

УЧЕТНАЯ ПОЛИТИКА И СТАТИСТИЧЕСКИЕ ИЗМЕРЕНИЯ

УДК 336. 004+519.7

РАЗРАБОТКА СТРУКТУРНО-ФУНКЦИОНАЛЬНОЙ СХЕМЫ ИНТЕЛЛЕКТУАЛЬНОЙ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКОЙ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПО НАЛОГОВОМУ АДМИНИСТРИРОВАНИЮ

DEVELOPMENT OF A STRUCTURAL-FUNCTIONAL DIAGRAM INTELLIGENT INFORMATION-ANALYTICAL SYSTEM SUPPORTING DECISION-MAKING ON TAX ADMINISTRATION

А.Н. Бирюков
A.N. Biryukov

*Федеральное государственное бюджетное образовательное учреждение высшего образования
Стерлитамакский филиал «Башкирский государственный университет»,
Россия, 453103, Республика Башкортостан, г. Стерлитамак, проспект Ленина, 49*

*Federal state budgetary educational institution of higher professional education Sterlitamak branch
"Bashkir state University", Russia, 453103, Republic of Bashkortostan, g. Sterlitamak, Lenin Avenue, 49*

e-mail: biryukov_str@mail.ru

Аннотация. Вопросы, рассматриваемые в публикации, возникли в связи с объективной необходимостью проведения исследований, направленных на повышение эффективности работы налоговых органов.

Государство не может тратить большие средства на сбор налогов, поэтому сама структура налогов и государственная система налогового администрирования, обеспечивающие их сбор, должны при минимальных затратах обеспечивать высокую эффективность работы, которая невозможна без хорошей информационно-аналитической системы поддержки принимаемых решений.

Система электронной обработки данных имеет одно узкое место – в ней слабо формализован аналитический блок, ядром которого должна служить математическая модель анализа финансово-экономического состояния налогоплательщиков, выявления нарушений налогового законодательства в декларациях, синтеза оптимального плана выездных налоговых проверок. Такие модели должны служить достаточно достоверной и объективной основой для поддержки принятия управленческих решений.

В системе налогового администрирования возникают задачи выявления «образа» нарушения налогового законодательства в декларациях. Для разработки решений по налоговому администрированию требуется также решение задачи кластеризации налогоплательщиков. Во всех этих случаях при моделировании необходимо решать задачу восстановления многомерных нелинейных экономических зависимостей, скрытых в данных.

В настоящее время существует мощный математический инструмент (универсальный аппроксиматор) – нейронные сети.

Предложенный системно – информационный подход создал методологическую основу новой компьютерной технологии и способность интеллектуальной нейросетевой модели извлекать новые знания о кластере налогоплательщиков для налогового администрирования.

Resume. At the macro level, tax administration is a large – scale sustainable functioning of the aggregate of the tax mechanism during the forecasting, regulation and control over tax revenues, unifying the legal framework of taxation and financial achievements of science and economic practices.

The actual problem is improvement of the technology of tax administration. Current activities of the taxpayers in the Russian economy is characterized by uncertainty external and internal environment. The result of these trends is the scatter of the output parameters of economic activities of organizations, which in many cases predetermines a high risk of inefficiency of tax audits. Therefore, the transition to a mathematical formalization of the stages of decision-making faces a number of difficulties associated with the problem of modeling poorly formalized systems.

The issues addressed in the publication, arose with the objective need for research aimed at improving the efficiency of tax authorities.



The state can't spend money on tax collection, therefore, the structure of taxes and the state tax administration, to ensure their collection, needs at minimum cost to provide a high efficiency, which is impossible without good information-analytical systems of support of decisions.

The data-processing system has one bottleneck – it is poorly formalised analysis unit, the core of which should serve as a mathematical model analysis of economic and financial condition of the taxpayer, of violations of the tax legislation in the declarations, the synthesis of an optimal plan of exit tax checks. Such model should be sufficiently reliable and objective basis to support management decisions.

In the system of tax administration arise with identifying "image" of violations of the tax legislation in the declarations. To develop solutions on tax administration also required the solution to the problem of clustering taxpayers. In all these cases, simulations are needed to solve the problem of reconstruction of multivariate nonlinear economic relationships hidden in the data.

Currently, there are powerful mathematical tools (universal approximator) is a neural network.

The proposed system – information approach has created a methodological basis of new computer technologies and the ability of intelligent neural network model to extract new knowledge about the cluster of taxpayers to tax administration.

Ключевые слова: система налогового администрирования; налогоплательщик; нейросетевая модель; база данных деклараций; экспертная система.

Keywords: the system of tax administration; taxpayer; neural network model; database declarations; expert system.

Введение

На макроуровне налоговое администрирование – это крупномасштабный процесс обеспечения устойчивого функционирования совокупного налогового механизма в ходе прогнозирования, регулирования и контроля за налоговыми поступлениями в бюджет, объединяющий правовые основы налогообложения, достижения финансовой науки и экономической практики.

Актуальной проблемой остается совершенствование технологии налогового администрирования. Современная деятельность налогоплательщиков в условиях кризиса российской экономики характеризуется неопределенностью внешней и внутренней среды. Результатом этих тенденций является разброс выходных параметров экономической деятельности организаций, что во многих случаях предопределяет высокий риск неэффективности налоговых проверок. Поэтому переход к математической формализации этапов принятия решений наталкивается на ряд трудностей, связанных с проблемой моделирования плохо формализуемых систем. Одна из таких трудностей в обратных задачах восстановления зависимостей (интерпретации), которые относятся к классу некорректных (или плохо определенных) – это обеспечение устойчивости решения в сложных условиях моделирования.

Объектом исследования в статье являются применение новых информационных технологий и математических методов в аспекте модернизации налогового администрирования на основе нейросетевого моделирования.

Результаты исследований

1. Комбинированный системно-синергетический-информационный подход к исследованию сложных математико-информационных систем

1.1. Условия моделирования

Прежде всего, поясним смысл термина «математико-информационная система». В самом общем виде нейросетевая модель может быть записана в виде:

$$\hat{Y} = F(\bar{X}, W); \bar{X} \in R^n \xrightarrow{F} \bar{Y} \in R^m, \quad (1)$$

где \hat{Y} – вектор выходных (расчетных) величин; \bar{X} – вектор входных сигналов нейросети (объединяющих переменных или факторов); W – матрица синаптических весов связей между нейронами; R^n , R^m – пространства вещественных чисел; n , m – размерность векторов входа и выхода сети; $F(\bullet)$ – оператор нейросетевого отображения пространства \bar{X} в пространстве \bar{Y} (рис. 1) [Эндрю, 1985].

Поскольку нейросеть отображает пространство одних функций в пространстве других функций, то нейросетевая модель – это математическая модель, заданная сложной композицией ряда операторов суммирования сигналов с синаптическими весами, нелинейного отображения результатов весового суммирования с помощью активационных функций, градиентного поиска при нахождении весов W в процессе обучения и др.

Следовательно, нейросетевая модель (НСМ) как инструмент отображения сути математическая интеллектуальная (обучаемая) модель [Хайкин, 2006].

С другой стороны, в НСМ согласно системе правил, предписанных программой для компьютера, происходит извлечение информации из данных, структурирование и ее дальнейшее преобразование. В этом смысле НСМ – это информационная система, состоящая из модулей преобразования информации. В нейроэмуляторах (программах) – это программные модули. В нейросетях типа реальных микросхем блоки – это выделенные электронные модули, содержащие нейросетевой блок (нейроплаты), выполняющие типовые нейрооперации.

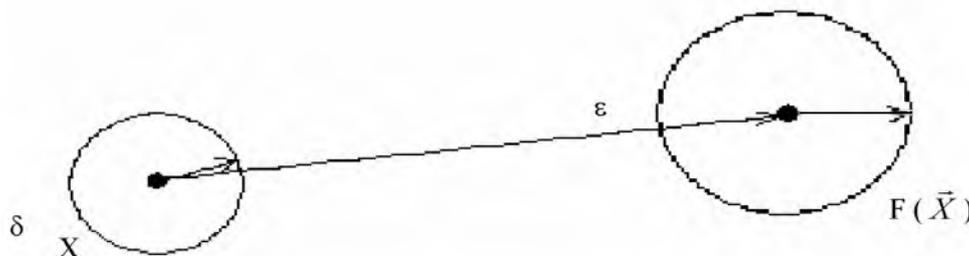


Рис. 1. Отображение входной области (шара δ) в выходной диапазон ε
 Fig. 1. The display of the input region (ball δ) in the output range ε

Поэтому в общем случае, т.е. без конкретизации контекста рассмотрения задачи, НСМ – это «математико-информационная система», причем довольно сложная (не по количеству модулей в ней, а по характеру связей между ними)[Ясницкий, 2016].

Для выработки управленческих решений в процессах налогового администрирования необходимы модели аппроксимации, кластеризации и ранжирования налогоплательщиков по показателям их финансового состояния.

Построение таких моделей является непростым делом, поскольку условия моделирования очень сложные:

- моделируемая система характеризуется наличием в базе данных сильного «зашумления» и даже сознательных искажений («приписок») в силу субъективных тенденций планирования и отчетности, т.е. стремления налогоплательщиков понизить налогооблагаемую базу;

- моделируемые системы налогового администрирования относятся к классу трудно формализуемых, особенно на этапе спецификации переменных и принятия решений с использованием моделей;

- моделируемая система, как впрочем, большинство экономических систем, отличается дефицитом наблюдений в базе данных, которые обновляются лишь в отчетные периоды (один раз в квартал), заметим, что для применяемого в работе инструментария – нейронных сетей это наиболее осложняющее обстоятельство;

- существенное влияние на деятельность систем налогового администрирования оказывают случайное (неконтролируемое) изменение внутренней структуры моделируемой системы, а также законодательства за период формирования базы данных;

- существует сильная зависимость показателей деятельности налогоплательщиков от случайных воздействий внешней экономической, политической и социальной среды.

Перечисленные условия потребовали комбинированного системного - синергетического - информационного подхода к разработке математических моделей с привнесением новой технологии в разработку алгоритма отдельных вычислительных операций с применением нейросетевого моделирования [Бирюков и др. 2012].

Детализируем сущность разработанных подходов.

1.2. Методические основы разработки системы налогового администрирования и место в ней гибридной нейросетевой модели

Под внутренней структурой, или просто структурой системы мы будем понимать для простоты организацию системы, т.е. пространственное расположение элементов, систему отношений элементов, совокупность устойчивых межэлементных связей, внутреннее устройство, закон взаимодействия [Hayes-Roth F., Jacobstein N.. 1994].

Под пространственной организацией понимается морфологический анализ базы данных деклараций (БДД) с образованием исходного однородного кластера в рамках налоговой инспекции. Эти границы очерчены требованием синтеза оптимального плана выездных проверок на уровне налоговых инспекций, т.е. по территориальному принципу.



В рассматриваемой системе налогового администрирования (СНА), показанной на рис.2, наименьший структурный элемент-фрагмент налоговой декларации, содержит весь набор компонентов вектора входных факторов \vec{X} согласно принятой спецификации переменных и взаимно однозначно соответствующее этим переменным значение выходной случайной величины Y .

При переходе от традиционной системы налогового контроля и управления (СНКУ) с электронной системой обработки данных (ЭОД) к новой, структурный элемент проходит эволюцию в системе, степень открытости которой возрастает. Связь между элементами (в информационном аспекте, т.е. в процессе порождения новых знаний) становится все более сложной и жесткой по мере усовершенствования информационной модели в системе налогового контроля и управления (СНКУ) в ходе научного исследования. В пределе, когда степень открытости достигает максимального значения (соответственно, адекватность математической модели достигнет требуемого уровня), поведение всех структурных элементов в алгоритме обучения нейросети – «ядро системы» – будет близким к детерминированному. Реально указанный предел недостижим, ибо сама природа системы налогового администрирования (СНА) стохастическая.

Отметим, что в кибернетике одной из важнейших задач оптимизации системы является достижение максимальной организованности при заданной сложности либо уменьшение сложности при данном уровне организованности.

В новой технологии налогового контроля, планирования и управления имеет место первый случай: из заданного количества элементов построить оптимальную систему, т.е. достичь максимальной организованности.

Структурные закономерности и свойства системы

При формулировке понятия «система», интегративные ее свойства обусловлены не только свойствами самих структурных элементов системы, но и зависят от взаимодействия внутренних элементов системы, так и внешней среды.

В этом смысле «внешний мир» формирует свойство системы. В самом деле, если во «внешнем мире» исчезает «потребитель» информации о каком-нибудь свойстве системы, то исчезает и само свойство системы, так как в отношении этого свойства система окажется изолированной (закрытой) и по закономерности возрастания энтропии в замкнутой системе в ней возникают процессы, дезорганизующие структуры, ответственные за данные свойства [Zimmermann H.-J. 1996].

С другой стороны, появление у системы некоторого нового свойства порождает информацию о свойстве, и, если во внешней среде имеется «потребитель» этой информации, в нашем случае экспертная система налогового администрирования (ЭСНА) (на рис.2), который способен воспринимать ее и тем самым взаимодействовать с системой, то по этому свойству система окажется открытой. Тогда закон убывания энтропии начнет преобладать над законом роста энтропии, и это свойство разовьется до некоторого оптимального уровня, которое определяется числовым значением степени открытости системы.

В системе налогового контроля и управления (СНКУ) упомянуто первое новое свойство – это способность интеллектуальной нейросетевой модели (НСМ) извлекать новые знания об усредненной «производственной функции» кластера налогоплательщиков из преднамеренно искаженных данных налоговых деклараций и на основе этих знаний идентифицировать нарушения налогового законодательства. Второе свойство – это оптимальным образом с заданной доверительной вероятностью производить отбор налогоплательщиков для выездных проверок.

Рассмотрим теперь вопрос о взаимосвязи новых свойств и структуры системы. Известно [Бирюков. 2011а, 2011б], что свойства системы порождаются структурными закономерностями системы. Следовательно, в системе образуются только такие структурные закономерности, которые через порождаемые свойства системы имеют внешнего потребителя. Структурные закономерности системы – это наиболее общие закономерности, которые порождают свойство системы как целого, т.е. свойства, по которым внешняя среда выделяет систему среди других.

Переработка информации приводит к сравнению и оценке действительного и желаемого «эталонного» состояния системы, описанного производственной функцией (1), в результате чего появляется новая информация, необходимая для выполнения управленческого акта, ведущего к полезному конечному результату в процессе управления. Команда управления требует ее исполнения, т.е. осуществления управленческих воздействий. Обычно в системах команды управления должны быть преобразованы в вещественно-энергетическое, финансовое или какое-то иное воздействие на объект управления \vec{U}_g , где g – номер предприятия – налогоплательщика.

В рассматриваемой системе налогового администрирования (СНА) мы определили границу системы на выходе блока экспертной системы налогового администрирования (ЭСНА), принимающая управленческие решения о налоговых проверках на базе синтезированного оптимального плана выездных проверок, а также по налоговому

планированию и регулированию. Эти решения и есть управляющие воздействия на субъекты налогообложения. Они возможны, в частности, в виде:

- реализации выездной проверки на данном субъекте;
- отправка «профилактического» уведомления о нарушении;
- предоставление налоговых льгот либо, наоборот, наложение санкций и др.;
- включения синтезированного плана выездных проверок в план работы системы налогового планирования (СНП).

Для этой цели исполнительный орган – лицо, принимающее решения (ЛПР), разрабатывающий управляющие воздействия на объекты управления (вещественно-энергетические, финансовые или информационные), должен располагать необходимыми законодательными ресурсами.

Именно в этом проявляется единство информационных и вещественно-энергетических процессов при управлении и возможна состоятельная постановка задачи оптимального управления, а именно: выработать такие управления, которые приводили бы к максимальной ожидаемой сумме доначислений при ограничениях на затраты на контролеобеспечивающие ресурсы.

2. Описание работы структурно-функциональной схемы системы налогового администрирования

Вначале укажем назначение всех подсистем и блоков, а затем опишем их работу во взаимодействии.

Цель функционирования системы налогового администрирования (СНА) – порождение интегративного эффекта – объективной классификации совокупности субъектов налогообложения на «нарушителей» и «не нарушителей» налогового законодательства, их ранжирование и синтез на этой основе управляющих воздействий \bar{U}_{N_3} и \bar{U}_{N_2} , где N_2 – количество субъектов, подвергаемых процедуре выездных проверок и штрафных санкций; $N_3=N_1-N_2$ – количество субъектов, подвергнутых профилактическим воздействиям. Другими словами, синергетический эффект в СНА состоит в восстановлении по совокупности исходных данных деклараций примерно однородных кластеров налогоплательщиков «обобщенной производственной функции» $\hat{y}(\bar{x})$ и, затем на ее основе, выявление среди субъектов налогообложения нарушителей налогового законодательства.

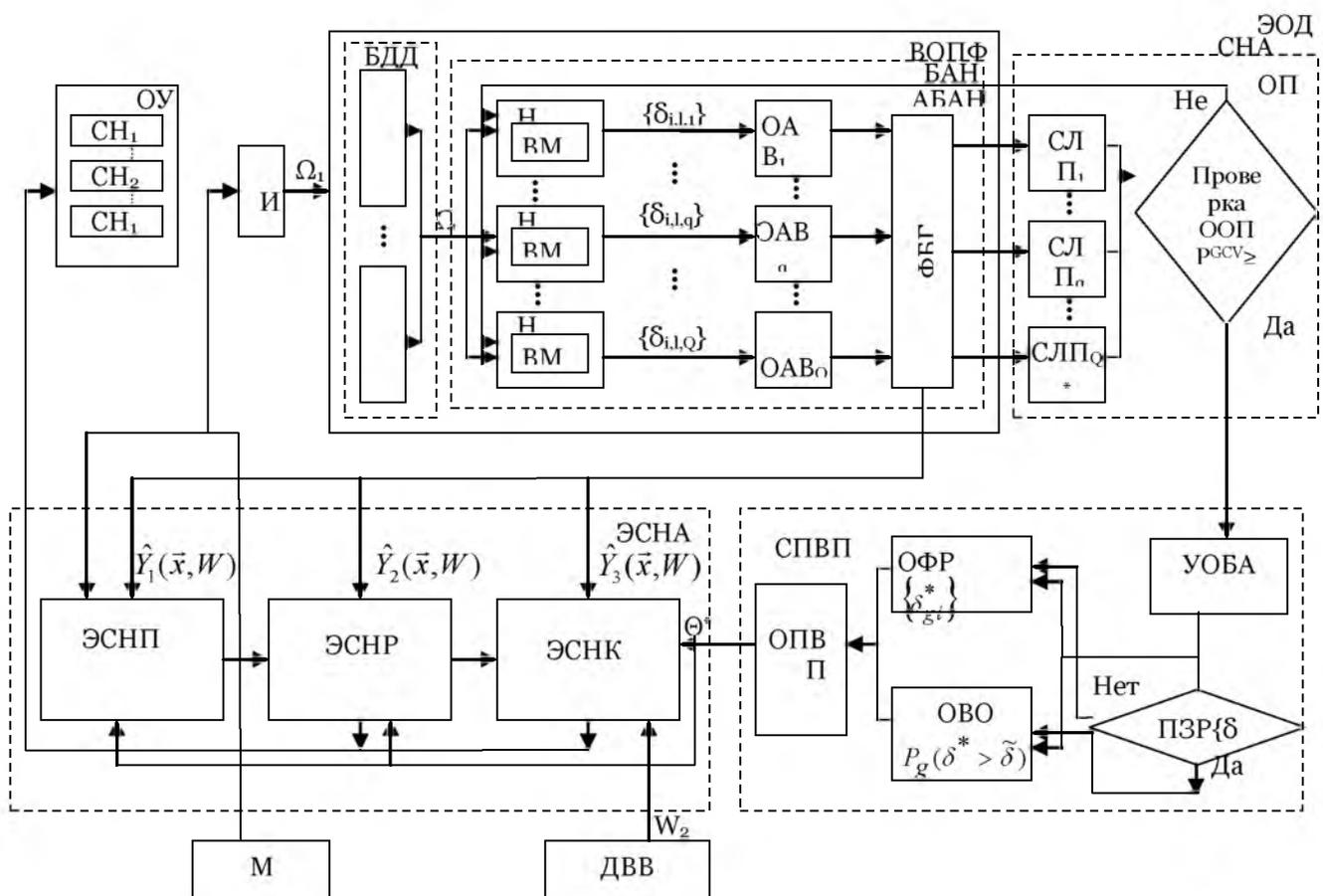


Рис. 2. Структурно-функциональная схема интеллектуальной информационно-аналитической системы поддержки принятия решений в СНА

Fig. 2. Structural-functional diagram of intelligent information - analytical system supporting decision-making in SLEAP



Опишем структурные связи и назначения над- и подсистем рассматриваемой интеллектуальной информационно-аналитической системы поддержки принятия решений по налоговому администрированию (ИСНА) (рис.2), а затем детализируем работу ее подсистем и блоков. На рисунке сплошными линиями показаны блоки алгоритмов, а пунктирными линиями – их условное объединение в подсистемы.

Потребителями структурированной информации, порождаемой на выходе ИСНА, являются соответствующие экспертные системы подсистем налогового контроля, налогового планирования и налогового регулирования, которые входят в состав системы налогового администрирования (СНА):

- ЭСНП – экспертная система налогового планирования;
- ЭСНР – экспертная система налогового регулирования;
- ЭСНК – экспертная система налогового контроля.

В структуре ИСНА на рис.2, подсистему, ЭСНК обслуживают все подсистемы ИСНА: ВОПФ, ОПП, СПВП, где введены обозначения:

- ВОПФ – подсистема восстановления «обобщенной производственной функции» кластера налогоплательщиков на основе данных деклараций с помощью нейросетевых моделей;

- ОПП – подсистема обобщенного перекрестного подтверждения, т.е. финишной (интегральной) проверки адекватности гибридной нейросетевой модели (ГНСМ) на байесовском ансамбле нейросетей;

- СПВП – подсистема синтеза планов выездных налоговых проверок.

Подсистемы ЭСНП и ЭСНР обслуживает одна подсистема – ВОПФ.

Надсистемой для ИСНА является система электронной обработки данных (ЭОД) [Бирюков, 2014], действующая в настоящее время в Федеральной налоговой службе РФ.

Назначением ИСНА в целом является порождение объективной и хорошо структурированной информации об экономических закономерностях, скрытых в зашумленных данных, для решения прикладных задач в подсистемах ЭСНП, ЭСНР, ЭСНК налоговых систем бюджетных уровней.

Отметим, что в статье исследованы не все функции подсистем налогового администрирования (планирования, контроля, регулирования) и не с одинаковой степенью детализации. Основное внимание уделено подсистеме налогового контроля.

В ИСНА также входят:

- ОУ – объект управления, который включает в себя совокупность субъектов налогообложения, т.е. предприятий – налогоплательщиков $\{CH_g\}, g = \overline{1, G}$.

- М – монитор системы ЭОД, который служит для сбора, хранения и структурирования информации о налогоплательщика с целью контроля и управления.

- И – интерфейс, который служит для формирования потока данных Ω_i , т.е. совокупности деклараций о налогоплательщиках.

В подсистему ВОПФ входят подсистемы ниже следующего уровня:

- БДД – база данных деклараций для множества $\{CH_g\}_{g=1}^G$ за некоторый ретроспективный отрезок времени $t_k \in [t_1; t_T]$, включая СУБД – систему управления базой данных.

- БАН - подсистема построения байесовского ансамбля нейросетей, которая предназначена для восстановления нелинейной многомерной «обобщенной производственной функции» $Y(\vec{x})$ скрытой в данных, и порождения отклонений $\{\delta_{g,i}\}, g = \overline{1, N_1}; i = \overline{1, N}$, где g – номер предприятия – налогоплательщика, между расчетными и декларированными данными.

- АБАН – подсистема регуляризации и оценки адекватности нейросетевой модели на байесовском отфильтрованном ансамбле нейросетей. Эта подсистема включает в себя блоки oAV_1, \dots, oAV_Q оценки апостериорной вероятности гипотез-нейросетей $\{q_q(\vec{x}, W) | H\}$ и блок фильтрации этих нейросетей, т.е. сужение априорного ансамбля.



- ОПП – подсистема обобщенного перекрестного подтверждения, которая служит для финишной (интегральной) оценки адекватности комплексной модели.

Блоки СЛП₁, ..., СЛП_Q, входящие в подсистему ОПП, служат для синтеза локальных (для каждой q-ой нейросети $\{t_q^*(\bar{x}, W)\}$ в отфильтрованном ансамбле) планов выездных налоговых проверок.

Логический блок «Проверка ОПП» служит для оценки условия взаимного подтверждения синтезированных локальных планов выездных проверок.

СПВП – подсистема синтеза единого плана выездных проверок Θ^* , усредненного на отфильтрованном и прошедшем ОПП байесовском ансамбле нейросетей.

В подсистему СПВП входят блоки:

- УОБА – блок усреднения отклонений $\{\delta_{g,t,q}\}, g = \overline{const}, t = \overline{1, T}; q = \overline{1, Q^*}$ на отфильтрованном и прошедшем процедуру ОПП байесовском ансамбле.

- ПЗР – блок проверки соответствия усредненных отклонений $\{\delta_{g,t}\}, g = \overline{const}, t = \overline{1, T}$ нормальному закону распределения плотности вероятности.

- ОВО – блок оценки вероятности $P_g(\bar{\delta}^* > \tilde{\delta})$ в критерии ранжирования налогоплательщиков Ψ_g по (2) на основе функции Лапласа:

$$\psi_g = \tilde{\delta}_{g,t} \Big|_{t=t_0} \cdot P(\delta_g \geq \tilde{\delta}) \Big|_{t_0} \cdot M_g, \tag{2}$$

где $\tilde{\delta}_{g,t} = \delta_{g,t} + U$ – значение верхней границы доверительного интервала для отклонения $\delta_{g,t}$ ($\delta_{g,t} = \delta_i$), т.е. в записи отклонения фиксируется номер налогоплательщика g и момент наблюдения t); $P(\delta_g \geq \tilde{\delta}) \Big|_{t_0}$ – вероятность события, что ожидаемое значение отклонение δ_g

моделируемой случайной величины Y будет больше выборочного среднего $\bar{\delta}$ с учетом его смещения на полуширину доверительного интервала для $\{\delta_g\}$; t_0 – момент времени, соответствующий последним данным декларации при ранжировании налогоплательщиков; M_g – экспертно задаваемый коэффициент масштаба g –го налогоплательщика.

- ОФР – блок оценки фрактальной размерности временного ряда остатков $\bar{\delta}_g^*(t), t = \overline{1, T}$ в критерии ранжирования налогоплательщиков Ψ_g .

- ДВВ – блок данных внешних ведомств, который служит для сбора дополнительной информации о налогоплательщиках (от таможенной службы, транспортных ведомств, администраций регионов и муниципальных образований, правоохранительных органов и др.). Эти дополнительные данные могут быть использованы в ЭСНК лицом, принимающим решения (ЛПР), для цензурирования синтезированных планов выездных проверок Θ^* , а также цензурирования вектора регулирующих (управляющих) воздействий $\bar{U}_g, (\bar{U} = U_1, \dots, U_m)$, прикладываемых к ОУ.

Схема рис.2 уже дает достаточно подробное представление о прикладных аспектах, разработанных в работе подходов, методов, алгоритмов. Остается дополнить описание схемы взаимодействием подсистем и блоков при обработке информации.

Поток данных деклараций Ω_1 совокупности налогоплательщиков $g = \overline{1, G}$ извлекается из интерфейса ЭОД и подается в подсистему БДД в исходном (не связанным с форматом конкретного программного продукта) формате.

В БДД в зависимости от используемого для построения нейросетей программного продукта (NeuroSolutions, BrainMaker Pro, MATLAB и др.) с помощью СУБД формируется поток данных Ω_2 в формате, требуемом в данном программном продукте. Например, в программном продукте Neuro Solutions – это таблица, вектор – строки которой образуют кортежи, соответствующие данному наблюдению с текущим сквозным номером i:

$$\langle \bar{x}_{g,t}, y_{g,t} \rangle_i; i = \overline{1, N}; g = \overline{1, G}; t \in \overline{1, T}. \tag{3}$$



В кортежах вектор входных факторов, $\vec{x}_{g,t}$ измеренный в момент времени t для предприятия – налогоплательщика с номером g взаимно однозначно соответствует значению выходного показателя $y_{g,t}$, для простоты изложения принятого скалярным. В работе использовались данные панельного типа [Романов и др., 2010], в которых кортежи упорядочены как по времени, так и по номеру. Следовательно, каждому фиксированному предприятию ($g=\text{const}$) соответствуют T временных отсчета ($i \in [1, N]$, $N = G \cdot T$).

В подсистеме БАН согласно методу вложенных математических моделей (МВММ), с помощью совокупности алгоритмов (I.1 – I.7 из рис.3) предпроцессорной обработки «сырых» данных, подступающих из БДД, строятся вспомогательные вложенные нейросетевые субмодели ВММ₁, ..., ВММ_q, ..., ВММ_Q соответственно для каждой q -ой априорной нейросети – гипотезы $h_q(\vec{x}, W | H)$ байесовского ансамбля. В основных нейросетевых моделях НСМ₁, ..., НСМ_q, ..., НСМ_Q ансамбля реализуется байесовская регуляризация обучения сетей в соответствии с приближенным методом байесовской регуляризации (алгоритмы II.1 – II.6 из рис.3).

При этом в соответствии с общесистемной фоновой закономерностью полезным сигналом, отраженным от фона, является относительное отклонение $\{\delta_t\}$, вычисленные по (4):

$$\delta_t = \left| \frac{y_t - \hat{y}_t}{y_t} \right| \cdot 100\% \quad , \quad (4)$$

и с выходов сетей ансамбля подаются в блоки оценки апостериорной вероятности ОАВ₁, ..., ОАВ_Q. Здесь вычисляются вероятности $\{P_q\}_{q=1}^{q=Q}$; $P_q = (N_q^* / N)$, где N_q^* - число хорошо объясненных точек множества N данных в q -ой сети ансамбля. Затем в блоке фильтрации байесовских гипотез (ФБГ) осуществляется отсев (фильтрация) нейросетей - гипотез с неприемлемым качеством объяснения данных на уровне заданного параметра ξ .

В блоке ФБГ хранятся необходимые выходные величины уже отфильтрованного ансамбля нейросетей, в частности оценки апостериорных вероятностей $\{P_q^*\}$, параметры регуляризации ξ , отклонений $\{\delta_{g,i,t}^*\}_{q=1, \dots, Q^*}$; $Q^* < Q$; $i = \overline{1, N}$; $g = \overline{1, G}$. Здесь и далее звездочки над величинами соответствуют отфильтрованному ансамблю.

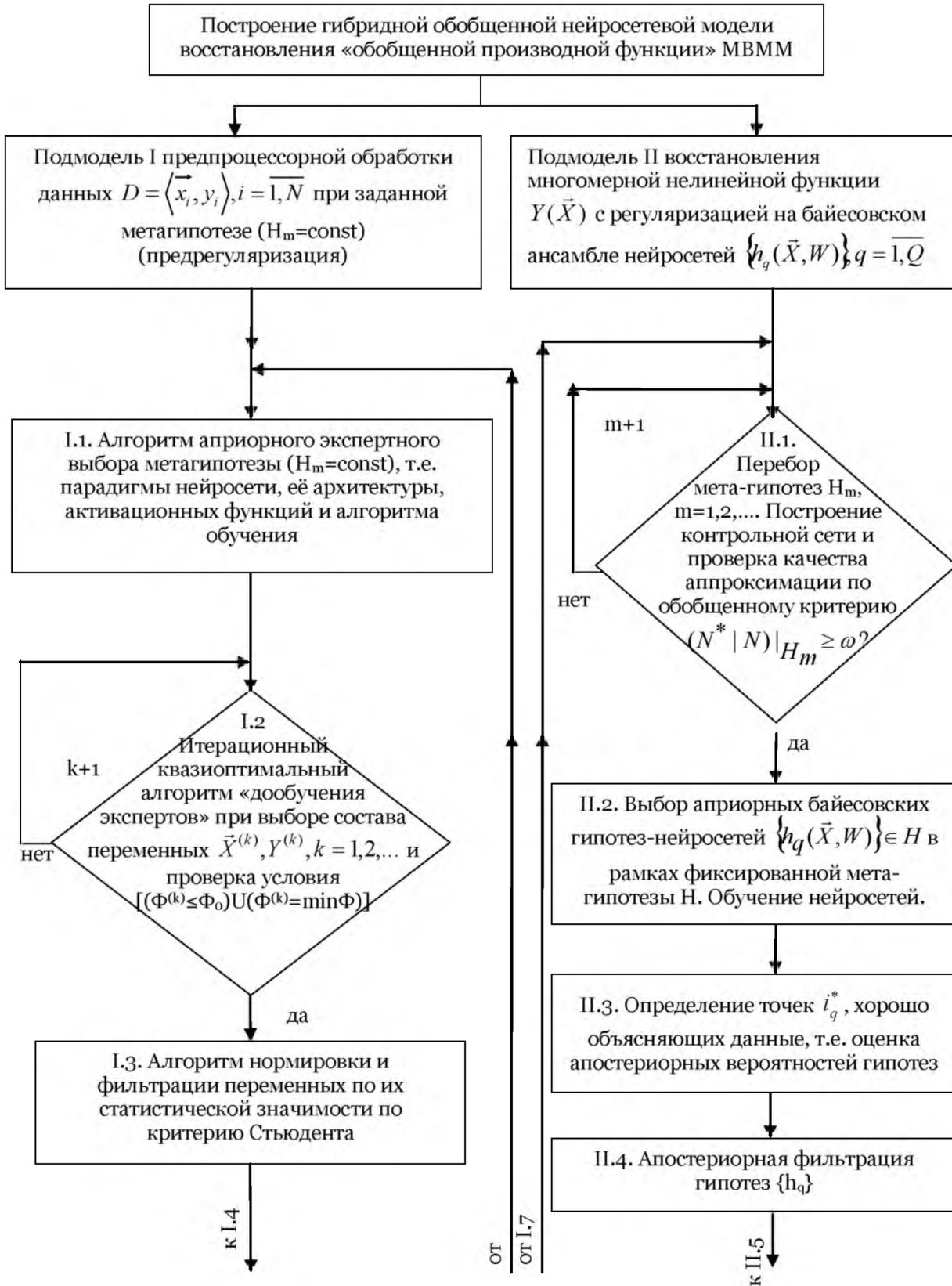
Отклонения $\{\delta_{t,q,g}^*\}$ извлекаются из блока ФБГ и подаются в подсистему ОПП. Здесь в блоках СЛП₁, ..., СЛП_q, ..., СЛП_{Q^*} производится синтез локальных планов выездных налоговых проверок. Термин «локальных» здесь означает, что план синтезируется для каждой отдельной q -ой нейросети – гипотезы, прошедшей процедуру фильтрации.

В блоке «Проверка ОПП» сравниваются локальные планы и степень совпадения в них множества налогоплательщиков, отобранных для первоочередной выездной налоговой проверки, т.е. реализуется процедура ОПП по финишному критерию P^{GCV} .

Сети, прошедшие процедуру ОПП, используются далее в подсистеме СПВП для получения, усредненного на ансамбле нейросетей рабочего плана выездных проверок. В блоке УОБА отклонения $\{\delta_{t,q,g}^*\}$ усредняются на отфильтрованном и прошедшем процедуру ОПП байесовском ансамбле нейросетей.

Полученное множество отклонений $\{\bar{\delta}_{t,g}^*\}$ сортируется с образованием временных рядов $\{\bar{\delta}_t^*\}_{g=1, 2, \dots, T}$ для каждого фиксированного налогоплательщика. В блоке ПЗР проверяется непротиворечие закона распределения плотности вероятности отклонений $\{\bar{\delta}_t^*\}_g$ нормальному закону.

В случае неподтверждения этой гипотезы в блоке ОВО производится оценка вероятности $P_g(\bar{\delta}^* > \bar{\delta})$, используемой в критерии отбора $\Psi(2)$.



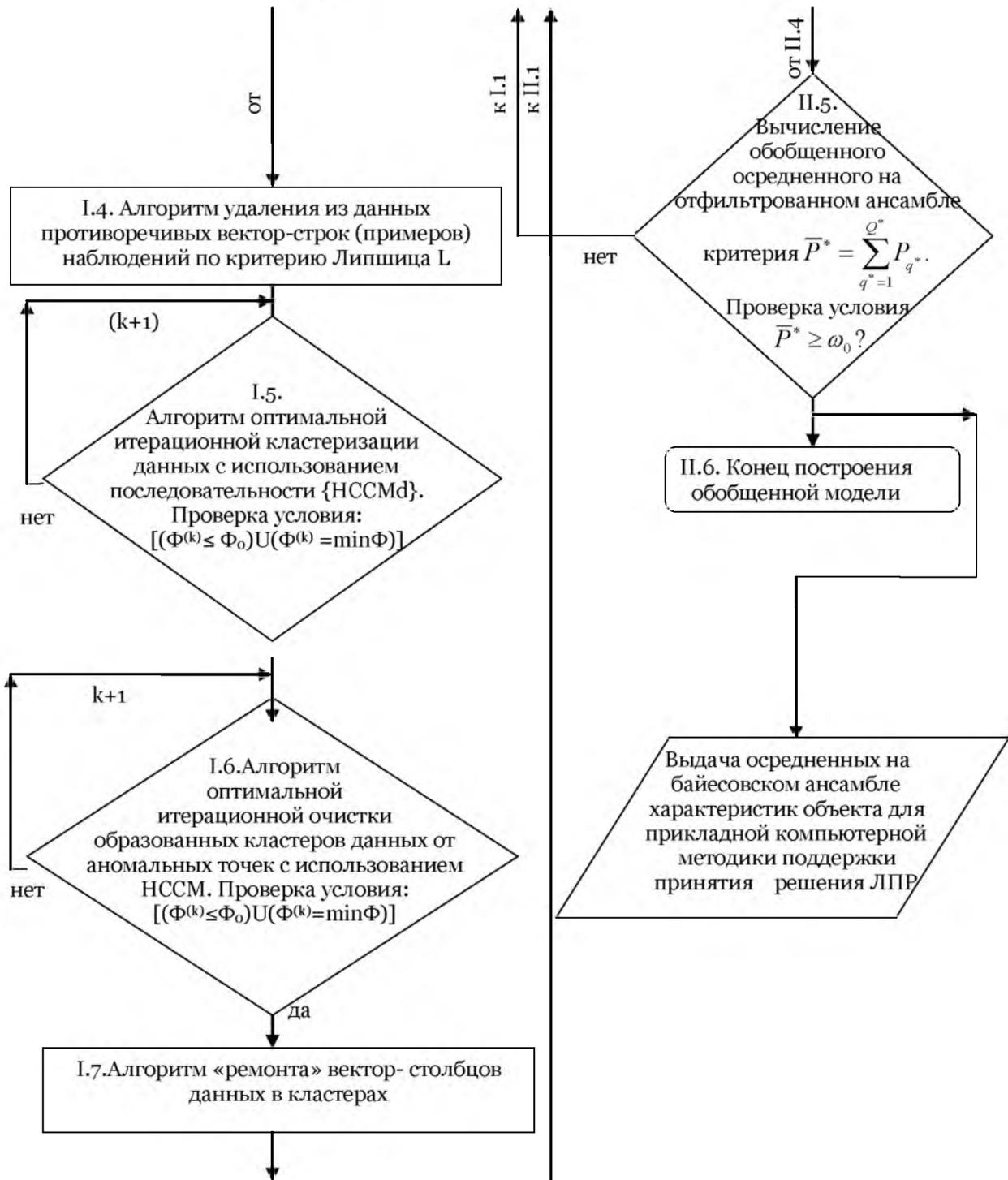


Рис 3. Логическая схема модифицированного метода вложения математических моделей (МВММ)
Fig. 3. The logical scheme of the modified method investments mathematical models (MUMM)

В случае подтверждения гипотезы о нормальности распределения в блоке ОФР производится оценка фрактальной размерности D^{-1} временных рядов $\{\bar{\delta}_t^*\}_g, g = \overline{1, G}$.

Затем по (2) вычисляются ранги $\{\Psi_g\}$ налогоплательщиков и синтезируется усредненный оптимальный план Θ^* выездных налоговых проверок.

Таким образом, в экспертную систему налогового контроля (ЭСНК) итоги переработки исходной сырой информации $\langle \bar{x}_{ig}, y_{ig} \rangle; i = \overline{1, N}; g = \overline{1, G}$ о налогоплательщиках поступают уже в глубоко структурированном и обобщенном виде оптимального плана выездных проверок Θ^* .



Остается в ЭСНК подвергнуть синтезированный план Θ^* цензурированию для выработки управляющих воздействий \vec{U}_g прикладываемых к ОУ.

Заключение

Информация в виде оптимальных планов выездных налоговых проверок Θ^* может быть использована также в ЭСНР и ЭСНП. Кроме того, в эти экспертные системы поступают промежуточные данные моделирования от блока ФБГ в виде нейросетевых восстановленных многомерных динамических зависимостей $\hat{Y}(\bar{x}, W)$, которые в ЭСНП используются для прогноза (планирования), а в ЭСНК – для кластеризации налогоплательщиков по их финансовому состоянию и выработки решений по налоговому регулированию.

Естественно, что вид регрессионных зависимостей $\hat{Y}_1(\bar{x}, W)$, $\hat{Y}_2(\bar{x}, W)$, $\hat{Y}_3(\bar{x}, W)$, поступающих в подсистемы ЭСНП, ЭСНР и ЭСНК, будет разным.

В целом, прикладная ГНСМ прошла успешно проверку адекватности на двух уровнях (фильтрации байесовского ансамбля и процедуру ОПП) для двух сильно зашумленных выборок Z^I и Z^{II} реальных данных деклараций. Это вселяет уверенность в продуктивности схемы интеллектуальной информационно-аналитической системы для реализации метода синтеза оптимального плана выездных проверок.

Список литературы References

- Методологические основы разработки нейросетевых моделей экономических объектов в условиях неопределенности / А.Н. Бирюков [и др.]. – М.: Издательский дом «Экономическая газета», 2012. – 494 с.
- Бирюков А.Н. Теоретические основы разработки нейросетевых моделей в системе налогового администрирования. – Уфа: Академия наук РБ, Издательство «Гилем», 2011. – 380 с.
- Бирюков А.Н. Байесовская регуляризация нейросетевых моделей ранжирования и кластеризации экономических объектов. – Уфа: Академия наук РБ, Издательство «Гилем», 2011. – 292 с.
- Бирюков А.Н. Бюджетно – налоговое администрирование: нейросетевые методы исследований. – Стерлитамак: РИО Стерлитамакский филиал БашГУ, 2014. – 152 с.
- Модернизация налогового контроля (модели и методы): Монография/ (Под ред. А.Н.Романова). – М.: Вузовский учебник, 2010. - 320 с.
- Хайкин С. Нейронные сети: полный курс = Neural Networks: A Comprehensive Foundation. 2-е изд. М –.: Вильямс, 2006. – 1104 с.
- Эндрю А. М. Искусственный интеллект / Под ред. Поспелова Д.А. - М.: Мир, 1985. – 264 с.
- Ясницкий, Л. Н. Интеллектуальные системы. – М.: Издательство «Лаборатория знаний», 2016. – 221 с.
- Hayes-Roth F., Jacobstein N. The State of Knowledge-Based Systems. Communications of the ACM, March, 1994, v.37, n.3, pp.27-39.
- Zimmermann H.-J. Fuzzy Set Theory and its Applications. 3rd ed.- Dordrecht: Kluwer Academic Publishers. – 1996. 315p.