

Интернет-журнал «Наукоедение» ISSN 2223-5167 <http://naukovedenie.ru/>

Том 9, №2 (2017) <http://naukovedenie.ru/vol9-2.php>

URL статьи: <http://naukovedenie.ru/PDF/32TVN217.pdf>

Статья опубликована 27.04.2017

Ссылка для цитирования этой статьи:

Горбатков С.А., Полупанов Д.В., Белолипцев И.И., Фархиева С.А. Нейросетевые методы отбора налогоплательщиков для проведения выездных проверок // Интернет-журнал «НАУКОВЕДЕНИЕ» Том 9, №2 (2017) <http://naukovedenie.ru/PDF/32TVN217.pdf> (доступ свободный). Загл. с экрана. Яз. рус., англ.

УДК 336.1

Горбатков Станислав Анатольевич

ФГБОУ ВО «Финансовый университет при Правительстве Российской Федерации»
Филиал в г. Уфа, Россия, Уфа¹
Профессор кафедры «Математика и информатика»
Доктор технических наук
E-mail: sgorbatkov@mail.ru
РИНЦ: http://elibrary.ru/author_items.asp?id=158740
SCOPUS: <https://www.scopus.com/authid/detail.uri?authorId=8646868800>

Полупанов Дмитрий Васильевич

ФГБОУ ВО «Башкирский государственный университет», Россия, Уфа
Доцент кафедры «Информационных технологий и компьютерной математики»
Кандидат технических наук
E-mail: demetrious@mail.ru
РИНЦ: http://elibrary.ru/author_profile.asp?id=700344

Белолипцев Илья Игоревич

ФГБОУ ВО «Финансовый университет при Правительстве Российской Федерации»
Филиал в г. Уфа, Россия, Уфа
Доцент кафедры «Математика и информатика»
Кандидат технических наук
E-mail: red7315@gmail.com
РИНЦ: http://elibrary.ru/author_items.asp?id=635807

Фархиева Светлана Анатольевна

ФГБОУ ВО «Финансовый университет при Правительстве Российской Федерации»
Филиал в г. Уфа, Россия, Уфа
Зав. кафедрой «Математика и информатика»
Кандидат технических наук
E-mail: ok-xi@yandex.ru
РИНЦ: https://elibrary.ru/author_items.asp?id=567037

Нейросетевые методы отбора налогоплательщиков для проведения выездных проверок

Аннотация. Данная статья обобщает исследования авторов по регуляризации обратных задач нейросетевого восстановления многофакторных нелинейных зависимостей, скрытых в данных. Одним из практических приложений служат, например, задачи налогового контроля, в том числе отбора налогоплательщиков для проведения выездных налоговых проверок. Большое

¹ 450015, г. Уфа, ул. Мустая Карима 69/1

внимание уделено практически малоисследованному вопросу обоснования адекватности создаваемых нейросетевых моделей. Предлагаемые теоретические идеи апробированы авторами в вычислительных экспериментах на реальных данных предприятий-налогоплательщиков. Авторы считают, что разработанные подходы, методы и нейросетевые модели могут быть использованы также и в ряде смежных областей: муниципального бюджетирования, диагностики и прогнозирования банкротства предприятий, мониторинга экономической безопасности корпораций в аспекте их финансово-экономического состояния и др.

Ключевые слова: налоговое законодательство; экономическая безопасность; нейросетевая модель; регуляризация; адекватность модели; нелинейные зависимости; байесовский подход

1. Введение

Основной целью настоящей работы является углубленное изучение вопроса об оценке вероятности нарушения налогового законодательства плательщиком при декларировании отчетности и создание адекватных нейросетевых моделей (НСМ) отбора корпораций-налогоплательщиков для выездных налоговых проверок. Актуальность использования продвинутых информационных интеллектуальных технологий, к которым относятся и нейросетевые технологии, в области налогового контроля подчеркивалась руководителями российских налоговых органов, в том числе Г.И. Букаевым [1], М.В. Мишустиним [2], где отмечалась важность совершенствования процедур проведения камеральных налоговых проверок и связанных с ними технологий отбора предприятий для проведения выездных налоговых проверок. Для рационального использования бюджетных средств и трудовых ресурсов проверки должны проводиться адресно, на основе оценки вероятности нарушения налогового законодательства конкретным налогоплательщиком. Для этого необходимо разработать математические модели, позволяющие надежно выявлять нарушителей налогового законодательства.

Перспективным инструментарием для построения указанных моделей являются нейронные сети (НС). В наших работах [3]-[8] предлагался подход к построению адекватных качественных нейросетевых моделей (НСМ) с хорошими аппроксимативными свойствами в условиях существенного зашумления данных, вплоть до сознательного искажения, отягощенного дефицитом наблюдений. В общую модель построения НСМ «вкладывается» совокупность субмоделей, с помощью которых выполняются операции по структурированию информационного пространства, такие как оптимальный итерационный процесс экспертных оценок с дообучением экспертов по выбору зависимых и независимых переменных НСМ; оптимальный итерационный процесс отбраковки аномальных, неинформативных и противоречивых вектор-строк из базы данных (БД) и др.

В рассматриваемом классе задач наблюдается высокий уровень неопределенности информационного пространства признаков - триада «НЕ-факторов» по терминологии А.С. Нариньяни [9] (неопределенность, неполнота, неточность). Предварительные эксперименты показывают, что при решении обратных задач восстановления зависимостей, скрытых в данных, необходима регуляризация - повышение устойчивости к возмущениям входных данных, о чем будет подробно сказано ниже.

В [3] было отмечено, что при построении плана выездных налоговых проверок на основе НСМ должны соблюдаться следующие принципы:

1. План проверок формируется после подачи налогоплательщиками последней налоговой декларации (квартального отчета) в момент времени t_0 . Появление значительных относительных отклонений $\delta_{gt} = (y - \hat{y})/y$ между фактическим y и рассчитанным с помощью НСМ \hat{y} значением выходной величины может свидетельствовать о нарушении g -м налогоплательщиком налогового законодательства и является основанием для включения его в план проверок;

2. Большие значения отклонений на момент составления плана проверок могут быть вызваны и объективными причинами (неудачная сделка, повлекшая за собой убытки, форс-мажор и т.д.). Поэтому необходимо учитывать предысторию появления у данного налогоплательщика таких отклонений и оценить вероятность нарушения данным налогоплательщиком налогового законодательства;

3. Проведение выездных проверок в первую очередь преследует цель получения максимальных налоговых доначислений. Очевидно, что величина возможных доначислений зависит от масштаба предприятия, который нужно учитывать при составлении плана проверок.

При составлении плана отбора налогоплательщиков для проведения выездных проверок авторами, в [3-5] предложен критерий вида

$$\psi_g = |\delta_g| \cdot M_g \cdot P_g, \quad (1)$$

где: $|\delta_g|$ - отклонение, полученное с помощью НСМ; M_g - коэффициент масштаба предприятия; P_g - вероятность нарушения налогового законодательства, которая может быть вычислена различными способами.

В [5] предложен подход к нахождению величины P_g как вероятности того, что текущее значение отклонения δ_g для g -го налогоплательщика будет больше его математического ожидания, смещенного вверх на полуширину доверительного интервала. Недостатком такого подхода является априорное задание закона нормального распределения отклонений, что зачастую не выполняется на практике.

Альтернативным подходом, предложенным авторами статьи [10], является оценка P_g на основе фрактальной теории, с помощью показателя Херста [11] и вычисления размерности минимального покрытия [12]. При этом осуществляется расчет величины фрактальной размерности D временных рядов отклонений $\{\delta_{g,t}\}$. Если уровни ряда независимы, то величина D будет стремиться к величине топологической размерности плоскости, то есть $D \rightarrow 2$. Если же значения ряда $\{\delta_{g,t}\}$ не являются независимыми, то величина D будет значительно меньше

2. Отношение $1/D$ трактуется как оценка вероятности больших отклонений $|\delta_g|$ в (1). Недостаток данного подхода в том, что для получения адекватной оценки нарушения налогового законодательства необходимо иметь достаточно большое количество данных (десятки, а то и сотни), что затруднительно на практике.

В настоящей статье авторами предлагается новая числовая мера оценки P_g : логистическая трансформация вероятности риска

$$P_g = \frac{1}{1 + e^{-\delta_g(x)}} \quad (2)$$

Показателем экспоненты в (2), по существу, является нелинейная нейросетевая свертка $\delta_g(x)$, зависящая от ряда производственных показателей (факторов) g -го предприятия.

Для оценки достоверности предлагаемого нового подхода к проблеме оценки вероятности нарушения налогового законодательства в данных деклараций проведено сравнение с вышеупомянутыми вероятностным и фрактальными методами.

В статье с учетом высокого уровня зашумленности данных большое внимание также уделено оценке эффекта регуляризации НСМ. В литературы известны подходы к регуляризации некорректных обратных задач восстановления зависимостей - разработанный академиком А.Н. Тихоновым [13] и байесовский [14]. Общая парадигма обоих подходов заключается в сужении класса искомых решений обратной задачи. Различия заключаются в следующем. Первый подход основан на принципе квазиминимизации, т.е. регуляризация осуществляется за счет введения стабилизаторов Тихонова с использованием априорных свойств искомых решений (чаще всего гладкости искомого решения). Байесовский подход к регуляризации НСМ, предложенный в работе С.А. Шумского [14] заключается в сужении класса искомых решений посредством трех процедур в алгоритме решения обратной задачи:

- априорного задания определенного класса нейросетей-гипотез, где аккумулируется опыт аналитика;
- апостериорной фильтрация нейросетей-гипотез в байесовском ансамбле на основе показателя качества аппроксимации данных;
- осреднения выходных характеристик на отфильтрованном ансамбле.

В работе [14] подход к байесовской регуляризации существенно опирается на задание априорного закона распределения шумов в данных. В данной статье авторы отказываются от стеснения алгоритма подобными допущениями, что приближает НСМ к практическим условиям.

Авторам не известны работы по байесовской регуляризации НСМ в условиях отсутствия каких-либо сведений о законе распределения шумов.

2. Постановка задачи

Для практического построения НСМ использовались данные квартальных налоговых деклараций 24 сельскохозяйственных предприятий Республики Башкортостан (формы 1 и 2 бухгалтерского баланса) на протяжении трех лет из [5]. В качестве моделируемой величины Y было взято отношение суммарных затрат к величине выручки предприятия. По мнению специалистов, эта величина точнее характеризует величину налогооблагаемой базы, чем собственно выручка или валовая прибыль предприятия. В качестве входных факторов, характеризующих финансовое состояние предприятия, были использованы 16 удельных показателей [15]. В итоге сформирована БД, состоящая из 276 наблюдений.

Необходимо построить нейросетевую модель

$$\hat{y} = F_1(x, W), x \in X \subset \mathfrak{R}^n, y \in Y \subset \mathfrak{R}, \hat{y} \in \hat{Y} \subset C[\mathfrak{R}] \quad (3)$$

где: $x = (x_1, \dots, x_n)$ конкретная численная реализация случайного n -мерного вектора входных факторов \vec{X} ; $y \in Y \subset \mathcal{R}^1$ - декларируемая налогоплательщиком конкретная числовая реализация наблюдаемой выходной случайной величины Y ; \hat{y} - объективный эталон - расчетное значение случайной величины Y ; $\{W\}$ - множество оцениваемых синаптических весов НС; $X \subset \mathcal{R}^n$ - множество значений вектора входных факторов, $Y \subset \mathcal{R}^1$ - множество декларированных значений выходной величины, $\hat{Y} \subset C[\mathcal{R}]$ - множество расчетных по (3) значений выходной величины, F_1 - оператор нейросетевого отображения. Отметим, что \hat{y} есть не что иное, как аппроксимация объясненной части условного математического ожидания случайной величины Y , при условии, что вектор входных факторов \vec{X} принял фиксированное значение x :

$$\hat{y} = M[Y|X = x] \quad (4)$$

НСМ дополняется вероятностной моделью ранжирования налогоплательщиков (ВМР) на основе критерия (1). Таким образом, необходимо построить гибридную нейросетевую модель (ГНСМ) отбора налогоплательщиков для проведения выездных проверок, которую можно записать в общем виде как

$$\Theta = F_2 \circ F_1, \quad (5)$$

где: F_2 - оператор ВМР; $\Theta = \left\{ g : \Psi = \sum_{v=1}^{G^*} \psi_{gv} \rightarrow \max_g \right\}$ - множество проранжированных на основе критерия (1) налогоплательщиков, т.е., по существу план отбора для проведения выездных проверок.

3. Оценка эффекта регуляризации

Первоначально план отбора с оценкой вероятности нарушения налогового законодательства по формуле (2) был синтезирован по результатам построения одной НСМ с заданной архитектурой многослойного персептрона с обратным распространением ошибки обучения, активационными функциями типа логистической сигмоиды в двух скрытых слоях и линейной активационной функцией в выходном слое. Для оценки эффекта регуляризации был создан байесовский ансамбль из 6 нейросетей-гипотез, принадлежащих к одному классу H - «многослойный персептрон с обратным распространением ошибки обучения и линейной активационной функцией в выходном слое». НСМ, вошедшие в ансамбль, различались архитектурой, видом активационных функций, количеством нейронов в слоях и смещениями в активационных функциях. Финальное сравнение всех планов отбора, синтезированных в данной статье, приведено в таблице 5.

Далее, для повышения качества и адекватности НСМ байесовский подход комбинируется с методом оптимальной очистки информационного пространства признаков от аномальных наблюдений [3, 4].

Первоначально (нулевая итерация очистки) строится байесовский ансамбль из Q НСМ, каждая из которых фильтруется согласно правилу:

$$(P_q^* = (N_q^* / N \geq \omega) \cap \left(\left| \frac{\bar{\delta}_i - \delta_i^q}{\bar{\delta}_i} \right| \leq \eta \right) \quad i = \overline{1, N}, \quad q = \overline{1, Q} \quad (6)$$

Здесь происходит сравнение каждой q -ой НСМ, $q = \overline{1, Q}$ в каждой точке наблюдения $i = \overline{1, N}$ с «эталоном», в качестве которого выбрано среднее значение отклонения в i -ой точке, рассчитанное по Q параллельным НСМ. В формуле (6) N_q^* - число наблюдений, для которых число отклонений от эталона не превышает экспертно задаваемого уровня ошибки η . Тем самым «хорошие» НСМ в байесовском ансамбле должны одинаково верно распознавать как большие, так и малые значения отклонений в каждом из наблюдений. Уровень вероятности «хорошего» объяснения данных ω был выбран экспертно: $\omega = 0,7$. НСМ, не прошедшие процедуру фильтрации, удаляются из ансамбля. Для НСМ, оставшихся в ансамбле, осуществляется осреднение значения выходной величины (3) и на его основе рассчитывается отклонение δ_{gr} . Наблюдения, в которых наблюдались аномально большие отклонения, удаляются из БД по правилу:

$$\delta_i = |\hat{y}_i - y_i| / y_i > \varepsilon, \quad (7)$$

где ε - допустимая погрешность.

На очищенной БД (первая итерация очистки) вновь строится байесовский ансамбль, производится фильтрация НСМ, удаляются аномальные наблюдения и т.д.

Обобщенный критерий качества НСМ Φ , уточненный нами в [2], представляет собой произведение двух частных критериев:

$$\Phi = E \cdot S \quad (8)$$

Первый критерий в (6)

$$E = \|\hat{y} - y\| / \|y\| \quad (9)$$

характеризует точность НСМ, т.е. является ошибкой обобщения. Второй критерий характеризует устойчивость НСМ, он вводится как аналог константы Липшица:

$$S = |\hat{y}_\alpha - \hat{y}_\beta| / \|\mathbf{x}_\alpha - \mathbf{x}_\beta\|_{\mathbb{R}^n}, \quad (10)$$

где: вектор-строки $\mathbf{x}_\alpha, \mathbf{x}_\beta$ близки по норме в \mathbb{R}^n на тестовом множестве НС, $\hat{y}_\alpha = F(\mathbf{x}_\alpha, W), \hat{y}_\beta = F(\mathbf{x}_\beta, W)$ - расчетные значения компонента выходной величины в точках наблюдений α, β .

Процесс очистки информационного пространства признаков останавливается, если критерий (8) достигает минимума, либо достигается допустимый уровень ошибки обобщения E (9).

В таблице 1 приведены значения P_q^* для каждой НСМ байесовского ансамбля на каждой k -ой итерации очистки кластера, жирным выделены НСМ, не попавшие в байесовский ансамбль.

Таблица 1

Результаты фильтрации гипотез (составлено авторами)

$k \backslash P_q^*$	НСМ1	НСМ2	НСМ3	НСМ4	НСМ5	НСМ6
0	0,75	0,78	0,60	0,77	0,84	0,87
1	0,72	0,65	0,66	0,83	0,69	0,72
2	0,84	0,79	0,55	0,81	0,76	0,77
3	0,87	0,82	0,43	0,80	0,79	0,82
4	0,89	0,81	0,56	0,86	0,82	0,84
5	0,80	0,68	0,63	0,81	0,79	0,85

Как видно из таблицы 1, процедуру фильтрации прошли пять НСМ из шести. НСМ 3 исключена из ансамбля как неудовлетворяющая критерию (10). На данном этапе можно считать, что адекватность НСМ подтверждается с осредненной апостериорной вероятностью

$\bar{P}^* = \frac{1}{Q^*} \left(\sum_{q=1}^{Q^*} P_q^* \right)$, где Q^* - количество НСМ, прошедших процедуру фильтрации. Таким образом, адекватность установлена с $\bar{P}^* = 0,8$.

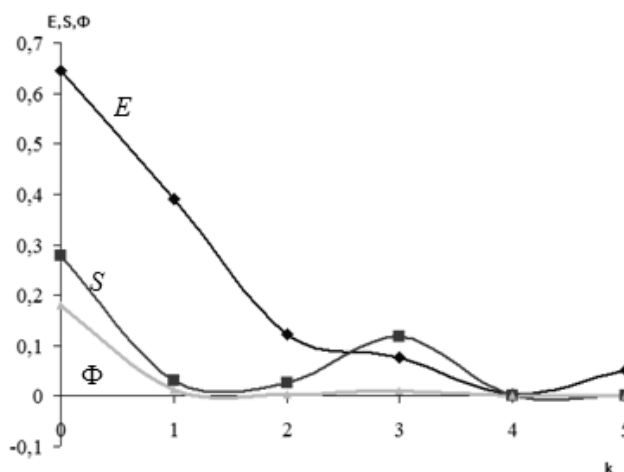


Рисунок 1. Зависимость частных критериев точности E, устойчивости S и обобщенного критерия Φ от номера итерации k (составлено авторами)

На рис. 1 и в таблице 2 представлены значения критериев качества НСМ (6)-(8) на каждой итерации. В таблице 2 также представлены: N - число наблюдений на каждой итерации, A - число аномальных по (9) наблюдений. Обобщенный показатель (6) достиг своего минимума на четвертой итерации, при этом ошибка обобщения E по (7) составила 0,3%, что говорит о хороших прогностических свойствах НСМ.

Таблица 2

Характеристики показателей качества НСМ на каждой итерации очистки кластера от аномальных наблюдений по (9)

Номер итерации, k	0	1	2	3	4	5
N	271	261	237	211	165	139
A	5	10	24	26	46	26
E	0,47	0,39	0,12	0,08	$3 \cdot 10^{-3}$	0,05

Номер итерации, k	0	1	2	3	4	5
S	0,04	0,03	0,03	0,11	$1 \cdot 10^{-3}$	$1 \cdot 10^{-3}$
Φ	0,02	0,01	$3 \cdot 10^{-3}$	0,01	$6,12 \cdot 10^{-7}$	$9,6 \cdot 10^{-5}$

Альтернативный подход по повышению однородности базы данных (БД) и качества НСМ основан на предложенной А.Л. Гусевым методики последовательного выбора части информационного пространства признаков [16]. Из БД, на каждом шаге построения НСМ, последовательно удаляются наблюдения, для которых выходная величина не удовлетворяет условию

$$m_y - l\sigma_y \leq y_i \leq m_y + l\sigma_y, \quad (11)$$

и входные факторы X_j , для которых не выполняется больше, чем M_r раз, условие

$$m_{x_j} - l\sigma_{x_j} \leq x_{ij} \leq m_{x_j} + l\sigma_{x_j}. \quad (12)$$

Здесь: m .- среднее значение, σ .- стандартное отклонение, параметры l и r последовательно уменьшаются на каждом шаге построения НСМ.

Данный подход так же использовался нами для построения НСМ с последующим синтезом плана отбора. Таблицы 3 и 4 и рис. 2 соответственно аналогичны таблицам 1 и 2 и рис. 1.

Таблица 3
Результаты фильтрации нейросетей-гипотез при предобработке данных по (11), (12)

$k \backslash P_q^*$	НСМ1	НСМ2	НСМ3	НСМ4	НСМ5	НСМ6
0	0,76	0,74	0,68	0,74	0,74	0,89
1	0,84	0,79	0,80	0,45	0,66	0,98
2	0,90	0,77	0,87	0,89	0,77	0,84
3	0,85	0,70	0,90	0,82	0,80	0,82
4	0,80	0,77	0,74	0,80	0,77	0,76
5	0,69	0,61	0,71	0,74	0,64	0,70

Как следует из таблицы 3, в байесовский ансамбль попадает 3 НСМ из 6. В таблице 4 приведено количество примеров в кластере N , количество аномальных по (11), (12) наблюдений A . Из таблицы 4 видно, что обобщенный показатель достиг своего минимума на пятой итерации, соответствующей значению параметра $l=1$ в формулах (11), (12). При этом ошибка обобщения (7) составляет 4,5%, что выше, чем для построения НСМ с удалением аномальных наблюдений по (9).

Таблица 4
Характеристики показателей качества НСМ на каждой итерации очистки кластера от аномальных наблюдений по (10), (11)

Номер итерации k	0	1	2	3	4	5
N	276,00	268,00	255,00	241,00	203,00	139,00
A	0,00	8,00	13,00	14,00	38,00	64,00
E	0,47	0,38	0,10	0,13	0,06	0,05
S	0,04	0,02	0,07	0,01	0,001	0,001

Номер итерации k	0	1	2	3	4	5
Φ	0,02	0,01	0,01	0,001	$5,16 \cdot 10^{-5}$	$3,77 \cdot 10^{-5}$

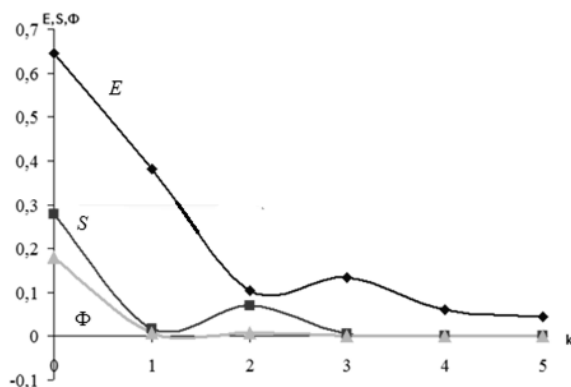


Рисунок 2. Зависимость частных критериев точности E , устойчивости S и обобщенного критерия Φ от номера итерации k при предобработке данных по (10), (11)

Таблица 5

Сравнение планов отбора налогоплательщиков для выездных проверок, синтезированных по разным моделям по критерию отбора Ψ из (1)

№ предприятия	ψ -критерий на основе одной HCM	№ предприятия	ψ -критерий на ансамбле HCM	№ предприятия	ψ -критерий на ансамбле HCM с очисткой (11)-(12)	№ предприятия	ψ -критерий на ансамбле HCM с очисткой (8)-(10)	№ предприятия	Доначисления, руб.
11	0,2259	11	0,0977	11	0,1731	11	0,1817	11	7002076
9	0,0483	9	0,0423	2	0,0796	2	0,0585	9	3140252
14	0,0251	2	0,0322	9	0,0612	9	0,0426	8	2780893
3	0,0166	14	0,0306	14	0,0417	22	0,0321	24	1911636
19	0,0151	22	0,0299	22	0,0353	21	0,0185	21	1700531
8	0,0150	21	0,0124	3	0,0345	24	0,017	22	1520399
10	0,0142	3	0,0117	8	0,0256	30	0,014	16	1427865
2	0,0130	16	0,0116	12	0,0181	14	0,0135	10	1384382
30	0,0110	8	0,0108	24	0,0176	3	0,0095	13	1371761
12	0,0088	30	0,0091	21	0,0165	8	0,0089	3	1366995
21	0,0073	24	0,0081	15	0,0151	15	0,0079	14	1041774
27	0,0047	15	0,0078	16	0,0098	16	0,0062	2	839696
18	0,0046	12	0,0070	27	0,0087	19	0,0054	15	604035
24	0,0032	19	0,0032	17	0,0060	17	0,0038	18	530011
13	0,0031	23	0,0011	13	0,0050	25	0,0021	17	522228
17	0,0011	17	0,0005	10	0,0043	12	0,0016	12	496236
22	0,0010	10	0,0004	23	0,0037	27	0,0008	27	428001
16	0,0006	26	0,0003	30	0,0036	18	0,0006	25	426376
15	0,0005	18	0,0003	19	0,0025	10	0,0005	30	411374
26	0,0005	27	0,0001	25	0,0021	23	0,0004	23	362864

№ предприятия	ψ -критерий на основе одной НСМ	№ предприятия	ψ -критерий на ансамбле НСМ	№ предприятия	ψ -критерий на ансамбле НСМ с очисткой (11)-(12)	№ предприятия	ψ -критерий на ансамбле НСМ с очисткой (8)-(10)	№ предприятия	Доначисления, руб.
25	0,0003	25	0,0001	26	0,0016	26	0,0002	19	259029
23	0,0003	7	$8,747 \cdot 10^{-5}$	18	0,0009	7	0,0002	26	83134
1	$4,178 \cdot 10^{-5}$	1	$1,967 \cdot 10^{-5}$	1	0,0002	1	0,0002	1	82674
7	$1,940 \cdot 10^{-5}$	13	$4,431 \cdot 10^{-5}$	7	$6,9320 \cdot 10^{-5}$	13	0,0001	7	49060

Составлено авторами

Результаты моделирования сравниваются с результатами выездных налоговых проверок, проводившихся на предприятиях, участвовавших в вычислительном эксперименте. В таблице 5 приводятся оптимальные планы отбора налогоплательщиков для проведения выездных налоговых проверок, синтезированные по различным НСМ. Также приводится величина налоговых доначислений, полученных по результатам выездных проверок. Из таблицы следует, что из пяти первых предприятий, допускающих крупное нарушение налогового законодательства, в план отбора, синтезированного по одной НСМ, попало только два (выделено жирным). В остальные же планы отбора попало по три предприятия.

В таблице 6 приводятся фактические доначисления, выплаченные первыми пятью предприятиями (последний столбец, первые пять строк), и доначисления, которые могли бы быть получены, если бы выездные проверки проводились по соответствующим планам отбора, а также отношения этих доначислений к фактическим.

Таблица 6

**Оценка прогностических свойств нейросетевых моделей
в аспекте прогноза доначислений**

Доначисления, руб.	Фактические	Одна НСМ	Ансамбль НСМ	Ансамбль НСМ с очисткой по (10)-(11)	Ансамбль НСМ с очисткой по (8)
	16535388	12810126	13544197	13544197	14202954
Отношение расчетных доначислений к фактическим		77,471%	81,910%	81,910%	85,894%

Составлено авторами

4. Сравнение планов выездных налоговых проверок

Таким образом, для синтеза плана выездных налоговых проверок была выбрана модель байесовского ансамбля НСМ с очисткой от аномальных наблюдений с учетом критериев точности и устойчивости (6)-(8).

Как уже было отмечено, для составления плана выездных проверок налогоплательщики ранжируются по ψ -критерию отбора (1). Коэффициент масштаба M_s в (1) вычисляется как отношение величины расходов каждого предприятия к максимальному значению по всей группе предприятий. Рассматривались различные способы вычисления критерия (1), упомянутые во введении:

1. С использованием логистической трансформации вероятности риска, согласно (2).

2. С использованием предложенного в [1] Ψ -критерия ранжирования налогоплательщиков на основе вероятностного подхода

$$\psi_g = |\delta_g| \cdot P(\delta_g > \tilde{\delta}_g) \cdot M_g, \quad \tilde{\delta}_g = M[\delta_{g,t}] + U_g, \quad (13)$$

где: g - номер предприятия; $P(\delta_g > \tilde{\delta}_g)$ - вероятность того, что текущее (по времени t) значение отклонения δ_g для g -го налогоплательщика будет больше его математического ожидания, смещенного вверх на полуширину доверительного интервала U_g ; $M[\cdot]$ - оператор математического ожидания временного ряда $\{\delta_{g,t}\}$, который считается стационарным процессом. В качестве U_g принимается половина размаха отклонения δ_g на отфильтрованном байесовском ансамбле:

$$U_g = \frac{1}{2} [\max\{\delta_{g,t,q^*}\} - \min\{\delta_{g,t,q^*}\}]_{t=t_0}; \quad q^* = \overline{1, Q^*}, \quad (14)$$

где: t_0 - момент последнего наблюдения, для которого синтезируется план выездных проверок. Вероятность $P(\delta_g > \tilde{\delta}_g)$ учитывает предысторию появления у g -го налогоплательщика больших отклонений от эталонной поверхности $\hat{Y}(\vec{X}, W, t)|_{t=t_0}$, задаваемой НС.

3. С использованием величины фрактальной размерности минимального покрытия D_μ [11, 13]:

$$\psi_g = |\delta_g| \cdot M_g / D_\mu, \quad D_\mu = 1 + \mu, \quad (15)$$

где: D_μ - размерность минимального покрытия; μ - индекс фрактальности.

4. С использованием величины фрактальной размерности на основе показателя Херста [10, 11]:

$$\psi_g = |\delta_g| \cdot M_g / D_H, \quad D_H = 2 - H, \quad (16)$$

где: D_H - величина фрактальной размерности; H - показатель Херста.

Сравним результаты моделирования с результатами выездных налоговых проверок, проводившихся на предприятиях, участвовавших в вычислительном эксперименте. В таблице 7, составленной по аналогии с таблицей 5, приведены оптимальные планы выездных проверок, согласно ψ -критериям (2), (13), (14), (15) и (16), а также величина налоговых доначислений, полученных по результатам выездных проверок. Из таблицы 7 следует, что из 5 предприятий, допускающих грубое нарушение налогового законодательства в окончательный план проверок по ψ -критериям (2), (13) и (14) попало 3, по ψ -критерию (16) попало 2 (выделены жирным).

Таблица 7

Сравнение результатов моделирования с итогами выездных проверок

№ предприятия	ψ-критерий по (2)	№ предприятия	ψ-критерий по (13)	№ предприятия	ψ-критерий по (14)	№ предприятия	ψ-критерий по (15)	№ предприятия	Доначисления, руб.
11	0,181741	11	0,184615	11	0,194917	22	0,129462	11	7002076
2	0,058505	9	0,061878	2	0,085407	11	0,072606	9	3140252
9	0,042647	2	0,037057	22	0,049059	3	0,067174	8	2780893
22	0,032078	14	0,023456	9	0,048288	9	0,059143	24	1911636
21	0,018489	24	0,023407	21	0,041217	14	0,055803	21	1700531
24	0,016981	8	0,015178	30	0,023516	2	0,039461	22	1520399
30	0,01395	21	0,014107	14	0,021583	16	0,034358	16	1427865
14	0,013479	30	0,012393	24	0,018483	21	0,018685	10	1384382
3	0,009459	3	0,010113	8	0,012293	8	0,018647	13	1371761
8	0,008948	22	0,008883	15	0,01183	15	0,018377	3	1366995
15	0,007898	15	0,008369	3	0,011197	12	0,014096	14	1041774
16	0,006152	16	0,004289	16	0,010687	30	0,013732	2	839696
19	0,005429	17	0,0033	19	0,009878	24	0,005781	15	604035
17	0,003769	25	0,0024	17	0,005314	19	0,004219	18	530011
25	0,00211	12	0,0011	25	0,002535	17	0,003694	17	522228
12	0,001581	27	0,001	12	0,002113	23	0,003388	12	496236
27	0,000765	19	0,0009	27	0,000985	10	0,001915	27	428001
18	0,00057	10	0,0008	18	0,000868	18	0,001749	25	426376
10	0,000539	18	0,0007	10	0,000819	25	0,001687	30	411374
23	0,000383	23	0,0005	23	0,000615	26	0,00128	23	362864
26	0,000214	1	0,0002	26	0,000312	13	0,000994	19	259029
7	0,000181	26	0,0002	1	0,00027	27	0,000793	26	83134
1	0,000177	13	0,0002	7	0,000247	7	$8,57 \cdot 10^{-5}$	1	82674
13	0,000104	7	0,0001	13	0,000146	1	$8,04 \cdot 10^{-5}$	7	49060

Составлено авторами

Таблица 8 составлена по аналогии с таблицей 6 и содержит значения фактических и возможных значений доначислений, которые могли бы быть полученными, если бы выездные проверки проводились по соответствующим планам отбора, а также отношения этих доначислений к фактическим.

Таблица 8

Доначисления

Доначисления, руб.	Фактические доначисления, руб.	ψ-критерий по (2)	ψ-критерий по (13)	ψ-критерий по (14)	ψ-критерий по (15)
		16535388	14202954	13935434	14202954
Отношение расчетных доначислений к фактическим		85,894%	84,276%	85,894%	85,099%

Составлено авторами

Таким образом, предложенная ГНСМ и синтезируемые на её основе планы отбора налогоплательщиков для проведения выездных проверок достаточно адекватно

идентифицируют нарушителей налогового законодательства и могут существенно повысить эффективность работы налоговой службы.

5. Выводы

1. Настоящая работа носит методологический характер. Предложенные ранее идеи, методы и алгоритмы получили подтверждение по вычислительным и натурным экспериментам.

2. При построении ГНСМ отбора налогоплательщиков для проведения выездных налоговых проверок недостаточно "лобового" построения НСМ. Требуется введение в нейросетевые модели специальных процедур, названных авторами процедурами предрегуляризации НСМ, таких как очистка БД от аномальных наблюдений, а также байесовской регуляризации. Так использование байесовского ансамбля НСМ вместо одной модели уже приводит к повышению доначислений. Синергетический эффект проявляется в комплексном использовании вышерассмотренных процедур. Таким образом, предлагаемый подход к регуляризации НСМ состоятелен.

3. Вероятность нарушения налогового законодательства налогоплательщиком можно оценить различными методами, рассчитав вероятность того, что текущее значение отклонения δ_g для g -го налогоплательщика будет больше его математического ожидания, смещенного вверх на полуширину доверительного интервала, а также вычислив фрактальную размерность ряда отклонений, или используя логистическую трансформация вероятности риска. Последний метод представляется наиболее перспективным, поскольку свободен от недостатков, связанных с априорным требованием задания закона распределения ряда отклонений и его размерности.

4. Результаты моделирования согласуются с результатами реальных налоговых проверок и свидетельствуют об адекватности построенных нейросетевых моделей. Дальнейшие исследования связаны с разработкой динамических логистических НСМ.

ЛИТЕРАТУРА

1. Букаев Г.И., Бублик Н.Д., Горбатков С.А., Саттаров Р.Ф. Модернизация системы налогового контроля на основе нейросетевых информационных технологий: Монография. - М.: Наука, 2001. - 344 с.
2. Мишустин М.В. Механизм государственного налогового администрирования в России: Дисс. ... к.э.н. 08.00.10. - М.: РЭА им Г.В. Плеханова, 2003. - 141 с.
3. Горбатков С.А., Полупанов Д.В. Методы нейроматематики в налоговом контроле: Монография. - Уфа: РИЦ БашГУ, 2008. - 136 с.
4. Горбатков С.А., Полупанов Д.В., Солнцев А.М., Фархиева С.А., Белолипцев И.И., Коротнева М.В., Рашитова О.Б. Нейросетевое математическое моделирование в задачах ранжирования и кластеризации в бюджетно-налоговой системе регионального и муниципального уровней: Монография. - Уфа: РИЦ БашГУ, 2011. - 224 с.
5. Горбатков С.А., Полупанов Д.В., Макеева Е.Ю., Бирюков А.Н. Методологические основы разработки нейросетевых моделей экономических

- объектов в условиях неопределенности: Монография. - М.: Издательский дом «Экономическая газета», 2012. - 494 с.
6. Белолипецв И.И., Горбатков С.А., Романов А.Н., Фархиева С.А. Моделирование управленческих решений в экономике в условиях неопределенности: Монография / Под ред. А.Н. Романова - М.: ИНФРА-М. - 2015. - 299 с.
 7. Горбатков С.А., Белолипецв И.И., Макеева Е.Ю. О моделях диагностики банкротств организаций // Менеджмент и бизнес-администрирование. - 2014. - №1. - С. 151-172.
 8. Горбатков С.А., Белолипецв И.И., Солнцев А.М. О состоятельности процедур регуляризации нейросетевых моделей налогового и финансового контроля // Информационные технологии. - 2013. - №11. - С. 64-68.
 9. Нариньяни А.С. Недоопределенные модели и операции с недоопределенными значениями // Препринт ВЦ СО АН СССР. - 1982. - №400. - 30 с.
 10. Горбатков С.А., Белолипецв И.И., Фархиева С.А. Использование методов фрактальной теории при ранжировании объектов налогового контроля // Научная сессия НИЯУ МИФИ - 2013. XV Всероссийская научно-техническая конференция «Нейроинформатика - 2013»: Сборник научных трудов. В 3-х частях. Ч.3. - М.: НИЯУ МИФИ, 2013. - С. 184-192.
 11. Петерс Э. Хаос и порядок на рынках капитала. - М.: Издательство «Мир», 2000. - 337 с.
 12. Дубовиков М.М., Крянев А.В., Старченко Н.В. Размерность минимального покрытия и локальный анализ фрактальных временных рядов // Вестник РУДН. - 2004. - Том 3. - №1. - С. 81-95.
 13. Тихонов А.Н., Леонов А.С., Ягола А.Г. Нелинейные некорректные задачи: Монография. - М.: Наука, 1995. - 312 с.
 14. Шумский С.А. Байесова регуляризация обучения // Лекции школы-семинара «Современные проблемы нейроинформатики». Часть 2. (23-25 января 2022 г., Москва). - М.: МИФИ, 2002. - с. 61-94.
 15. Шевченко И.В., Халафян А.А., Васильева Е.Ю. Создание виртуальной клиентской базы для анализа кредитоспособности российских предприятий // Финансы и кредит. - 2010. - №1 (385). - С. 13-18.
 16. Гусев А.Л. Выбор информационного пространства для построения нейронной сети как модели управления в условиях зашумления и неполноты данных // Нейрокомпьютеры: разработка, применение. - 2011. - №7. - С. 55-77.

Gorbatkov Stanislav Anatol'evich

Financial university under the government of the Russian Federation
Ufa (branch), Russia, Ufa
E-mail: sgorbatkov@mail.ru

Polupanov Dmitrii Vasil'evich

Bashkir state university, Russia, Ufa
E-mail: demetrious@mail.ru

Beloliptsev Iliа Igorevich

Financial university under the government of the Russian Federation
Ufa (branch), Russia, Ufa
E-mail: red7315@gmail.com

Farkhieva Svetlana Anatolievna

Financial university under the government of the Russian Federation
Ufa (branch), Russia, Ufa
E-mail: ok-xi@yandex.ru

Neuronet methods of selection of taxpayers for conducting on-site inspections

Abstract. This article summarizes the authors' research on the regularization of inverse problems of neural network reconstruction of multifactor nonlinear dependencies hidden in data. One of the practical applications is, for example, the tasks of tax control, including the selection of taxpayers for on-site tax audits. Much attention is paid to the almost inadequate question of justifying the adequacy of the neural network models being created. The proposed theoretical ideas were approved by the authors in computational experiments on real data of enterprises-taxpayers. The authors consider that the developed approaches, methods and neural network models can be used also in a number of related areas: municipal budgeting, diagnostics and forecasting of bankruptcy of enterprises, monitoring of economic security of corporations in the aspect of their financial and economic state, etc.

Keywords: tax legislation; economic security; neural network model; regularization; model adequacy; nonlinear dependencies; Bayesian approach