

*Е.С. Алдошкина, А.Д. Кузнецов, А.А. Пугачев,
О.С. Сероухова, Т.Е. Симакина, В.В. Чукин*

**ОПЫТ ИСПОЛЬЗОВАНИЯ АППАРАТА НЕЙРОННЫХ СЕТЕЙ
ДЛЯ АНАЛИЗА И ПРОГНОЗА ВРЕМЕННОГО РЯДА
ТЕМПЕРАТУРЫ ВОЗДУХА**

*E.S. Aldoshkina, A.D. Kuznetsov, A.A. Pugachev,
O.S. Seroukhova, T.E. Simakina, V.V. Chukin*

**EXPERIENCE OF STAFF NEURAL NETWORKS
FOR ANALYSIS AND FORECAST OF THE TIME SERIES
OF AIR TEMPERATURE**

В работе рассматривается применение аппарата нейронных сетей для анализа и сверхкраткосрочного прогноза приземной температуры с использованием созданной базы метеорологических данных, получаемых от автоматической метеорологической станции РГГМУ.

Ключевые слова: нейронная сеть, температура воздуха, база метеорологических данных.

The application of the apparatus of the neural networks for the analysis and super-short-term forecast of surface temperature using the established base of meteorological data obtained from automatic weather station RSHU is considered.

Key words: neural network, air temperature, meteorological database.

Нейронные сети – исключительно мощный метод численного моделирования, позволяющий воспроизводить чрезвычайно сложные зависимости. В частности, нейронные сети нелинейны по своей природе. На протяжении многих лет линейное моделирование было основным методом моделирования в большинстве областей, поскольку для него хорошо разработаны процедуры оптимизации. В задачах, где линейная аппроксимация неудовлетворительна, линейные модели работают с большими ошибками и не позволяют моделировать линейные зависимости в случае большого числа переменных.

Нейронные сети учатся на примерах. Пользователь нейронной сети подбирает представительные данные, а затем запускает алгоритм обучения, который автоматически воспринимает структуру данных. При этом от пользователя, конечно, требуется набор эвристических знаний о том, как следует отбирать и подготавливать данные, выбирать нужную архитектуру сети и интерпретировать результаты, однако уровень знаний, необходимый для успешного применения нейронных сетей, гораздо скромнее, чем, например, при использовании традиционных методов статистики.

Нейронные сети сегодня находят все большее применение для анализа различных зависимостей в технике, экономике и других областях.

База метеорологических данных

На основе данных измерений основных параметров состояния атмосферы (температура и влажность воздуха, атмосферное давление, скорость и направление ветра) с помощью автоматической станции Vaisala, установленной в РГГМУ, создана специализированная база данных. Исходные данные представляют собой измерения значений пяти параметров атмосферы через интервал времени 10 с:

- 1) температура воздуха, °С;
- 2) относительная влажность, %;
- 3) атмосферное давление, гПа;
- 4) скорость ветра, м/с;
- 5) направление ветра, градусы.

При записи этих значений в базу данных с помощью разработанных программных средств производится их преобразование в следующие параметры:

- 1) температура воздуха, К;
- 2) парциальное давление водяного пара, Па;
- 3) атмосферное давление, Па;
- 4) скорость ветра, м/с;
- 5) направление ветра, градусы.

Для хранения и доступа к информации используется система управления базами данных (СУБД) MySQL, разрабатываемая фирмой Sun. Данная СУБД является реляционной, что предусматривает представление хранимых данных в виде таблиц. Для работы с данными используются специальные команды структурированного языка запросов SQL.

База данных размещена на сервере meteolab.rshu.ru, и для доступа к данным используется web-интерфейс, реализованный на языке программирования Java. Для обеспечения взаимодействия web-пользователей и виртуальной Java-машины используется контейнер сервлетов Apache Tomcat. Внешний вид пользовательского интерфейса представлен на рис. 1.

При отправке через сеть Интернет запроса по адресу: <http://meteolab.rshu.ru:8080> пользователь получает в ответ от сервера html-страницу (см. рис. 1), с помощью которой возможно отправить запрос к базе метеорологических данных и получить требуемые значения в текстовом виде. С помощью представленной на html-странице формы можно указать время начала блока данных. Затем пользователь указывает требуемую продолжительность блока данных. Далее следует список из десяти параметров, которые пользователь имеет возможность получить в итоговом текстовом файле.

Так, например, есть возможность выбора двух параметров влажности воздуха: относительной влажности воздуха и парциального давления водяного пара. Кроме того, представлены четыре характеристики ветра: скорость ветра, направление ветра, зональная скорость ветра, меридиональная скорость ветра. Значения температуры воздуха представляются в К, атмосферного давления и

парциального давления водяного пара – в ГПа, относительной влажности воздуха – в %.

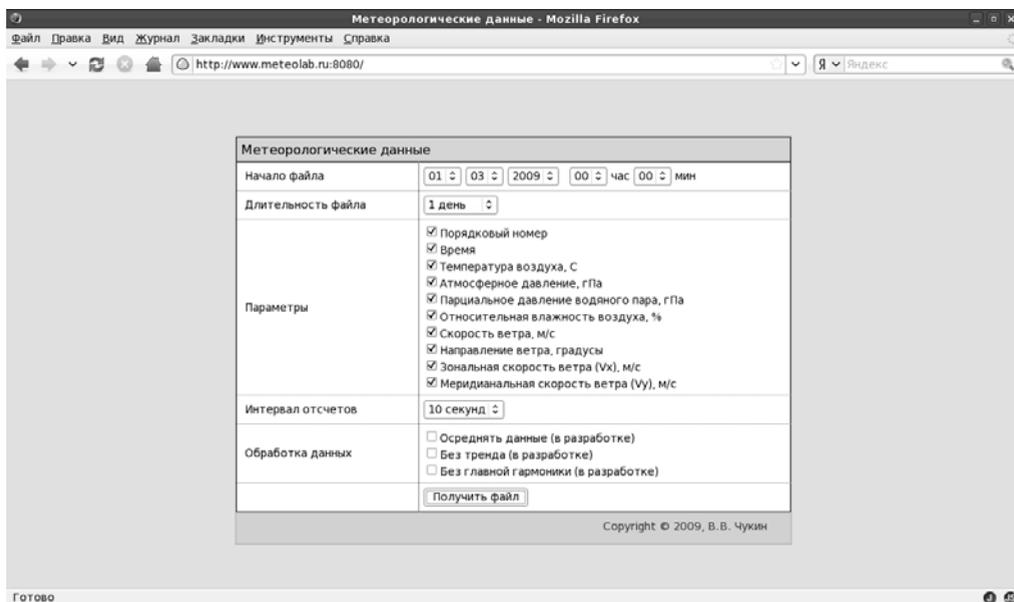


Рис. 1. Внешний вид пользовательского web-интерфейса

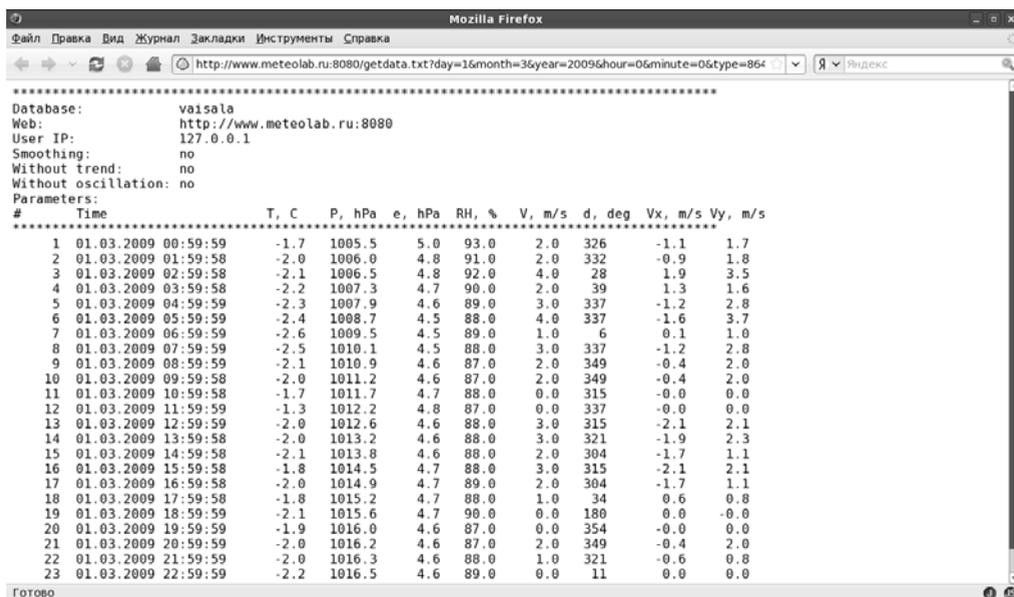


Рис. 2. Текстовый файл с метеорологическими данными

Для целей уменьшения объема создаваемого текстового файла предусмотрена возможность выбора интервала между отдельными отсчетами значений метеорологических параметров.

При нажатии на кнопку «Получить файл» осуществляется запуск программы *GetData*, которая в соответствии с заданными пользователем параметрами осуществляет обращение к базе данных посредством SQL-запросов и генерирует текстовый файл с данными, получаемыми в ответ из базы данных. Текстовый файл отображается в отдельном окне браузера и имеет вид, представленный на рис. 2.

В первых строках текстового файла передается служебная информация с указанием названия базы данных, адреса ресурса в сети Интернет, IP-адресе пользователя и дополнительных параметрах предварительной обработки данных. Затем следует заголовок с указанием названий параметров. Далее расположены строки со значениями метеорологических данных.

Для сохранения текстового файла на диске компьютера пользователя необходимо в меню браузера «Файл» выбрать пункт «Сохранить как...». В появившемся окне указать каталог для сохранения файла, задать желаемое имя файла и нажать на кнопку «Сохранить».

Анализ временного ряда температуры воздуха

В качестве исходных данных для анализа и сверхкраткосрочного прогноза выбран временной ряд значений температуры в срок с 01.04 по 04.04.2009 г. с 15-минутным интервалом, представленный на рис. 3:

$$T = (t_1, t_2, \dots, t_n),$$

где n – размер ряда, равный 385 наблюдениям.

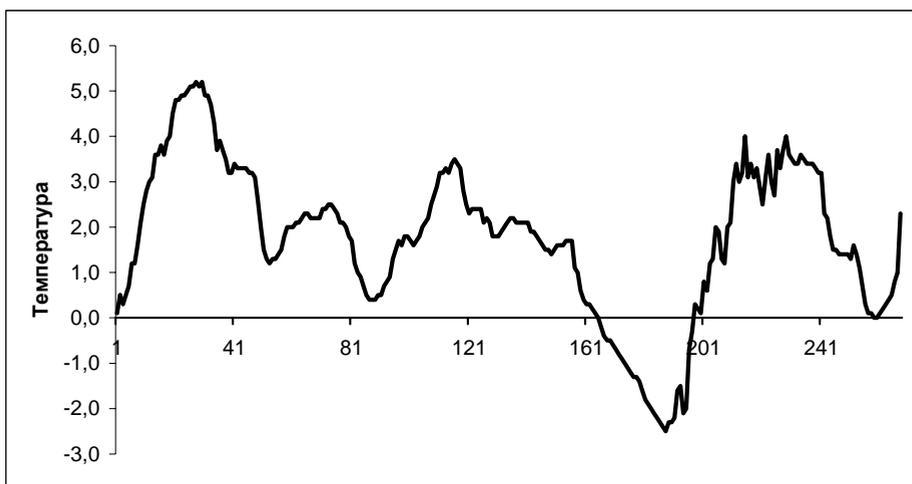


Рис. 3. Исходный ряд температуры

Существенная особенность нейронных сетей состоит в том, что зависимость между входом и выходом находится в процессе обучения сети. В задаче анализа временных рядов обучающее множество данных представлено значениями одной переменной, которая является для сети и входом, и выходом.

Для анализа и прогноза временного ряда используем наиболее широко применяемую для задач такого типа архитектуру сети – многослойный персептрон. Каждый элемент данной сети строит взвешенную сумму своих входов с поправкой в виде слагаемого и затем пропускает эту величину через передаточную функцию, и таким образом получается выходное значение этого элемента. Элементы организованы в послойную топологию с прямой передачей сигнала. Такую сеть можно интерпретировать как модель вход-выход, в которой веса и пороговые значения являются свободными параметрами модели. Данная сеть может моделировать функцию практически любой степени сложности, причем число слоев и число элементов в каждом слое определяют сложность функции.

Важным вопросом при конструировании многослойного персептрона является определение числа промежуточных слоев и числа элементов в них.

Сеть с небольшим числом весов может оказаться недостаточно гибкой, чтобы смоделировать имеющуюся зависимость. Например, сеть без промежуточных слоев на самом деле моделирует обычную линейную функцию. Более сложная сеть способна моделировать более сложные функции, однако она склонна к переобучению.

После того как определено число слоев и число элементов, в каждом из них нужно найти значения для весов и порогов сети, которые бы минимизировали ошибку прогноза, выдаваемого сетью. Ошибка для конкретной конфигурации сети определяется путем прогона через сеть всех имеющихся наблюдений и сравнения реально выдаваемых выходных значений с желаемыми значениями. Полученные разности суммируются в так называемую *функцию ошибок*, значение которой и есть ошибка сети. В качестве функции ошибок чаще всего берется сумма квадратов ошибок. Именно для этого служат *алгоритмы обучения*.

Для обучения многослойного персептрона будем применять наиболее известный метод – метод обратного распространения, при котором имеющиеся данные используются для корректировки весов и пороговых значений сети таким образом, чтобы минимизировать ошибку прогноза на обучающем множестве. Если сеть обучена хорошо, она приобретает способность моделировать неизвестную функцию, связывающую значения входных и выходных переменных, и впоследствии такую сеть можно использовать для прогнозирования в ситуации, когда выходные значения неизвестны.

Алгоритм действует итеративно, и его шаги принято называть эпохами. На каждой эпохе на вход сети поочередно подаются все обучающие наблюдения, выходные значения сети сравниваются с целевыми значениями и вычисляется

ошибка. Значение ошибки используется для корректировки весов, после чего все действия повторяются.

Для независимого контроля результата мы резервируем часть обучающих наблюдений и не используем их в обучении по алгоритму обратного распространения. Следующий рисунок представляет собой зависимость ошибок сети на обучающем и контрольном множестве от числа итераций алгоритма обучения. Как видно из рис. 4, сначала обе ошибки одинаковые. По мере того как сеть обучается, ошибка обучения (нижняя кривая) убывает, а контрольная ошибка (верхняя кривая) перестает убывать и даже растет, что указывает на чересчур точную аппроксимацию данных в процессе обучения. Это явление называется переобучением. В случае переобучения уменьшим число скрытых элементов, так как, по-видимому, сеть является слишком мощной для данной задачи.

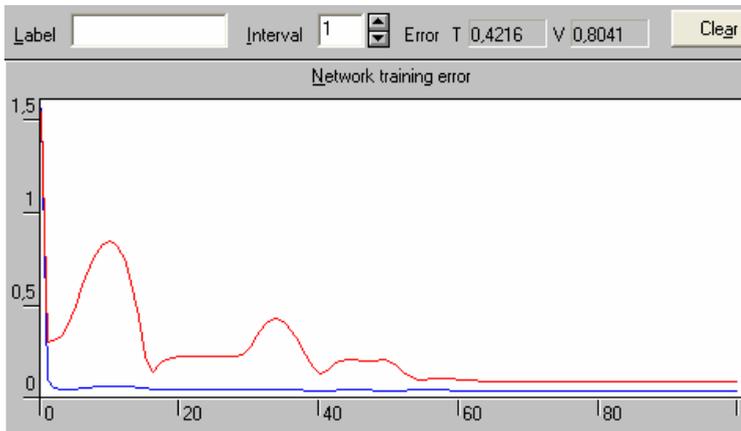


Рис. 4. Ошибки сети при обучении на обучающем (нижняя кривая) и контрольном (верхняя кривая) множестве данных свидетельствуют о переобучении

Проделав эксперименты с различным количеством нейронов в скрытом слое, можно выбрать несколько следующих вариантов, представленных в табл. 1. Критерием выбора было минимальное значение контрольной ошибки.

Рис. 5 представляет собой трехслойный персептрон с семью нейронами в скрытом слое. Входной слой служит для ввода значений входных переменных. Каждый из скрытых и выходных нейронов соединен со всеми элементами предыдущего слоя.

Таблица 1

Зависимость значений контрольной ошибки от количества нейронов

Контрольная ошибка	Количество нейронов в скрытом слое
0,3562097	4
0,3524528	2
0,3474641	8
0,3449294	4
0,3208528	7

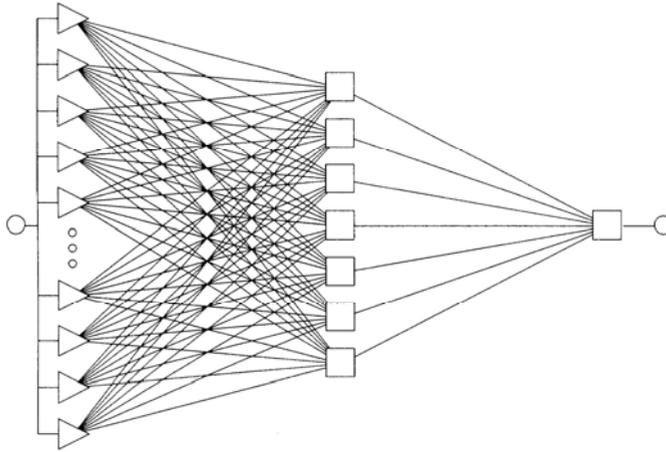


Рис. 5. Структура трехслойного перцептрона с семью нейронами в скрытом слое

В этом случае при обучении добились уменьшения ошибок как на обучающем множестве, так и на контрольном, что демонстрирует рис. 6.

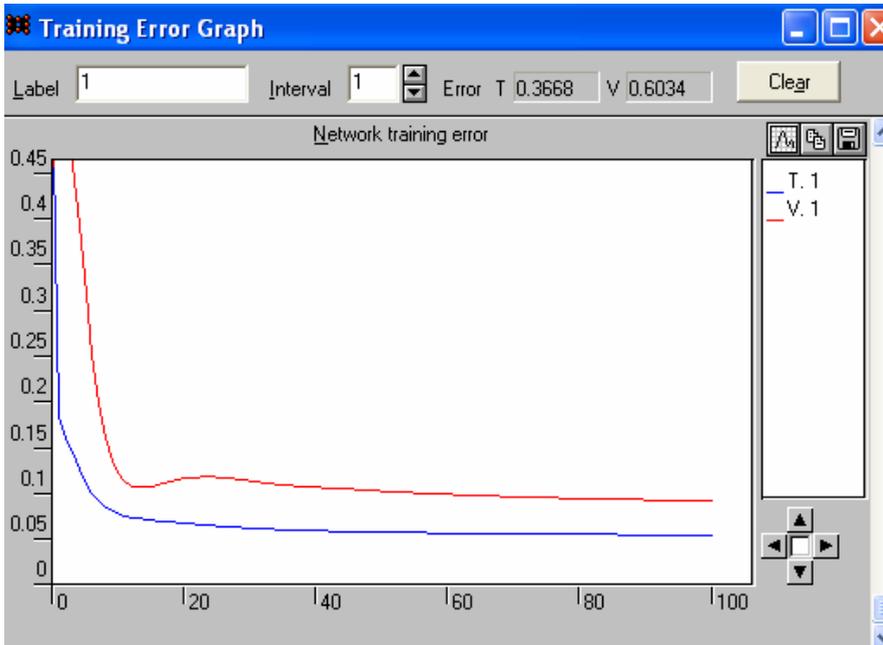


Рис. 6. Ошибки сети при обучении на обучающем (нижняя кривая) и контрольном (верхняя кривая) множестве данных в отсутствии переобучения

Следующие рис. 7–9 представляют собой постепенное улучшение аппроксимации выхода сети исходных данных при изменении количества итераций процесса обучения от 200 до 1800.

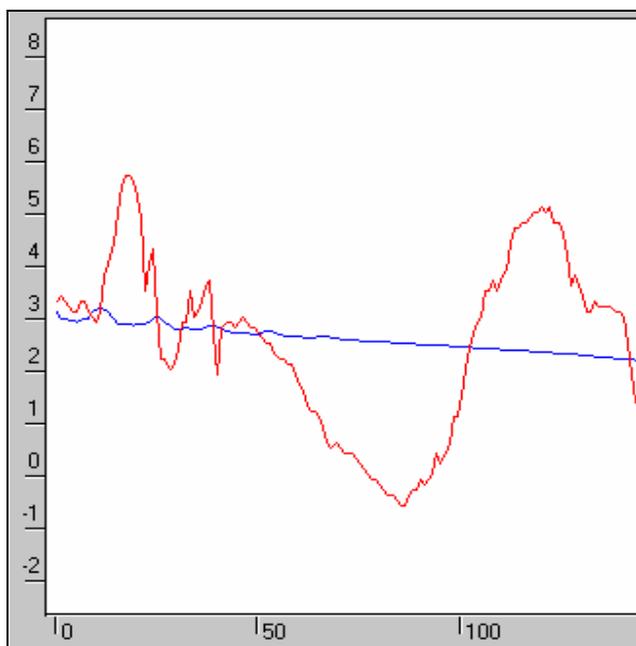


Рис. 7. Исходный ряд и аппроксимационная кривая (сглаженная линия) после 200 итераций

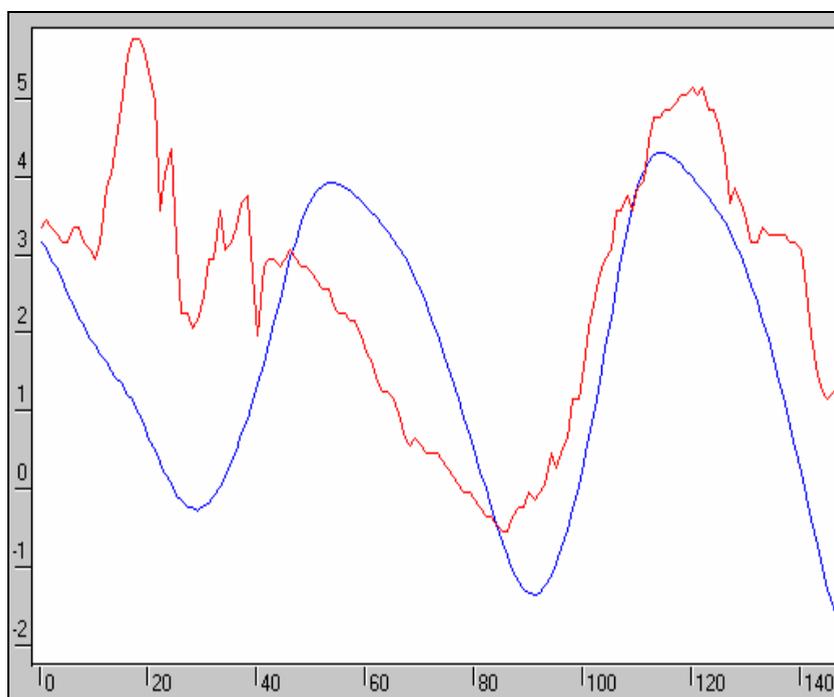


Рис. 8. Исходный ряд и аппроксимационная кривая (сглаженная линия) после 1000 итераций

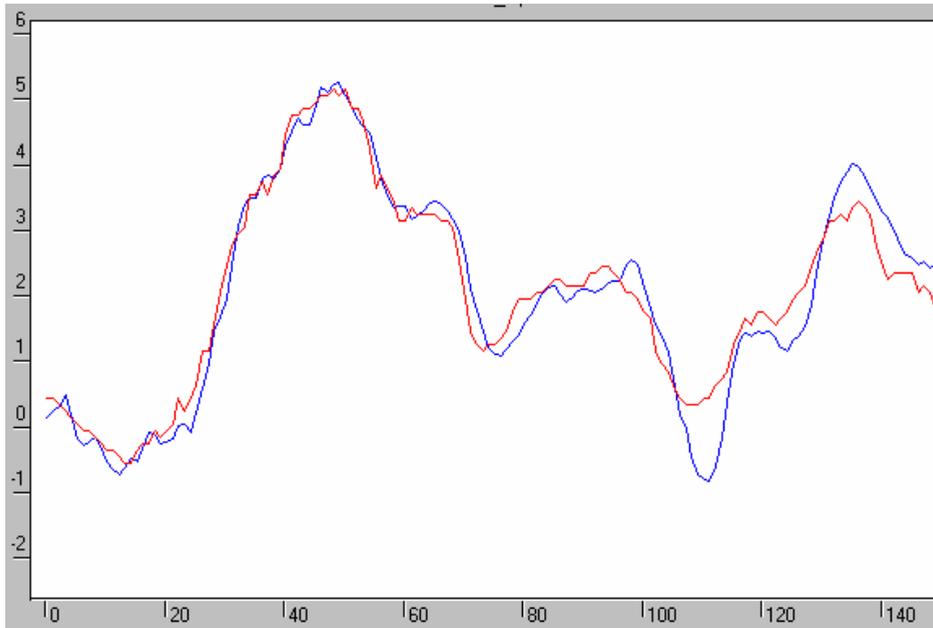


Рис. 9. Исходный ряд и аппроксимационная кривая (сглаженная линия) после 1800 итераций

Прогноз временного ряда осуществляется следующим образом. Сеть, обработав начальный ряд значений, делает прогноз на один шаг. Затем первое наблюдение из начального набора отбрасывается, а оставшиеся значения дополняются прогнозом сети. Получается новый набор входных значений, по которому строится следующий прогноз, и т.д.

Для анализа полученных результатов резервируем часть исходного множества данных. Эта тестовая выборка не будет участвовать в обучении сети. В результате работы нейронной сети при величине обучающей выборки (ОВ) 140, контрольной выборки (КВ) 140 и тестовой выборки (ТВ) 10 был осуществлен прогноз на восемь шагов вперед (2 ч) и получены итоговые статистики, представленные в табл. 2.

Таблица 2

Итоговые статистики

Статистические параметры	ОВ	КВ	ТВ
Data Mean	10,71581	8,622389	9,904598
Data S.D.	3,561492	3,413686	3,616957
Error Mean	-0,02495	0,03557	0,1231606
Error S.D.	0,5065186	0,5299668	0,4435808
Abs E. Mean	0,4088119	0,4434047	0,3795447
S.D. Ratio	0,1422209	0,1552477	0,1226392

В табл. 2 Data Mean – среднее значение обучающих данных; Data S.D. – стандартное отклонение обучающих данных; Error Mean – среднее значение

ошибки прогноза; Error S.D. – стандартное отклонение ошибки прогноза; S.D. Ratio – отношение стандартного отклонения ошибки прогноза к стандартному отклонению обучающих данных.

Из данных табл. 2 следует, что среднее значение ошибки прогноза близко к нулю. Стандартное отклонение ошибки прогноза оказалось существенно меньше стандартного отклонения обучающих данных. Значение отношения стандартного отклонения ошибки к стандартному отклонению данных существенно меньше единицы, что говорит о хорошем качестве регрессии. Это регрессионное отношение (точнее, величину единица минус это отношение) иногда называют долей объясненной дисперсии модели.

Удовлетворительные результаты проведенных численных экспериментов позволяют говорить о перспективности использования аппарата нейронных сетей для решения задач анализа и прогноза временных рядов приземной температуры.

Литература

1. Горбань А.Н. Обучение нейронных сетей. – М.: СП Пара Граф, 1991. – 235 с.
2. Трикоз Д.В. Нейронные сети: как это делается? // Компьютеры + программы, 1993, № 4(5), с. 14–20.
3. Boudouris G. On the index of refraction of air, the absorption and dispersion of centimeter waves by gases // J. Res. National Bureau of Standards, 1963, 67D(6), p. 631–684.
4. Smith E.K., Weintraub S. The constants in the equation for atmospheric refractive index at radio frequencies. in Proceedings of Proc. IRE 41, 1953, p. 1035–1037.
5. Thayer G.D. An improved equation for the radio refractive index of air // Radio Science, 1974, 9(10), p. 803–807.