4(\bar{X}, t)$  – показатель состояния внешней экономической среды.

Перемножение частных критериев  $\{\Phi_x\}$  в (1) отражает взаимосвязь всех показателей в общей оценке  $\Phi$  эффективности работы предприятия: показатель эффективности будет максимальным, если одновременно будут максимальными все агрегаты  $\Phi_1, \dots, \Phi_2$  (умножение в булевой алгебре соответствует логической операции «И», т.е. пересечению множеств).

Агрегат  $\Phi_1$  определим как линейную свертку из показателей эффективности производства и ресурсных переменных:

$$\Phi_1 = \sum_{r=1}^{q_1} C_r \Phi_{1r}, \quad C_r \geq 0, \quad \sum_{r=1}^{q_1} C_r = 1, \quad (2)$$

где  $\Phi_{11}$  – рентабельность производства (чистая прибыль на единицу затрат);

$\Phi_{12}$  – трудовой потенциал (отношение фонда оплаты труда к выручке);

$\Phi_{13}$  – вооруженность труда (отношение (фондоёмкость) основных средств к выручке).

Показатель финансового состояния, по аналогии с Z-моделью Альтмана [7] определим как аддитивную свертку.

$$\Phi_2 = \sum_{r=1}^{q_2} d_r \Phi_{2r}, \quad d_r \geq 0, \quad \sum_{r=1}^{q_2} d_r = 1, \quad (3)$$

где  $\Phi_{21}$  – чистая маржа (отношение чистого дохода к обороту);

$\Phi_{22}$  – коэффициент покрытия при обслуживании долга (отношение прибыли до уплаты процентов и налогов к затратам на выплату процентов по займу);

$\Phi_{23}$  – степень ликвидности (отношение ликвидности фирмы к сумме долговых обязательств);

$\Phi_{24}$  – рост объема продаж (отношение текущего объема к среднему объему наблюдаемого периода);

$\Phi_{25}$  – дебиторский показатель (отношение среднего срока платежей в днях к 365 (количество дней в году));

$\Phi_{26}$  – доля заемных средств в активах;

$\Phi_{27}$  – показатель динамики прибыли.

Качественные переменные включены в состав организационного агрегированного показателя:

$$\Phi_3 = \sum_{r=1}^{q_3} b_r \Phi_{3r}, \quad b_r \geq 0; \quad \sum_{r=1}^{q_3} b_r = 1, \quad (4)$$

где

$\Phi_{31}$  – удельные коммерческие расходы (отношение коммерческих расходов к выручке);

$\Phi_{32}$  – связь управляющий – владелец (единица – управляющий является владельцем, ноль – в противном случае);

$\Phi_{33}$  – оперативность принятия решения (единица – принимает ЛПР на местах; ноль – централизованное принятие решения);

$\Phi_{34}$  – благонадежность управляющего (единица – при высокой вероятности возврата долга по экспертным сведениям из налоговой инспекции и других структур; ноль – в противном случае);

$\Phi_{35}$  – стаж работы управляющего на руководящей работе;

$\Phi_{36}$  – стаж работы управляющего на данной фирме.

Последние два показателя должны быть нормированы, например делением на пенсионный горизонт (60 лет).

$$\Phi_4 = \sum_{r=1}^{q_4} g_r \Phi_{4r}; \quad \sum_{r=1}^{q_4} g_r = 1; \quad g_r \geq 0. \quad (5)$$

где

$\Phi_{41}$  – индекс инфляции (средний на промышленную продукцию и товары народного потребления);

$\Phi_{42}$  – зависимость прибыли от таможенных сборов (отношение выручки к таможенным сборам).

Далее в предлагаемой концепции факторы  $\{\Phi_x\}$  привязываются к конкретному сектору экономики и определяются методом экспертных оценок с «дообучением» экспертов [9].

В качестве примера для решения задачи кластеризации 24 заемщиков по шести признакам использовались опубликованные данные налоговых деклараций по группе торговых предприятий из [7, 10]. Кластеризация проводилась методом k-средних [1] с помощью программы «СтартЭксперт». Результаты кластеризации показаны в табл. 1-4. В качестве меры расстояния элемента (заемщика) от центра кластера использовалось евклидово расстояние:

$$d(\bar{X}_k, \bar{X}_j) = \sqrt{\sum_{i=1}^6 (x_{ki} - x_{ji})^2}, \quad (6)$$

где

$\bar{X}_k, \bar{X}_j$  – векторы кластерообразующих признаков k-го и j-го элементов;

i – номер компоненты вектора  $\bar{X} = (X_1, X_2, \dots, X_6)$ .

Из табл. 2 видно, что во второй кластер вошло одно предприятие 7, которое имеет максимальный показатель, характеризующий финансовое состояние ( $\Phi = 2,129618$ ). Значит предприятие является наиболее кредитоспособным и платежеспособным и «выживет, по крайней мере, в течении года». Поэтому кредитные учреждения при выдаче кредитов отдадут предпочтение предприятиям, входящим в кластер 2, так как в этом случае кредитный риск будет минимальным. Целесообразна также поддержка предприятий из кластера 2 при налоговом регулировании.

Все торговые предприятия, попавшие в кластер 1 примерно однородны в отношении кредитоспособности, но риск непогашения кредита у них выше, чем у предприятий, входящих в кластер 2. Кластер 1 образует «пограничную зону». Среднее по кластеру значение  $\Phi = 1.144884$ . В него вошло 22 предприятия и кредитные организации при решении вопроса о выдаче кредита должны применять индивидуальный подход к каждому предприятию. Скорее всего предпочтение будет отдаваться тем предприятиям, чьи обобщенные показатели ближе к показателям второго кластера.

Самый максимальный кредитный риск у предприятий, попавших в кластер 3. В данный кластер входит одно предприятие 17, которое по своим обобщенным показателям находится на грани банкротства ( $\Phi = -0,01084$ ). А значит, предприятие не является кредитоспособным и не имеет возможности получения заемных средств в кредитных организациях. Кластеризация на основе самоорганизующихся карт Кохонена (СОК) [8] дала близкие результаты (табл. 5) за исключением предприятий 17 и 24.

## ВЫВОДЫ

Разработанным алгоритм ранжирования налогоплательщиков в задаче кластеризации с использованием как количественных финансовых показателей, так и качественных (переменных Арженти) представляется автору работоспособным и перспективным.

В дальнейших исследованиях в качестве инструмента кластеризации представляет интерес нейросети Кохонена [8] и их обобщение – упругие карты [13].

Таблица 1

### ИСХОДНЫЕ ДАННЫЕ К РЕШЕНИЮ ЗАДАЧИ КЛАСТЕРИЗАЦИИ ЗАЕМЩИКОВ

Код предприятия	$\Phi_{11}$	$\Phi_{13}$	$\Phi_{21}$	$\Phi_{24}$	$\Phi_{25}$	$\Phi_{31}$
1	0,674775	3,469699	0,147356	0,734811	0,571909	0,120102
2	-0,00116	4,097098	0,140368	0,909369	33,37182	0,141439
3	-0,01965	3,029453	0,149007	1,124958	512,0149	0,165814
4	0,020874	4,719943	0,136775	0,81118	0	0,118851
5	0,000875	17,817	0,121488	1,151437	5,428571	0,119701
6	-0,01603	7,459294	0,110385	1,797353	76,8	0,131973
7	3,087263	12,58268	0,210486	2,822438	1348,242	0,00061
8	0,517365	4,656689	-0,5451	1,457635	161,6636	0,118821
9	1,044762	10,19989	0,219898	1,253768	0	0,131103
10	-0,11691	7,209096	0,154997	0,766042	169,333	0,259315
11	0,061235	11,49239	-0,10726	0,741007	136,26	0,155933
12	0,040408	5,689122	0,156164	1,466179	8,246545	0,122713
13	0,090687	5,777054	0,365016	0,750813	16,94	0,38265
14	-0,00523	0	0,070319	1,085416	24,47	0,075356
15	0,987311	6,254143	0,18673	1,816608	0	0,189232
16	-0,00078	1,36599	0,091554	1,302559	62,54545	0,091312
17	4,34077	0	-0,82783	0,750707	0	0,196319

Код предприятия	$\Phi_{11}$	$\Phi_{13}$	$\Phi_{21}$	$\Phi_{24}$	$\Phi_{25}$	$\Phi_{31}$
18	0,004861	1,12325	0,178928	3,386772	182,28	0,174576
19	-0,00977	1,977682	0,114798	0,732977	61,62214	0,123529
20	-0,40534	0	-0,5946	0,783374	0	0,168495
21	-0,41496	7,502189	0,106245	1,00342	21,77	0,476843
22	-0,0215	12,61556	0,125	0	0	0,143815
23	-0,07429	0,876862	0,058524	2,755212	11,80017	0,124175
24	1,666226	25,7034	0,277557	2,573148	149,1967	0,131065

Таблица 2

### РАСЧЕТНЫЕ ЗНАЧЕНИЯ ОБОБЩЕННЫХ ПОКАЗАТЕЛЕЙ

Код предприятия	$\Phi_1$	$\Phi_2$	$\Phi_3$	$\Phi$
1	2,072237	0,479845	0,120102	0,119424
2	2,047967	11,35911	0,141439	3,290304
3	1,504902	169,3853	0,165814	42,26732
4	2,370409	0,312825	0,118851	0,088131
5	8,908935	2,211494	0,119701	2,358355
6	3,721632	25,97355	0,131973	12,75704
7	7,834973	445,9208	0,00061	2,129618
8	2,587027	53,65014	0,118821	16,49168
9	5,622325	0,48631	0,131103	0,35846
10	3,546092	56,18383	0,259315	51,66409
11	5,776813	45,17494	0,155933	40,69331
12	2,864765	3,256733	0,122713	1,144884
13	2,93387	5,958423	0,38265	6,689193
14	-0,00262	8,456493	0,075356	-0,00167
15	3,620727	0,661102	0,189232	0,452958
16	0,682602	21,10006	0,091312	1,315163
17	2,170385	-0,02545	0,196319	-0,01084
18	0,564056	61,32908	0,174576	6,039099
19	0,983954	20,61507	0,123529	2,505699
20	-0,20267	0,062297	0,168495	-0,00213
21	3,543616	7,550289	0,476843	12,7581
22	6,297027	0,04125	0,143815	0,037356
23	0,401285	4,822588	0,124175	0,240308
24	13,68483	50,17563	0,131065	89,99499

Таблица 3

### ПРОТОКОЛ КЛАСТЕРНОГО АНАЛИЗА (РЕЗУЛЬТАТЫ КЛАСТЕРИЗАЦИИ)<sup>1</sup>

Код предприятия	Кластер	Расстояние от центра	Координата X	Координата Y	Координата Z
1	1	1,147	-0,327	0,154	0,187
2	1	0,749	-0,428	-0,161	0,403
3	1	1,719	0,230	-0,154	0,591
4	1	0,919	-0,423	-0,132	0,374
5	1	1,97	0,176	-0,900	-0,953
6	1	0,781	0,143	-0,390	0,503
8	1	2,343	-0,014	1,465	0,171
9	1	1,213	0,218	-0,348	-0,504
10	1	1,297	-0,566	-0,583	-0,506
11	1	1,286	-0,202	-0,014	-0,729
12	1	0,663	-0,099	-0,315	0,573
13	1	2,614	-0,084	-1,111	-0,946
14	1	1,489	-0,336	0,315	1,238
15	1	1,122	0,118	-0,245	-0,040

<sup>1</sup> Алгоритм  $k$ -средних ( $k = 3$ ), метрика – Евклидова, количество переменных = 6.

Код предприятия	Кластер	Расстояние от центра	Координата X	Координата Y	Координата Z
16	1	1,179	-0,182	0,139	1,128
18	1	2,825	0,645	-0,446	1,852
19	1	1,143	-0,495	0,070	0,637
20	1	2,762	-1,101	1,507	0,0496
21	1	3,366	-1,409	-0,015	-1,365

Код предприятия	Кластер	Расстояние от центра	Координата X	Координата Y	Координата Z
22	1	1,955	-0,580	-0,508	-1,036
23	1	2,128	0,217	-0,040	1,76
24	1	3,896	1,717	-1,306	-1,846
7	2	0	3,806	0,382	-0,113
17	3	0	-0,023	3,635	-1,876

Таблица 4

МАТРИЦА РАССТОЯНИЙ

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	0,000	0,710	2,065	0,640	2,503	1,645	6,252	2,704	1,360	1,859	1,809	1,144	2,965	1,101	1,638	1,090	4,926	3,508	0,717	2,928	3,987	1,896	2,673	4,482
2	0,710	0,000	1,771	0,313	2,299	1,260	6,348	2,637	1,489	1,453	1,574	0,779	2,705	1,030	1,590	0,877	5,335	3,206	0,475	2,755	3,626	1,814	2,397	4,448
3	2,065	1,771	0,000	1,969	3,087	1,960	5,215	2,907	2,442	1,786	2,205	1,968	3,081	2,080	2,313	1,846	5,650	3,088	1,764	3,314	3,836	2,818	2,798	4,658
4	0,640	0,131	1,969	0,000	2,204	1,354	6,404	2,667	1,450	1,668	1,553	0,840	2,922	1,001	1,736	0,920	5,353	3,397	0,522	2,800	3,856	1,674	2,532	4,434
5	2,503	2,299	3,087	2,204	0,000	1,913	6,171	3,325	1,625	2,422	1,546	2,044	3,570	2,985	2,396	2,748	6,058	4,025	2,674	3,973	4,176	1,697	3,450	2,785
6	1,645	1,260	1,96	1,354	1,913	0,000	5,776	2,492	1,354	1,907	1,705	0,597	3,122	1,651	1,182	1,260	5,547	2,336	1,613	3,131	3,813	2,424	1,648	3,576
7	6,252	6,348	5,215	6,404	6,171	5,776	0,000	6,098	5,738	6,443	6,119	6,071	7,424	6,368	5,821	6,122	7,341	5,744	6,412	7,478	8,032	6,814	6,098	5,201
8	2,704	2,637	2,907	2,667	3,325	2,492	6,098	0,000	2,989	3,118	2,203	2,604	4,428	2,512	2,858	2,434	3,945	3,662	2,631	1,630	4,620	3,379	2,872	4,872
9	1,360	1,489	2,442	1,450	1,625	1,354	5,738	2,989	0,000	2,003	1,713	1,235	3,013	2,107	1,148	1,865	5,169	3,306	1,840	3,678	3,961	1,923	2,698	3,150
10	1,859	1,453	1,786	1,668	2,422	1,907	6,443	3,118	2,003	0,000	1,618	1,813	1,633	2,399	1,924	2,183	5,573	3,548	1,733	3,152	2,409	1,899	3,128	4,373
11	1,809	1,574	2,205	1,553	1,546	1,705	6,119	2,203	1,713	1,618	0,000	1,720	3,114	2,254	2,176	2,082	5,086	3,873	1,813	2,657	3,615	1,358	3,178	3,854
12	1,144	0,799	1,968	0,840	2,044	0,597	6,071	2,604	1,235	1,813	1,720	0,000	2,994	1,208	1,207	0,869	5,477	2,654	1,130	3,031	3,840	2,177	1,834	3,934
13	2,965	2,705	3,081	2,922	3,570	3,122	7,424	4,428	3,013	1,633	3,114	2,994	0,000	3,581	2,659	3,398	6,161	4,140	2,963	4,242	1,498	3,053	3,960	5,052
14	1,101	1,030	2,08	1,001	2,985	1,651	6,368	2,512	2,107	2,399	2,254	1,208	3,581	0,000	2,088	0,422	5,273	3,147	0,778	2,624	4,454	2,599	2,159	4,988
15	1,638	1,590	2,313	1,736	2,396	1,182	5,821	2,858	1,148	1,924	2,176	1,207	2,659	2,088	0,000	1,767	5,030	2,415	1,940	3,488	3,480	2,719	1,947	3,515
16	1,090	0,877	1,846	0,920	2,748	1,260	6,122	2,434	1,865	2,183	2,082	0,869	3,398	0,422	1,767	0,000	5,316	2,804	0,801	2,706	4,25	2,545	1,866	4,657
17	4,926	5,335	5,650	5,353	6,058	5,547	7,341	3,945	5,169	5,573	5,086	5,477	6,161	5,273	5,030	5,316	0,000	6,319	5,273	4,418	6,357	5,737	5,754	6,742
18	3,508	3,206	3,088	3,397	4,025	2,336	5,744	3,662	3,306	3,548	3,873	2,654	4,140	3,147	2,415	2,804	6,319	0,000	3,403	4,340	4,564	4,704	1,216	4,488
19	0,717	0,475	1,764	0,522	2,674	1,613	6,412	2,631	1,840	1,733	1,813	1,130	2,963	0,778	1,940	0,801	5,273	3,403	0,000	2,631	3,892	2,005	2,552	4,839
20	2,928	2,755	3,314	2,800	3,973	3,131	7,478	1,63	3,678	3,152	2,657	3,031	4,242	2,624	3,488	2,706	4,418	4,340	2,631	0,000	4,309	3,475	3,442	6,058
21	3,987	3,626	3,836	3,856	4,176	3,813	8,032	4,62	3,961	2,409	3,615	3,840	1,498	4,454	3,480	4,250	6,357	4,564	3,892	4,309	0,000	3,861	4,486	5,525
22	1,896	1,814	2,818	1,647	1,697	2,124	6,814	3,379	1,923	1,899	1,358	2,177	3,053	2,599	2,719	2,545	5,737	4,704	2,005	3,475	3,861	0,000	3,967	4,243

Таблица 5

РЕЗУЛЬТАТЫ КЛАСТЕРИЗАЦИИ С ПОМОЩЬЮ СОК

Код предприятия	Номер ячейки	Расстояние до центра ячейки	Номер кластера	Расстояние до центра кластера
1	176	0,0000000101	0	0,2053507546
2	133	0,0000109952	0	0,144332265
3	141	0,0000001826	0	0,3477317391
4	130	0,0000051546	0	0,1759720415
5	182	0,0000125312	0	0,5011927633
6	55	0,0002735934	0	0,2469991918
8	3	0,0000005955	0	0,5216649438
9	186	0,0000085836	0	0,2961076075
10	191	0,000000188	0	0,2751367332
11	180	0,0185488556	0	0,2981309211
12	103	0,000188006	0	0,1864024867
13	8	0,0396027184	0	0,5250595523
14	48	0,0005990427	0	0,2952312544
15	137	0,0000083266	0	0,2925941652
16	96	0,0006357914	0	0,2427887337
17	6	0,0000000004	0	1,1458648288
19	67	0,0000063857	0	0,2145350948
20	0	0,000005508	0	0,6103594929
21	10	0,0399850902	0	0,661994717
22	178	0,0182419681	0	0,4277996
18	60	0,0000217523	1	0,1828000125
23	74	0,0000167314	1	0,1353561119
7	14	0,0000000006	2	0,582767984
24	95	0,0000000023	2	0,5502415001

Литература

1. Айвазян С.А. Прикладная статистика и основы эконометрики [Текст] : учеб. для вузов / С.А. Айвазян, В.С. Мхитарян. – М. : ЮНИТИ, 1998. – 1001 с.
2. Ахмадеев А.М. Оценка банком кредитоспособности заемщика [Текст] : учеб. пособие / А.М. Ахмадеев, Э.А. Комарова. – Уфа : РИЦ БашГУ, 2009. – 108 с.
3. Бирюков А.Н. Агрегирование переменных для нейросетевой модели оценки эффективности работы учреждений здравоохранения с позиций системного подхода [Текст] / А.Н. Бирюков // Препринт №1. – Уфа : Гилем, 2010. – 75 с.
4. Бирюков А.Н. Обобщение метода вложенных математических моделей на основе байесовского подхода к регуляризации задач нейросетевого моделирования налогового и финансового контроля [Текст] / А.Н. Бирюков // Экономика и управление. – 2010. – №1. – С. 85-89.
5. Бублик Н.Д. и др. Управление финансовыми и банковскими рисками [Текст] : учеб. пособие / Н.Д. Бублик, С.В. Попенов, А.Б. Секерин. – Уфа : Альтернатива РИЦ, 1998. – 190 с.
6. Бэстен Д.-Э. и др. Нейронные сети и финансовые рынки: принятие решений в торговых операциях [Текст] / Д.-Э. Бэстен, В.-М. Ван дер Берг, Д. Вуд. – М. : ТВП, 1998. – 235 с.
7. Горбатов С.А. Методы нейроматематики в налоговом контроле [Текст] / С.А. Горбатов, Д.В. Полупанов. – Уфа : РИЦ БашГУ, 2008. – 136 с.
8. Дебок Г. Анализ финансовых данных с помощью самоорганизующихся карт [Текст] / Г. Дебок, Т. Кохонен. – М. : Альпина бизнес бук, 2001. – 278 с.
9. Моделирование бюджетных процессов на муниципальном уровне на основе нейросетей [Текст] / Гатауллин Р.Ф.,

- Горбатков С.А., Бирюков А.Н., Глуценко О.Н. – Уфа : Восточный ун-т, 2008. – 216 с.
10. Модернизация системы налогового контроля на основе нейросетевых информационных технологий [Текст] / Букаев Г.И., Бублик Н.Д., Горбатков С.А., Сатаров Р.Ф. – М. : Наука, 2001. – 344 с.
  11. Петерс Э. Хаос и порядок на рынках капитала. Новый аналитический взгляд на циклы, цены и изменчивость рынка [Текст] : пер. с англ. / Э. Петерс. – М. : Мир, 2000. – 356 с.
  12. Питенко А.А. Нейросетевой анализ в геоинформационных системах [Текст] : автореф. дис. ... канд. техн. наук по специальности 05.13.16 / А.А. Питенко. – Красноярск : Ин-т вычислительного моделирования СО РАН, 2000. – 28 с.
  13. Старт-Эксперт. Программные продукты серии ОЛИМП [Текст]. – М., 1996.
  14. Argenti J. Corporate collapse. – England: Mc Craw – Hill, 1976.

### Ключевые слова

Нейросетевая модель; лицо, принимающее решение (ЛПР); муниципальное учреждение (МУ); экономико-математическая концепция моделирования; самоорганизующиеся карты Кохонена (СОК); алгоритм ранжирования; метод множественного дискриминантного анализа (MDA); модель множественной линейной регрессии (LPM).

*Бирюков Александр Николаевич*

### РЕЦЕНЗИЯ

Статья посвящена вопросу объективной количественной кластеризации налогоплательщиков – юридических лиц на базе нейросетей самоорганизующихся карт Кохонена (СОК). Необходимость разработки такой модели кластеризации часто возникает в аспекте модели поддержки принятия решений в процедурах налогового регулирования (налоговых льгот, каникул, налоговых кредитов, налоговых вычетов и др.)

Основная идея статьи заключается в следующем. Из показателей налоговых деклараций целесообразно образовывать несколько обобщенных критериев – агрегатов (линейных сверток), характеризующих:

- эффективность производства и ресурсы налогоплательщика;
- финансовое состояние (аналог Z-индикатора финансовой устойчивости Альтмана);
- организационной структуры предприятия (этот критерий включает в себя, в основном, качественные переменные);
- состояние внешней экономической среды.

Образованные критерии – агрегаты (линейные свертки) затем сворачиваются мультипликативно, образуя главную полезную функцию  $\Phi$ , положенную в основу кластеризации.

Приведены данные вычислительных экспериментов на реальных данных деклараций группы закодированных торговых предприятий города Уфы. Для оценки адекватности модели кластеризации сравнивались результаты кластеризации для двух методов – СОК и *k*-средних. Совпадение результатов можно считать хорошим (~83%).

Публикация статьи целесообразна.

*Карлов В.Г., д.э.н., профессор кафедры экономики и управления на предприятии нефтяной и газовой промышленности Уфимского государственного нефтяного технического университета*

## 1. TAXATION AND THE FINANCIAL SCREENING

### 1.1. NEURAL NETWORK MODEL OF CLUSTERING IN PROBLEMS OF TAXPAYERS OF TAX REGULATION

A.N. Biryukov, Candidate of Economic Sciences,  
Senior Lecturer

*Bashkir State University; Sterlitamak*

Work is devoted to objective quantitative clustering taxpayer – legal persons based on neural networks – Koho-

nen self-organizing maps (SOC). Need to develop a clustering model frequently arises in the aspect of the model to support decision-making procedures in the tax regulations (tax breaks, holidays, tax credits, tax deductions, etc.) The basic idea is as follows. Indicator of appropriate tax returns to form several generalized criteria – aggregates (linear convolutions), describing:

- production efficiency and resources of the taxpayer;
- financial condition (similar to *Z* – an indicator of financial stability Altman);
- organizational structure (this criterion includes mainly qualitative variables);
- state of the external economic environment.

Educated criteria – units (linear convolution), then fold multiplicatively to form the main useful function  $F$ , form the basis of clustering. Data are presented computational experiments on real data declarations of coded trading enterprises in the city of Ufa. To assess the adequacy of the clustering model compared the results of clustering for the two methods – the juice and the *k*-means. The work is intended for scientific workers and tax authorities, graduate students, undergraduates and students concerned with improving the tax and financial control, neural network modeling and automation of economic objects.