-156606

На правах рукописи

Яковлева Галина Леонтьевна

УДК 004.4

ИССЛЕДОВАНИЕ И РАЗРАБОТКА АВТОМАТИЗИРОВАННЫХ СИСТЕМ ПРОГНОЗИРОВАНИЯ НА ОСНОВЕ МЕТОДОВ ТЕОРИИ НЕЙРОННЫХ СЕТЕЙ

Специальность 05.13.06 - Автоматизированные системы управления

Автореферат

диссертации на соискание ученой степени
кандидата технических наук

Москва - 2000

Работа выполнена в Московском Государственном Техническом Университете им. Н.Э. Баумана

Научный руководитель: кандидат технических наук, доцент Ревунков Г.И.

Официальные оппоненты: доктор технических наук, профессор Карпов В.И., Московский государственный университет пищевых производств;

кандидат технических наук, доцент Федоров Б.С., Центральный Банк Российской Федерации.

Ведущее предприятие: Институт проблем управления им.В.А.Трапезникова РАН.

Защита диссертации состоится "2" июня 2000 г. на заседании диссертационного совета Д 053.15.03 в Московском Государственном Техническом Университете им. Н.Э. Баумана по адресу: 107005, г. Москва, 2-ая Бауманская ул., д.5.

С диссертацией можно ознакомиться в библиотеке МГТУ им. Н.Э. Баумана.

Автореферат разослан "23" апреля 2000 г.

Ученый секретарь диссертационного совета
Иванов С.Р.

НТБ МГТУ им. Н.Э. Баумана
1568606
Яковлева Г.П. Исследование

НА ДОМ НЕ ВЫДАЕТСЯ

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность:

Представляемая диссертационная работа посвящена вопросам создания автоматизированных систем прогнозирования (АСП). Данный класс систем включает в себя системы, использующие при прогнозе различный математический аппарат, для реализации которого применяется широкий спектр методов и алгоритмов. Поэтому задача была сужена, и в данной работе рассматриваются только АСП, решающие задачи регрессионного анализа с использованием элементов теории нейронных сетей

Одной из наиболее часто встречающихся проблем, встающих перед учеными различных специальностей, является проблема нахождения зависимости между некоторым набором величин, описывающих поведение объекта (прогнозируемые величины и аргументы прогноза). Эта зависимость может быть выведена из теории и (или) может быть получена на основании экспериментальных исследований. В любом случае, при построении математической модели должны использоваться данные об исследуемом объекте, на основании которых мог бы быть сделан вывод о достаточной точности описания объекта моделью.

Несмотря на идентичность математического обеспечения, существующие АС, решающие задачи регрессионного анализа при прогнозировании, ориентированы на решение проблем конкретной предметной области. Поэтому *определение обобщенной методики построения данных АС является актуальной задачей.*

При определении методики построения АСП были определены следующие задачи, подлежащие решению в рамках работы:

1. Разработка набора функций для формирования функции, аппроксимирующей искомую зависимость прогнозируемой величины от аргументов прогноза. В существующих АС либо реализованы функции, адаптированные для решения конкретных задач прогнозирования, либо, как в случае нейропакетов, считается, что предлагаемые функции могут использоваться для решения самого широкого круга задач, учет особенностей предметной области происходит при проектировании архитектуры нейросети. Поэтому *проблема разработки набора функций, позволяющего создавать эффективно работающие модели для описания различных нелинейных зависимостей, является актуальной.*

2. Определение правил формирования аппроксимирующей функции. При ее реализации нейронной сетью - определение архитектуры нейромодели.

3. Разработка методов и алгоритмов обучения нейронных сетей выбранного класса, позволяющих улучшить качество аппроксимации. На сегодняшний день в АС для решения задач нахождения параметров аппроксимирующей функции успешно применяются численные методы, методы Гаусса-Ньютона и Марквардта, безусловной минимизации, использующие производные функции, обратного распространения ошибки, обучения Коши и Больцмана и т.д. Все перечисленные методы имеют свои достоинства и недостатки, и *создание на их основе методов и алгоритмов, умножающих достоинства и исключающих наиболее существенные недостатки - является актуальной задачей.*

4. Определение методов оценки значимости аргументов и параметров, анализа результатов применения моделей и методов Для многих объектов прогноза характерно изменение с течением времени найденных зависимостей, т.е. возникает необходимость в переопределении условий проведения прогноза. Поэтому в системе должны быть реализованы методы, сигнализирующие об "устаревании" используемых наборов и моделей.

МГТ

Большой вклад в развитие теории нейронных сетей и ее применении к решению задач регрессионного анализа внесли Хебб Д., Минский М., Розенблат Ф., Уидроу Б., Андерсон, Кохонен Т., Гроссберг С., Коско Б., Галушкин А.И., Гобань А.Н., Мкртчян С.О. и многие другие.

Рассматриваемые АС используются в различных предметных областях, например, в промышленности, финансовой деятельности, социологии, медицине и т.д. В данной работе в качестве примера АСП, иллюстрирующего эффективность применения разработанных моделей, методов и алгоритмов, рассматривается автоматизированная дилинговая система (АДС), основной задачей которой является прогноз тенденций финансовых рынков.

Целью настоящей работы является разработка для автоматизированных систем прогнозирования, решающих задачи регрессионного анализа, методов и алгоритмов, позволяющих повысить качество прогноза, с использованием элементов теории нейронных сетей.

Для достижения поставленной цели в работе предпринята попытка решения комплекса задач, в частности: научно-технических (проведена классификация существующих АСП, определена обобщенная методика построения АС данного класса, определены операции над временными рядами данных, посредством которых происходит предварительная обработка исходных данных, предложены математические модели, методы и алгоритмы для решения задач прогноза, на их основе разработаны модели, методы и алгоритмы АСП); прикладных (разработан информационно-программный комплекс, реализующий разработанные методы и алгоритмы; данный комплекс является ядром АСП, решающих задачи регрессионного анализа, на его основе разработана автоматизированная дилинговая система, предназначенная для использования в организациях, работающих на финансовых рынках); экспериментальных (проведены экспериментальные исследования эффективности работы реализованных моделей, методов и алгоритмов, разработаны рекомендации).

Методы исследования. При решении поставленных задач в работе использованы элементы теории детерминированного хаоса, математический аппарат линейного регрессионного анализа, линейной алгебры, теории нейронных сетей, теории оптимизации, нелинейного программирования, теории вероятностей, теории информационных систем и обработки данных.

Научная ценность работы состоит в следующем:

- Для формирования функции, аппроксимирующей зависимость прогнозируемой величины от множества аргументов прогноза, разработан набор функций на основе моделей, полученных в рамках теории детерминированного хаоса.
- Обоснованы и развиты методы теории нейронных сетей применительно для решения задачи аппроксимации искомой зависимости. Разработан обобщенный метод обучения двухслойной неоднородной искусственной нейросети с произвольным числом нейронов в первом слое путем разбиения ее на потоки.
- Введено понятие потока, где поток - это структура, являющаяся составной частью нейронной сети и пересекающаяся с другими подобными структурами только при формировании выходного сигнала сети.
- Разработаны методы и алгоритмы обучения многослойных искусственных нейросетей, основанные на методе обратного распространения ошибки, методах безусловной минимизации, использующих производные функции, и методе обучения Коши, применяющемся для выхода из областей локальных минимумов.

Достоверность полученных научных результатов, выводов и рекомендаций диссертационной работы подтверждена результатами экспериментальных исследований разработанной АДС, проведенных на базе Казначейства Промстройбанка России. Полученные в работе результаты наглядно демонстрируют эффективность использования разработанных моделей, методов и алгоритмов для решения задач прогноза.

Естественно, что погрешность прогноза во многом зависит от значимости выбранных аргументов прогноза. При эксплуатации данной автоматизированной системы на базе Казначейства Промстройбанка России была проведена работа по выявлению наиболее значимых аргументов для финансовых рынков и их инструментов. Так, с момента внедрения системы в опытную эксплуатацию в ноябре 1998 года средняя погрешность краткосрочного прогноза цен (горизонт 1-3 дня) для рынка облигаций составила 0.21%, для рынка акций - 0.88%, для валютного рынка -1.01%, средняя погрешность долгосрочного прогноза доходности с горизонтом не более 10 дней для рынка ценных бумаг и валюты составила 4.5%. Для сравнения, в аналогичных системах средняя погрешность краткосрочного прогноза равна 3-6%, средняя погрешность более долгосрочного прогноза (до 10 дней) достигает 15%.

Положения, выносимые на защиту:

- Обобщенная методика построения автоматизированных систем прогнозирования, решающих задачи регрессионного анализа.
- Набор функций, предоставленный для формирования функции, аппроксимирующей зависимость прогнозируемой величины от аргументов. Результаты экспериментальных исследований эффективности применения данных функций при решении задачи прогнозирования тенденций финансовых рынков.
- Развитие, обоснование и применение методов теории нейронных сетей для решения задачи аппроксимации искомой зависимости.
- Метод поточного обучения двухслойной неоднородной искусственной нейронной сети с произвольным числом нейронов в первом слое.
- Методы и алгоритмы обучения многослойных нейронных сетей, основанные на методе обратного распространения ошибки, методах безусловной минимизации, использующих производные функции, и методе обучения Коши. Результаты экспериментальных исследований эффективности использования разработанных методов и алгоритмов при решении задач прогнозирования тенденций финансовых рынков.
- Программная реализация предложенных и разработанных моделей, методов и алгоритмов для решения задач прогноза в виде библиотек функций, являющихся ядром автоматизированных систем прогнозирования, и программная реализация на их основе автоматизированной дилинговой системы.

Практическая ценность работы состоит в том, что предложенные и разработанные модели, методы и алгоритмы позволяют реализовывать автоматизированные системы, эффективно решающие задачи прогноза.

Разработанные библиотеки ядра АСП содержат функции, решающие задачи регрессионного анализа согласно определенным условиям прогноза. Благодаря предоставленным и разработанным моделям, методам и алгоритмам, обеспечены наилучшие показатели прогноза по критерию качество/время. Созданный программный комплекс АДС позволяет: работать с различными источниками данных; проводить визуальный анализ рынка посредством индикаторов технического анализа; выполнять прогноз поведения финансовых рынков и их инструментов; определять оптимальное распределение средств инвестора на основе прогноза ценовой динамики.

Реализация результатов

Полученные в работе результаты: математические модели, методы, алгоритмы и программы используются при прогнозировании тенденций финансовых рынков и определении структуры банковского портфеля в Казначействе Промстройбанка России. Копия акта о внедрении прилагается.

Внедрение результатов работы продолжается, они также могут быть использованы при создании иных автоматизированных систем прогнозирования, решающих задачи регрессионного анализа.

Апробация работы

Результаты работы были представлены на V-ой Всероссийской конференции "Нейрокомпьютеры и их применение" (Россия, Москва, 17-19 февраля 1999), конференции "Информатика и системы управления в XXI веке" (Россия, Москва, март 1999), международной конференции стран СНГ "Молодые ученые - науке, технологиям и профессиональному образованию для устойчивого развития: проблемы и новые решения" (Россия, Москва, декабрь 1999) Докладывались на семинарах "Современные информационные системы" и "Наукоемкие технологии и интеллектуальные системы" в МГТУ им. Н.Э. Баумана.

Публикации

Основные результаты работы опубликованы в 7 печатных работах.

Структура и объем работы

Диссертационная работа состоит из введения, пяти глав, заключения, списка литературы и двух приложений. Общий объем диссертации 219 страниц, 36 рисунков, список использованных источников из 126 наименований.

КРАТКОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении описана задача прогноза поведения объектов, решаемая автоматизированными системами прогнозирования рассматриваемого класса, и определены этапы ее решения.

Имеется объект прогноза, обладающий следующими свойствами. Совокупность числовых характеристик Z, описывающих или влияющих на поведение объекта прогноза, можно представить как множество векторов характеристик z_q (q = 1, ..., l), каждый элемент z_q^i которых - значение q-й характеристики в t_i момент времени, т.е. значения характеристик получены в дискретные моменты времени. Совокупность возможных прогнозируемых величин W является подмножеством Z.

Пусть на прогнозируемую величину w_θ оказывают влияние аргументы прогноза z_q (q = 1, ..., m) $m \le l$. Естественно, что даже в самых простых случаях невозможно учесть все факторы, влияющие на изменение состояния объекта прогноза, поэтому в данной работе считается, что определены наиболее значимые аргументы z_q (q = 1, ..., m) для рассматриваемой прогнозируемой величины w_θ. Значения w_θ известны в точках $t_i + \Delta t$ (i = 1, ..., n), они образуют вектор известных значений прогнозируемой величины $Y = (w_\theta^1, ..., w_\theta^n)$, элементы которого удобнее обозначать y_i (i = 1, ..., n). Значения аргументов прогноза $\{z_1, ..., z_m\}$ известны в точках t_i (i = 1, ..., n + 1), они образуют матрицу аргументов прогноза $X = \|z_q^i\|$ (i = 1, ..., n), (q = 1, ..., m), элементы которой удобно обозначать x_q^i (i = 1, ..., n), (q = 1, ..., m), и вектор $x^{(n+1)} = (x_1^{(n+1)}, ..., x_m^{(n+1)})$.

В задаче необходимо определить $y_{(n+1)}$, т.е. значение прогнозируемой величины w_θ в точке $t_{(n+1)}+\Delta t$

Описанная выше задача является задачей регрессионного анализа. Можно выделить следующие этапы ее решения:

1) сформировать аппроксимирующую функцию F(x, a), где a - вектор неизвестных параметров, x - вектор аргументов прогноза размерностью m;

2) найти параметры a, обеспечивающие наилучшую аппроксимацию согласно определенному критерию качества; в данной работе предложен метод наименьших квадратов (МНК), тогда целевая функция:

$$Q(a) = \sum_{i=1}^{n} (y_i - F(x^i, a))^2 \rightarrow \min; \quad (1)$$

3) вычислить $y_{(n+1)} = F(x^{(n+1)}, a)$ и погрешность прогноза ε,

Далее во введении обоснована актуальность темы диссертации, сформулирована цель настоящей работы. Представлена научная и практическая значимость полученных результатов.

В первой главе проведен анализ и классификация АСП (раздел 1.1), решающих задачи регрессионного анализа, определены их особенности (раздел 1.2). АСП характеризуются следующими основными особенностями: во-первых, данные АС должны функционировать в составе АСУ организаций, во-вторых, системы работают с большими объемами данных, поступающими из различных источников, и, в-третьих, реализуют многоитерационные алгоритмы обработки данных. Разработана классификация существующих АСП в зависимости от вида результатов анализа; способа работы с выборкой данных X; функций, предоставляемых для формирования F(x, a); типа функционирования и состава нейронной сети, реализующей F(x, a). Классификация и особенности исследуемых АС определяют требования к функциональным возможностям АСП, изложенным в разделе 1.3. Последний раздел главы посвящен описанию обобщенной методики построения АСП, решающих задачи регрессионного анализа. Несмотря на различность объектов прогноза, сказывающуюся на режим и характер работы с источниками данных, способ работы с выборкой данных, требования к временным и точностным характеристикам прогноза, а также наличие дополнительных задач, АС рассматриваемого класса имеют идентичное математического обеспечение. Поэтому целесообразно реализовать все программные и информационные модули, решающие задачи прогноза, в виде общего ядра (рис. 1). Возможно наличие других программных и информационных модулей, решающих дополнительные задачи, определяемые потребностями автоматизируемой предметной области. Дополнительные задачи решаются при реализации конкретной АСП. Т.к. пользовательский интерфейс должен быть адаптирован для работы с предметной областью, он также индивидуален для каждой АС. Вызов функций библиотек ядра и функций дополнительных модулей осуществляется из приложении, реализующих пользовательский интерфейс.



Рис.1 Обобщенная схема построения АСП

Вторая глава посвящена исследованию и разработке математических моделей, методов и алгоритмов решения задач прогноза, т.е. разработке математического обеспечения ядра АСП, решающих задачи регрессионного анализа.

В разделе 2.1 описаны операции над временными рядами, определенные в системе для преобразования исходных данных прогнозируемой величины и аргументов прогноза при подготовке данных. Необходимость предобработки данных обусловлена в первую очередь тем, что, как правило, значения аргументов прогноза и прогнозируемой величины должны находиться внутри определенных отрезков. Кроме того, с помощью преобразований можно получить более значимые аргументы прогноза, чем при использовании исходных данных.

В разделе 2.2 определены следующие функции и модели, предлагаемые для формирования аппроксимирующей функции F(x, a). Здесь $x^i(x_1^i, x_2^i, ..., x_m^i)$ - i-ый вектор аргументов прогноза (i = 1,..., n) (входные сигналы в нейросетевой терминологии); $y_{вых}$ - результат преобразований функцией i-го вектора входных сигналов (i = 1,... , n) (выходной сигнал в нейросетевой терминологии); a, a^i - векторы параметров размерностью m; $a_{вых}$ - выходные параметры; v, w, w_1, w_2 - числовые константы, причем v - натуральная и четная, w > 0, w_2 > 0; r, u, ε, a, b - внешние параметры моделей детерминированного хаоса.

1. Линейная функция:

$$y_{вых} = (x, a). \quad (2)$$

2. Нелинейные функции.

Требования к нелинейным функциям:

1) везде дифференцируемы и непрерывны;

2) являются гладким отображением (-∞, +∞) на ограниченный отрезок.

2. 1. Нелинейные функции, используемые в нейросетевом анализе для решения широкого круга задач (3). Данные функции можно назвать универсальными, т.к. они хорошо зарекомендовали себя при аппроксимации различных нелинейных зависимостей.

$$y_{вых} = a_{вых}f(x, a), \quad (3)$$

где в качестве f(x, a) предложены

$$f(x, a) = 1/[1 + EXP-(x, a)], \quad (4)$$

$$f(x, a) = [EXP(x, a) - w_1]/[EXP(x, a) + w_2], \quad w_1 > 0 \quad (5)$$

$$f(x, a) = COS(x, a), \quad (6)$$

$$f(x, a) = SIN(x, a), \quad (7)$$

$$f(x, a) = 1/[w + (x, a)^v], \quad (8)$$

$$f(x, a) = 1/[LN(w_1 + (x, a)^v) + w_2]. \; w_1 \geq 1 \quad (9)$$

2.2. Нелинейные функции, полученные в рамках теории детерминированного хаоса. Представленные ниже модели могут с успехом применяться при аппроксимации ряда нелинейных зависимостей (в случае, когда система проявляет детерминированный хаос). Экспериментальные исследования показали их эффективность при прогнозировании тенденций финансовых рынков. В работе предложено внешние параметры моделей детерминированного хаоса определять как нелинейные функции от скалярного произведения векторов аргументов прогноза и неизвестных параметров. К функциям и выходным параметрам предъявлены следующие дополнительные требования:

1) $y_{вых} \in [0, 1]$;

2) $f_1(x, a_1) > 0$ и $f_2(x, a_2) > 0$, $a_{вых1} > 0$ и $a_{вых2} > 0$.

$$y_{вых(i+1)} = \sum_{j=1}^{3} f_j(x^{(i+1)}, a^j)a_{вых}c_j, \text{ где} \quad (10)$$

2.2.1 Модель 1, в основе которой лежит бифуркация удвоения $y_{(i+1)} = ry_i(1 - y_i)$, $r > 0$:

$$y_{вых(i+1)} = f(y_{вых}, x, a_1, a_2, a_{вых}) = y_{вых}a_{вых1}f_1(x, a_1) - y_{вых}^2a_{вых2}f_2(x, a_2), \quad (11)$$

т.е. $c_1 = y_{вых}$, $c_2 = -y_{вых}^2$, $c_3 = 0$. (12)

2.2.2 Модель 2, в основе лежит касательная бифуркация $y_{(i+1)} = y_i + uy_i^2 + \varepsilon$, $u > 0$.

$$y_{вых(i+1)} = f(y_{вых}, x, a_1, a_2, a_3, a_{вых}) = y_{вых}a_{вых1}f_1(x, a_1) + y_{вых}^2a_{вых2}f_2(x, a_2) + a_{вых3}f_3(x, a_3), \quad (13)$$

т.е. $c_1 = y_{вых}$, $c_2 = y_{вых}^2$, $c_3 = 1$. (14)

2.2.3 Модель 3, в основе лежит бифуркация Хопфа $y_{(i+1)} = 1 - ay_i^2 + by_{i-1}$, $a > 0$, $|b| < 1$:

$$y_{вых(i+1)} = f(y_{вых}, x, a_1, a_2, a_3, a_{вых}) = a_{вых1}f_1(x, a_1) - y_{вых}^2a_{вых2}f_2(x, a_2) + y_{вых(i-1)}a_{вых3}f_3(x, a_3), \quad (15)$$

т.е. $c_1 = 1$, $c_2 = -y_{вых}^2$, $c_3 = y_{вых(i-1)}$, $|f_3(x, a_3)| < 1$, $|a_{вых3}| < 1$. (16)

В качестве функций $f_1(x, a_1)$ и $f_2(x, a_2)$ предложены (4), (8), (9); в качестве $f_3(x, a_3)$ второй модели - (5), (6), (7); третьей модели - (5) при $w_1 = w_2 = 1$.

Для реализации аппроксимирующей функции, состоящей не более, чем из одной линейной функции и из нескольких нелинейных, предложена двухслойная неоднородная искусственная нейронная сеть с произвольным числом нейронов в первом слое (раздел 2.3). Она представлена на рис. 2. Как правило, для получения множества входных и требуемых выходных сигналов выполняется предобработка значений аргументов и прогнозируемой величины (нормировка, масштабирование), соответственно, после получения выходных сигналов нейросети выполняется их интерпретация.

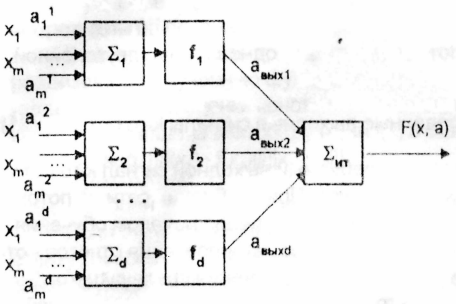


Рис.2 Двухслойная нейросеть.

Функция, реализуемая данной сетью.

$$F(x, a) = \sum_{j=1}^{d}(a_{вых}f_j(x, a^j)c_j), \quad (17)$$

где d - число нейронов,
f_j - функция j-го нейрона,
$x^j(x_1^j,...,x_m^j)$ - i-ый вектор вх. сигналов,
$a^j(a_1^j,...,a_m^j)$ - вектор параметров j-го нейрона,
$a_{вых}$ - вектор вых. параметров,
c_j используется для хаотических моделей.

В разделе 2.4 предложен метод обучения двухслойной, неоднородной нейронной сети с произвольным числом нейронов в первом слое, заключающийся в последовательном обучении потоков сети. Понятие "потока" введено в данной работе для обозначения структуры, являющейся составной частью нейросети и пересекающейся с другими подобными структурами только при формировании выходного сигнала сети. В зависимости от реализуемой функции потоки делятся на три класса:

1. Поток, реализующий линейную функцию (1 класс)

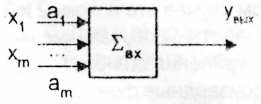
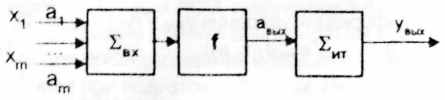


Рис. 3 Поток 1 класса.

Выходной сигнал потока, реализующего линейную функцию:

$$y_{вых} = (x, a). \quad (18)$$

2. Поток, реализующий нелинейную функцию, из числа традиционно использующихся в нейросетевом анализе (2 класс)



Выходной сигнал данного нелинейного потока:

$$y_{вых} = a_{вых} f(x, a). \quad (19)$$

Рис. 4 Поток 2 класса.

3. Поток, реализующий нелинейную функцию, полученную в рамках теории детерминированного хаоса

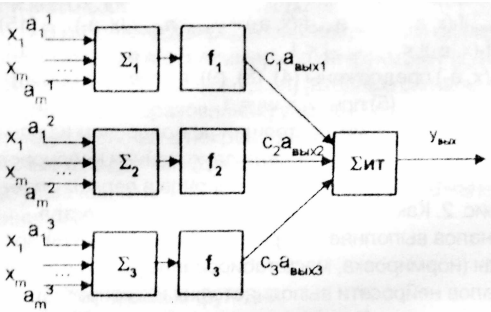


Рис. 5 Поток 3 класса.

Выходной сигнал потока, реализующего нелинейную функцию, полученную в рамках теории детерминированного хаоса:

$$y_{вых(i+1)} = \sum_{j=1}^{3} f_j(x^{(i+1)}, a^j) a_{вых} c_j, \quad (20)$$

где для первой модели:
$c_1 = y_{вых}$, $c_2 = -y_{вых}^2$, $c_3 = 0$;
для второй модели:
$c_1 = y_{вых}$, $c_2 = y_{вых}^2$, $c_3 = 1$;
для третьей модели:
$c_1 = 1$, $c_2 = -y_{вых}^2$, $c_3 = y_{вых(i-1)}$.

Пусть нейронная сеть содержит I потоков, тогда выходной сигнал двухслойной нейронной сети $Y_{вых}$ определяется:

$$Y_{вых} = \sum_{k=1}^{I} y_{вых}^k, \text{ где } y_{вых}^k \text{ - интерпретированные выходные сигналы k-го потока.} \quad (21)$$

Т.е. каждый поток обучается отдельно, и требуемый выходной сигнал k-го потока есть погрешность обучения предыдущих k - 1 потоков. При разбиении сети на потоки преобразование входных сигналов производится один раз перед началом обучения сети, а преобразование требуемых выходных сигналов и интерпретация ответов - отдельно для каждого потока после завершения процесса обучения предыдущего.

В разделе 2.5 рассмотрены методы и алгоритмы обучения многослойных нейронных сетей. Критерием оценки качества обучения определен метод наименьших квадратов, исключение составляет случай использования гребневой оценки для линейной аппроксимации вырожденной модели. Обучение может проводиться для одного примера, постранично и для всей выборки.

Для обучения потоков, выполняющих линейную аппроксимацию, применяются алгоритм решения полученного в рамках линейного регрессионного анализа матричного уравнения и методы безусловной минимизации, использующие производные функции (раздел 2.5.3). Установлено, что применение методов безусловной минимизации целесообразно, если число входных сигналов выше восьми.

Для обучения многослойных нейронных сетей (в данном случае это потоки 2 и 3 классов), чьи нейроны реализуют нелинейные функции, разработаны обучающие методы и алгоритмы, основанные на методе обратного распространения ошибки (раздел 2.5.2), методах безусловной минимизации, использующих производные функции, и методе обучения Коши. К недостаткам метода обратного распространения ошибки можно отнести то, что он достаточно медленно сходится к точке минимума целевой

функции, плохо справляется с локальными минимумами, глубокими и узкими местами. Недостатками методов безусловной минимизации, использующих производные функции, являются: во-первых, необходимость знать по крайней мере первые частные производные (а для некоторых методов и вторые) целевой функции по параметрам, но при обучении потоков двухслойной нейросети получить их в общем виде, зная первые и вторые частные производные функций нейронов, не составляет труда, при обучении многослойных сетей произвольной архитектуры используется метод двойственного функционирования; во-вторых, данные методы в общем случае находят точку локального минимума. Т.е. метод обратного распространения ошибки и методы безусловной минимизации, имеют один общий серьезный недостаток - они находят локальный минимум целевой функции, исключение составляет случай, когда целевая функция - квадратичная положительно определенная форма (поток 1 класса). Для борьбы с этим недостатком в работе предложено использовать метод обучения Коши, который, как и метод Больцмана (их описание приведено в разделе 2.5.4), в своем первоначальном виде выполняет случайные преобразования параметров нейронной сети и не предлагается для реализации в данной работе по причине очень медленной сходимости. Идея разработанных методов состоит в использовании случайного шага изменения параметров метода Коши a_c для выхода из локальных минимумов. Описание разработанных методов приведено в разделе 2.5.5, их алгоритмы приведены в разделах 2.5.6 и 2.5.7. В рамках работы реализовано три варианта, основанных на методах безусловной минимизации и методе обучения Коши.

В разделе 2.6 определены методы оценки значимости входных сигналов и параметров сети на основании функции оценки и по изменению выходных сигналов.

В качестве метода проверки возможности с помощью заданной нейронной сети реализовать полученную функцию предложен анализ выборочной оценки константы Липшица снизу и оценки константы Липшица сверху по архитектуре сети (раздел 2.7). Константа Липшица вектор-функции f(x) в области D (оценка сверху) определяется:

$$L_f = \sup_{X1, X2 \in D,\ X1 \neq X2} [\|f(x_1) - f(x_2)\| / \|x_1 - x_2\|], \qquad (22)$$

для оценки L_f используется эвклидова норма.

Если функция задана на конечной выборке, то можно получить выборочную константу Липшица L_{sam}, заменяя всю область D на конечное множество, это - оценка L_{sam} снизу. Функцию можно реализовать с помощью данной сети, когда

$$L_{sam} < L_f. \qquad (23)$$

Пусть l - число потоков нейронной сети, h - число нейронов потока, a_v^k - вектор входных параметров v-го нейрона (v = 1, ..., h) k-го потока, $a_{вых_v}^k$ - выходной параметр v-го нейрона k-го потока, $f_{kv}(x, a_v^k)$ - функция, реализуемая преобразователем v-го нейрона k-го потока. Если поток содержит один нейрон, индекс v опускается.

Для потока, реализующего линейную функцию, константа Липшица сверху равна:

$$L_k \le \|a^k\|. \qquad (24)$$

Для k-го потока, реализующего нелинейную функцию, из числа традиционно использующихся в нейросетевом анализе, константа Липшица сверху равна:

$$L_k \le \|a_{вых}^k\| \|a^k\| \max_x |f_k'(x, a^k)|. \qquad (25)$$

Для k-го потока, реализующего нелинейную функцию, из числа хаотических моделей, константа Липшица сверху равна:

$$L_k \le [\sum_{v=1}^{h} (\|a_{вых_v}^k\| \|a_v^k\| \max_x |f_{kv}'(x, a_v^k)|)^2]^{1/2}. \qquad (26)$$

Практика показала, что вместо максимума функции лучше использовать ее среднее значение по интервалу изменения входных сигналов.

Константа Липшица сверху по архитектуре сети равна:

$L_F \le [\sum_{k=1}^{i} L_k^2]^{1/2}$, где L_k - константа Липшица сверху для k-ого потока. (27)

Третья глава посвящена созданию программного комплекса, являющегося ядром АСП, решающих задачи регрессионного анализа.

В разделе 3.1 приведены требования к проектированию и реализации, учитывающие особенности АС данного класса. Данные требования явились определяющими при проектировании размещения типовых компонент АСП (раздел 3.2). Можно выделить следующие типовые функциональные компоненты, достаточные для формирования любого приложения на основе БД: PS (Presentation Services) - средства представления, PL (Presentation Logic) - логика представления, BL (Business or Application Logic) - прикладная логика, DL (Data Logic) - логика управления данными, DS (Data Services) - операции с базой данных, FS (File Services) - файловые операции. Размещение типовых компонент представлено на рис. 6.

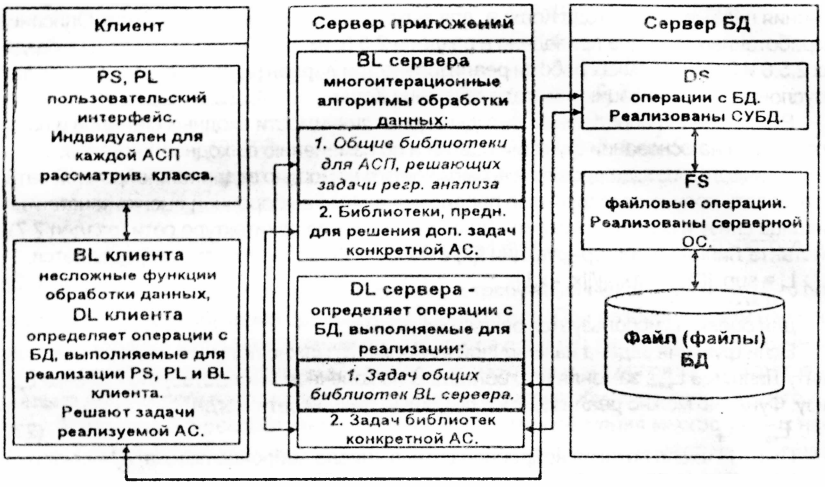


Рис. 6 Схема размещения типовых компонент исследуемых АСП (цветом выделены компоненты ядра; стрелками показаны пути обмена данными между приложениями и БД)

Клиентская часть, состоящая из пользовательского интерфейса, клиентских прикладной логики и логики управления данными, решает задачи, стоящие перед конкретной АС. В сервер приложений выделены серверные прикладная логика и логика управления данными, которые можно разделить на общие для всех АСП данного класса (составляют ядро системы, реализуют математическое обеспечение, описанное в гл. 2) и решающие дополнительные задачи АС. Операции с БД и файловые операции выполняются на сервере БД, они реализуются, соответственно, сервером СУБД и серверной ОС.

Далее в разделе 3.3 рассмотрены вопросы разработки информационного, программного и методического обеспечения модулей, являющихся общими для исследуемых АСП, на рис. 6 они выделены цветом. Библиотеки сервера приложений долж-

ны: во-первых, как можно меньше зависеть от аппаратно-программной платформы системы; во-вторых, оперативно реализовывать многоитерационные алгоритмы обучения нейросети, поэтому в качестве средств для разработки серверной логики приложений, входящих в ядро системы, определен ANSI-C; для разработки серверной логики управления данными - библиотеки программного интерфейса (C-API), предназначенные для обеспечения доступа к данным выбранного сервера СУБД. Поэтому выбор возможных серверов СУБД ограничен серверами СУБД, предоставляющими C-API, выбор возможных серверных ОС ограничен ОС, для которых реализован компилятор ANSI-C или C-SDK.

В разделе 3.3 приведена классификация таблиц БД, с которыми работают функции библиотек сервера приложений. Классификация проведена в зависимости от назначения данных, хранящихся в этих таблицах, и связей между таблицами, описаны потоки данных, поступающие в БД системы.

Последний раздел главы посвящен краткому описанию методического обеспечения АСП. Оно включает в себя электронную справочную систему и документацию, определенную требованиями ГОСТ.

В четвертой главе рассмотрены вопросы разработки автоматизированной дилинговой системы (АДС), в которой задача прогноза тенденций финансовых рынков решается с использованием функций ядра АСП.

В разделе 4.1 рассмотрено функционирование АДС в составе автоматизированной банковской системы (АБС) и описана основная задача, решаемая АДС - прогноз тенденций финансовых рынков. Данная задача является задачей регрессионного анализа и для ее решения в работе предложено использовать методы и алгоритмы, приведенные в главе 2. В разделе 4.1 также выделены четыре подсистемы АДС: ввода данных, технического анализа рынка, прогноза тенденций финансовых рынков, оптимального распределения свободных средств инвестора между финансовыми рынками и их инструментами, и разработаны требования к функциональным возможностям каждой из подсистем.

Раздел 4.2 посвящен описанию функциональных возможностей перечисленных подсистем. В разделе 4.2.1 определены механизмы работы АДС с внешними источниками данных (подсистема ввода): для работы с источниками, покрывающими основные информационные потребности системы, разрабатываются индивидуальные конвертеры, считывающие данные из формата внешнего источника и управляющие записью в БД системы, для работы с прочими источниками предоставляются конвертеры из файлов текстового и DBF форматов установленной структуры и ручной ввод посредством соответствующего приложения.

Кроме задачи прогноза тенденций финансовых рынков (раздел 4.2.3) АДС должна решать задачи визуального анализа рынка посредством технического анализа (подсистема технического анализа, раздел 4.2.2) и определения стратегии поведения в прогнозируемой рыночной ситуации (подсистема оптимального распределения свободных средств инвестора, раздел 4.2.4). Индикаторы технического анализа реализуются посредством операций над временными рядами, определенных для преобразования исходных данных прогнозируемой величины и аргументов прогноза. Для решения второй дополнительной задачи АДС в разделе приведены разработанные математические модели задач минимизации риска инвестиций и минимизации риска при фиксированном уровне доходности, т.е. данная задача является двухкритериальной. При создании моделей использовались элементы теории оптимального портфеля Марковица. В предлагаемых моделях исходными данными являются прогнозируемые доходности финансовых инструментов. В случае, если составляющие портфеля могут

быть как положительными, так и отрицательными (модель Блека), задачи минимизации решаются в общем виде методом неопределенных множителей Лагранжа. Если составляющие портфеля могут быть только положительными (модель Марковица), задачи минимизации решаются методом проекции градиента, на каждой итерации которого предусмотрена процедура возврата очередного приближения градиентного спуска на допустимое множество, если найденная точка ему не принадлежит. Такой возврат производится проектированием найденной точки на допустимое множество.

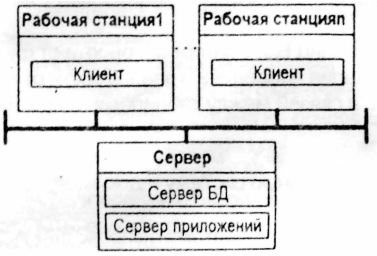


Рис. 7 Схема конфигурации АДС

Последний раздел главы посвящен реализации АДС. Разработка АДС и экспериментальные исследования проведены на базе Казначейства Промстройбанка России. При этом использовалась конфигурация системы, приведенная на рис.7, где сервер: SUN Server Ultra Enterprise 4500: 4 процессора, арх-ра Superscalar SPARC™ v 9, основная память 6ГБ; серверная ОС: SUN Solaris 2.6; сервер СУБД: Oracle 7.3 Server; рабочая станция: Intel Pentium 90, RAM - 32МБ, HDD - 1ГБ; клиентская ОС: Windows NT Workstation 4.0.

С учетом особенностей АДС и требований к проектированию и реализации в разделе 4.3 обоснованы рекомендуемые и возможные варианты выбора в зависимости от масштаба организации сервера СУБД, серверной и клиентской ОС, аппаратной платформы, определены средства разработки клиентских приложений. Для организаций любого масштаба рекомендуемым сервером СУБД является сервер семейства Oracle, рекомендуемой серверной ОС - UNIX-подобная ОС и клиентской ОС - Windows NT Workstation. В качестве средства разработки приложений клиента определен программный продукт Borland Delphi 3.0 (это ограничивает выбор клиентских ОС Microsoft Windows 9x и Microsoft NT Workstation). Далее в разделе 4.3 рассмотрены реализации АДС, взаимодействие приложений и описание потоков данных.

Пятая глава посвящена описанию экспериментальных исследований, иллюстрирующих эффективность применения предложенных и разработанных моделей, методов и алгоритмов на примере задачи прогноза тенденций финансовых рынков, решаемой АДС. В разделе 5.1 определены задачи, ставящиеся перед экспериментальными исследованиями, результаты которых приводятся в последующих разделах данной главы.

Раздел 5.2 посвящен сравнению качества аппроксимации зависимости прогнозируемой величины от аргументов с помощью двухслойной и трехслойной нейронных сетей. В экспериментах, описанных в этом разделе, рассматривались нейронные сети с одинаковым количеством нейронов в каждом слое. Обучение проводилось методами обратного распространения ошибки и безусловной минимизации, усовершенствованными с помощью метода обучения Коши. Установлено, что незначительное улучшение целевой функции при использовании трехслойной сети приводит к существенному увеличению времени обучения. Целевая функция, полученная при аппроксимации нелинейной составляющей с помощью трехслойной нейронной сети, не более чем на 7.5% лучше целевой функции, полученной с помощью двухслойной сети. Число итераций при обучении трехслойной сети методами безусловной минимизации, использующими производные функции, на 11-25% больше, чем при обучении двухслойной сети. Время обучения трехслойной сети методом обратного распространения

ошибки на 15-22% больше, чем время обучения этим же методом двухслойной сети. Эффект уменьшения целевой функции при использовании трехслойной нейронной сети легко достигается с помощью двухслойной путем увеличения числа нейронов в первом слое (или количества потоков).

В разделе 5.3 проведена оценка величины линейной составляющей зависимости прогнозируемой величины от некоторого множества аргументов прогноза для различных финансовых рынков. При проведении экспериментальных исследований выявлено, что величина линейной составляющей зависимости прогнозируемой величины от аргументов, полученной путем обучения линейного потока (поток 1 класса), достаточно велика. За время эксплуатации системы средняя погрешность прогноза цен, полученная при реализации линейной аппроксимации, составила: для рынка облигаций 1.3%, для рынка акций 2.6%, для валютного рынка 3.7%.. Поэтому рекомендуется в состав функций, реализуемых нейронами сети, включать линейную функцию.

В разделе 5.4 проиллюстрированы исследования качества аппроксимации нелинейной составляющей посредством различных функций из предоставляемого системой набора. В результате разработаны следующие рекомендации по выбору нелинейных функций для реализации потоками сети, относящихся к двум различным классам: хорошо зарекомендовавших себя в нейросетевом анализе и полученных в рамках теории детерминированного хаоса. Внутри каждого класса установлены наборы функций, обеспечивающие наиболее высокое качество нелинейной аппроксимации, отдельные улучшения целевой функции при переходе от одного набора функций к другому достигали 20%. Получено, что в подавляющем числе экспериментов (из 30 экспериментов в 27) использование функций, полученных в рамках теории детерминированного хаоса, обеспечивает улучшение целевой функции (сравнивались наборы функций, наиболее хорошо зарекомендовавшие себя внутри классов), в среднем для различных прогнозируемых величин оно составило 15.3%. Данное утверждение проиллюстрировано на рис.8 на примере аппроксимации нелинейной составляющей средневзвешенной цены облигации ОФЗ 27010. Использование третьей модели, полученной в рамках теории детерминированного хаоса (длинный пунктир), позволило улучшить целевую функцию на 15.5% по сравнению с сетью, содержащей нейроны, реализующие функции (8), (5), (6) (короткий пунктир), сплошной линией на рисунке показаны требуемые значения выходного сигнала.

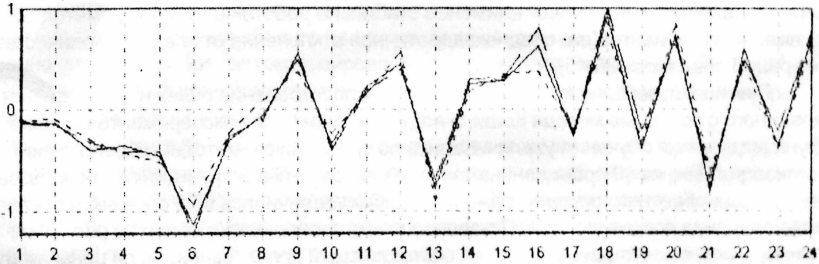


Рис. 8 Иллюстрация аппроксимации нелинейной составляющей с помощью нейронных потоков, реализующих различные нелинейные модели.

Улучшение целевой функции не наблюдалось при реализации потоками одной сети функций, принадлежащих различным классам.

В разделе 5.5 изучена работа методов, предлагаемых системой для обучения потоков сети. При обучении потоков, реализующих линейные функции, рекомендуется

алгоритм решения полученного в рамках линейного регрессионного анализа матричного уравнения, если количество аргументов прогноза не более восьми, иначе - рекомендуются методы безусловной минимизации, использующие производные функции. Результат линейной аппроксимации проиллюстрирован на рис. 9 на примере средневзвешенной цены акции РАО "ЕЭС". В качестве аргументов использовались: цена закрытия за предыдущий день, оценка общественно-политической ситуации для акций, курс доллара, предоставляемый Центробанком России, индекс "REUTERS" по АДР российских компаний, бразильский фондовый индекс BOVESPA, индекс РТС на торгуемые российские акции, индекс RTX OHKL. Здесь средняя погрешность прогноза по всей выборке равнялась 2.03%, погрешность прогноза цены на 12.05.1999 - 3.24%.

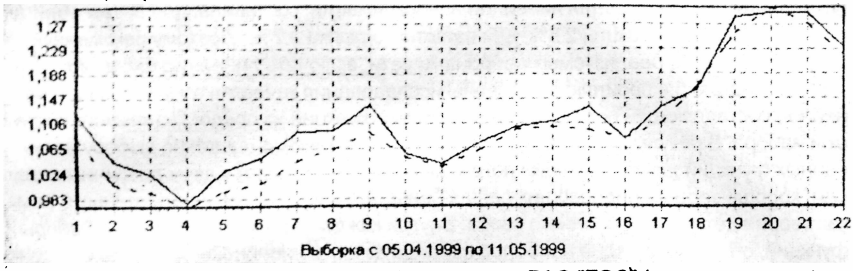


Рис. 9 Графики средневзвешенной цены акции РАО "ЕЭС" (сплошная линия) и результат линейной аппроксимации (пунктирная линия).

При обучении потоков, реализующих нелинейные функции, методами безусловной минимизации, усовершенствованными с помощью метода Коши, установлено:

- качество аппроксимации нелинейной составляющей при обучении с помощью методов наискорейшего спуска, сопряженных градиентов и ДФП идентично;

- наиболее быструю сходимость к точке локального минимума из трех рассматриваемых методов обеспечивает метод ДФП (для нахождения точки локального минимума методом ДФП в среднем требовалось 59 итераций, а в случае использования метода наискорейшего спуска - 82 итерации);

- наиболее стабильно работающий метод - метод наискорейшего спуска (его среднеквадратичное отклонение от среднего количества итераций равнялось 15.29), он же и наиболее медленный; наименее стабильно работающий метод - метод сопряженных градиентов(его среднеквадратичное отклонение от среднего количества итераций равнялось 39.45).

При использовании метода обратного распространения ошибки, усовершенствованного с помощью метода Коши, в подавляющем числе экспериментов целевая функция и время обучения увеличивались по сравнению с методами безусловной минимизации. Так, из 40 проведенных экспериментов целевая функция при использовании метода обратного распространения ошибки увеличилась в 38 случаях, в среднем ее увеличение составило 18%. Сказанное выше проиллюстрировано на рис. 10 на примере аппроксимации нелинейной составляющей средневзвешенной цены акции РАО "ЕЭС" (сама нелинейная составляющая показана сплошной линией). Использование метода наискорейшего спуска (короткий пунктир) позволило уменьшить целевую на 16.8% по сравнению с результатом обучения методом обратного распространения ошибки (длинный пунктир).

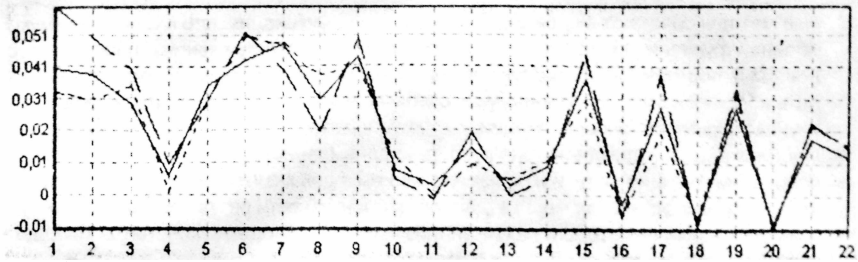


Рис. 10 Иллюстрация улучшения целевой функции при обучении методом, основанным на методах безусловной минимизации и Коши, по сравнению с обучением методом, основанным на методах обратного распространения ошибки и Коши.

Поэтому для обучения потоков, реализующих нелинейные функции, рекомендуются методы наискорейшего спуска, сопряженных градиентов и ДФП, усовершенствованные с помощью метода Коши.

Погрешность прогноза во многом зависит от значимости выбранных аргументов прогноза, т.е. от профессионализма аналитика. При эксплуатации системы на базе Казначейства Промстройбанка России была проведена работа по выявлению наиболее значимых аргументов для различных финансовых инструментов. За время опытной эксплуатации АДС средняя погрешность краткосрочного прогноза цен для рынка облигаций составила 0.21%; для рынка акций - 0.88%; для валютного рынка -1.01%. В аналогичных системах средняя погрешность краткосрочного прогноза равна 3-6%.

В разделе 5.6 проиллюстрировано решение задач оптимизации распределения средств для различных классов портфеля. Показана состоятельность рыночных моделей, использующих прогнозируемую доходность элементов портфеля вместо вероятностных характеристик их доходности за предшествующие периоды. Так, за время опытной эксплуатации АДС средняя погрешность прогнозируемой доходности портфеля с горизонтом 10 дней равнялась 4.5%, в аналогичных АС она достигает 15%.

Основные результаты диссертации

1. Разработано информационное, математическое, программное и методическое обеспечение программного комплекса, автоматизирующего процесс прогнозирования будущего состояния объекта на основе текущих данных, что позволило качественно улучшить работу автоматизированных систем прогнозирования.
2. Определена обобщенная методика построения автоматизированных систем прогнозирования, решающих задачи регрессионного анализа.
3. Разработан набор функций, предоставленный для формирования функции, аппроксимирующей зависимость прогнозируемой величины от аргументов прогноза. Их использование позволило создавать наиболее эффективно работающие модели.
4. Разработан метод поточного обучения двухслойной неоднородной искусственной нейронной сети с произвольным числом нейронов в первом слое, позволивший упростить процесс обучения нейронных моделей этого класса. В работе введено понятие потока - структуры, являющейся составной частью нейросети и пересекающейся с другими подобными структурами только при формировании выходного сигнала сети.
5. Разработаны методы и алгоритмы обучения многослойных нейронных сетей, основанные на методе обратного распространения ошибки, методах безусловной минимизации, использующих производные функции, и методе обучения Коши. Дан-

ные методы сочетают в себе относительно быструю сходимость к точке минимума целевой функции, присущую методам обратного распространения ошибки и безусловной минимизации, использующим производные функции, и способность выходить из точек локальных минимумов, свойственную методу Коши, что позволило улучшить качество аппроксимации.

6. Предложены методы оценки значимости входных сигналов и параметров сети на основании функции оценки и по изменению выходных сигналов, а также метод проверки возможности с помощью заданной нейронной сети реализовать полученную функцию, в основе которого лежит анализ выборочной оценки константы Липшица снизу и оценки константы Липшица сверху по архитектуре сети.
7. Проведены экспериментальные исследования, наглядно демонстрирующие эффективность применения предложенных и разработанных моделей, методов и алгоритмов на примере задачи прогноза тенденций финансовых рынков, разработаны рекомендации. С момента внедрения системы в опытную эксплуатацию в ноябре 1998 года средняя погрешность краткосрочного прогноза цен для рынка облигаций составила 0.21%, для рынка акций - 0.88%, для валютного рынка -1.01%. Для сравнения, в аналогичных системах средняя погрешность краткосрочного прогноза равна 3-6%.

Основные положения диссертации изложены в следующих работах:

1. Яковлева Г.Л., Лисицкий Л.А. Автоматизированное определение стоимости недвижимости // Материалы межрегиональной научно-практической конференции. - Саратов, 1999. - Ч. 2.- С. 146-150.
2. Яковлева Г.Л., Яковлев В.Л., Малиевский Д.А. Нейросетевая экспертная система управления портфелем банка // Нейрокомпьютеры и их применение: Сборник докладов V Всероссийской конференции. - М., 1999. - С. 291-294.
3. Яковлева Г.Л., Яковлев В.Л., Лисицкий Л.А. Применение нейросетевых алгоритмов к анализу финансовых рынков // Информационные технологии. - 1999.- № 8. - С. 25-30.
4. Яковлева Г.Л., Яковлев В.Л. Нейросетевые математические модели прогнозирования тенденций финансовых рынков // Молодые ученые - науке, технологиям и профессиональному образованию для устойчивого развития: проблемы и новые решения: Сборник докладов международной конференции стран СНГ. - М., 1999. - Ч. 2. - С. 74-78.
5. Яковлева Г.Л., Яковлев В.Л. Хаотические модели в виде двухслойных нейронных сетей и их применение в менеджменте инноваций // Молодые ученые - науке, технологиям и профессиональному образованию для устойчивого развития: проблемы и новые решения: Сборник докладов международной конференции стран СНГ. - М., 1999. - Ч. 2. - С. 78-81.
6. Яковлева Г.Л., Яковлев В.Л. Методология построения интегрированных финансово-экономических экспертно-аналитических систем // Молодые ученые - науке, технологиям и профессиональному образованию для устойчивого развития: проблемы и новые решения: Сборник докладов международной конференции стран СНГ. - М., 1999. - Ч. 2. - С. 81-85.
7. Яковлева Г.Л., Яковлев В.Л. Математические модели прогнозирования тенденций финансовых рынков, полученные в рамках теории детерминированного хаоса и реализуемые при помощи нейросетевых алгоритмов // Нейроинформатика-2000: Сборник докладов 2-ой Всероссийской научно-технической конференции. - М., 2000. - Ч. 2. - С. 156-162.

Подписано к печати "24" 04 2000 г. Зак.аз 59т. Объем 1 п.л. Тираж 100 экз.
Типография МГТУ им. Н.Э.Баумана, 107005, г. Москва, 2-ая Бауманская ул., д.5.