Учреждение образования
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

УДК 004.032.26

МАНЬЯКОВ Николай Владимирович

АЛГОРИТМЫ ОБУЧЕНИЯ НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧАХ ПРОГНОЗИРОВАНИЯ

05.13.17 – теоретические основы информатики

Автореферат диссертации на соискание ученой степени
кандидата технических наук

Минск 2005

Библиотека БГУИР

Работа выполнена в учреждении образования «Брестский государственный технический университет».

Научный руководитель - д.т.н., профессор Головко В.А. (Брестский государственный технический университете, кафедра интеллектуальных информационных технологий)

Официальные оппоненты: д.т.н., профессор Птичкин В.А. (Белорусский государственный университет информатики и радиоэлектроники, кафедра информационных технологий автоматизированных систем)

к.ф.-м. н., доцент Краснопрошин В.В. (Белорусский государственный университет, кафедра математического обеспечения автоматизированных систем управления)

Оппонирующая организация – НИИ прикладных физических проблем им. А.Н. Севченко при Белорусском государственном университете

Защита состоится « 22 » сентября 2005 года в 14.00 часов на заседании совета по защите диссертаций Д 02.15.04 при учреждении образования «Белорусский государственный университет информатики и радиоэлектроники» по адресу: 220013, г. Минск, ул. П. Бровки, 6, БГУИР, 1 уч. корпус, тел. 239-89-89.

С диссертацией можно ознакомиться в библиотеке учреждения образования «Белорусский государственный университет информатики и радиоэлектроники.

Автореферат разослан « 8 » августа 2005г.

Библиотека БГУИР

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы диссертации. Одной из актуальных задач является разработка методов прогнозирования временных рядов. Наиболее развитыми из них являются статистические прогнозирующие классы моделей регрессии, авторегрессии и скользящего среднего, а также их смешанные типы, такие как модель Бокса-Дженкинса, АРИСС (обобщенная интегральная модель авторегрессии – скользящего среднего) и др. При этом процедуры выбора и адаптации статистических прогнозных моделей являются сложными и во многих случаях требующими эмпирического подбора большого количества параметров. С другой стороны модели на основе нейронных сетей характеризуются способностью сети к обучению и самоорганизации, а также к обобщению результатов, что способствует их широкому применению при прогнозировании различных временных процессов.

При этом следует отметить, что при нейросетевом решении различных классов задач основной проблемой является выбор адекватной архитектуры сети и методов ее функционирования. При прогнозировании временных последовательностей такими проблемами являются: количество слоев, частота дискретизации подаваемых на вход сигналов, количество элементов в слоях.

Другой важной задачей является создание эффективных алгоритмов обучения данного класса нейронных сетей. Большое количество работ посвящено созданию различных алгоритмов обучения. Но многие из предложенных алгоритмов из-за своей сложности используются только для однослойных сетей. Для многослойных сетей наиболее широкое применение получил алгоритм обратного распространения ошибки, основанный на методе градиентного спуска. Тем не менее, отмечается его трудоемкость, большие временные затраты, неустойчивость, необходимость эмпирического задания параметров обучения, что часто не приводит к оптимальному обучению сети. Во многом это зависит от определения констант обучения, инициализация и изменение которых не всегда приводит к хорошим результатам адаптации синаптических связей и не дает быстрого схождения к точке минимума функции ошибки.

Связь работы с крупными научными программами, темами. Работа выполнялась в рамках госбюджетных тем: №02/202 «Методы анализа и прогнозирования хаотических процессов» (2002-2003, № госрегистрации 20022950), финансируемой и утвержденной МО РБ, и ГПОФИ «Природные комплексы». Задание 19. «Оценка состояния, прогноз изменения водного режима, степени деградации почв осушенных земель, научное обоснование устойчивого развития экосистем Белорусского Полесья» (2001-2005, № госрегистрации 20031263) в разделе прогноза изменения водного режима Белорусского Полесья.

Цель и задачи исследования. Целью диссертационной работы является разработка и совершенствование алгоритмов обучения нейронных сетей без

Библиотека БГУИР

обратных связей и их применение в задачах прогнозирования временных процессов. Для достижения этой цели решаются следующие задачи:

- разработать эффективные алгоритмы обучения персептронных нейронных сетей с различными функциями активации нейронных элементов, характеризующиеся быстродействием;
- разработать матричную алгоритмизацию процесса обучения многослойных нейронных сетей на основе градиентных методов, позволяющую упростить программную реализацию алгоритмов обучения;
- разработать методику построения базовой архитектуры прогнозирующей нейронной сети на основе сохранения динамических свойств временного ряда, проявляемых в схождении к идентичному аттрактору;
- исследовать эффективность алгоритмов обучения персептронных нейронных сетей для решения задач прогнозирования
- осуществить программную реализацию алгоритмов обучения для прогнозирования временных последовательностей.

Объект и предмет исследования. Объектом исследования являются многослойные гетерогенные нейронные сети прямого распространения. Предметом исследования являются алгоритмы обучения сетей данной архитектуры и их применение в задачах прогнозирования рядов расходов воды рек Беларуси.

Методология и методы проведенного исследования. Теоретические результаты диссертационной работы были получены с использованием методов теории нейронных сетей, методов оптимизации, теории вероятностей и математической статистики, математического моделирования, теории хаоса и нелинейной динамики. Экспериментальные исследования проводились с использованием персональных ЭВМ.

Научная новизна и значимость полученных результатов.

1. Разработаны алгоритмы обучения однослойных нейронных сетей, базирующиеся на использовании ковариационной матрицы входных образов и определении адаптивного шага обучения. Показана сходимость алгоритма обучения со строго монотонной функцией активации к точке глобального минимума при использовании алгоритма на основе ковариационной матрицы. Предложенные алгоритмы позволяют увеличить быстродействие и стабильность обучения по сравнению с известными.

2. Предложен ряд алгоритмов обучения для многослойных нейронных сетей, которые базируются на обобщенном методе наискорейшего спуска, послойном и двухпараметрическом обучении. Получены аналитические выражения вычисления адаптивного шага для различных алгоритмов обучения. В отличии от известных предложенные подходы позволяют снизить вычислительную сложность и увеличить точность обучения.

3. Предложена матричная алгоритмизация процесса обучения многослойной сети, позволяющая свести процесс обучения к матричным операциям, что приводит к упрощению программной реализации процессов обучения.

Библиотека БГУИР

4. Предложен способ построения архитектуры прогнозирующей нейронной сети на основе многослойного персептрона, который позволяет сохранять динамические свойства временного ряда и повысить точность прогнозирования.

5. Предложен нейросетевой подход для определения длины прогноза хаотической временной последовательности, что позволяет оценить адекватность полученных результатов экстраполяции.

Практическая значимость полученных результатов состоит в использовании новых и эффективных способов обработки информации для обучения нейронных сетей и прогнозирования рядов расходов воды рек Беларуси. Применение их позволяет повысить точность прогнозирования и улучшить эффективность обучения нейронных сетей. Основные результаты диссертационной работы были использованы при выполнении госбюджетных НИР, ориентированных на прогнозирование различных процессов. Предлагаемые методики и алгоритмы могут быть использованы при проектировании интеллектуальных систем обработки информации для решения задач прогнозирования временных последовательностей. Базовые результаты диссертационной работы применяются в Отделе проблем Полесья АН РБ, а также используются в учебном курсе «Методы и системы принятия решений».

Основные положения диссертации, выносимые на защиту. На защиту выносятся следующие научные положения:

- алгоритмы обучения нейронных сетей с одним обрабатывающем слоем, которые в отличии от известных позволяют снизить вычислительную сложность и обеспечить устойчивость обучения;
- алгоритмы обучения многослойных нейронных сетей, которые базируются на вычислении адаптивного шага обучения и различных подходах к обучению. Предложенные алгоритмы позволяют повысить быстродействие и точность обучения;
- матричная алгоритмизация процесса обучения на основе градиентных методов для многослойных нейронных гетерогенных сетей прямого распространения без обратных связей;
- методика построения базовой архитектуры прогнозирующей нейронной сети, которая обеспечивает сохранение динамических свойств временной последовательности.

Личный вклад соискателя. Основные положения диссертации получены автором лично и отражены в 36 публикациях. Соавтором в некоторых из них является научный руководитель Головко В.А. [1-3, 7, 9, 15, 16, 24, 26, 27].

Апробация результатов диссертации. Основные результаты работы докладывались автором на IV республиканской научной конференции студентов и аспирантов «Новые математические методы и компьютерные технологии в проектировании, производстве и научных исследованиях» (Гомель, 19-22 марта 2001г.); международной математической конференции «Еругинские чтения – VII» (Гродно, 28-30 мая 2001г.); Six International

Библиотека БГУИР

Conferences on Pattern Recognition and Information Processing (PRIP'2001) (Minsk, May 20-22 2001); Workshop of Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2001) (Foros, Ukraine, July 1-4, 2001); 2nd International Conference on Neural Networks and Artificial Intelligence (ICNNAI'2001) (Minsk, October 2-5, 2001г.); II региональной конференции молодых ученых "Современные проблемы математики и вычислительной техники"(Брест, 28-30 ноября 2001г.); V республиканской научной конференции студентов и аспирантов «Новые математические методы и компьютерные технологии в проектировании, производстве и научных исследованиях» (Гомель, 18-20 марта 2002г.); международной математической конференции «Еругинские чтения – VIII», (Брест, 20-23 мая 2002г.); VII международном семинаре «Устойчивость и колебания нелинейных систем управления» (Москва, 22-24 мая 2002г.); втором международном Конгрессе «Нелинейный динамический анализ (NDA'2)» (Москва, 3-8 июня 2002г.); VI республиканской научной конференции студентов и аспирантов «Новые математические методы и компьютерные технологии в проектировании, производстве и научных исследованиях» (Гомель, 19-21 марта 2003г.); первом научном семинаре «Танаевские чтения» (Минск, 28 марта 2003г.); VI международной научно-методической конференции «Наука и образование в условиях социально-экономической трансформации общества» (Минск, 15-16 мая 2003 г.); международной математической конференции «Еругинские чтения – IX» (Витебск, 20-23 мая 2003г.); International Conferences on Pattern Recognition and Information Processing (PRIP'2003) (Minsk, May 21-23 2003); международной конференции «Аналитические методы анализа и дифференциальных уравнений» AMADE'2003 (Минск, 4-9 сентября 2003); Workshop of Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2003) (Lviv, Ukraine, September 8-10, 2003); 1-ой Международной научной конференции «Математические методы экономических процессов переходного периода» (Минск, 29-31 октября 2003); 3nd International Conference on Neural Networks and Artificial Intelligence (ICNNAI'2003) (Minsk, November 12-14, 2003); VII международной научно-методической конференции «Наука и образование в условиях социально-экономической трансформации общества» (Брест, 13-14 мая 2004 г.); VI International conference on Artificial Intelligence (AI-19'2004) (September 21-23, 2004, Siedlce, Poland); III Международная научно-практическая конференция «Динаміка наукових досліджень '2004» (Днепропетровск, Украина, 21-30 июня 2004); Международная конференция «IX Белорусская математическая конференция» (Гродно, 3-6 ноября 2004 г.); 4-ая Международная конференция «Обработка информации и управление в чрезвычайных и экстремальных ситуациях» (ОИУЧЭС'2004) (ОИПИ, Минск, 29 ноября – 1 декабря 2004 г.).

Опубликованность результатов. По теме диссертации опубликовано 36 работ, в том числе: глава в книге – одна, статей – двадцать одна, тезисов

Библиотека БГУИР

докладов – четырнадцать. Общее количество страниц опубликованных материалов составляет 174 страницы.

Структура и объем диссертации. Диссертация состоит из введения, общей характеристики, четырех глав, заключения, списка используемых источников и приложений. Работа изложена на 150 страницах машинописного текста и содержит: 62 рисунка; 6 таблиц; список литературы, включающий 137 наименований, размещенный на 12 страницах; 3 приложения, объемом 17 листов.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении дана оценка состояния современного развития теории нейронных сетей и ее места в развитии проблематики искусственного интеллекта; указаны этапы развития данного направления и его связь с открытиями в области знаний о нервной деятельности человеческого мозга; приведены предпосылки для разработки темы диссертации.

В первой главе рассмотрена задача обучения многослойной нейронной сети, состоящей из N нейронных блоков (рис.1), каждый из которых имеет структуру, представленную на рис. 2.

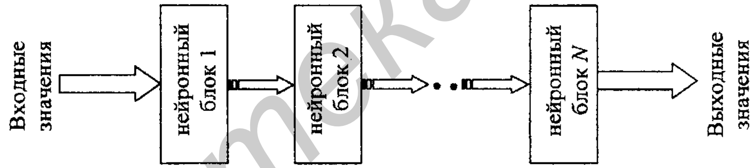

Рис.1. Блочное представление многослойной нейронной сети

Входными значениями для каждого нейронного блока являются выходы предыдущего; для первого – последовательность входных образов $\overline{x^k} = \left(x_1^k, ..., x_{m_0}^k\right), \left(k = \overline{1, L}\right)$. Выходное значение i_n-ого нейрона n-ого блока сети для k-ого образа определяется рекуррентным соотношением:

$$y_{i_n}^{(n),k} = F_n\left(S_{i_n}^{(n),k}\right),$$

где $S_{i_n}^{(n),k} = \sum_{i_{n-1}=1}^{m_{n-1}} w_{i_{n-1}i_n}^{(n)} y_{i_{n-1}}^{(n-1),k} - T_{i_n}^{(n)}, \; i_n = \overline{1, m_n}, \; k = \overline{1, L}$.

При этом формируется вектор $Y^{(n),k} = \begin{pmatrix} y_1^{(n),k} & y_2^{(n),k} & \dots & y_{m_n}^{(n),k} & -1 \end{pmatrix}^T$.

Задача обучения данной многослойной гетерогенной нейронной сети состоит в нахождении матриц весовых коэффициентов

Библиотека БГУИР

$$W^{(n)} = \begin{pmatrix} w_{11}^{(n)} & w_{21}^{(n)} & \dots & w_{m_{n-1}1}^{(n)} \\ w_{12}^{(n)} & w_{22}^{(n)} & \dots & w_{m_{n-1}2}^{(n)} \\ \dots & \dots & \dots & \dots \\ w_{1m_n}^{(n)} & w_{2m_n}^{(n)} & \dots & w_{m_{n-1}m_n}^{(n)} \end{pmatrix}_{m_n \times m_{n-1}}$$

и столбцов порогов $\overline{T^{(n)}} = \left(T_1^{(n)}, T_2^{(n)}, \dots, T_{m_n}^{(n)}\right)^T$, $n = \overline{1, N}$, которые минимизируют некоторую ошибку сети E_s, как отклонение выходных значений сети $y_{i_N}^{(N),k}$ от эталонных $t_{i_N}^{k}$ – i_N-ого нейрона сети для k-ого образа. В качестве ошибки рассматривается усредненное по количеству образов «квадратичное отклонение» $E_S = \frac{1}{2L}\sum\limits_{k=1}^{L}\sum\limits_{i_N=1}^{m_N}\left(y_{i_N}^{(N),k} - t_{i_N}^{k}\right)^2$.

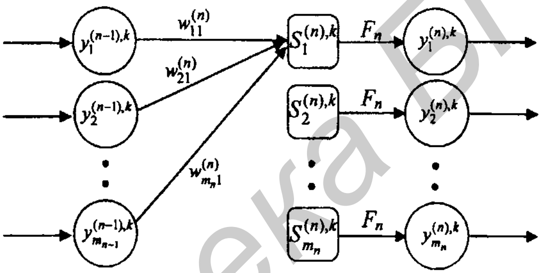

Рис.2. Архитектура n-ого блока многослойной нейронной сети

Для данного типа гетерогенных многослойных сетей прямого распространения рассмотрены различные известные методы обучения. Отмечено, что большинство из них применимы в основном к однослойным нейронным сетям из-за сложности вычислений. К таким относятся метод градиентного спуска, метод Ньютона, псевдоньютоновский метод, метод Левенберга-Марквардта, алгоритм экспоненциального затухания. Для этих методов приведены соотношения для изменения весовых коэффициентов и порогов сети.

Отмечено, что для обучения многослойных сетей используется в основном метод обратного распространения ошибки, основанный на методе градиентного спуска. Главным его преимуществом является рекуррентное вычисление изменений синаптических связей. Приведен вывод основных соотношений для данного метода. Однако этот метод не всегда приводит к приемлемому результату за достаточно небольшой промежуток времени, что влияет на его устойчивость. На это в той или иной степени влияют параметры обучения. Показаны различные подходы к их определению, но отмечена их невысокая эффективность.

Приведен ряд теорем, доказывающих способность двухслойных нелинейных нейронных сетей быть универсальными аппроксиматорами. Даны соотношения, определяющие количество нейронных элементов в слоях

Библиотека БГУИР

для достижения произвольной, напередзаданной точности ошибки аппроксимации.

Во второй главе рассматриваются однослойные нелинейные нейронные сети, состоящие из n нейронных элементов распределительного слоя и m – выходного слоя.

Для данной сети каждый нейрон распределительного слоя имеет синаптические связи w_{ij} $\left(i=\overline{1,n}, j=\overline{1,m}\right)$ со всеми нейронами обрабатывающего слоя. В качестве нейронов выходного слоя используются элементы с некоторой строго монотонной функцией активации F (строгая монотонность необходима для существования обратной функции). На вход сети подаются входные образы – векторы $\overline{x^k}=\left(x_1^k,...,x_n^k,-1\right)^T$, $\left(k=\overline{1,L}\right)$ или, что то же самое, на вход сети подается вектор $\overline{\xi}=\left(\xi_1^T,...,\xi_n^T,\xi_{n+1}^T\right)$, где вектора $\xi_q=\left(x_q^1,...,x_q^L\right)$, $\left(q=\overline{1,n}\right)$ и $\xi_{n+1}=(-1,...,-1)$.

Выходное значение j-ого нейрона сети для k-ого образа определяется соотношением:

$$y_j^k=F\left(S_j^k\right),$$

где $S_j^k=\sum\limits_{i=1}^{n}w_{ij}x_i^k-T_j$, $j=\overline{1,m}$, $k=\overline{1,L}$.

Задача обучения нейронной сети с фиксированной функцией активации F состоит в нахождении весовых коэффициентов w_{ij} $\left(i=\overline{1,n}, j=\overline{1,m}\right)$ и порогов нейронных элементов T_j $\left(j=\overline{1,m}\right)$, которые минимизируют некоторую ошибку сети E_S как отклонение выходных значений y_j^k от эталонных значений t_j^k – j-ого нейрона сети для k-ого образа. В качестве ошибки рассматривается усредненное по количеству образов «квадратичное отклонение» $E_S=\frac{1}{2L}\sum\limits_{k=1}^{L}\sum\limits_{j=1}^{m}\left(y_j^k-t_j^k\right)^2$.

Для нахождения решения исходной задачи, рассмотрим вспомогательную, решение которой получается из системы:

$$\sum_{i=1}^{n}w_{ij}x_i^k-T_j=F^{-1}\left(t_j^k\right),\ j=\overline{1,m},\ k=\overline{1,L},$$

для чего необходимо минимизировать усредненное по количеству образов «квадратичное отклонение», задаваемое выражением:

$$E_S^*=\frac{1}{2L}\sum_{k=1}^{L}\sum_{j=1}^{m}\left(S_j^k-F^{-1}\left(t_j^k\right)\right)^2=\frac{1}{2L}\sum_{k=1}^{L}\sum_{j=1}^{m}\left(\sum_{i=1}^{n}w_{ij}x_i^k-T_j-F^{-1}\left(t_j^k\right)\right)^2 \quad (1)$$

Данная взаимосвязанность справедлива в силу однозначности и монотонности функции активации F.

Теорема 2.1. Если входные образы $\xi_1,\xi_2,...,\xi_n,\xi_{n+1}$ линейно независимы, то весовые коэффициенты однослойной нейронной сети с монотонной

Библиотека БГУИР

функцией активации определяются однозначно из соотношения $W_j = K^{-1}\left[\overline{\xi},\overline{\xi}\right]\cdot K\left[\overline{\xi},\eta_j\right]$, где $K\left[\overline{\xi},\overline{\xi}\right]$ – ковариационная матрица случайного вектора $\overline{\xi}$, а $K\left[\overline{\xi},\eta_j\right]$ – ковариация векторов $\overline{\xi}$ и $\eta_j = \left(F^{-1}\left(t_j^1\right),...,F^{-1}\left(t_j^L\right)\right)$, $j=\overline{1,m}$.

Следствие 1. Ковариационная матрица $K\left[\overline{\vartheta_k},\overline{\vartheta_k}\right]$, где $\overline{\vartheta_k} = (\xi_1 \quad \xi_2 \cdot \quad ... \quad \xi_k)$ – случайный вектор с элементами, являющимися последовательностями входных образов k нейронов распределительного слоя, невырождена.

Следствие 2. Для оптимального прогнозирования функции методом скользящего окна необходимо выбирать количество входных нейронов таким образом, чтобы ковариационная матрица входных образов была невырождена.

Замечание 1. Решение исходной задачи обучения нейронной сети существует, и при её обучении можно достичь ошибку, не худшую, чем $\varepsilon_m = \left(F'(0)\right)^2 \cdot \varepsilon^*$, где ε^* – решение в соответствии с теоремой 1.1.

Замечание 2. Количество входных образов должно быть большим количества нейронов в распределительном слое для однозначного обучения вспомогательной сети.

Теорема 2.2. Функция E_S эквивалентной задачи, определяемая соотношением (1), является сильно выпуклой функцией.

Это подтверждает факт, что функция E_S имеет единственную точку минимума, т.к. она сильно выпукла.

Исходя из вышеизложенного приведен алгоритм решения вспомогательной задачи на основании метода с использованием ковариационной матрицы.

При обучении однослойной нейронной сети на основе градиентных методов модификация весовых коэффициентов и порогов сети производится в соответствии с выражениями:

$$w_{ij}(t+1) = w_{ij}(t) - \alpha \frac{\partial E_s(t)}{\partial w_{ij}}, \quad i=\overline{1,n}, \;\; j=\overline{1,m}$$

$$T_j(t+1) = T_j(t) - \alpha \frac{\partial E_s(t)}{\partial T_j}, \quad j=\overline{1,m}.$$

При этом, если обучать однослойную нейронную сеть в соответствии с методом наискорейшего спуска, выбор шага обучения описывается следующей теоремой:

Теорема 2.3. Для нейронной сети величина адаптивного шага обучения с использованием метода наискорейшего спуска в момент времени t определяется соотношением:

Библиотека БГУИР

$$\alpha^{(1)} = \frac{\sum_{i=1}^{n+1}\sum_{j=1}^{m}\left(G_{ij}\right)^2}{\sum_{i_1,i_2=1}^{n+1}\sum_{j_1,j_2=1}^{m} G_{i_2 j_2}\cdot S_{i_1 j_1}^{i_2 j_2}\cdot G_{i_1 j_1}} \qquad (2)$$

при этом модификации синаптических связей определяются соотношениями:

$$w_{i_1 j_1}(t+1) = w_{i_1 j_1}(t) - \alpha^{(1)}\cdot G_{i_1 j_1},\ j_1 = \overline{1,m}, i_1 = \overline{1,n} \qquad (3)$$

$$T_{j_1}(t+1) = T_{j_1}(t) - \alpha^{(1)}\cdot G_{(n+1) j_1},\quad j_1 = \overline{1,m}, \qquad (4)$$

где $G_{ij} = \sum_{k=1}^{L}\varepsilon^k \cdot MF' \cdot M_{ji}\cdot \overline{x^k}$, $\varepsilon^k = \left(\left(y_1^k - t_1^k\right)\ \left(y_2^k - t_2^k\right)\ \ldots\ \left(y_m^k - t_m^k\right)\right)$,

$MF' = diag\left(F'\left(S_1^k\right)\ \ F'\left(S_2^k\right)\ \ \ldots\ \ F'\left(S_m^k\right)\right)$ – матрица размерности $m\times m$, а матрица $M_{j_1 i_1}$ размерности $m\times(n+1)$ состоит из числа 1 на позиции $j_1 i_1$ и нулей в качестве остальных элементов матрицы; а $S_{i_1 j_1}^{i_2 j_2} = \sum_{k=1}^{L}\left(\left(M_{j_2 i_2}\cdot\overline{x^k}\right)^T\cdot\left(\left(MF'\right)^2 + DE^k\cdot MF''\right)\cdot\left(M_{j_1 i_1}\cdot\overline{x^k}\right)\right)$, при

$$DE^k = diag\left(y_1^k - t_1^k\quad y_2^k - t_2^k\quad \ldots\quad y_m^k - t_m^k\right),$$

$$MF'' = diag\left(F''\left(S_1^k\right)\ \ F''\left(S_2^k\right)\ \ \ldots\ \ F''\left(S_m^k\right)\right).$$

В третьей главе рассматривается двухслойная гетерогенная нейронная сеть, состоящая из m_0 нейронных элементов распределительного слоя, m_1 нейронов скрытого слоя и m_2 – выходного слоя. Для данной сети каждый нейрон распределительного слоя имеет синаптические связи $w_{i_0 i_1}^{(1)}\left(i_0 = \overline{1,m_0}\right)$ со всеми нейронами скрытого слоя, а каждый нейрон скрытого слоя синаптические связи $w_{i_1 i_2}^{(2)}\left(i_1 = \overline{1,m_1}, i_2 = \overline{1,m_2}\right)$ – со всеми нейронами выходного слоя. В качестве нейронов скрытого слоя используются элементы с функцией активации F_1, в качестве нейронов выходного слоя – с функцией активации F_2. На вход сети подаются входные образы – векторы $\overline{x^k} = \left(x_1^k,\ldots,x_{m_0}^k\right)$, $\left(k = \overline{1,L}\right)$. Входами распределительного слоя являются значения $y_{i_0}^{(0),k} = x_{i_0}^k$. При этом формируется вектор

$$Y^{(0),k} = \left(y_1^{(0),k}\quad y_2^{(0),k}\quad \ldots\quad y_{m_0}^{(0),k}\quad -1\right)^T.$$

Выходное значение i_l-ого нейрона l-ого слоя $(l = 1,2)$ сети для k-ого образа определяется соотношением $y_{i_l}^{(l),k} = F_l\left(S_{i_l}^{(l),k}\right)$, где

$$S_{i_l}^{(l),k} = \sum_{i_{l-1}=1}^{m_l} w_{i_{l-1} i_l}^{(l)} y_{i_{l-1}}^{(l),k} - T_{i_l}^{(l)},\ i_l = \overline{1,m_l},\ k = \overline{1,L}.$$

При этом формируется вектор $Y^{(l),k} = \left(y_1^{(l),k}\quad y_2^{(l),k}\quad \ldots\quad y_{m_1}^{(l),k}\quad -1\right)^T$

Библиотека БГУИР

Задача обучения нейронной сети с фиксированными функциями активации состоит в нахождении весовых коэффициентов $w_{i_0 i_1}^{(1)}$, $w_{i_1 i_2}^{(2)}$ и порогов нейронных элементов $T_{i_1}^{(1)}$, $T_{i_2}^{(2)}$, которые минимизируют некоторую ошибку сети E_S, как отклонение выходных значений $y_{i_2}^{(2),k}$ от эталонных значений $t_{i_2}^{k}$ – i_2-ого нейрона сети для k-ого образа. В качестве ошибки будем рассматривать усредненное по количеству образов «квадратичное отклонение» $E_S = \frac{1}{2L}\sum_{k=1}^{L}\sum_{i_2=1}^{m_2}\left(y_{i_2}^{(2),k} - t_{i_2}^{k}\right)^2$.

Матрицы $W^{(l)} = \left\{w_{i_{l-1} i_l}^{(l)}\right\}_{m_{l-1}\times m_l}$ и $\overline{T^{(l)}} = \left(T_1^{(l)}, T_2^{(l)}, \dots, T_{m_l}^{(l)}\right)^T$ будем называть приближенным решением или просто решением (по методу наименьших квадратов) системы:

$$F_2\left(\sum_{i_1=1}^{m_1} w_{i_1 i_2}^{(2)}\cdot F_1\left(\sum_{i_0=1}^{m_0} w_{i_0 i_1}^{(1)} y_{i_0}^{(0),k} - T_{i_1}^{(1)}\right) - T_{i_2}^{(2)}\right) = t_{i_2}^{k},\ i_2 = \overline{1, m_2},\ k = \overline{1, L},$$

если «квадратичное отклонение»

$$E_S = \frac{1}{2L}\sum_{k=1}^{L}\sum_{i_2=1}^{m_2}\left(F_2\left(\sum_{i_1=1}^{m_1} w_{i_1 i_2}^{(2)}\cdot F_1\left(\sum_{i_0=1}^{m_0} w_{i_0 i_1}^{(1)} y_{i_0}^{(1),k} - T_{i_1}^{(1)}\right) - T_{i_2}^{(2)}\right) - t_{i_2}^{k}\right)^2 = \frac{1}{L}\sum_{k=1}^{L} E_s^{(k)}$$

достигает своего наименьшего значения.

В качестве метода обучения приведенной сети предлагается использовать следующую матричную алгоритмизацию, значительно упрощающую его программную реализацию.

Используя обозначения $\alpha^{(l)}$- шаг смещения на вектор антиградиента, $MF_i' = diag\left(F_l'\left(S_1^{(l),k}\right), F_l'\left(S_2^{(l),k}\right), \dots, F_l'\left(S_{m_l}^{(l),k}\right)\right)$ – диагональная матрица и $M_{j_l j_{l-1}}^{(l)}$ – матрица размерности $m_l \times (m_{l-1}+1)$ с единицей на позиции $j_l j_{l-1}$ и остальными нулевыми элементами имеем:

$$w_{j_{l-1} j_l}^{(l)}(t+1) = w_{j_{l-1} j_l}^{(l)}(t) - \alpha^{(l)}\cdot\frac{1}{L}\cdot G_{j_{l-1} j_l}^{(l)},\ T_{j_l}^{(l)}(t+1) = T_{j_l}^{(l)}(t) - \alpha^{(l)}\cdot\frac{1}{L}\cdot G_{(m_l+1) j_l}^{(l)},$$

где $G_{j_1 j_2}^{(2)} = \sum_{k=1}^{L} C^{(2),k}\cdot K_{j_1 j_2}^{(2),k}$, $G_{j_0 j_1}^{(1)} = \sum_{k=1}^{L} C^{(1),k}\cdot K_{j_0 j_1}^{(1),k}$

и $K_{j_0 j_1}^{(1),k} = M_{j_1 j_0}^{(1)}\cdot Y^{(0),k}$, $K_{j_1 j_2}^{(2),k} = M_{j_2 j_1}^{(2)}\cdot Y^{(1),k}$, $C^{(2),k} = \varepsilon_2^{k}\cdot MF_2'$,

$$C^{(1),k} = C^{(2),k}\cdot W^{(2)}\cdot MF_1' = C^{(2),k}\cdot MW^{(1)}.$$

При обучении сети с использованием метода наискорейшего спуска, можно применить один из следующих методов:

1. <u>Послойное обучение.</u> Данный метод заключается в том, что сначала, после подачи всех элементов обучающего множества, изменяем синаптические связи последнего (или первого) слоя, а затем, после того как на измененную сеть были опять поданы все элементы обучающего

множества, изменяем синаптические связи оставшегося слоя. После переходим к началу алгоритма. Процедура обучения останавливается после того, как ошибка не превышает заданную после двух последовательных итераций.

2. Двухпараметрическое обучение. Данный метод заключается в том, что после подачи всех элементов обучающего подмножества ищутся такие адаптивные шаги обучения $\alpha^{(1)}$ и $\alpha^{(2)}$, чья комбинация минимизирует функцию ошибки обучения сети.

3. Обобщенный метод наискорейшего спуска. При данном методе после подачи всех элементов обучающего подмножества ищется такой адаптивный шаг обучения $\alpha = \alpha^{(1)} = \alpha^{(2)}$, который минимизирует ошибку сети в направлении антиградиента.

Приведем теоремы для выбора адаптивных шагов в каждом из предложенных методов. Будем использовать следующие обозначения

$$\left(S^{(2)}\right)_{l_1 l_2}^{j_1 j_2} = \sum_{k=1}^{L}\left(\left(K_{l_1 l_2}^{(2),k}\right)^T \cdot \left(\left(MF_2'\right)^2 + DE^k \cdot MF_2''\right) \cdot K_{j_1 j_2}^{(2),k}\right),$$

$$\left(S^{(1)}\right)_{l_0 l_1}^{j_0 j_1} = \sum_{k=1}^{L}\left(\left(K_{l_0 l_1}^{(1),k}\right)^T \cdot \left(\left(MW^{(1)}\right)^T \cdot \left[\left(MF_2'\right)^2 + DE^k \cdot MF_2''\right] \cdot MW^{(1)} + \right.\right.$$

$$\left.\left. DE^k \cdot \left(MF_2' \cdot W^{(2)}\right) \cdot MF_1''\right) \cdot K_{j_0 j_1}^{(1),k}\right),$$

$$\left(S^{(1,2)}\right)_{l_1 l_2}^{j_0 j_1} = \sum_{k=1}^{L}\left(\left(K_{l_1 l_2}^{(2),k}\right)^T \cdot \left(\left(MF_2'\right)^2 + DE^k \cdot MF_2''\right) \cdot MW^{(1)} \cdot K_{j_0 j_1}^{(1),k}\right),$$

$$A = \sum_{j_0, l_0 = 1}^{m_0+1} \sum_{j_1, l_1 = 1}^{m_1} G_{l_0 l_1}^{(1)} \cdot \left(S^{(1)}\right)_{l_0 l_1}^{j_0 j_1} G_{j_0 j_1}^{(1)}, B = \sum_{j_0=1}^{m_0+1} \sum_{j_1=1}^{m_1} \sum_{l_1=1}^{m_1+1} \sum_{l_2=1}^{m_2} G_{l_1 l_2}^{(2)} \cdot \left(S^{(1,2)}\right)_{l_1 l_2}^{j_0 j_1} \cdot G_{j_0 j_1}^{(1)}$$

$$C = \sum_{j_1, l_1 = 1}^{m_1+1} \sum_{j_2, l_2 = 1}^{m_2} G_{l_1 l_2}^{(2)} \cdot \left(S^{(2)}\right)_{l_1 l_2}^{j_1 j_2} G_{j_1 j_2}^{(2)}, \ D_1 = \sum_{j_0=1}^{m_0+1} \sum_{j_1=1}^{m_1} \left(G_{j_0 j_1}^{(1)}\right)^2, D_2 = \sum_{j_1=1}^{m_1+1} \sum_{j_2=1}^{m_2} \left(G_{j_1 j_2}^{(2)}\right)^2.$$

Теорема 3.1. При послойном обучении двухслойной гетерогенной нейронной сети адаптивный шаг для каждого слоя определяется в соответствии с формулами:

$$\alpha_m^{(2)} = \frac{D_2}{C}, \quad \alpha_m^{(1)} = \frac{D_1}{A}$$

Теорема 3.2. При двухпараметрическом обучении двухслойной гетерогенной нейронной сети адаптивный шаг для каждого слоя определяется в соответствии с формулами:

$$\alpha_m^{(1)} = \frac{D_1 C - D_2 B}{AC - B^2}, \ \alpha_m^{(2)} = \frac{D_2 A - D_1 B}{AC - B^2}.$$

Теорема 3.3. При обучении обобщенным методом наискорейшего спуска двухслойной гетерогенной нейронной сети адаптивный шаг для каждого слоя определяется в соответствии с формулами:

Библиотека БГУИР

$$\alpha_m = \frac{D_1 + D_2}{A + 2B + C}.$$

Для многослойной нейронной сети приведена теорема, описывающая матричную алгоритмизацию процесса обучения, значительно упрощающая программную реализацию процедуры обучения.

Теорема 3.4. Модификация синаптических связей и порогов многослойной гетерогенной нейронной сети производится в соответствии с формулами:

$$w_{j_{n-1}j_n}^{(n)}(t+1) = w_{j_{n-1}j_n}^{(n)}(t) - \alpha^{(n)} \cdot \frac{1}{L} \cdot \sum_{k=1}^{L} C^{(n)} \cdot M_{j_n j_{n-1}}^{(n)} \cdot Y^{(n-1),k}$$

$$T_{j_n}^{(n)}(t+1) = T_{j_n}^{(n)}(t) - \alpha^{(n)} \cdot \frac{1}{L} \cdot \sum_{k=1}^{L} C^{(n)} \cdot M_{j_n(m_{n-1}+1)}^{(n)} \cdot Y^{(n-1),k}$$

где $C^{(n)}$ вычисляется рекуррентно:

$$C^{(n)} = C^{(n+1)} \cdot W^{(n+1)} \cdot MF_n', \; C^{(N)} = \varepsilon_N^k \cdot MF_N',$$

$$\varepsilon_N^k = \left(\left(y_1^{(N),k} - t_1^k\right) \; \left(y_2^{(N),k} - t_2^k\right) \; \ldots \; \left(y_{m_2}^{(N),k} - t_{m_2}^k\right) \right),$$

а $MF_n' = diag\left(F_n'\left(S_1^{(n),k}\right) \; F_n'\left(S_2^{(n),k}\right) \; \ldots \; F_n'\left(S_{m_n}^{(n),k}\right) \right)$ – матрица размерности $m_n \times m_n$, а матрица $M_{j_n j_{n-1}}^{(n)}$ размерности $m_n \times (m_{n-1}+1)$ состоит из числа 1 на позиции $j_n j_{n-1}$ и нулей в качестве остальных элементов матрицы.

Изменение синаптических связей и порогов сети производится начиная с последнего N-ого до первого блока сети.

Данная теорема позволяет привести к алгоритму изменение весов и порогов сети в процессе обучения градиентным методом, сведением их модификации к матричным операциям. При этом шаг обучения $\alpha^{(n)}$ может быть как постоянным, так и адаптивным. В случае адаптивного шага можно привести алгоритм обучения на основе послойного обучения, как расширение данной методики с двухслойной сети. Алгоритмы двухпараметрического метода и метода на основе обобщенного наискорейшего спуска дают довольно сложные вычислительные формулы.

При послойном обучении сначала, после подачи всех элементов обучающего множества, изменяются синаптические связи последнего слоя. Затем, после того как на модифицированную сеть были поданы все элементы обучающего множества, изменяются синаптические связи предшествующего слоя. И так далее, до достижения первого слоя сети. Затем осуществляется переход к началу алгоритма. Процедура обучения останавливается после того, как ошибка сети не превышает заданную после двух последовательных итераций. Шаги обучения $\alpha^{(n)}$ выбираются для наилучшей минимизации ошибки сети при изменении синаптических связей соответствующих слоев.

Теорема 3.5. При послойном обучении многослойной гетерогенной нейронной сети прямого распространения без обратных связей адаптивный шаг для каждого слоя определяется в соответствии с формулами:

Библиотека БГУИР

$$\alpha^{(n)} = \frac{L \cdot \sum\limits_{j_{n-1}=1}^{m_{n-1}} \sum\limits_{j_n=1}^{m_n} \left(\sum\limits_{k=1}^{L} C^{(n)} \cdot M^{(n)}_{j_n j_{n-1}} \cdot Y^{(n-1),k} \right)}{\sum\limits_{j_{n-1}=1}^{m_{n-1}} \sum\limits_{j_n=1}^{m_n} \sum\limits_{l_{n-1}=1}^{m_{n-1}} \sum\limits_{l_n=1}^{m_n} \left(\sum\limits_{k=1}^{L} \left(\left(K^{(n),k}_{l_{n-1} l_n} \right)^T \cdot U^{(n),k} \cdot \left(K^{(n),k}_{j_{n-1} j_n} \right) \right) \right)},$$

где $K^{(n),k}_{j_{n-1} j_n} = M^{(n)}_{j_n j_{n-1}} \cdot Y^{(n-1),k}$,

а $U^{(n),k} = \left(W^{(n+1)} \cdot MF_n' \right)^T \cdot U^{(n+1),k} \cdot \left(W^{(n+1)} \cdot MF_n' \right) + CD^{j_n}_{m_n m_{n+1}} \cdot W^{(n+1)} \cdot MF_n''$

вычисляется рекуррентно с начальным условием

$$U^{(N),k} = \left(MF_N' \right)^2 + DE^{(N),k} \cdot MF_N'',$$

при этом $CD^{j_n}_{m_n m_{n+1}}$ – матрица размерности $m_n \times m_{n+1}$, все элементы которой равны нулю, кроме стоящих в j_n строке, равных единице.

Разработан также эвристический метод обучения сети с вычислением адаптивного шага на основе условной минимизации ошибок каждого слоя. При этом каждый слой нейронной сети рассматривается как однослойная нейронная сеть, обучение которой производится путем настройки градиентным методом выходных значений к полученным «эталонным». Таким образом, в ходе обучения пересчитываются эталонные значения для каждого слоя.

Теорема 3.6. При использовании вышепредложенного алгоритма вычисление «эталонных» значений выходов, при некотором заданном параметре β, производится в соответствии с формулой:

$$t^{(n),k}_{j_n} = y^{(n),k}_{j_n} - \alpha \cdot C^{(n+1)} \cdot W^{(n+1)} \cdot \Delta^n_{j_n}, \; j_n = \overline{1, m_n},$$

с последующей коррекцией:

$$t^{(n),k}_{i_n} := \begin{cases} a + \beta, & if \;\; t^{(n),k}_{i_n} < a + \beta \\ t^{(n),k}_{i_n}, & if \;\; t^{(n),k}_{i_n} \in [a + \beta, b - \beta]; \\ b - \beta, & if \;\; t^{(n),k}_{i_n} > b - \beta \end{cases}$$

адаптивный шаг при этом вычисляется в соответствии с равенством:

$$\alpha = \frac{\sum\limits_{j_n=1}^{m_n} \left(C^{(n+1)} \cdot P^n_{j_n} \right)^2}{\sum\limits_{j_n=1}^{m_n} \sum\limits_{l_n=1}^{m_n} \left(C^{(n+1)} \cdot P^n_{j_n} \right) \cdot \left(\left(P^n_{j_n} \right)^T \cdot U^{(n+1),k} \cdot \left(P^n_{l_n} \right) \right) \cdot \left(C^{(n+1)} \cdot P^n_{l_n} \right)},$$

где

$$U^{(n),k} = \left(W^{(n+1)} \cdot MF_n' \right)^T \cdot U^{(n+1),k} \cdot \left(W^{(n+1)} \cdot MF_n' \right) + CD^{l_n}_{m_n m_{n+1}} \cdot W^{(n+1)} \cdot MF_n'',$$

$$U^{(N),k} = \left(MF_N' \right)^2 + DE^{(N),k} \cdot MF_N'', \; P^n_{j_n} = W^{(n+1)} \cdot \Delta^n_{j_n},$$

$$DE^{(N),k} = diag\left(\left(y^{(N),k}_1 - t^k_1 \right) \;\; \left(y^{(N),k}_2 - t^k_2 \right) \;\; \dots \;\; \left(y^{(N),k}_{m_2} - t^k_{m_2} \right) \right),$$

Библиотека БГУИР

а $\Delta^n_{j_n}$ - вектор столбец длины m_n с нулями в качестве всех элементов, кроме стоящего на позиции j_n, равного 1, где $CD^{l_n}_{m_n m_{n+1}}$ – матрица размерности $m_n \times m_{n+1}$, все элементы которой равны нулю, кроме стоящих в l_n строке, равных единице.

При этом настройка весов и порогов сети производится по формулам:

$$w^{(n)}_{j_{n-1}j_n}(t+1) = w^{(n)}_{j_{n-1}j_n}(t) - \alpha^{(n)} \cdot G^{(n),layer}_{j_{n-1}j_n},\ \ j_{n-1} = \overline{1, m_{n-1}},\ \ j_n = \overline{1, m_n}$$

$$T^{(n)}_{j_n}(t+1) = T^{(n)}_{j_n}(t) - \alpha^{(n)} \cdot G^{(n),layer}_{j_{n-1}j_n},\quad j_n = \overline{1, m_n},$$

с адаптивным шагом, вычисляемым в соответствии с

$$\alpha^{(n)} = \frac{\sum\limits_{j_{n-1}=1}^{m_{n-1}+1} \sum\limits_{j_n=1}^{m_n} \left(G^{(n),layer}_{j_{n-1}j_n}\right)^2}{\sum\limits_{j_{n-1}, l_{n-1}=1}^{m_{n-1}+1} \sum\limits_{j_n, l_n=1}^{m_n} G^{(n),layer}_{l_{n-1}l_n} \cdot \left(S_{(n)}\right)^{j_{n-1}j_n}_{l_{n-1}l_l} \cdot G^{(n),layer}_{l_{n-1}l_n}},$$

где $G^{(n),layer}_{j_{n-1}j_n} = \sum\limits_{k=1}^{L} C^{(n),k}_{layer} \cdot K^{(n),k}_{j_{n-1}j_n}$, $C^{(n),k}_{layer} = \varepsilon^k_n \cdot MF'_n$, $K^{(n),k}_{ij} = M^{(n)}_{ji} \cdot Y^{(n-1),k}$,

и

$$\left(S_{(n)}\right)^{j_{n-1}j_n}_{l_{n-1}l_l} = \sum_{k=1}^{L} \left(\left(K^{(n),k}_{l_{n-1}l_n}\right)^T \cdot \left(\left(MF'_n\right)^2 + DE^{(n),k} \cdot MF''_n \right) \cdot K^{(n),k}_{j_{n-1}j_n} \right),$$

$$K^{(n),k}_{j_{n-1}j_n} = M^{(n)}_{j_n j_{n-1}} \cdot Y^{(n-1),k}.$$

В четвертой главе описано построение прогнозирующей нейронной сети. Для аргументации ее построения временной ряд рассматривается как изменение некой фазовой переменной системы. В соответствии с теоремой Такенса такой временной ряд $x(t)$ можно вложить в пространство задержек $\left(x(t), x(t+\tau), ..., x(t+(N-1)\tau)\right)$ размерности $N = 2[m]+1$, где m – фрактальная размерность временного ряда, а $[\,.\,]$ – целая часть числа. Для определения размерности пространства вложения, кроме предварительного вычисления корреляционной размерности D_2 методом, разработанным Грассбергером и Прокаччиа, можно использовать методы на основе анализа главных компонент и метод ложных соседей. Для определения временной задержки τ используются метод автокорреляционной функции и метод, использующий меру взаимной информации. После вложения, временной ряд представляет собой многообразие в N-мерном пространстве, где по $(N-1)$ координате, можно однозначно определить оставшуюся. Таким образом, задача прогнозирования сводится к задаче аппроксимации поверхности в N-мерном пространстве. Для ее решения используем двухслойную нелинейную нейронную сеть с как минимум $(N-1)$ нейроном во входном слое. Процесс обучения и прогнозирования осуществляется методом скользящего окна, где входные образы подаются в соответствии с задержкой, полученной при построении. При таком прогнозе сохраняется динамика временного ряда, выраженная в сходимости к подобному аттрактору. Приведено

Библиотека БГУИР

использование предлагаемой прогнозирующей сети для предсказания годовых расходов воды рек Беларуси. Полученные результаты и их сравнение с методами прогнозирования, используемыми в гидрологии, позволяют сделать вывод об эффективности предложенной методики (так, при использовании нейросетевого метода в 5% ошибки при прогнозировании с заблаговременностью в 1 год попало 45%, при использовании динамико-статистического метода Алехина – 15%, при использовании метода на основе простых цепей Маркова – 0%, сложных цепей Маркова – 8.3%). На Рис.3 представлен результат сравнения нейросетевого прогноза с заблаговременностью в один год и реальных данных для реки Припять (Мозырь).

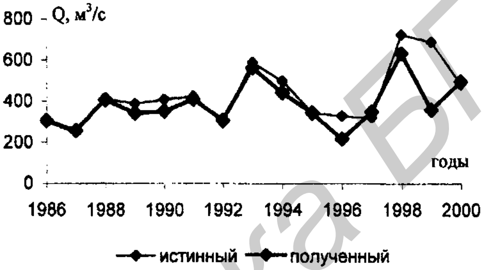

Рис.3. Результаты численного эксперимента по прогнозированию годовых расходов воды реки Припять с заблаговременностью 1 год

Предложенные методы для прогноза годовых расходов воды рек Беларуси использовались при выполнении ГПОФИ «Природные комплексы». Задание 19. «Оценка состояния, прогноз изменения водного режима, степень деградации почв осушенных земель, научное обоснование устойчивого развития экосистем Белорусского Полесья» (2001-2005, № госрегистрации 20031263) в разделе изменения водного режима Белорусского Полесья.

В главе также представлен нейросетевой алгоритм вычисления длины достоверности прогноза на основе вычисления показателей Ляпунова.

В заключении сформулированы основные полученные результаты.

В приложениях приведены выводы формул для вычисления частных производных функции ошибки по переменным синаптическим связям, численные данные временных рядов для экспериментов прогнозирования, а также документы, подтверждающие внедрение.

ЗАКЛЮЧЕНИЕ

Основные результаты диссертационной работы можно сформулировать следующим образом:

1. Для многослойных гетерогенных нейронных сетей прямого распространения предложена матричная алгоритмизация градиентных

Библиотека БГУИР

методов обучения. Данная методика упрощает программную реализацию процедуры обучения [4, 7, 18, 19, 30].

2. Разработаны алгоритмы обучения многослойных нейронных сетей. Для нейронных сетей с одним обрабатывающим слоем представлены алгоритмы на основе ковариационной матрицы и на основе вычисления адаптивного шага обучения. Для нейронных сетей со многими обрабатывающими слоями предложены алгоритмы адаптивного обучения, которые базируются на методах послойного обучения, двухпараметрического обучения, обобщенного метода наискорейшего спуска и условной оптимизации ошибки каждого слоя. Предложена процедура вычисления шага обучения для каждого из этих методов и проведен их сравнительный анализ с известными процедурами обучения персептронных сетей данной архитектуры. Предложенные алгоритмы позволяют повысить быстродействие и скорость обучение, что дает возможность за меньшее время достигать точки локального минимума целевой функции среднеквадратичной ошибки [5-9, 12-14, 18, 20, 21, 29, 31-33].

3. Предложен алгоритм синтеза архитектуры прогнозирующей двухслойной гетерогенной нейронной сети в соответствии с динамическими свойствами временного ряда и свойством сети являться универсальным аппроксиматором. Аргументирован выбор количества элементов в распределительном слое и правило подачи образов на вход сети. Представлено построение экстраполирующей нейросетевой модели для упреждающего прогнозирования. Применение разработанных нейросетевых моделей позволяет повысить точность и долгосрочность прогноза по сравнению с известными методами [1, 3, 10, 11, 15, 19, 22-24, 28, 34-36].

4. Для временных рядов диссипативных систем, характеризующихся притяжением к аттрактору, предложен подход для нахождения длины достоверности прогноза на основе нейросетевых прогнозирующих моделей. Данный подход позволяет использовать обучающие выборки малого объема и является эффективным средством определения состояния динамики системы [1, 2, 16, 17, 25-28].

5. Разработана нейросетевая методика прогнозирования рядов расходов воды. Эксперименты для рек Беларуси показали, что предложенная методика позволяет экстраполировать данный тип временных рядов с большей точностью по сравнению с известными методами [10, 11, 22, 36].

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ ПО ТЕМЕ ДИССЕРТАЦИИ

Глава в книге:

1. V. Golovko, Y. Savitsky, N. Maniakov. Neural Network for Signal Processing in Measurement Analysis and Industrial Applications: the Case of Chaotic Signal Processing - Chapter of NIMIA Book. – Amsterdam: IOS Press, 2003 – P. 119-144.

Библиотека БГУИР

Статьи:

2. Головко В.А., Савицкий Ю.В., Маньяков Н.В., Рубанов В.С. Методы анализа хаотических процессов // Вестник Брестского государственного технического университета. Сер. Физика, математика, химия. – 2001. – № 5(11) – С. 75-78.
3. V. Golovko, Y. Savitsky, N. Maniakov. Modeling Nonlinear Dynamic using Multilayer Neural Networks // Proceedings of the Workshop Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2001), July 1-4, 2001, Foros, Ukraine. – Ternopil: Lileya, 2001. – P.197-202.
4. Маньяков Н.В., Махнист Л.П. Матричная алгоритмизация обучения многослойных нейронных сетей с использованием градиентных методов // Вестник Брестского государственного технического университета. Сер. Физика, математика, химия – 2002. – №5 (17). – С. 60-64.
5. Маньяков Н.В., Махнист Л.П. Обучение нейронных сетей с использованием ковариационной матрицы // Вестник Брестского государственного технического университета. Сер. Физика, математика, химия. – 2002. – №5 (17). – С. 67-72.
6. Гладкий И.И., Маньяков Н.В., Махнист Л.П. Теоретические основы алгоритмов обучения гетерогенных нейронных сетей двухслойной архитектуры // Вестник Брестского государственного технического университета. Сер. Машиностроение, автоматизация, ЭВМ. – 2003. – № 4 (22). – С. 47-54.
7. V. Golovko, N. Maniakov, L. Makhnist. Multilayer Neural Networks Training Methodic // Proceedings of the Workshop Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2003). – Lviv, Ukraine, September 8-10, 2003. – PP. 185-190.
8. Гладкий И.И., Маньяков Н.В., Махнист Л.П. Об одном алгоритме обучения многослойных нейронных сетей на основе условной оптимизации ошибки // Вестник Брестского государственного технического университета. Сер. Физика, математика, химия - 2003. – № 5 (23). – С. 25-30.
9. Головко В.А., Маньяков Н.В. Матричный нейросетевой метод обучения многослойной сети с использованием адаптивного шага // Вестник Брестского государственного технического университета. Сер. Физика, математика, химия - 2003. – № 5 (23). – С. 56-58.
10. Маньяков Н.В. К вопросу прогнозирования временных рядов реальной природы // Вестник Брестского государственного технического университета. Сер. Физика, математика, химия. – 2003. – № 5 (23). – С. 20-25.
11. Логинов В.Ф., Волчек А.А., Маньяков Н.В. Прогнозирование годовых расходов воды рек с помощью нейронных сетей // Природопользование. Сборник научных трудов. Выпуск 10. Ин-т пробл. Использования природ. Ресурсов и экологии НАН ~~Беларуси.~~ ~~Мн.~~ ОДО «Топник», 2004. – С.16-21.

Библиотека БГУИР

12. L. Makhnist, N. Maniakov, V. Rubanov. Some Methods of Adaptive Multilayer Neural Networkd Training // Computing. International Scientific Journal of Computing. Vol. 3, Issue 1. – Ternopil, Ukraine, 2004. – PP. 99-106.
13. N. Manyakov. Use of gradient methods for feedforward neural network training // Artificial Intelligence Study, Vol. 1 (23)/2004 – Siedlce: Publishing House of University of Podlasie – PP. 39-50.
14. Маньяков Н.В., Махнист С.Л. Сравнение эффективности алгоритмов обучения нейронных сетей на основе градиентных методов // Вестник Брестского государственного технического университета. Сер. Физика, математика, информатика. – 2004. – № 5 (29). – С. 40-44.

Тезисы и статьи в сборниках трудов конференций:

15. V. Golovko, Y. Savitsky, N. Maniakov, V. Shut. Some Aspects of Identification of Nonlinear System // Proceedings of Six International Conferences on Pattern Recognition and Information Processing (PRIP'2001). – Minsk: IES NASB, 2001. – P. 110-117.
16. V. Golovko, Y. Savitsky, N. Maniakov, V. Rubanov. Some Aspects of Chaotic Time Series Analysis // Proceedings of the 2nd International Conference on Neural Networks and Artificial Intelligence. (ICNNAI'2001), October 2-5, 2001 – Minsk: BSU, 2001. – P. 66-69.
17. Маньяков Н.В. Вычисление характеристик хаотической диссипативной динамической системы // Материалы II региональной конференции молодых ученых "Современные проблемы математики и вычислительной техники", Брест, 28-30 ноября 2001г. – Брест: БГТУ, 2001. – С.75-78.
18. Маньяков Н.В., Махнист Л.П. Обучение нейронных сетей с использованием градиентных методов // Материалы первого научного семинара «Танаевские чтения», Минск, 28 марта 2003г. – Минск: ОИПИ, 2003. – С. 114-118.
19. N. Maniakov, L. Makhnist, V. Rubanov. Training Algorithm for Forecasting Multilayer Neural Network // Proceedings of Seven International Conferences on Pattern Recognition and Information Processing (PRIP'2003), Vol. 1, Minsk, May 21-23, 2003. – Minsk: UIIP of NASB, 2003. – P. 26-30.
20. N. Maniakov, L. Makhnist, V. Rubanov. Some Methods of Adaptive Multilayer Neural Networks Training // Proceedings of the 3nd International Conference on Neural Networks and Artificial Intelligence (ICNNAI'2003). – Minsk, Belarus, November 12-14, 2003. – PP. 109-115.
21. Маньяков Н.В., Махнист Л.П. Оптимизация процесса обучения экстраполирующей нейронной сети // Материалы V Международной научно-методической конференции «Наука и образование в условиях социально-экономической трансформации общества», часть 2, Брест, 13-14 мая 2004 г. – Брест: ИСЗ, 2004. – С. 359-362.
22. Волчек А.А., Маньяков Н.В., Махнист Л.П. Планирование колебаний среднегодового стока рек Беларуси // Материалы 4-ой Международной научной конференции «Обработка информации в чрезвычайных и

Библиотека БГУИР

экстремальных ситуациях» (ОИУЧЭС'2004), ОИПИ, Минск, 29 ноября – 1 декабря 2004 г. – С. 61-65.

23. Маньяков Н.В. Прогнозирование хаотических временных рядов с использованием нейронных сетей // Материалы IV республиканской научной конференции студентов и аспирантов «Новые математические методы и компьютерные технологии в проектировании, производстве и научных исследованиях», Гомель, 19-22 марта 2001г. – Гомель: ГГУ, 2001. – С. 35-36.
24. Головко В.А., Маньяков Н.В., Рубанов В.С. Применение нейронных сетей к прогнозированию хаотических временных рядов // Тезисы докладов международной математической конференции «Еругинские чтения – VII», Гродно, 28-30 мая 2001г. – Гродно: ГрГУ, 2001. – С. 34-35.
25. Маньяков Н.В. Нейросетевой подход к вычислению фрактальной размерности Каплан-Йорке хаотических процессов // Материалы V республиканской научной конференции студентов и аспирантов «Новые математические методы и компьютерные технологии в проектировании, производстве и научных исследованиях», Гомель, 18-20 марта 2002г. – Гомель: ГГУ, 2002. – С.178-179.
26. Головко В.А., Маньяков Н.В., Рубанов В.С. Численные оценки хаотичности системы // Тезисы докладов международной математической конференции «Еругинские чтения – VIII», Брест, 20-23 мая 2002 г. – Брест: БрГУ, 2002. – С. 39-40.
27. Головко В.А., Маньяков Н.В., Рубанов В.С. Численное вычисление показателей Ляпунова // Тезисы докладов VII международного семинара «Устойчивость и колебания нелинейных систем управления», Москва, 22-24 мая 2002г. – Москва: ИПУ РАН, 2002. – С.23-24.
28. Маньяков Н.В. Использование нейронных сетей в нелинейном анализе // Тезисы докладов второго международного Конгресса «Нелинейный динамический анализ (NDA'2)», Москва, 3-8 июня 2002г. – Москва: МАИ, 2002. – С.80.
29. Маньяков Н.В. К вопросу обучения двухслойных гетерогенных нейронных сетей // Материалы VI республиканской научной конференции студентов и аспирантов «Новые математические методы и компьютерные технологии в проектировании, производстве и научных исследованиях». – Гомель: ГГУ, 2003. – С.27-28.
30. Маньяков Н.В. Методика рекуррентного обучения многослойной нейронной сети // Материалы VI республиканской научной конференции студентов и аспирантов «Новые математические методы и компьютерные технологии в проектировании, производстве и научных исследованиях». – Гомель: ГГУ, 2003. – С.28-29.
31. Маньяков Н.В., Махнист Л.П., Рубанов В.С. Нейросетевой подход к одной задаче управления движением // Тезисы докладов международной математической конференции «Еругинские чтения – IX», Витебск, 20-22 мая, 2003 г. – Витебск: ВГУ, 2003. – С. 115-116.

Библиотека БГУИР

32. Маньяков Н.В., Махнист Л.П. Применения нейронных сетей в эконометрическом моделировании. // Материалы VI Международной научно-методической конференции «Наука и образование в условиях социально-экономической трансформации общества» часть 2, Минск, 15-16 мая 2003 г. – Минск: ИСЗ, 2003. – С. 158-159.
33. Маньяков Н.В., Махнист Л.П., Рубанов В.С. Нейросетевые алгоритмы решения одного класса разностно-дифференциальных уравнений // Материалы международной математической конференции AMADE'2003. – Минск, 2003. – С.118-119.
34. Маньяков Н.В., Махнист Л.П., Рубанов В.С. Моделирование динамики временных рядов в нейросетевом базисе // Материалы 1 Международной научной конференции «Математическое моделирование экономических процессов переходного периода», Минск, 29-31 октября 2003 г. – Мн.: БГЭУ, 2003. – С. 114-115.
35. Маньяков Н.В. Математическое моделирование гидрологических процессов // Матеріали III міжнародної науково-практичної конференції «Динаміка наукових досліджень '2004», Дніпропетровськ, 21-30 червня 2004 року. – Дніпропетровськ: Наука і освіта, 2004. – С. 63-64.
36. Гладкий И.И., Маньяков Н.В., Махнист Л.П. К задаче гидрологического моделирования // Материалы Международной конференции «IX Белорусская математическая конференция», Гродно, 3-6 ноября 2004 г. Т. 3.– С. 21-22.

Библиотека БГУИР

РЭЗЮМЭ

МАНЬЯКОЎ Мікалай Уладзіміравіч

Метады навучання шматслойных нейронных сетак

Ключавыя словы: нейронныя сеткі, градыентныя алгарытмы навучання, апраксімацыя функцый, часовыя шэрагі, хаатычныя працэсы, прадказанне.

Праца прысвечана распрацоўцы эфектыўных метадаў навучання шматварстввых гетэрагенных нейронных сетак прамога распаўсюджвання з рознымі актывацыйнымі функцыямі.

Распрацаваны алгарытмы адаптыўнага пошуку мінімум мэтавай функцыі памылкі сеткі на аснове градыентных метадаў. Дадзена матрычная алгарытмізацыя працэсу навучання, якая спрашчае праграмную рэалізацыю працэсу. Для аднаслойных сетак даследавана магчымасць адназначнага навучання пры пэўных умовах да функцый актывацыі і прапанаванай навучаючай выбарцы.

Дадзена методыка сінтэзу базавай архітэктуры прадказваючай нейроннай сеткі, з захаваннем асноўных дынамічных уласцівасцяў працэсу. Прапанаваны нейрасетачныя метады вызначэння гарызонта прагнозу, што дазваляе рабіць вывад аб магчымасці прадказання.

Прадказваючая мадэль з выкарыстаннем адаптыўнага навучання атрымала ўкараненне.

Галіна выкарыстання вынікаў дысертацыйнай працы – тэорыя нейрасеткавага кіравання.

Библиотека БГУИР

РЕЗЮМЕ

МАНЬЯКОВ Николай Владимирович

Методы обучения многослойных нейронных сетей

Ключевые слова: нейронные сети, градиентные алгоритмы обучения, аппроксимация функции, временные ряды, хаотические процессы, прогнозирование.

Работа посвящена разработке эффективных алгоритмов обучения многослойных гетерогенных нейронных сетей прямого распространения с различными активационными функциями.

Разработаны алгоритмы адаптивного поиска минимума целевой функции ошибки сети на основе градиентных методов. Дана матричная алгоритмизация процесса обучения, упрощая его программную реализацию. Для однослойных сетей исследована возможность однозначного обучения при определенных условиях к функциям активации и предлагаемой обучающей выборке.

Дана методика синтеза базовой архитектуры прогнозирующей нейронной сети, с сохранением основных динамических свойств процесса. Предложены нейросетевые методы определения длины прогноза, что позволяет делать вывод о возможности предсказания.

Прогнозирующая модель с использованием адаптивного обучения получила внедрение.

Область применения результатов диссертационной работы – теория нейросетевого управления.

Библиотека БГУИР

SUMMARY

Nikolaj Maniakov
Multilayer neural networks training methods

Keywords: neural networks, gradient descent training algorithms, approximation of function, time series, chaotic process, forecasting.

This work deals with construction of effective multilater heterogeneous feedforward neural networks training algorithms with different type of activation functions in networks' architecture.

Are developed adaptive algorithms of search of error criterion function minimum based on gradient descent methods. Is given matrix algorithmization of training process, which simplifies its program realization. For onelayer networks is investigated possibility of unique training for a certain conditions on activation functions and offered learning sample.

In dissertation is given method of construction of forecasting neural network architecture, which preserves basic dynamical properties of process. Also is developed neural networks' methods of forecasting horizon's estimation. Such characteristic is helpful to draw a conclusion of possibility of prediction.

Forecasting model based on adaptive learning rule is applied to production.

The sphere of dissertations' work application is neural networks management theory.

Библиотека БГУИР

МАНЬЯКОВ Николай Владимирович

АЛГОРИТМЫ ОБУЧЕНИЯ НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧАХ ПРОГНОЗИРОВАНИЯ

05.13.17 – теоретические основы информатики

Автореферат диссертации на соискание ученой степени кандидата технических наук

Подписано в печать	07.07.2005.	Формат 60x84 1/16.	Бумага офсетная.
Гарнитура «Таймс».	Печать ризографическая.		Усл. печ. л. 1,63.
Уч.-изд. л. 1,4.	Тираж 60 экз.		Заказ 479.

Издатель и полиграфическое исполнение: Учреждение образования «Белорусский государственный университет информатики и радиоэлектроники»
Лицензия на осуществление издательской деятельности №02330/0056964 от 01.04.2004.
Лицензия на осуществление полиграфической деятельности №02330/0131518 от 30.04.2004.
220013, Минск, П. Бровки, 6.

Библиотека БГУИР