

На правах рукописи



Лузгин Александр Николаевич

**КОМПЛЕКСНОЕ ИССЛЕДОВАНИЕ ИНТЕРВАЛЬНОГО
ПРОГНОЗИРОВАНИЯ НЕСТАЦИОНАРНЫХ ПОКАЗАТЕЛЕЙ С
ПРИМЕНЕНИЕМ КЛАСТЕРНЫХ И НЕЙРОННЫХ МОДЕЛЕЙ**

Специальность 05.13.18 – Математическое моделирование, численные методы и
комплексы программ

АВТОРЕФЕРАТ

диссертация на соискание ученой степени
кандидата технических наук

Иркутск – 2015

Работа выполнена на кафедре информатики и математического моделирования
ФГБОУ ВО «Иркутский государственный аграрный университет имени А.А. Ежевского»

Научный руководитель: доктор технических наук, профессор
Краковский Юрий Мечеславович

Официальные оппоненты: **Куцкий Николай Николаевич**,
доктор технических наук, профессор, профес-
сор кафедры автоматизированных систем
ФГБОУ ВО «Иркутский национальный иссле-
довательский технический университет»

Массель Людмила Васильевна,
доктор технических наук, профессор, главный
научный сотрудник ФГБУН «Институт систем
энергетики им. Л.А. Мелентьева» Сибирского
отделения Российской академии наук

Ведущая организация: ФГБОУ ВПО «Иркутский государственный
университет»

Защита состоится 24 декабря 2015 г. в 15.00 часов на заседании диссертационного со-
вета Д 212.070.07 при ФГБОУ ВПО «Байкальский государственный университет экономики
и права» по адресу: 664003, г. Иркутск, ул. Карла Маркса, д. 24, корп. 3, ауд. 402.

С диссертацией можно ознакомиться в библиотеке и на сайте ФГБОУ ВПО «Байкаль-
ский государственный университет экономики и права» по адресу 664003, г. Иркутск, ул.
Ленина, 11, БГУЭП, корпус 2, аудитория 101, (www.isea.ru).

Объявление о защите и автореферат диссертации размещены 23 октября 2015 года на
официальном сайте ВАК Минобрнауки РФ (www.vak.ed.gov.ru) и на официальном сайте
ФГБОУ ВПО «Байкальский государственный университет экономики и права»
(www.isea.ru).

Отзывы на автореферат направлять по адресу: 664003, г. Иркутск, ул. Ленина, 11, уче-
ному секретарю диссертационного совета Д 212.070.07

Автореферат разослан ___ ноября 2015 г.

Ученый секретарь диссертационного совета
кандидат технических наук, доцент



Т.И. Ведерникова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. В настоящее время большинство организаций и предприятий осуществляют свою деятельность в условиях неопределённости, где принятие эффективных решений по управлению производственными, технологическими и финансовыми процессами зависит от точности прогнозирования различных динамических показателей. Проблемы прогнозирования показателей широко освещены в работах отечественных и зарубежных ученых. Можно отметить следующих авторов: Х. Волд, Н. Винер, Дж. Бокс, Г. Дженкинс, Т. Андерсон, М. Кендэл, А. Колмогоров, Э. Тихонов, Ю. Лукашин и др.

В постоянно усложняющихся социально-экономических процессах и явлениях, особенно трудно добиваться приемлемой точности прогнозирования нестационарных динамических показателей (НДП). Это обусловлено многими причинами, основной из которых является развивающаяся, но все ещё плохо проработанная методологическая база в части анализа и прогнозирования НДП. В тоже время тенденция накопления ретроспективных значений НДП в базах данных, бурное развитие комплексов программ и средств информатизации в целом, способствуют разработке новых, более сложных методов и алгоритмов для решения задачи по прогнозированию НДП.

Существует множество распространённых методов и классов математических моделей прогнозирования НДП, таких как регрессионные, авторегрессионные, кластерные, нейронные и т.п. К моделям, демонстрирующим наилучшую точность прогнозирования НДП (в сравнении с моделями других классов) следует отнести нейронные и кластерные модели, каждая из которых имеет свои преимущества и недостатки и позволяют прогнозировать с определённой точностью будущие значения НДП. В тоже время проводить оценку непосредственно будущих значений НДП нужно далеко не всегда. На практике для выработки управленческих решений часто достаточно знать превысит или не превысит будущее значение НДП некоторое заранее заданное значение. Так как при таком прогнозировании оценивается не само будущее значение НДП, а то, в каком интервале оно будет находиться, в работе такое прогнозирование названо интервальным. Следует подчеркнуть, что в данном случае речь идет не об интервальном прогнозе, когда оценивается попадание будущего значения в доверительный интервал с заданной доверительной вероятностью, а об интервальном прогнозировании.

В связи с этим необходимо разработать методы и алгоритмы, а также реализовать их в виде программного обеспечения для комплексного исследования интервального прогнозирования НДП с применением кластерных и нейронных моделей, как наиболее перспективных для достижения приемлемой точности.

Все выше сказанное обосновывает актуальность выбранной темы диссертационной работы и позволяет сформулировать её цель и задачи.

Целью диссертационной работы является разработка методов и алгоритмов, а также их реализация в виде программного обеспечения для комплексного исследования интервального прогнозирования нестационарных динамических показателей с помощью адаптивной вероятностно-статистической кластерной и вероятностной нейронной моделей.

Для реализации поставленной цели необходимо решить следующие **задачи**:

1. Обосновать необходимость применения кластерных и вероятностных нейронных моделей для комплексного исследования интервального прогнозирования нестационарных динамических показателей.

2. Создать усовершенствованное алгоритмическое обеспечение интервального прогнозирования нестационарных динамических показателей с помощью кластерной модели.

3. Разработать модифицированное алгоритмическое обеспечение интервального прогнозирования нестационарных динамических показателей с помощью вероятностной нейронной модели.

4. Создать программный комплекс для исследования интервального прогнозирования нестационарных динамических показателей на основе предложенного алгоритмического обеспечения.

5. Разработать технологию и провести комплексное исследование интервального прогнозирования нестационарных динамических показателей по тестовым и реальным исходным данным.

Объект и предмет исследования. Объектом исследования являются нестационарные динамические показатели, используемые организациями, при осуществлении своей деятельности в условиях неопределённости и применяющими методы прогнозирования НДП. Предмет исследования – алгоритмическое и программное обеспечение интервального прогнозирования НДП на основе адаптивной вероятностно-статистической и вероятностной нейронной моделях.

Методы исследования и достоверность результатов. При решении поставленных задач в работе использованы методы математического моделирования, численные и кластерные методы, методы искусственного интеллекта, а также методы объектно-ориентированного программирования. Для реализации программного обеспечения интервального прогнозирования НДП использовался свободно распространяемый программный продукт «R» и его пакеты расширения. Достоверность результатов, полученных в ходе проведения комплексного исследования интервального прогнозирования НДП на реальных выборках различного объёма, подтверждена их сравнением с фактическими, заранее известными результатами.

Научную новизну диссертации представляют следующие положения, которые выносятся на защиту:

1. Усовершенствованное алгоритмическое обеспечение интервального прогнозирования на основе адаптивной вероятностно-статистической кластерной модели с использованием численных методов проверки подобия кластеров на основе коэффициента «линейного сопряжения» и алгоритма оценки «интервальных» вероятностей.

2. Алгоритмическое обеспечение интервального прогнозирования на основе вероятностной нейронной модели с использованием усовершенствованного алгоритма обучения и модифицированной функции классификации входных векторов.

3. Вычислительный алгоритм псевдослучайного увеличения объема выборки нестационарных динамических показателей на основе компьютерного моделирования с использованием генератора «Вихрь Мерсенна».

4. Постановка, технология и программное обеспечение комплексного исследования интервального прогнозирования нестационарных динамических показателей на основе адаптивной вероятностно-статистической кластерной и вероятностной нейронной моделей.

Практическая значимость. Практическая значимость результатов исследования заключается в создании и применении алгоритмического и программного обеспечения интервального прогнозирования НДП. Созданный программный комплекс «Интервальное прогнозирование НДП» апробирован на различных типах данных. Результаты диссертационной работы используются в Иркутском областном гарантийном фонде, ООО «Сибпрофкосметик», территориальном органе Федеральной службы государственной статистики по Иркутской области, о чем имеются соответствующие акты и справка о внедрении. Разработанное алгоритмическое и программное обеспечение, позволяющее оперативно решать задачи интервального прогнозирования и компьютерного моделирования НДП, может найти применение и в других предметных областях.

Апробация работы. Основные положения проведённых исследований докладывались на международных и всероссийских конференциях: всероссийская научно-практическая

конференция «Новые информационные технологии в экономике, управлении, образовании», Самара, 2012; всероссийская научно-практическая конференция «Аспирантские чтения в ИГЛУ - 2012», Иркутск, 2012; всероссийская научно-практическая конференция «Аспирантские чтения в ИГЛУ - 2013», Иркутск, 2013; IV международная практическая конференция «Безопасность регионов – основа устойчивого развития», Иркутск, 2014; XX всероссийская Байкальская конференция «Информационные и математические технологии в науке и управлении», Иркутск, 2015.

Публикации. По результатам исследований опубликовано 10 научных работ в виде статей и докладов, включая 2 свидетельства о государственной регистрации программы для ЭВМ. Из них 3 публикации в изданиях, рекомендованных ВАК, число публикаций без соавторов – 6.

Структура и объем работы. Диссертационная работа состоит из введения, трех глав, заключения, списка сокращений и условных обозначений, списка литературы из 136 наименований и приложения. Общий объем работы без приложения составляет 151 страница.

СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность работы, определена научная и практическая новизна, представлено краткое содержание диссертационной работы по главам.

В первой главе формализовано понятие НДП и рассмотрены особенности их прогнозирования. При этом: а) проведен обзор математических моделей прогнозирования НДП, среди которых определены наиболее перспективные, с точки зрения достижения лучшей точности прогнозирования, – это кластерные и нейронные модели прогнозирования; б) проведен обзор программного обеспечения для прогнозирования НДП, среди которого осуществлен выбор программного обеспечения под названием «R», как наиболее подходящего для реализации алгоритмов интервального прогнозирования; в) проведен обзор генераторов псевдослучайных чисел, среди которых выбран генератор под названием «Вихрь Мерсенна», как наиболее подходящий для последующей разработки вычислительного алгоритма псевдослучайного увеличения выборки НДП с использованием компьютерного моделирования.

Во второй главе описано созданное алгоритмическое и программное обеспечение интервального прогнозирования НДП на основе адаптивной вероятностно-статистической кластерной модели и модели вероятностной нейронной сети.

Введем интервал возможных значений НДП $(c_1; c_2)$, где $c_1 > 0$, $c_2 < \infty$, и внутреннюю точку c : $c_1 < c < c_2$. Это позволяет создать два интервала:

$$I^\downarrow = (c_1; c], I^\uparrow = (c; c_2). \quad (1)$$

Интервальное прогнозирование (ИП) НДП заключается в оценке прогнозных значений НДП на основе вероятностей их принадлежности интервалам I^\uparrow и I^\downarrow (1). Так как при таком прогнозировании оценивается не само будущее значение НДП, а то, в каком интервале оно будет находиться, такое прогнозирование названо интервальным, а соответствующие вероятности «интервальными».

Формализуем понятие НДП следующим образом:

$$Q = \{q_t; t \in T\}. \quad (2)$$

Здесь q_t – значения НДП доступные в дискретные моменты времени t , где t принимает значения из множества $T = \{0, \dots, n-1\}$, а n – количество доступных значений НДП, $q_t > 0$.

Введём формулу сглаживания НДП (2) простым скользящим средним:

$$s_t = (q'_t + q'_{t-1} + \dots + q'_{t-d+1}) / d = \sum_{\xi=0}^{d-1} q'_{t-\xi} / d, \quad q'_{t-\xi} = \begin{cases} q_{t-\xi}, & (t-\xi) \geq 0, \\ q_0, & (t-\xi) < 0. \end{cases} \quad (3)$$

где s_t – сглаженные значения НДП, $d \geq 1$ – период сглаживания, q_t – исходные значения НДП (2). После проведения процедуры сглаживания новые сглаженные значения НДП:

$$\mathbf{S} = \{s_t : t \in \mathbf{T}\}. \quad (4)$$

Здесь s_t – сглаженные по формуле (3) значения НДП (2), $s_t > 0$.

Подчеркнём, что для применения любого из алгоритмов основанном на кластерной или нейронной модели необходимо выполнить одно важное требование – обеспечить достаточно большой объём выборки значений НДП (желательно в несколько сотен значений). При попытке использовать подобные алгоритмы на небольших выборках возникает проблема неполноты и неточности исходной информации. Принимая во внимание данное требование в работе рассматриваются НДП двух типов: 1) валютные курсы, сглаженные простым скользящим средним (получены через Интернет, используются для тестирования алгоритмов ИП НДП); 2) еженедельные цены на сливочное масло, пшеничную муку (получены в территориальном органе Федеральной службы государственной статистики по Иркутской области) и ежемесячные объёмы выданных поручительств (получены в Иркутском областном гарантийном фонде) в городе Иркутске, объём выборки которых был увеличен с помощью разработанного вычислительного алгоритма псевдослучайного увеличения объёма выборки НДП на основе компьютерного моделирования (используются для апробации алгоритмов ИП НДП).

Для осуществления ИП НДП предлагается разработать алгоритм ИП НДП на основе кластерной модели и алгоритм ИП НДП на основе вероятностной нейронной модели, в рамках которых будет происходить оценка и использование интервальных вероятностей (рисунок 1). Так как данные алгоритмы будут независимы друг от друга, и каждый, в свою очередь, иметь индивидуальный набор параметров, влияющих на точность ИП НДП, далее предлагается оценить влияние значений этих параметров на точность ИП НДП, выработать рекомендации по выбору параметров каждого из алгоритмов, при которых алгоритмы демонстрируют наилучшую точность ИП на НДП первого типа. Исходя из выработанных рекомендаций, выбрав значения параметров каждого из алгоритмов ИП НДП, далее предлагается произвести апробацию ИП НДП на НДП второго типа.

Постановка задачи комплексного исследования ИП НДП (рис. 1).

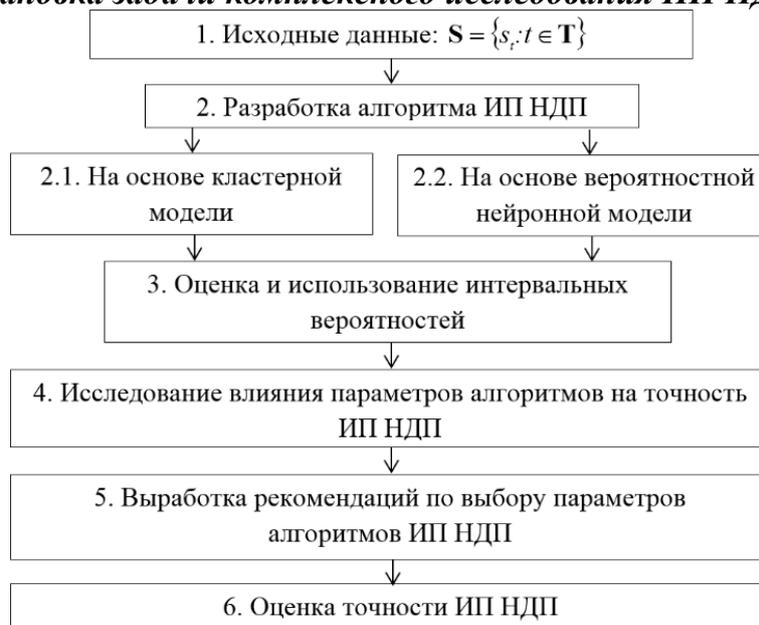


Рис. 1. Постановка задачи комплексного исследования ИП НДП

Формализуем ИП НДП. Введем величину Δ , рассчитывать которую будем так:

$$\Delta = \alpha \cdot \bar{s}, \quad (5)$$

где $\alpha \in [-1, 1]$ – коэффициент, задается заранее, \bar{s} – определяется по формуле (6).

$$\bar{s} = \left(\sum_{t=1}^{n-1} |s_t - s_{t-1}| \right) / (n-1). \quad (6)$$

При ИП НДП (4) в момент времени $t = n - 1$ необходимо провести оценку интервальной вероятности ρ_{t+p}^+ того, что будущее значение НДП $s_{t+p} \in \Gamma^\uparrow$ или оценку интервальной вероятности ρ_{t+p}^- того, что будущее значение НДП $s_{t+p} \in \Gamma^\downarrow$, где Γ^\uparrow и Γ^\downarrow определяется по формуле (1), $p = 1, \dots, r$ есть время упреждения, а $\rho_{t+p}^+ + \rho_{t+p}^- = 1$. Если оценки интервальных вероятностей $\rho_{t+p}^+ = \rho_{t+p}^- = 1/2$ (ситуация неопределённости), то прогнозирование не делается. Величина $c = (s_t + \Delta)$ – разделительная граница интервалов, а Δ (5) – смещение относительно значения s_t .

Алгоритм интервального прогнозирования на основе адаптивной вероятностно-статистической кластерной модели

Новизна: предложено усовершенствованное алгоритмическое обеспечение интервального прогнозирования НДП на основе адаптивной вероятностно-статистической кластерной модели с использованием численных методов проверки подобия кластеров на основе коэффициента «линейного сопряжения» и алгоритма оценки «интервальных» вероятностей.

Под кластером $\mathbf{S}_i^f = \{s_i, \dots, s_{i+f-1}\}$ будем понимать выборку последовательных значений из \mathbf{S} (4) с позиции $i \in \mathbf{T}$ и с количеством значений $f = 1, \dots, n$, $(i + f - 1) \in \mathbf{T}$.

1) Выберем из \mathbf{S} (4) кластер \mathbf{S}_i^f , при $i = n - f$, $f \in [1; n]$ и назовем его базовым.

2) Выберем из \mathbf{S} (4) кластер \mathbf{S}_k^g , для которого выполняется условие:

$$k \leq n - g - p. \quad (7)$$

3) Если кластеры содержат одинаковое количество значений, то есть $f = g$, то переходим на этап 4; иначе кластеры не подобны.

4) Если вычисленное значение коэффициента «линейного сопряжения» $R(\mathbf{S}_i^f, \mathbf{S}_k^g) \geq R_d$, где $R_d \in [0; 1]$ – допустимое значение (это значение определяется заранее), то кластеры подобны и переходим на этап 5; иначе не подобны.

Коэффициент «линейного сопряжения» (основан на коэффициенте линейной корреляции Пирсона) предложено определять следующим образом:

$$R(\mathbf{S}_i^f, \mathbf{S}_k^g) = \begin{cases} N(\mathbf{S}_i^f, \mathbf{S}_k^g) / D(\mathbf{S}_i^f, \mathbf{S}_k^g), & D(\mathbf{S}_i^f, \mathbf{S}_k^g) \neq 0, \\ \begin{cases} 1, & \dot{D}(\mathbf{S}_i^f) = \ddot{D}(\mathbf{S}_k^g) \\ 0, & \dot{D}(\mathbf{S}_i^f) \neq \ddot{D}(\mathbf{S}_k^g) \end{cases}, & D(\mathbf{S}_i^f, \mathbf{S}_k^g) = 0. \end{cases}$$

$$N(\mathbf{S}_i^f, \mathbf{S}_k^g) = f \cdot \sum_{\xi=0}^{f-1} s_{i+\xi} \cdot s_{k+\xi} - \sum_{\xi=0}^{f-1} s_{i+\xi} \cdot \sum_{\xi=0}^{g-1} s_{k+\xi}, \quad \dot{D}(\mathbf{S}_i^f) = \left(f \cdot \sum_{\xi=0}^{f-1} s_{i+\xi}^2 - \left(\sum_{\xi=0}^{f-1} s_{i+\xi} \right)^2 \right), \quad (8)$$

$$\ddot{D}(\mathbf{S}_k^g) = \left(g \cdot \sum_{\xi=0}^{g-1} s_{k+\xi}^2 - \left(\sum_{\xi=0}^{g-1} s_{k+\xi} \right)^2 \right), \quad D(\mathbf{S}_i^f, \mathbf{S}_k^g) = \sqrt{\dot{D}(\mathbf{S}_i^f) \cdot \ddot{D}(\mathbf{S}_k^g)}.$$

Способ расчета и название данного коэффициента предложено в рамках данной работы. Фактически значение коэффициента «линейного сопряжения» можно интерпретировать как численную «меру» подобия двух кластеров. Допустимое значение такой «меры» определяется через параметр R_d .

Значения величин $\dot{D}(\mathbf{S}_i^f)$ или $\ddot{D}(\mathbf{S}_k^g)$ в (8) равны нулю только в том случае, если все значения соответствующего кластера \mathbf{S}_i^f или \mathbf{S}_k^g равны между собой.

5) Аппроксимируем значения базового кластера \mathbf{S}_i^f посредством значений кластера \mathbf{S}_k^g следующим образом:

$$\tilde{s}_{i+\xi} = a_{i,k} \cdot s_{k+\xi} + b_{i,k}. \quad (9)$$

Здесь $s_{k+\xi}$ – значения кластера \mathbf{S}_k^g , $\tilde{s}_{i+\xi}$ – аппроксимированные значения базового кластера \mathbf{S}_i^f , $\xi \in [0; f-1]$, $a_{i,k}$, $b_{i,k}$ – коэффициенты уравнения (9).

6) Для нахождения коэффициентов уравнения (9) воспользуемся численным методом, в качестве которого выберем метод наименьших квадратов. Тогда используя обозначения, данные в формуле (8) коэффициенты уравнения (9) можно вычислить так:

$$a_{i,k} = \begin{cases} N(\mathbf{S}_i^f, \mathbf{S}_k^g) / \ddot{D}(\mathbf{S}_k^g), \ddot{D}(\mathbf{S}_k^g) \neq 0, \\ 1, \ddot{D}(\mathbf{S}_k^g) = 0, \end{cases} \quad b_{i,k} = \left(\sum_{\xi=0}^{f-1} s_{i+\xi} - a_{i,k} \cdot \sum_{\xi=0}^{g-1} s_{k+\xi} \right) / f. \quad (10)$$

Когда величина $\ddot{D}(\mathbf{S}_k^g)$ в (10) равна нулю, то коэффициент $a_{i,k} = 1$ (в этом случае все значения кластера \mathbf{S}_k^g равны между собой и коэффициент $b_{i,k} = 0$). Такое уточнение метода наименьших квадратов предлагается в рамках данной работы с целью сохранения возможности вычисления коэффициентов уравнения (9) в случае выявления подобных кластеров, значения которых не варьируются.

7) Вычислив коэффициенты $a_{i,k}$ и $b_{i,k}$, найдем оценку $\tilde{s}_{i+f-1+p}$ будущего (неизвестного) значения s_{i+p} следующим образом:

$$\tilde{s}_{i+f-1+p} = a_{i,k} \cdot s_{k+f-1+p} + b_{i,k}. \quad (11)$$

Напомним, что $i = n - f$, а момент прогнозирования $t = n - 1$. Тогда $t = i + f - 1$ и (11) можно переписать в виде:

$$\tilde{s}_{t+p} = a_{i,k} \cdot s_{k+f-1+p} + b_{i,k}. \quad (12)$$

8) Сравнение величин $s_{n-1} + \Delta$ и \tilde{s}_{t+p} будем проводить следующим образом:

$$\Psi_t = \begin{cases} 1, \tilde{s}_{t+p} > (s_{n-1} + \Delta), \\ -1, \tilde{s}_{t+p} \leq (s_{n-1} + \Delta). \end{cases} \quad (13)$$

Здесь Ψ_t – величина, характеризующая результат сравнения. При $\Psi_t = 1$ оценка \tilde{s}_{t+p} будущего значения s_{t+p} больше $s_{n-1} + \Delta$, а при $\Psi_t = -1$ меньше, либо равна.

9) Рассчитав значения коэффициентов (10) и величины (13) для всех кластеров, найденных в \mathbf{S} (4) и подобных базовому кластеру \mathbf{S}_i^f , можно провести оценку искомых интервальных вероятностей ρ_{t+p}^+ и ρ_{t+p}^- по формуле:

$$\rho_{t+p}^+ = \begin{cases} N_t^+ / (N_t^+ + N_t^-), (N_t^+ + N_t^-) \neq 0, \\ 1/2, (N_t^+ + N_t^-) = 0, \end{cases} \quad \rho_{t+p}^- = \begin{cases} N_t^- / (N_t^+ + N_t^-), (N_t^+ + N_t^-) \neq 0, \\ 1/2, (N_t^+ + N_t^-) = 0, \end{cases} \quad (14)$$

где N_t^+ – количество случаев, когда $\Psi_t = 1$, а N_t^- – количество случаев, когда $\Psi_t = -1$.

10) Проведение прогноза:

а) будущее значение s_{t+p} больше $s_{n-1} + \Delta$, если $\rho_{t+p}^+ > \rho_{t+p}^-$;

б) будущее значение s_{t+p} меньше, либо равно $s_{n-1} + \Delta$, если $\rho_{t+p}^- < \rho_{t+p}^+$;

в) прогноз не делается, если $\rho_{t+p}^+ = \rho_{t+p}^-$.

Рекомендации по выбору значений R_d , α , p , f даются в третьей главе диссертации, основываясь на результатах изучения влияния этих параметров на точность ИП НДП на основе АВСКМ.

Алгоритмическое обеспечение интервального прогнозирования на основе вероятностной нейронной модели

Новизна: предложено алгоритмическое обеспечение интервального прогнозирования НДП на основе вероятностной нейронной модели с использованием усовершенствованного алгоритма обучения и модифицированной функции классификации векторов.

Архитектура ВНС была предложена в 1988г. Д. Спехтом для проведения классификации векторов (образов). Перед проведением такой классификации ВНС должна быть обучена на множестве векторов с известной классификацией. Модель ВНС представляет собой инструмент, который несложно адаптировать для осуществления ИП НДП.

1) Выберем из S (4) кластер S_k^g для которого выполняется ранее введенное условие (7): $k \leq n - g - p$. Будем рассматривать такой кластер в качестве обучающего вектора ВНС с известной классификацией, класс которого определяется по формуле:

$$\Psi'_{k+g-1} = \begin{cases} 1, s_{k+g-1+p} > (s_{k+g-1} + \Delta), \\ -1, s_{k+g-1+p} \leq (s_{k+g-1} + \Delta). \end{cases} \quad (15)$$

Здесь Ψ'_{k+g-1} – величина, характеризующая результат сравнения, $s_{k+g-1+p}$ – оценка будущего (неизвестного) значения s_{t+p} , s_{k+g-1} – последнее значение кластера S_k^g , Δ – величина, определённая по формуле (5). Кластер S_k^g относится к классу a , если $\Psi'_{k+g-1} = 1$ и к классу b , если $\Psi'_{k+g-1} = -1$.

2) Перед обучением ВНС кластер S_k^g предварительно нормируется:

$$\bar{S}_k^g = \begin{cases} \left\{ \bar{s}_k = \frac{s_k - s_{\min}^{g,k}}{s_{\max}^{g,k} - s_{\min}^{g,k}}, \dots, \bar{s}_{k+g-1} = \frac{s_{k+g-1} - s_{\min}^{g,k}}{s_{\max}^{g,k} - s_{\min}^{g,k}} \right\}, s_{\min}^{g,k} \neq s_{\max}^{g,k}, \\ \left\{ \bar{s}_k = \frac{s_k - s_{\min}^{g,k}}{s_{\min}^{g,k}}, \dots, \bar{s}_{k+g-1} = \frac{s_{k+g-1} - s_{\min}^{g,k}}{s_{\min}^{g,k}} \right\}, s_{\min}^{g,k} = s_{\max}^{g,k}, \end{cases} \quad (16)$$

где $s_{\min}^{g,k}$ – минимальное значение кластера S_k^g , $s_{\max}^{g,k}$ – максимальное значение кластера S_k^g .

3) Затем проведём дополнительную корректировку по формуле:

$$\tilde{S}_k^g = \begin{cases} \left\{ \tilde{s}_k = \bar{s}_k / \tilde{s}, \dots, \tilde{s}_{k+g-1} = \bar{s}_{k+g-1} / \tilde{s} \right\}, \tilde{s} \neq 0, \\ \left\{ \tilde{s}_k = \bar{s}_k, \dots, \tilde{s}_{k+g-1} = \bar{s}_{k+g-1} \right\}, \tilde{s} = 0. \end{cases} \quad (17)$$

Здесь $\tilde{s} = \sqrt{\sum_{j=k}^{k+g-1} \bar{s}_j^2}$; нормированные значения дополнительно не корректируются, если значения (16) равны нулю.

4) Базовый кластер S_i^f , рассматривается в качестве кластера с неизвестной классификацией. Предварительно проводится его нормировка (16) и корректировка (17), за тем на основе обученной ВНС с модифицированной функцией выходного нейрона определяется его класс:

$$\Omega'_h = \begin{cases} \rho_h^a > \rho_h^b, 1, \\ \rho_h^a = \rho_h^b, 0, \\ \rho_h^a < \rho_h^b, -1. \end{cases} \quad (18)$$

Здесь ρ_h^a и ρ_h^b – оценки апостериорных вероятностей принадлежности кластера к определённому классу, рассчитанные ВНС, $h = i = n - f$, Ω'_h – признак: если $\Omega'_h = 1$, то базовый кластер принадлежит к классу a , если $\Omega'_h = -1$, то к классу b , если то класс не определяется.

5) Проведение прогноза:

- а) будущее значение s_{t+p} больше $s_t + \Delta$, если $\Omega'_h = 1$;
- б) будущее значение s_{t+p} меньше, либо равно $s_t + \Delta$, если $\Omega'_h = -1$;
- в) прогноз не делается, если $\Omega'_h = 0$.

Рекомендации по выбору значений σ , α , p , f даются в третьей главе диссертации, основываясь на результатах изучения влияния этих параметров на точность прогнозирования алгоритма ИП НДП на основе ВНС.

Фактически оценки вероятностей ρ_h^a и ρ_h^b , произведённые в рамках алгоритма ИП НДП на основе ВНС, по своему смыслу эквивалентны оценкам интервальных вероятностей ρ_{t+p}^+ и ρ_{t+p}^- , произведённым в рамках алгоритма ИП НДП на основе АВСКМ. Для сохранения общности дальнейших рассуждений будем обозначать оценки вероятностей ρ_h^a и ρ_h^b также через ρ_{t+p}^+ и ρ_{t+p}^- с уточнением, в рамках какого алгоритма ИП НДП они были сделаны.

Вычислительный алгоритм псевдослучайного увеличения объёма выборки нестационарного динамического показателя на основе компьютерного моделирования

Новизна: Обоснована необходимость и разработан вычислительный алгоритм псевдослучайного увеличения объёма выборки НДП на основе компьютерного моделирования с использованием генератора «Вихрь Мерсенна».

В данной работе предлагается вычислительный алгоритм псевдослучайного увеличения объёма выборки НДП (4) на основе компьютерного моделирования. Данный алгоритм предполагает использование генератора псевдослучайных чисел (ГПСЧ). Для реализации предлагаемого алгоритма был выбран ГПСЧ под названием «Вихрь Мерсенна», разработанный в 1997г. японскими учёными М.Мацумото и Т.Нисимура. Вихрь Мерсенна обеспечивает быструю генерацию высококачественных псевдослучайных чисел, лишен многих недостатков, присущих другим ГПСЧ, таких как малый период, предсказуемость, легко выявляемая статистическая зависимость.

1) Зададим требуемый объём увеличенной выборки $x > n$, где n – количество доступных значений в S .

2) Сгенерируем три псевдослучайной величины: $a \in [0; n - 1]$, $b \in [0; n - 1]$, $c \in [-1; 1]$. Значения данных величин из заданных диапазонов будем генерировать с помощью Вихря Мерсенна.

3) Сформируем из S (4) дополнительную выборку по следующему правилу:

$$S' = \begin{cases} \{s'_0 = s_a, s'_1 = s_{a+1}, \dots, s'_{m-1} = s_b\}, a \leq b, \\ \{s'_0 = s_a, s'_1 = s_{a-1}, \dots, s'_{m-1} = s_b\}, a > b. \end{cases} \quad (19)$$

Здесь $S' = \{s'_0, \dots, s'_{m-1}\}$ – дополнительная выборка из S с количеством значений $m = |a - b + 1|$. Если $a \leq b$ значения S берутся в прямом порядке, если $a > b$, то в обратном (выборка инвертируется).

4) Вычислим величину:

$$\delta = s'_0 - s_{n-1} + c \cdot \bar{s}. \quad (20)$$

Здесь s'_0 – первое значение новой выборки S' , s'_{n-1} – последнее значение исходной выборки S , c – ранее введённая псевдослучайная величина, \bar{s} – величина, вычисленная по формуле (6).

5) Скорректируем новую выборку S' по формуле:

$$S'' = \{s''_0 = s'_0 - \delta, \dots, s''_{m-1} = s'_{m-1} - \delta\}. \quad (21)$$

6) Если $n + m < x$, то добавляем в конец выборки S новую выборку S'' , при этом $n = n + m$ и возвращаемся на этап 2; иначе добавляем в конец выборки S только $x - n$ первых значений выборки S'' . Если в исходной выборке S все значения $s_i \geq 0$ и необходимо, чтобы данное условие сохранялось и после увеличения данной выборки, то добавление выборки S'' в конец выборки S происходит только тогда, когда все значения полученной выборки S'' отвечают условию $s''_i \geq 0$.

Программное обеспечение интервального прогнозирования нестационарных динамических показателей

Новизна: создано программное обеспечение интервального прогнозирования НДП – ПК «ИП НДП», основанное на разработанном алгоритмическом обеспечении адаптивной вероятностно-статистической кластерной и вероятностной нейронной моделях.

Разработанные и описанные алгоритмы были реализованы посредством создания прикладного программного обеспечения – Программного комплекса «Интервальное прогнозирование нестационарных динамических показателей» (ПК «ИП НДП»). Для реализации ПК «ИП НДП» был выбран свободно распространяемый интерпретируемый язык программирования «R». Язык программирования «R» представляет собой консоль последовательного ввода команд и не удобен для разработки программного обеспечения. Для этой цели существуют различные оболочки (среды разработки) к нему. В качестве подходящей среды разработки на языке «R» была выбрана свободно распространяемая среда под названием «R-studio». Программное обеспечение, реализованное на языке «R» является кроссплатформенным и способно работать в 32/64-битных средах. При запуске ПК «ИП НДП», появляется главное окно, приведённое на рисунке 2, посредством которого пользователь имеет возможность доступа ко всем функциям и настройкам программы.

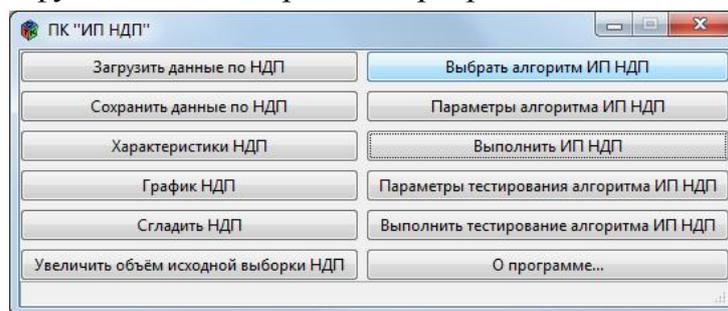


Рис 2. Главное окно программы

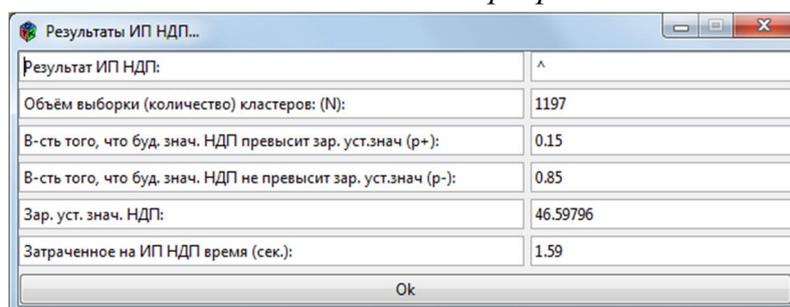


Рис. 3. Окно вывода результатов ИП НДП

На рисунке 3 приведен пример вывода результатов ИП НДП.

В третьей главе проведено тестирование и апробация алгоритмического и программного обеспечения интервального прогнозирования НДП, предложена технология комплексного исследования интервального прогнозирования НДП. Тестирование алгоритмов проводилось по схеме, показанной на рисунке 4.

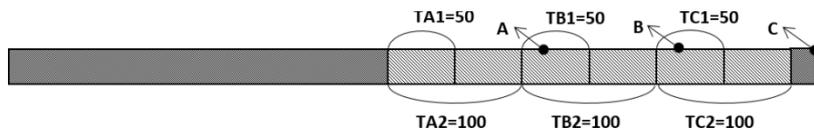


Рис.4. Схема тестирования алгоритмов

В каждой выборке НДП первого типа выбиралось три «точки» А, В и С и шесть тестовых периодов ТА1, ТВ1, ТС1 (содержат 50 значений НДП), ТА2, ТВ2, ТС2 (содержат 100 значений НДП). В «точках» А, В, С исследовалось влияние значений параметров алгоритмов ИП НДП на оценки интервальных вероятностей ρ_{t+p}^+ и ρ_{t+p}^- , на тестовых периодах проводилось тестирование алгоритмов ИП НДП, исследовалось влияние значений их параметров на точность ИП НДП и осуществлялась выработка рекомендация по выбору значений параметров алгоритмов.

При тестировании и оценке точности алгоритмов ИП НДП использовались показатели: L – число оправдавшихся прогнозов; M – число ошибочных прогнозов; $PL = L \times 100 / (L + M)$ – процент оправдавшихся прогнозов; $PM = M \times 100 / (L + M)$ – процент ошибочных прогнозов, PS – число случаев, прогноз по которым не делался; $PPS = PS \times 100 / (L + M + PS)$ – процент случаев, прогноз по которым не делался. В работе считается, что алгоритм ИП НДП имеет приемлемую точность прогнозирования, если выполняются условия: $PL \geq 60$, $PPS \leq 1$.

По завершению тестирования были выработаны следующие рекомендации: 1) для алгоритма на основе АВСКМ: $f = 2$, $R_d = 0,9$, $\alpha \in [-1; 1]$, $p = 1, 2, 3, 4$; 2) для алгоритма на основе ВНС: $f = 3$, $\alpha \in [-1; 1]$, $p = 1, 2, 3, 4$, $\sigma = 0,1$.

Полученные результаты тестирования далее используются для апробации. Пример результатов тестирования при вариации параметра p для алгоритма на основе АВСКМ (тестовый период ТС2, курс евро к рублю) приведен в таблице 1, а для алгоритма на основе ВНС в таблице 2 (тестовый период ТВ1, курс евро к рублю).

Таблица 1

EUR/RUR (тестовый интервал ТС2)						
p	1	2	3	4	5	6
L	83	75	67	65	58	52
M	17	25	33	35	42	48
PS	0	0	0	0	0	0
PL	83,0	75,0	67,0	65,0	58,0	52,0
PM	17,0	25,0	33,0	35,0	42,0	48,0
PSS	0,0	0,0	0,0	0,0	0,0	0,0

Таблица 2

EUR/RUR (тестовый интервал ТВ1)						
p	1	2	3	4	5	6
L	44	42	42	41	38	36
M	6	8	8	9	12	14
PS	0	0	0	0	0	0
PL	88,0	84,0	84,0	82,0	76,0	72,0
PM	12,0	16,0	16,0	18,0	24,0	28,0
PSS	0,0	0,0	0,0	0,0	0,0	0,0

Апробация интервального прогнозирования нестационарных динамических показателей на основе разработанных алгоритмов

Для апробации ИП НДП было выбрано шесть вариантов НДП второго типа, обозначенных как MS1, MS2, MP1, MP2, POR1, POR2. Прежде всего, проведено сравнение точности разработанных алгоритмов ИП НДП при $\alpha = 0$, $p = 1$ (таблица 3).

Таблица 3

Алгоритм на основе АВСКМ						
НДП	MS1	MS2	MP1	MP2	POR1	POR2
Номер НДП	1	2	3	4	5	6
<i>L</i>	205	227	184	190	180	191
<i>M</i>	95	72	116	109	120	109
<i>PS</i>	0	1	0	1	0	0
<i>PL</i>	68,3	75,9	61,3	63,5	60,0	63,7
<i>PM</i>	31,7	24,1	38,7	36,5	40,0	36,3
<i>PSS</i>	0,0	0,3	0,0	0,3	0,0	0,0
Алгоритм на основе ВНС						
НДП	MS1	MS2	MP1	MP2	POR1	POR2
Номер НДП	1	2	3	4	5	6
<i>L</i>	193	201	201	202	218	212
<i>M</i>	107	99	99	98	82	88
<i>PS</i>	0	0	0	0	0	0
<i>PL</i>	64,3	67,0	67,0	67,3	72,7	70,7
<i>PM</i>	35,7	33,0	33,0	32,7	27,3	29,3
<i>PSS</i>	0,0	0,0	0,0	0,0	0,0	0,0

Согласно полученным результатам, в 4-х из 6-ти случаев алгоритм ИП НДП на основе ВНС продемонстрировал лучшую точность ИП НДП, в отличие от алгоритма ИП НДП на основе АВСКМ, где только в 2-х из 6-ти случаев точность была выше. Далее проведено сравнение точности ИП НДП при $\alpha = 1$, $p = 1$ (таблица 4) и при $\alpha = -1$, $p = 1$ (таблица 5). Согласно полученным результатам, во всех случаях алгоритм на основе АВСКМ продемонстрировал лучшую точность ИП НДП, в отличие от алгоритма на основе ВНС. В таблице 4 приведены результаты апробации алгоритмов ИП НДП при $\alpha = 0$, $p = 3$. Согласно полученным результатам, в 5-ти из 6-ти случаев при более отдалённом прогнозе алгоритм на основе ВНС продемонстрировал лучшую точность ИП НДП, в отличие от алгоритма на основе АВСКМ, где в 1-ом из 6-ти случаев точность была выше.

Таблица 4

Алгоритм на основе АВСКМ						
НДП	MS1	MS2	MP1	MP2	POR1	POR2
Номер НДП	1	2	3	4	5	6
<i>L</i>	246	270	249	248	242	245
<i>M</i>	54	30	51	52	58	55
<i>PS</i>	0	0	0	0	0	0
<i>PL</i>	82,0	90,0	83,0	82,7	80,7	81,7
<i>PM</i>	18,0	10,0	17,0	17,3	19,3	18,3
<i>PSS</i>	0,0	0,0	0,0	0,0	0,0	0,0
Алгоритм на основе ВНС						
НДП	MS1	MS2	MP1	MP2	POR1	POR2
Номер НДП	1	2	3	4	5	6
<i>L</i>	193	204	224	208	194	192
<i>M</i>	107	96	76	92	106	108
<i>PS</i>	0	0	0	0	0	0
<i>PL</i>	64,3	68,0	74,7	69,3	64,7	64
<i>PM</i>	35,7	32,0	25,3	30,7	35,3	36
<i>PSS</i>	0,0	0,0	0,0	0,0	0,0	0,0

Таблица 5

Алгоритм на основе АВСКМ						
<i>НДП</i>	MS1	MS2	MP1	MP2	POR1	POR2
<i>Номер НДП</i>	1	2	3	4	5	6
<i>L</i>	283	255	254	255	265	248
<i>M</i>	17	45	46	45	35	52
<i>PS</i>	0	0	0	0	0	0
<i>PL</i>	94,3	85	84,7	85	88,3	82,7
<i>PM</i>	5,7	15	15,3	15	11,7	17,3
<i>PSS</i>	0,0	0,0	0,0	0,0	0,0	0,0
Алгоритм на основе ВНС						
<i>НДП</i>	MS1	MS2	MP1	MP2	POR1	POR2
<i>Номер НДП</i>	1	2	3	4	5	6
<i>L</i>	202	218	202	187	214	224
<i>M</i>	98	82	98	113	86	76
<i>PS</i>	0	0	0	0	0	0
<i>PL</i>	67,3	72,7	67,3	62,3	71,3	74,7
<i>PM</i>	32,7	27,3	32,7	37,7	28,7	25,3
<i>PSS</i>	0,0	0,0	0,0	0,0	0,0	0,0

Таблица 6

Алгоритм на основе АВСКМ						
<i>НДП</i>	MS1	MS2	MP1	MP2	POR1	POR2
<i>Номер НДП</i>	1	2	3	4	5	6
<i>L</i>	180	206	187	190	198	192
<i>M</i>	120	92	113	110	102	108
<i>PS</i>	0	2	0	0	0	0
<i>PL</i>	60,0	69,1	62,3	63,3	66,0	64,0
<i>PM</i>	40,0	30,9	37,7	36,7	34,0	64,0
<i>PSS</i>	0,0	0,7	0,0	0,0	0,0	0,0
Алгоритм на основе ВНС						
<i>НДП</i>	MS1	MS2	MP1	MP2	POR1	POR2
<i>Номер НДП</i>	1	2	3	4	5	6
<i>L</i>	183	195	198	194	226	223
<i>M</i>	117	105	102	106	74	77
<i>PS</i>	0	0	0	0	0	0
<i>PL</i>	61,0	65,0	66,0	64,7	75,3	74,3
<i>PM</i>	39,0	35,0	34,0	35,3	24,7	25,7
<i>PSS</i>	0,0	0,0	0,0	0,0	0,0	0,0

Таким образом, при $\alpha = 0$, точность алгоритма на основе ВНС выше, а при $\alpha \neq 0$ точность выше у алгоритма на основе АВСКМ. При этом во всех случаях алгоритм на основе АВСКМ демонстрировал лучшее быстродействие в отличие от алгоритма на основе ВНС при ИП НДП. Для повышения надежности ИП НДП на практике рекомендуется последовательно использовать оба метода. Совпадение результатов ИП НДП повышает его точность. Апробация созданного алгоритмического и программного обеспечения ИП НДП показала его работоспособность и практическую значимость.

В заключении приведены основные результаты работы.

В приложениях приведены экспериментальные данные, полученные по результатам комплексного исследования интервального прогнозирования НДП, два акта внедрения и справка о внедрении результатов диссертационной работы.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ ДИССЕРТАЦИИ

1. Разработана постановка задачи комплексного исследования интервального прогнозирования нестационарных динамических показателей с использованием кластерных и вероятностных нейронных моделей.

2. Предложено усовершенствованное алгоритмическое обеспечение интервального прогнозирования на основе адаптивной вероятностно-статистической кластерной модели с использованием численных методов проверки подобия кластеров на основе коэффициента «линейного сопряжения» и алгоритма оценки «интервальных» вероятностей;

3. Создано алгоритмическое обеспечение интервального прогнозирования на основе вероятностной нейронной модели с использованием усовершенствованного алгоритма обучения и модифицированной функции классификации.

4. Обоснована необходимость и разработан вычислительный алгоритм псевдослучайного увеличения объема выборки нестационарных динамических показателей на основе компьютерного моделирования с использованием генератора «Вихрь Мерсенна».

5. Создано программное обеспечение интервального прогнозирования нестационарных динамических показателей – программный комплекс «Интервальное прогнозирование нестационарных динамических показателей», основанное на разработанном алгоритмическом обеспечении адаптивной вероятностно-статистической кластерной и вероятностной нейронной моделях. Получено два свидетельства о государственной регистрации программ для ЭВМ.

6. Предложена технология комплексного исследования интервального прогнозирования нестационарных динамических показателей в качестве которых были использованы показатели двух типов: а) валютные курсы, сглаженные простым скользящим средним (полученные через открытые источники Интернет); б) еженедельные цены на сливочное масло, пшеничную муку (получены в территориальном органе Федеральной службы государственной статистики по Иркутской области) и ежемесячные объёмы выданных поручительств (получены в «Иркутском областном гарантийном фонде» поддержки субъектов малого и среднего предпринимательства) в городе Иркутске, объём выборки которых был увеличен с помощью разработанного алгоритма псевдослучайного увеличения объема выборки нестационарных динамических показателей на основе компьютерного моделирования», выработаны рекомендации по выбору значений параметров разработанных алгоритмов для прогнозирования НДП.

7. Проведена апробация созданного программного комплекса «Интервальное прогнозирование нестационарных динамических показателей». Программный комплекс используется при осуществлении и планировании финансово-экономической деятельности как государственными, так и коммерческими организациями. Апробация подтвердила его работоспособность и практическую значимость. Получены акты внедрения результатов диссертационного исследования в ООО «Сибпрофкосметик» и в «Иркутском областном гарантийном фонде» поддержки субъектов малого и среднего предпринимательства, а также получена справка о внедрении результатов диссертационного исследования в территориальном органе Федеральной службы государственной статистики по Иркутской области.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ:

В изданиях, рекомендованных ВАК:

1. Лузгин, А.Н. Теории и перспективы прогнозирования валютных и фондовых рынков как социально-экономических явлений [Электронный ресурс] / А.Н. Лузгин // Известия ИГЭА. – 2013. – №2. – Режим доступа: <http://eizvestia.isea.ru/pdf.aspx?id=17010>.

2. Лузгин, А.Н. Адаптивная вероятностно-статистическая кластерная модель интервального прогнозирования нестационарных динамических показателей / А.Н. Лузгин, Ю.М. Краковский // Современные технологии. Системный анализ. Моделирование. – 2015. – № 1(45). – С. 80 – 84.

3. Лузгин, А.Н. Программное обеспечения интервального прогнозирования нестационарных динамических показателей / А.Н. Лузгин, Ю.М. Краковский // Вестник ИрГТУ. – 2015. – № 4. – С. 12 – 16.

Свидетельства о государственной регистрации программы для ЭВМ:

4. Свидетельство о государственной регистрации программы для ЭВМ №2014611270 от 29.01.2014. Программный информационно-аналитический модуль «Альтернатива» / А.Н. Лузгин // Федеральная служба по интеллектуальной собственности и товарным знакам. – 2012.

5. Свидетельство о государственной регистрации программы для ЭВМ №2015617751 от 22.07.2015. Программный комплекс «Интервальное прогнозирование нестационарных динамических показателей» / А.Н. Лузгин // Федеральная служба по интеллектуальной собственности и товарным знакам. – 2015.

В других изданиях:

6. Лузгин, А.Н. Методика проведения анализа валютных и фондовых рынков на основе матричных временных рядов / А.Н. Лузгин // Материалы всероссийской научно-практической заочной конференции «Новые информационные технологии в экономике, управлении, образовании». – 2012. – С. 182-188.

7. Лузгин, А.Н. Адаптивные статистические модели, как инструмент анализа и прогнозирования синергетических систем [Электронный ресурс] / А.Н. Лузгин // Аспирантские чтения ИГЛУ: сборник научных статей. – 2012. – Режим доступа: 1 электрон. опт. диск (CD-R).

8. Лузгин, А.Н. Прогнозирование стохастических нестационарных динамических показателей на основе математических моделей / А.Н. Лузгин, Ю.М. Краковский // Вопросы естествознания. – 2014. – № 2(3). – С. 42 –50.

9. Лузгин, А.Н. Исследование влияния параметров вероятностно-статистической кластерной модели на результаты интервального прогнозирования / А.Н. Лузгин // Материалы IV Международной практической конференции «Безопасность регионов – основа устойчивого развития». – 2014. – С. 208-212.

10. Лузгин, А.Н. Сравнение точности интервального прогнозирования нестационарных динамических показателей на основе кластерной модели и нейронной сети / А.Н. Лузгин, Ю.М. Краковский // Труды XX Байкальской Всероссийской конференции «Информационные и математические технологии в науке и управлении». – 2015. – Т.3. – С. 36-43.

Подписано в печать 21.10.2015 г.

Формат бумаги 60 × 84 1/16. Бумага офсетная.

Печать трафаретная. Усл. печ. л. 1,0. Заказ 5233. Тираж 100 экз.

Отпечатано с готового оригинал-макета в ИПО ФГБОУ ВПО
«Байкальский государственный университет экономики и права»
664003, г. Иркутск, ул. Ленина, 11.
