

Анализ эффективности применения искусственных нейронных сетей для решения задач распознавания, сжатия и прогнозирования¹

Аннотация. В статье обобщен опыт, накопленный в результате создания прототипа программной системы «ПС ИНС» для распознавания графических образов на основе искусственных нейронных сетей (ИНС). Показаны реальные возможности нейронных сетей на примерах решения задач распознавания, сжатия данных и экстраполяции. Приведены результаты экспериментальных исследований по ускорению решения задач на кластерном вычислителе.

Ключевые слова: искусственная нейронная сеть, алгоритм, распознавание, графический образ, эффективность, параллельные вычисления, кластерное вычислительное устройство.

Введение

Существуют две крайние точки зрения на эффективность практического применения ИНС для решения сложных задач - от признания того, что они могут все, до полного отрицания их перспективности. Теоретически, в рамках новой науки - нейроматематики, показано, что ИНС могут решать задачи распознавания, прогнозирования, оптимизации, сжатия информации, аппроксимации и т.д. [1]. Однако вопрос насколько хорошо они это делают, остается открытым и лежит в практической плоскости. Как правило, в научных статьях освещаются только положительные факты целенаправленного использования ИНС, что не дает возможности объективно оценить их эффективность. Существуют десятки видов ИНС и методов их обучения, каждая из которых может успешно выполнить некоторую частную задачу. Однако записать решение сложной практической задачи, к которой можно отнести, например, задачу обнаружения и распознавания объектов на космических снимках, на некотором едином «нейросетевом языке» не удастся в виду его отсутствия. Имеющиеся ней-

ропакеты, по сути, представляют наборы подпрограмм на универсальных языках программирования, моделирующих работу отдельных ИНС. Скомпоновать же решение практически значимой задачи удастся только с привлечением множества других полезных алгоритмов обработки информации. В 2005 году силами лаборатории интеллектуального управления ИПС РАН были начаты соответствующие исследования в рамках одного из проектов Программы Союзного государства «Разработка новых алгоритмов, принципов создания систем обработки изображений и другой информации от средств наблюдения, ориентированных на применение многопроцессорных вычислительных кластеров повышенной вычислительной мощности» (шифр «Триада»). Общим результатом двухлетней работы явилось создание экспериментальной программной системы «ПС ИНС», ориентированной на решение практических задач обработки изображений [2]. Цель настоящей статьи – показать некоторые реальные возможности ИНС по результатам экспериментальных исследований, выполненных в рамках данного проекта.

¹ Работа выполнена в рамках программы Союзного государства «Триада» при частичной финансовой поддержке РФФИ (проекты № 06-07-89083, № 07-07-12029-офи).

1. Проект «ПС ИНС»

«ПС ИНС» - это прототип программной системы распознавания графических образов, обеспечивающей решение задач пользователя путем гибкого объединения средствами графического интерфейса в систему необходимого числа модулей-обработчиков (ИНС, фильтры, специальные модули) с привязкой к отечественной суперкомпьютерной платформе «СКИФ». В рамках проекта проведен детальный анализ основных нейронных сетей на предмет их включения в библиотеку системы. Выполнена программная реализация ИНС на кластерных вычислительных устройствах (КВУ) «СКИФ Первенец-М» и «СКИФ K1000» с использованием стандарта MPI, проведены экспериментальные исследования на задачах распознавания текстовой информации, человеческих лиц и летательных аппаратов, экстраполяции и интерполяции временных рядов, сжатия снимков. Реализованы модули типовой и специальной обработки изображений, включая: фильтрацию, выделение интегральных инвариантных моментов, нормализацию положения объекта, выделение контуров, «скелетов» и других информативных признаков. Основная цель предварительной обработки – подготовка информации к виду, необходимому для последующего решения задачи с применением ИНС.

Для удобства формирования задачи и ее параллельного выполнения на КВУ был разработан специальный графический интерфейс, средствами которого можно выбирать методы предварительной обработки, типы и способы обучения ИНС, запускать на выполнение алгоритмы, сохранять настройки ИНС и повторно использовать полученные результаты.

2. Анализ эффективности ИНС на задачах распознавания графических образов

В результате экспериментов установлено, что нейронные сети могут работать эффективно в области распознавания изображений только в сочетании с другими методами [3]. Так для построения системы «ПС ИНС» потребовалась интеграция:

1) вероятностной нейронной сети с методом сжатия входных данных по принципу выделения инвариантных моментов;

2) персептрона и сети Хемминга с методами геометрической нормализации полутоновых и бинарных входных изображений для их последующего распознавания;

3) сети Кохонена с обобщенной метрикой для решения задачи раскраски регионов на панорамных снимках;

4) практически всех типовых ИНС с комплексом алгоритмов стандартной предварительной обработки (фильтрация, логическая обработка (утонышение, устранение выбросов, «скелетизация», выделение границ), адаптивная бинаризации и т.д.).

Для решения других задач на ИНС потребовались:

1) методы преобразования цветного изображения в яркостное представление;

2) алгоритмы компрессии и корреляционные коэффициенты;

3) методы анализа гистограмм с целью удаления фона, ДПФ и др. преобразования.

Возможности интерфейса позволяют выполнять интеграцию ИНС в комитет большинства, объединяющий различные сети для выработки единого решения. Общий список всевозможных модулей, потребовавшихся для реализации системы, способной гибко формировать задачу пользователя, превысил 60 наименований. В силу открытости архитектуры, список может быть расширен. Рассмотрим два характерных примера, демонстрирующих реальные возможности ИНС.

2.1. Распознавание человеческих лиц

Проведем эксперимент с непосредственным распознаванием лиц по фотографиям без специальной дополнительной обработки. Цель исследования - выявить степень пригодности нейронных сетей к распознаванию биометрической информации. Схема данной задачи, выполненная средствами графического интерфейса «ПС ИНС», показана на Рис.1.

В данном тесте применялись двухслойная сеть прямого распространения с обучением по методу обратного распространения ошибки, однослойный персептрон с обучением по методу Видроу-Хоффа и сеть Хемминга. Структура первых двух ИНС показана на Рис.2.

Заметим, что структура классической сети Хемминга в отличие от схемы персептрона содержит обратные связи, однако в эксперимен-

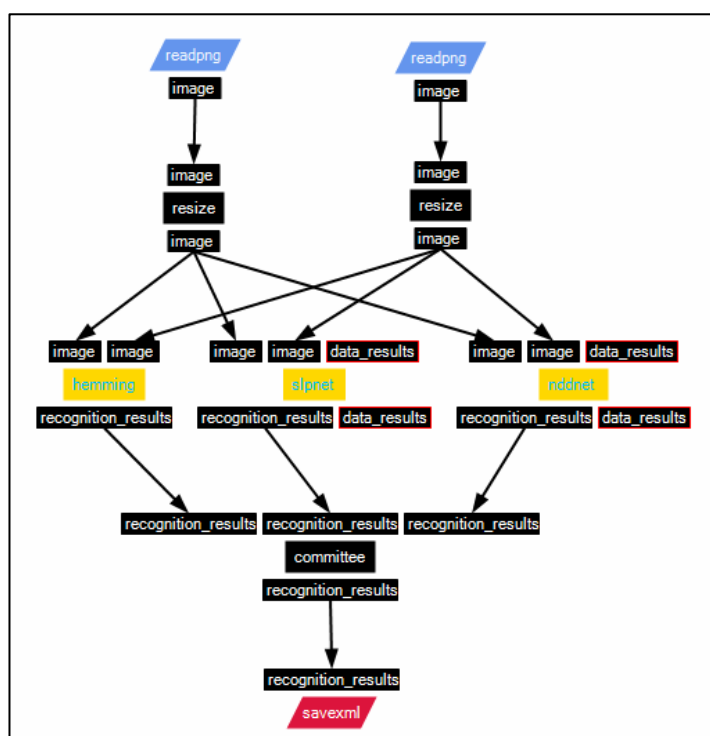


Рис 1. Пример формирования задачи средствами интерфейса

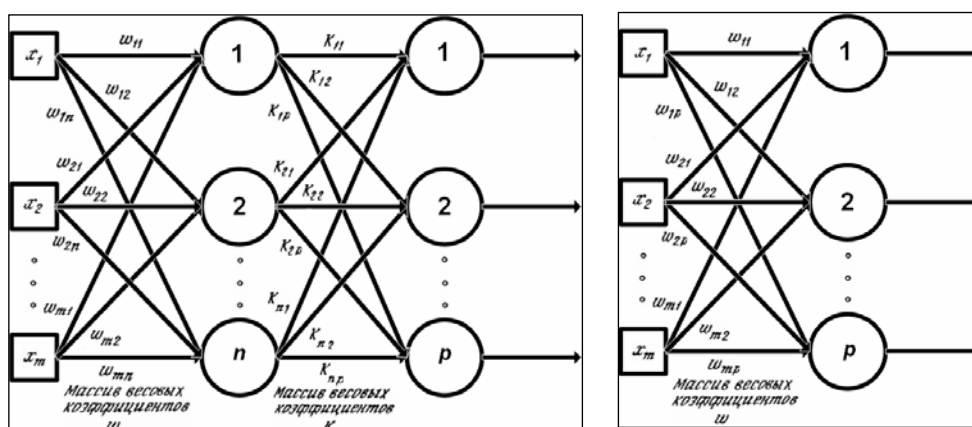


Рис.2. Схемы двухслойной сети прямого распространения и персептрона

тах от них пришлось отказаться с целью ускорения и единообразия процессов анализа выходов ИНС.

Обучающая выборка состояла из 133 фотографий лиц мужчин и женщин, в общей сложности 14 разных людей (классов). Конфигурация сетей задавалась параметрами: $n=26$, $p=14$. Исходная информация представляла собой цветные изображения с размерами от 59x76 до 968x1376. В соответствии со схемой решения задачи (Рис.1) входные изображения преобразовывались в полутоновые и нормализовались.

Непосредственно на входы сетей подавались изображения размером 100x150. Каждый выходной нейрон у сети настраивался на определение вероятности соответствующего класса.

После обучения каждой ИНС для распознавания предъявлялись 28 фотографий, по две на каждого человека (Рис.3).

Сравнительная оценка полученных результатов тестирования по критерию правильно распознанных лиц с учетом уверенности сети дана в Табл.1.



Рис. 3. Тестовая выборка

Табл. 1. Статистика правильно распознанных фотографий

Тип ИНС	Степень уверенности нейронной сети (%)			
	≥ 0	≥ 35	≥ 50	≥ 75
Сеть прямого распространения	24	20	16	14
Персептрон	23	14	7	1
Сеть Хемминга	15	15	13	2

В Табл. 2 приведены ошибки ИНС. Сеть прямого распространения, как видно из Табл. 1 и Табл.2, показала лучшие результаты по сравнению с однослойным персептроном и ИНС Хемминга. Заметим, что результаты могут варьироваться путем изменения стратегии обучения и выбора системы эталонов. В целом же они свидетельствуют о достаточно низкой эффективности нейросетевого подхода при решении биометрической задачи без использования дополнительных инструментов анализа. В этой связи следует указать на наличие других методов, возможно, более успешного решения данной задачи, включая структурный подход, основанный на лингвистическом анализе изображений [4,5].

2.2. Выделение и распознавание летательных объектов на панорамном снимке

Решение данной задачи строится по технологической цепочке, в которой нейронная сеть

вступает в действие только на последнем этапе. Задача явилась одним из главных тестов для изучения способности системы «ПС ИНС» находить и идентифицировать объекты на космических снимках. Технологическая цепочка решения задачи выглядит как совокупность взаимосвязанных модулей (чтение объекта исследования, проведение предварительно обработки, выбор нейронной сети, ее обучение, распознавание, запись результата).

В свою очередь предварительная обработка – это устранение фона, выделение объекта по определенному критерию, сборка объектов из отдельных частей и т.д. Графическая схема задачи распознавания приведена на Рис. 4. Вся конструкция запускается на КВУ «СКИФ» и функционирует в параллельно-конвейерном режиме.

В качестве обучающей выборки использовался набор из 142 эталонных изображений самолетов, часть из которых представлена на Рис.5.

Табл. 2. Характеристики ошибок ИНС

Сеть прямого распространения			Персептрон		
Исходное изображение	Распознано как	Уверенность (%)	Исходное изображение	Распознано как	Уверенность (%)
		0.25			0.01
		0.35			0.27
		0.25			0.23
		0.27			0.01
					0.09

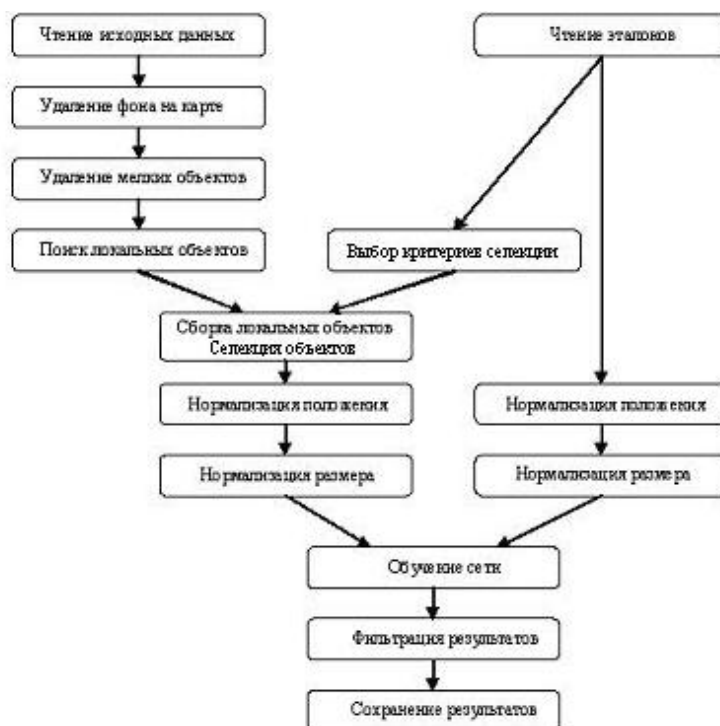


Рис.4. Графическая схема задачи

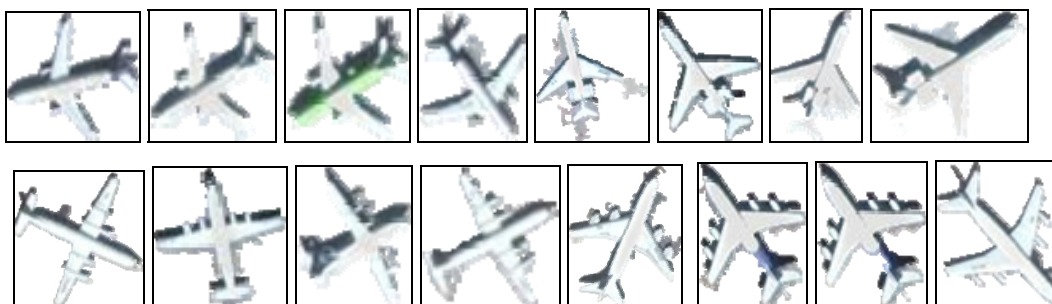


Рис.5. Примеры эталонов

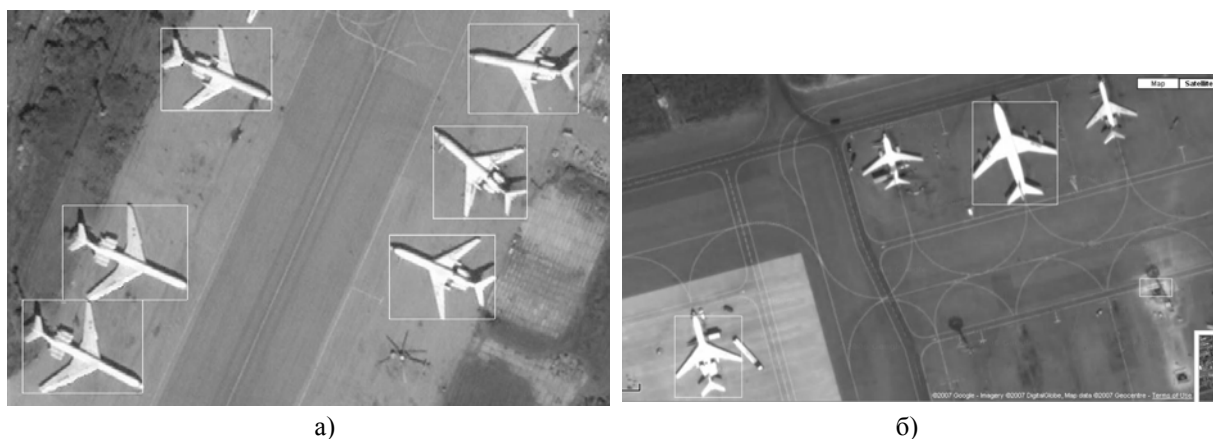


Рис. 6. Результаты выделения распознавания ЛА

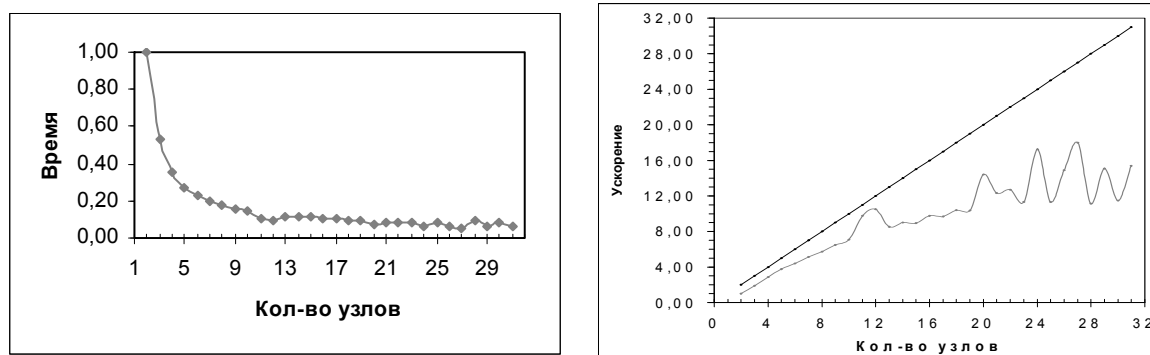


Рис.7. Эффективность распараллеливания

Результаты распознавания летательных аппаратов (ЛА) для двух панорамных снимков приведены на Рис.6, где распознанные объекты выделены прямоугольниками.

Качество решения данной задачи в целом более высокое по сравнению с распознаванием лиц. Тем не менее, видно, что на Рис. 6 б имеются нераспознанные самолеты и объект, ошибочно принятый за самолет, что объясняется несовершенством модуля отвечающего за селекцию. Для улучшения результатов приходится использовать комитет ИНС, включающий сеть прямого распространения, сеть Хемминга и однослойный персептрон.

Эффективность работы ПС ИНС при запуске на различном количестве узлов кластерной установки большой мощности «СКИФ К1000» можно оценить по графикам, приведенным на Рис.7. Здесь за единицу отсчета принято время решения задачи на двух узлах кластера, составившее 215.87 сек.

На Рис.7. слева показано изменение относительного времени решения задачи с увеличением числа узлов кластера (каждый узел кластерной установки имеет по два процессора типа AMD Opteron 275 Dual-Core 2.2 Ghz, однако, при проведении тестирования, на каждом узле реально использовался лишь один из доступных процес-

соров). Справа показано теоретическое (прямая линия) и реальное ускорение вычислений.

3. Анализ эффективности ИНС на задаче сжатия изображений

Много неточностей связано с оценкой возможности ИНС сжимать информацию. При подсчете коэффициента компрессии зачастую не учитываются реальные затраты на хранение и передачу весовых коэффициентов обученной ИНС, что дает искаженное представление о степени сжатия [6-9]. Рассмотрим, например, задачу сжатия на основе сети Кохонена. Технология основывается на векторном квантовании – способе кластеризации, в котором пространство входов делится на ряд областей, для каждой из которых определяется вектор восстановления [9]. Архитектура самоорганизующейся карты признаков для векторного квантования определяется размером словаря кодовых векторов. Каждый кодовый вектор – это матрица весов в соревновательном слое. В эксперименте блок, содержащий 16 пикселей, подавался на слой Кохонена, состоящий из 256 узлов-кластеров, размещенных в двумерном массиве 16x16. Веса, связывающие j -й нейрон слоя Кохонена и входы, представлены матрицей $[w_{ji}]$, $j = 0, 1, \dots, 255$, $i = 0, 1, \dots, 15$.

Схема сжатия изображения с помощью словаря:

- 1) изображение делится на блоки, подаваемые в случайном порядке на вход сети;
- 2) выбирается нейрон с минимальным евклидовым расстоянием до поданного блока;
- 3) веса нейронов, не победивших хотя бы у одного блока, исключаются из словаря;
- 4) сохраняется множество индексов соответствия нейронов-победителей блокам сжатого изображения;
- 5) сохраняется словарь кодовых векторов.

Схема восстановления изображения:

- 1) для каждого индекса блока сжатого изображения находим соответствующий кодовый вектор нейрона-победителя;
- 2) найденный кодовый вектор формирует блок результирующего изображения.

Для сжатия использовались файлы изображений размером 2112 на 2816 пикселей, то есть площадью 5947392 пикселей. Коэффициент сжатия рассчитывается по формуле

$$G = \frac{O}{N \cdot S \cdot \frac{|\log_2 N|}{8} \cdot B},$$

где O – площадь сжимаемого изображения, N – размер словаря, S – длина кодовых векторов (площадь блока), B – число блоков в изображении. Результаты эксперимента при $N=25$ представлены в Табл. 3.

Табл. 3. Результаты восстановления изображений блоками 8x8



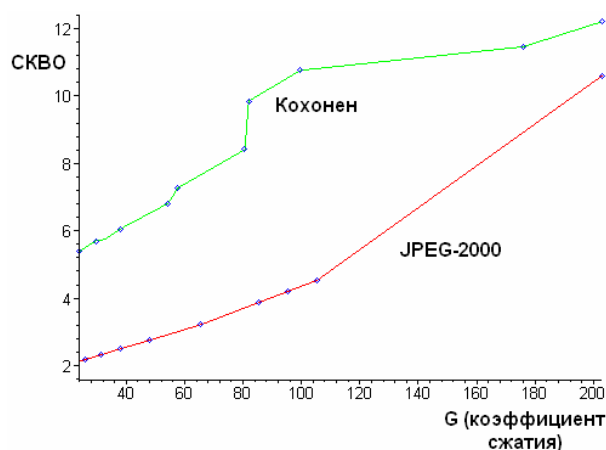


Рис. 8. Сравнение алгоритмов сжатия

Эксперимент показывает, что восстановленные изображения обладают приемлемым качеством даже при высоких степенях сжатия. На Рис. 8 приведены результаты сравнения качества сжатия изображений по Кохонену и широко известного алгоритма JPEG-2000 на основе оценки СКВО при различных значениях коэффициента G .

Для визуальной оценки качества сжатия с помощью ИНС Кохонена на Рис. 9 показаны фрагменты оригинального и восстановленного изображений для $G=33$.

При сильном увеличении фрагмента восстановленного изображения хорошо заметны последствия ошибок сети — «лестничный» эффект, вызванный визуализацией границ пересекающихся блоков.

4. Анализ эффективности ИНС на задаче экстраполяции температурного ряда

Нейронные сети являются удобным инструментом для работы со статистическими данными. В частности они позволяют экстраполировать временные ряды. Для проведения эксперимента были взяты некоторые реальные данные среднесуточных температур с января 2006-го года по май 2008-го (Рис. 10). Каждый отсчет $P(t)$ здесь определяет среднюю температуру за один день, соответствующий моменту времени t на графике.

Воспользуемся трехслойной ИНС прямого распространения с функцией активации сигмоид (Рис. 11). Непосредственно для прогнозирования значений температур используется единственный нейрон третьего слоя сети.



Рис. 9. Демонстрация качества восстановления изображения

Обучение выполняется с использованием скользящего окна по методу обратного распространения ошибки. На каждом шаге обучения берут k значений ряда (по размеру окна) и обучают сеть, предъявляя в качестве эталона на выходе ИНС $(k+1)$ значение ряда. Затем окно перемещается на шаг вправо и цикл обучения возобновляется. Данный процесс повторяют, начиная с первой позиции окна до тех пор, пока не будут получены приемлемые результаты по точности. При экстраполяции ряда более чем на один шаг используются данные, полученные в результате предшествующих этапов экстраполяции. Для эксперимента применялась ИНС с 10 нейронами входного слоя и 5 скрытого. Проведем следующий эксперимент: обучим нейронную сеть на данных с 1 января 2006-го по 21 апреля 2008-го включительно и спрогнозируем температуру воздуха с 22-го апреля 2008-го года по 6-е мая включительно (15 дней, 15 отсчетов). Результаты экстраполяции представлены на Рис. 12, причем нулевой точке отсчета соответствует 12-е апреля 2008-го года.

Размер скользящего окна составил $k=90$, критерием останова служило приемлемое значение СКВО. Время обучения ИНС на одном процессоре AMD AthlonMP 1800+ составило 280 сек, а СКВО — 3.01. Из графика (Рис. 10) видно, что ИНС экстраполирует ряд с некоторыми ошибками, но при этом достаточно хорошо прогнозирует тенденцию. Данный эксперимент, разумеется, не может служить примером реального прогноза температуры, зависящего от множества внешних факторов, но демонстрирует способность ИНС к обобщению информации.

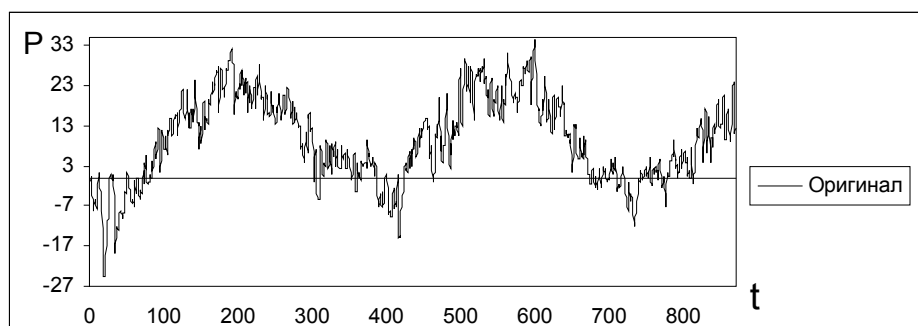


Рис.10. График температурного ряда $P(t)$

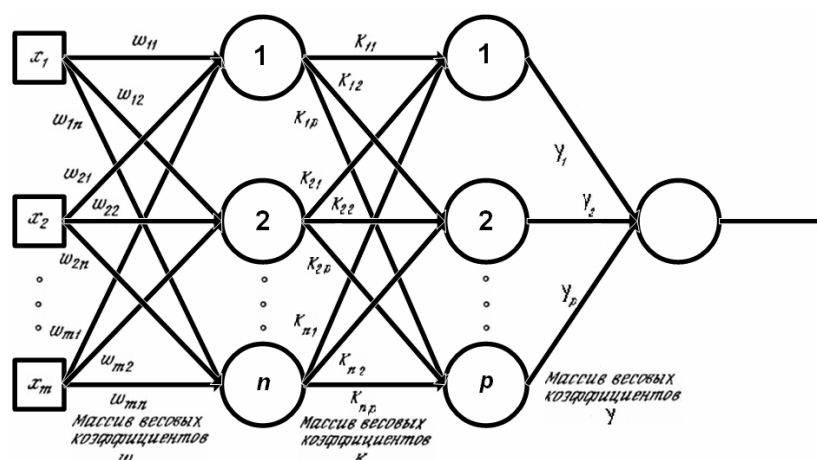


Рис. 11. Структура ИНС прямого распространения

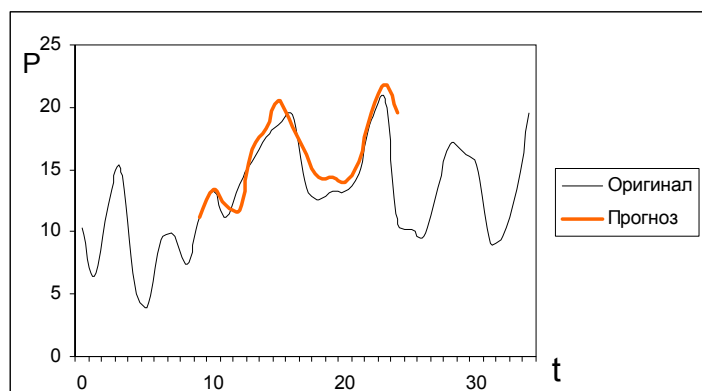


Рис. 12. Результат экстраполяции температурного ряда

Заключение

Экспериментальные исследования с ИНС, проведенные в настоящей работе на основе программной системы «ПС ИНС», показали следующее.

1) ИНС могут быть успешно использованы для распознавания графических образов только при условии их интеграции с дополнительными

методами обработки изображений, осуществляющими выделение из фона и нормализацию изображений, которые могут составить большую часть от общего объема вычислений.

2) Качество сжатия и восстановления изображений на ИНС уступает при прочих равных условиях известному алгоритму JPEG-2000. Однако при этом ИНС обеспечивает уникаль-

ность для установленной серии изображений набора весовых коэффициентов, что может быть использовано при шифровании передаваемой и восстанавливаемой информации.

3) Возможности распараллеливания задач с использованием ИНС достаточно высоки, что обеспечивается аппаратно-программной поддержкой отечественных КВУ «СКИФ», обеспечивающих масштабируемое ускорение вычислений.

4) Эксперименты с экстраполяцией временных рядов показывают, что нейросетевые алгоритмы могут быть использованы в прикладных задачах прогнозирования потоков данных.

Разумеется, описанными в работе экспериментами не исчерпывается исследование возможностей и эффективности применения ИНС. В настоящее время продолжается работа над совершенствованием модулей и алгоритмов решения прикладных задач на основе ИНС. Возможные области применения разрабатываемой программной системы «ПС ИНС» аэрокосмического назначения: обработка информации ДЗЗ, решение задач обнаружения и распознавания локальных объектов, прогнозирование динамических процессов и решение траекторных задач в системах управления.

Литература

1. Галушкин А.И. Нейроматематика (проблемы развития).- М.: Радиотехника, 2003. – 40 с. (серия «Нейрокомпьютеры и их применение»)
 2. Виноградов А.Н., Калугин Ф.В., Недев М.Д., Погодин С.В., Талалаев А.А., Тищенко И.П., Фраленко В.П., Хачумов В.М. Выделение и распознавание локальных объектов на аэрокосмических снимках. – Авиакосмическое приборостроение, № 9, 2007, с.39-45.
 3. Хачумов В.М., Виноградов А.Н. Разработка новых методов непрерывной идентификации и прогнозирования состояния динамических объектов на основе интеллектуального анализа данных. – Сб. докладов 13-ой Всероссийской конференции «Математические методы распознавания (Ленинградская обл., г. Зеленогорск, 30-сентября-6 октября, 2007 г).- М.: МАКС Пресс, 2007, с.548-550.
 4. Глазунов А. Компьютерное распознавание человеческих лиц. <http://www.osp.ru/os/2000/03/177945/>
 5. Русын Б.П. Структурно-лингвистические методы распознавания изображений в реальном времени. - Киев: Наук. думка, 1986. -128 с.
 6. Verma B., Blumenstein M., Kulkarni S. A New Compression Technique Using an Artificial Neural Network, Journal of Intelligent Systems, 1999, vol. 9, № 1, pp.39-54. <http://www.avaye.com/files/articles/imgcompr/imagecompression.pdf>
 7. Verma B., Blumenstein M., Kulkarni S. A Neural Network Based Technique for Data Compression, Proceedings of the IASTED International Conference on Modelling and Simulation, MSO '97, Singapore, pp. 12-16. <http://www98.griffith.edu.au/dspace/bitstream/10072/15827/1/7038.pdf>
 8. Vilovic I. An Experience in Image Compression Using Neural Networks.- 48th International Symposium ELMAR-2006, 07-09 June 2006, Zadar, Croatia, pp. 95-98
 9. Durai A.S, Saro A.E., Phil M. An Improved Image Compression approach with Self-Organizing Feature Maps using Cumulative Distribution Function.- GVIP Journal, Volume 6, Issue 2, September, 2006, pp. 122-128. <http://www.icgst.com/gvip/Volume6/Issue2/P1150630002.pdf>
- Талалаев Александр Анатольевич.** Институт программных систем РАН, инженер ИЦИИ. В 2006 году окончил Университет имени А.К. Айламазяна города Переславля. 7 печатных работ. Область научных интересов: искусственный интеллект, машинная графика.
- Тищенко Игорь Петрович.** Институт программных систем РАН, аспирант. В 2005 году окончил Университет имени А.К. Айламазяна города Переславля. 2 печатные работы. Область научных интересов: искусственный интеллект, машинная графика, распознавание образов, нейронные сети, параллельные и разрядно-параллельные вычисления.
- Фраленко Виталий Петрович.** Институт программных систем РАН, аспирант. В 2007 окончил Университет имени А.К. Айламазяна города Переславля. 8 печатных работ. Область научных интересов: искусственный интеллект, машинная графика, распознавание образов, нейронные сети, параллельные и разрядно-параллельные вычисления.
- Хачумов Вячеслав Михайлович.** Главный научный сотрудник Института системного анализа РАН. В 1971 окончил Ленинградский кораблестроительный институт, приборостроительный факультет. Д.т.н., профессор. 120 печатных работ, 2 монографии, 1 учебное пособие. Область научных интересов: искусственный интеллект, распознавание образов, нейронные сети, параллельные и разрядно-параллельные вычисления.