

В. П. Фраленко

## Прогнозирование, сжатие и фильтрация данных на нейронных сетях<sup>\*)</sup>

Научный руководитель: д.т.н. В. М. Хачумов

Аннотация. Приведены результаты некоторых экспериментов по использованию нейронных сетей для прогнозирования временных рядов, сжатия изображений с потерями и фильтрации изображений. В частности показано, что нейронная сеть прямого распространения может быть достаточно успешно применена для прогнозирования временных рядов с использованием средств распараллеливания. Для сжатия изображений применены сети прямого распространения и Кохонена. Выполнен более корректный учет степени компрессии информации. Предложены и экспериментально исследованы новые схемы фильтрации изображений, показана их эффективность при удалении помех на сильно зашумленных изображениях.

### 1. Введение

В настоящее время на передачу данных тратятся большие временные ресурсы. В связи с этим актуальными являются вопросы сжатия (компрессии) информации перед ее передачей. Компрессия позволяет значительно увеличить пропускную способность линий связи и является дополнительной мерой обеспечения защиты конфиденциальной информации. На сегодняшний день имеется ряд перспективных схем сжатия DVI, JPEG, MPEG и DWT. Однако в последнее время возник интерес к альтернативным способам сжатия, в том числе с помощью искусственных нейронных сетей (ИНС).

Многообразие недостоверной и противоречивой информации о возможностях искусственных нейронных сетей побудило автора провести некоторые собственные исследования и эксперименты в наиболее значимых и критичных приложениях (прогнозирование, сжатие, фильтрация). Если работы в области прогнозирования выглядят относительно убедительно [1–9], то этого нельзя сказать об алгоритмах сжатия на ИНС [10–19]. В целом эти алгоритмы не являются универсальными. Кроме того, зачастую при подсчете коэффициента сжатия

---

<sup>\*)</sup>Представлено по тематике: *Программное обеспечение для суперЭВМ, Распознавание образов и нейронные сети.*

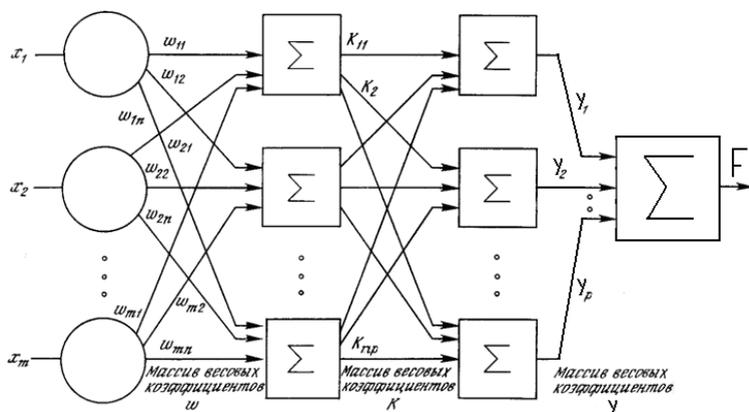


Рис. 1. Блок-схема нейронной сети

не учитываются реальные затраты на хранение и передачу весовых коэффициентов обученной ИНС, что дает искаженное представление о качестве сжатия [10–13, 15, 18].

Заметим, что прогнозирование появления символов может быть использовано и как один из этапов для решения задачи эффективного сжатия информации с помощью алгоритма Хаффмана и его аналогов [14]. Наиболее мощные методы компрессии используют комбинация аппаратных средств и программного обеспечения. Это направление стало наиболее актуальным в связи с появлением многоядерных микропроцессоров, кластерных вычислительных устройств (КВУ) с распределенной памятью и многопроцессорных вычислительных устройств (МВС) с общей памятью. Далее остановимся подробнее на рассмотрении принципов прогнозирования и сжатия изображений, проведенных исследованиях в области фильтрации изображений на основе ИНС.

## 2. Применение ИНС для прогнозирования и интерполяции временного ряда

Известно, что функцию многих переменных можно приблизить с заданной точностью с помощью трехслойной ИНС прямого распространения с достаточным количеством нейронов в скрытом слое и правильно подобранными синаптическими коэффициентами [1]. Данная сеть (рис. 1) с функцией активации типа сигмоид и обучением по

методу обратного распространения ошибки была выбрана для экспериментов с временными рядами.

Пусть задан временной ряд в виде последовательности  $n + 1$  отсчетов  $p(0), p(1), \dots, p(n)$  дискретной функции надежности  $p(t)$ , называемой также функцией вероятности безотказной работы. Значения  $p(n + 1), \dots, p(n + m)$  в рассматриваемом контексте называются прогнозными. Задача состоит в построении функции  $F(t)$ , экстраполирующей  $p(t)$  на интервале  $[(n + 1), \dots, (n + m)]$ .

Для повышения качества обучения ИНС следует стремиться к статистической независимости входных данных, то есть к отсутствию корреляций [3]. В качестве входных переменных были выбраны изменения значений ряда  $\Delta p(t) = p(t + 1) - p(t)$ . Выполним нормирование  $\Delta p(t)$  в диапазоне от 0.25 до 0.75 следующим образом:  $\Delta p(t) = \frac{\Delta p(t) - \Delta p_{\min}}{2 \cdot \Delta p_{\max - \min}} + 0.25$ , где  $\Delta p_{\min} = \min \{\Delta p(t)\}$ ,  $\Delta p_{\max} = \max \{\Delta p(t)\}$ ,  $t = 0, \dots, n - 1$ ,  $\Delta p_{\max - \min} = \Delta p_{\max} - \Delta p_{\min}$ .

## 2.1. Эксперименты по прогнозированию данных

Для экспериментов с прогнозированием использовалась типовая кривая вероятности безотказной работы, представленная своими отсчетами в табл. 1. Требовалось решить задачу обучения ИНС на указанном ряде и прогнозирования ее поведения (экстраполяции) на некоторое число шагов вперед. Эталонные экстраполированные значения представлены в табл. 2.

Схема решения задачи выглядит следующим образом:

- обучить ИНС на нормированных отсчетах табл. 1 методом «скользящего окна», используя в качестве входов  $q$ -мерные входные вектора;
- взять в качестве входного вектора первые  $q$  значений  $\Delta p(t)$  и восстановить значения табл. 2.

Результаты работы ИНС, содержащей 24 нейрона входного слоя и 12 скрытого при  $q = 7$ , представлены в табл. 3. Графики оригинальной кривой надежности и результатов прогнозирования представлены на рис. 2. Видно, что спрогнозированная кривая практически полностью совпадает с оригинальной кривой.

Для оценки эффективности параллельной реализации процесса прогнозирования было сделано десять долгосрочных прогнозов по миллиону прогнозируемых значений в каждом. За единицу масштаба

ТАБЛИЦА 1. Известные значения временного ряда

$t$	$p(t)$	$\Delta p(t)$
0	1	0.750000
1	0.99	0.750000
2	0.98	0.719697
3	0.96	0.734848
4	0.945	0.704545
5	0.92	0.719697
6	0.9	0.704545
7	0.875	0.674242
8	0.84	0.659091
9	0.8	0.553030
10	0.725	0.250000
11	0.55	0.628788
12	0.5	0.553030
13	0.425	0.250000
14	0.25	0.628788
15	0.2	0.734848
16	0.185	0.583333
17	0.12	-----

ТАБЛИЦА 2. Восстанавливаемые значения временного ряда

$t$	$p(t)$
18	0.1000
19	0.0875
20	0.0500
21	0.0140
22	0.0040

взято время счета всех прогнозов на одном процессоре (без сохранения результатов на жесткий диск), равное 183.41 секунды.

Результаты эксперимента в виде графика ускорения приведены на рис. 3, который свидетельствует о хорошей масштабируемости времени прогнозирования с ростом числа вычислительных узлов при равномерном распределении задач между узлами.

ТАБЛИЦА 3. Вычисленные значения временного ряда

$t$	$F(t)$
8	0.842385
9	0.800702
10	0.730786
11	0.564739
12	0.486919
13	0.454956
14	0.234814
15	0.219667
16	0.172118
17	0.133719
18	0.074754
19	0.064802
20	0.058627
21	0.025216
22	0.003959

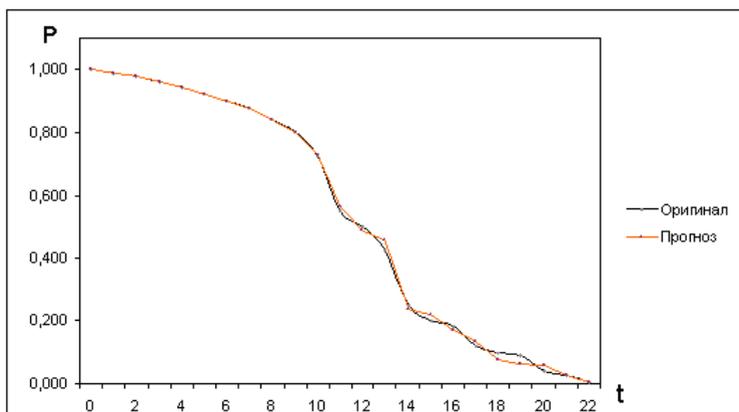


Рис. 2. Прогноз поведения функции надежности

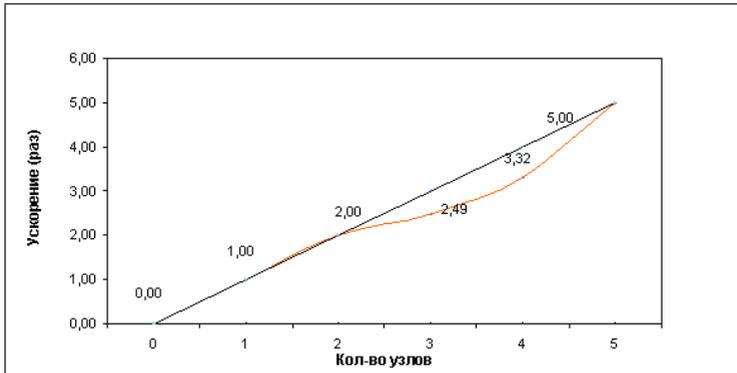


Рис. 3. График эффективности параллельной постановки прогнозов

## 2.2. Эксперименты по интерполяции

Рассмотрим временной ряд в виде последовательности  $2n + 1$  отсчетов  $p(0), \dots, p(n-1), p(n), p(n+1), \dots, p(2n)$  дискретной функции  $p(t)$ . Будем полагать, что известны первые  $n$  значений функции до  $p(n)$  и  $n$  значений после. Необходимо определить неизвестное нам значение  $p(n)$ .

Для решения поставленной задачи построим интерполяционную функцию  $F(t)$ , которая будет принимать в узлах те же значения, что и  $p(t)$ . Схема действий по построению такой функции выглядит следующим образом:

- обучить нейронную сеть на двух группах отсчетов по  $n$  отсчетов в каждой, используя в качестве входных данных  $q$ -мерные вектора нормированных значений  $\Delta p(t)$  из табл. 4 при  $q < n$ ;
- взять в качестве входного вектора  $q$  последних значений  $\Delta p(t)$  из первой группы и восстановить  $p(n)$  с помощью одношагового прогнозирования.

В результате применения ИНС, содержащей 12 нейронов входного слоя и шесть скрытого при  $n = 9$ ,  $q = 7$ , получено  $F(9) = 0.808710$  при  $p(9) = 0.8$ . Качество интерполяции можно повысить, взяв для нейронной сети более представительные обучающие выборки с большим числом эталонных векторов.

ТАБЛИЦА 4. Заданные узлы интерполяции

$t$	$p(t)$	$\Delta p(t)$
0	1.000	0.750000
1	0.990	0.750000
2	0.980	0.719697
3	0.960	0.734848
4	0.945	0.704545
5	0.920	0.719697
6	0.900	0.704545
7	0.875	0.674242
8	0.840	-----
9	$F(9)$	$\Delta F(9)$
10	0.725	0.250000
11	0.550	0.628788
12	0.500	0.553030
13	0.425	0.250000
14	0.250	0.628788
15	0.200	0.734848
16	0.185	0.583333
17	0.120	0.719697
18	0.100	-----

### 3. Сжатие и фильтрация изображений с помощью ИНС

Обычно для сжатия используют ИНС прямого распространения с обучением по Хеббу [15] или методом обратного распространения ошибки [10–13, 15, 17–19] и ИНС с обучением по методу векторного квантования (LVQ, Learning Vector Quantization) [16, 19]. Задача фильтрации изображений на ИНС ранее практически не рассматривалась.

#### 3.1. Эксперименты по сжатию и восстановлению изображений

Известные подходы к сжатию изображений на ИНС предполагают разбиение изображений на непересекающиеся блоки 4x4, 8x8 или 16x16 пикселей [10–13, 15–19].

### 3.1.1. Сжатие с применением ИНС прямого распространения

Целью обучения ИНС прямого распространения является получение на выходе изображений, мало отличающихся от входных. Архитектура, применяемая для сжатия изображений, представлена на рис. 4.

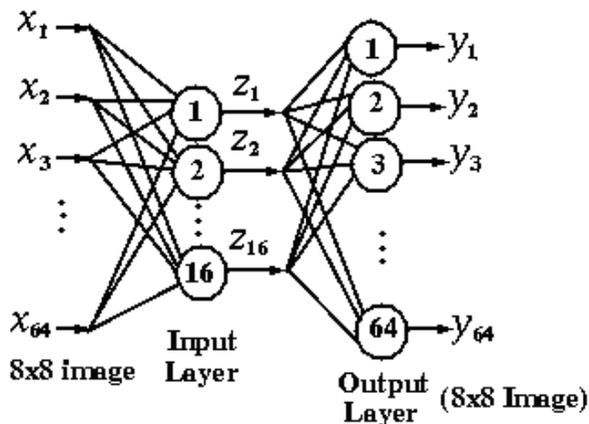


Рис. 4. Архитектура двуслойной ИНС прямого распространения

В качестве функции активации используется сигмоид. Первая половина этой схемы — кодер, находящийся на передающей стороне, а вторая — декодер, который находится на приемной стороне. Исходное изображение восстанавливается (с потерями) с помощью весовых коэффициентов декодера. При этом необходимо хранить лишь выходы кодера для каждого блока и весовые коэффициенты декодера.

Для сжатия были взяты файлы изображений размером 2112 на 2816 пикселей, то есть площадью 5947392 пикселей. Коэффициент сжатия вычислялся по формуле  $G_1 = \frac{O}{4 \cdot (S+B) \cdot M}$ , где  $O$  — площадь сжимаемого изображения,  $M$  — число нейронов,  $S$  — площадь блока изображения, а  $B$  — число блоков в изображении. Для хранения выходов кодера и весовых коэффициентов декодера требуется по 4 байта на каждый элемент данных. Результаты эксперимента представлены в табл. 5, табл. 6 и табл. 7.

Восстанавливаемые изображения обладают хорошим качеством только в случае использования большого числа нейронов кодера, что

ТАБЛИЦА 5. Сжимаемые на ИНС прямого распространения изображения



ТАБЛИЦА 6. Результаты декомпрессии изображений на ИНС прямого распространения

<i>M</i>	<i>S</i>	Восстановленные изображения			
1	4x4				
16	8x8				

сильно снижает коэффициент сжатия и делает данный подход неэффективным.

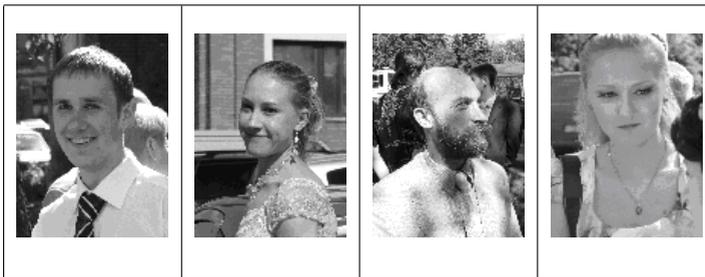
ТАБЛИЦА 7. Коэффициенты сжатия на ИНС прямого распространения

Число нейронов, $M$	Размер блока, $S$	Число блоков, $B$	Коэффициент сжатия, $G_1$
1	4x4	371712	4.00
4	4x4	371712	1.00
8	4x4	371712	0.50
16	4x4	371712	0.25
1	8x8	92928	16.0
4	8x8	92928	4.00
8	8x8	92928	2.00
16	8x8	92928	1.00
1	16x16	23232	63.3
4	16x16	23232	15.83
8	16x16	23232	7.91
16	16x16	23232	3.96

### 3.1.2. Сжатие с применением ИНС Кохонена

Возможность применения сети Кохонена для сжатия изображений основывается на классическом векторном квантовании (vector quantization) — способе кластеризации, в котором пространство входов делится на ряд областей, для каждой из которых определяется вектор восстановления [16, 19].

ТАБЛИЦА 8. Сжимаемые на ИНС Кохонена изображения



Настройка сети подробно описана в работе [16]. Архитектура самоорганизующейся карты признаков (рис. 5) для векторного квантования определяется размером словаря кодовых векторов. Каждый

кодový вектор — это матрица весов в соревновательном слое. В эксперименте 256 кодовых векторов представляют изображение как выход нейронной сети.

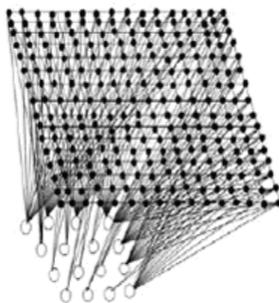


Рис. 5. Самоорганизующаяся карта признаков

Блок, содержащий 16 пикселей, подается на слой Кохонена, состоящий из 256 нейронов, размещенных в двумерном массиве 16x16. Весовые коэффициенты, связывающие  $j$ -ый нейрон слоя Кохонена и входы сети, представлены матрицей  $[\varpi_{ji}]$ ,  $j = 0, \dots, 255$ ,  $i = 0, \dots, 15$ . После того, как словарь создан (ИНС обучена), он может использоваться для сжатия и восстановления изображения по описанным ниже схемам.

*Схема сжатия изображения с помощью словаря:*

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

*Схема восстановления изображения:*

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

ТАБЛИЦА 9. Результаты декомпрессии изображений на ИНС Кохонена (блоки 8x8)

M	Восстановленные изображения			
25				
64				
256				
1024				

ТАБЛИЦА 10. Коэффициенты сжатия на ИНС Кохонена

Размер словаря, $M$	Размер блока, $S$	Число блоков, $B$	Коэффициент сжатия, $G_2$
25	4x4	371712	7.99
64	4x4	371712	7.99
256	4x4	371712	7.96
1024	4x4	371712	7.83
25	8x8	92928	31.73
64	8x8	92928	31.31
256	8x8	92928	29.41
1024	8x8	92928	23.66
25	16x16	23232	112.50
64	16x16	23232	94.63
256	16x16	23232	53.10
1024	16x16	23232	19.27

Для сжатия использовались файлы изображений размером 2112 на 2816 пикселей, то есть площадью 5947392 пикселей. Формула расчета коэффициента сжатия:  $G_2 = \frac{O}{M \cdot S + 2 \cdot B}$ , где  $O$  — площадь сжимаемого изображения,  $M$  — размер словаря,  $S$  — длина кодовых векторов (площадь блока), а  $B$  — число блоков в изображении. Для хранения каждого индекса блока требуется два байта. Результаты эксперимента представлены в табл. 8, табл. 9 и табл. 10.

Эксперимент показывает, что восстановленные изображения обладают высоким качеством даже в случае использования словаря малого размера.

### 3.2. Эксперименты с фильтрацией изображений

Задача заключается в построении ИНС, удаляющей шумы на известных ей изображениях с помощью системы кодер/декодер [12, 19] по аналогии с операцией шумоподавления, которая сводится к вычитанию спектра шума из спектра изображения. Важно отметить, что ИНС обучается на всех изображениях эталонной выборки. Обученная сеть стремится для каждого входного изображения сформировать выход декодера так, чтобы он соответствовал наиболее близкому эталону.

Особенность предложенного подхода — использование наборов сетей-экспертов, каждая из которых представляет комбинацию кодирующего и декодирующего слоев. Входные изображения разбиваются на области, для чего применяется кольцо вычетов. Основание кольца  $N$ , отвечающее за число классов кольца, выбирается исходя из характера сжимаемых изображений. Для каждого класса кольца создается своя нейронная сеть, ответственная за кодирование и декодирование тех пикселей, чьи номера имеют одинаковый остаток при делении на  $N$ . Для примера приведем матрицу изображения размером  $7 \times 7$  пикселей. Пусть  $N = 3$ . Обозначим двоичными кодами 00, 01 и 10 подобласти изображения (классы). В соответствии с выбранным основанием кольца получим табл. 11.

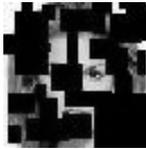
ТАБЛИЦА 11. Формирование областей на изображении

00	01	10	00	01	10	00
01	10	00	01	10	00	01
10	00	01	10	00	01	10
00	01	10	00	01	10	00
01	10	00	01	10	00	01
10	00	01	10	00	01	10
00	01	10	00	01	10	00

Для обучения нейронной сети использовались 13 полутоновых изображений размером  $64$  на  $64$  пикселя ( $4096$  признаков). Число нейронов входного слоя (кодера) взято равным  $32$ . Обучение ИНС велось до тех пор, пока среднеквадратичное отклонение от желаемых выходов декодера не становилось меньшим  $0.01$  для каждой из полученных областей. Эксперимент демонстрирует возможности восстановления поврежденных изображений на кольце вычетов по модулям  $3$  и  $4$  (табл. 12).

Видно, что при  $N = 4$  возникает эффект появления «полос». Это связано с тем, что области, по которым обучались нейронные сети, не пересекались между собой. В случаях, когда шумы заполняют изображение более чем на  $50\%$ , нейронная сеть пытается объединить несколько известных ей образов в один.

ТАБЛИЦА 12. Фильтрация зашумленных изображений

Оригинальное изображение	Кодируемое изображение	Результат декодирования (по модулю 3)	Результат декодирования (по модулю 4)
			
			
			
			
			

#### 4. Заключение

Эксперименты показали, что трехслойные ИНС прямого распространения способны решать задачи прогнозирования и интерполяции временного ряда с высокой точностью. Алгоритмы сжатия, основанные на использовании аппарата сети Кохонена, вполне пригодны для практического применения. В ходе полученных экспериментов удалось получить реальные степени сжатия изображений порядка 63 на сети прямого распространения и 112 на сети Кохонена. Предложенный механизм удаления шумов обладает высокой эффективностью, т. к. позволяет восстанавливать изображения даже при 50% заполнении шумами с заменой поврежденных областей на полноценные.

#### Список литературы

- [1] Дубынин И.А. О некоторых теоретических и практических аспектах применения искусственных нейронных сетей для обработки психологических данных. — Лаборатория математических моделей в психологии и педагогике МГППУ, 2006, Эл. ресурс: <http://www.matlab.mgppu.ru/work/0027.htm>.
- [2] Пауков Д. Прогнозирование с помощью искусственных нейронных сетей, 2001, Эл. ресурс: <http://paukoff.fromru.com/neuro/wneuro/index.html>.
- [3] Мицель А.А., Ефремова Е.А. Прогнозирование динамики цен на фондовом рынке. — Известия Томского политехнического университета. — Т. 309, № 8, 2006. — 197 — 201 с.
- [4] Шумилова Г.П., Готман Н.Э., Старцева Т.Б. Краткосрочное прогнозирование электрических нагрузок с использованием искусственных нейронных сетей. — Электричество, № 10, 1999.
- [5] Кириченко Л.М. Статистика промышленности. Обзор и анализ статистических методов прогнозирования. — Сборник трудов магистров Донецкого Национального Технического Университета, 2005.
- [6] Панфилов П. Прогнозирование курсов валют на рынке Forex. — Современный трейдинг, № 1, 2001. — 16 — 18 с.
- [7] Сидоркин К.В., Костюхин М.Н. Прогнозирование на основе аппарата нейронных сетей (дипломная работа), Эл. ресурс: <http://www.compref.ru/refs/121/1>.
- [8] Глазунова А.М. Повышение точности результатов нейросетевого прогнозирования нагрузки. — Институт систем энергетики им. Мелентьева СО РАН, г. Иркутск, Эл. ресурс: <http://www.anares.ru/DOC/Seminar3/a3isem2.doc>.
- [9] Бережная Е.В., Порохня Т.А. Методы прогнозирования временных рядов в экономических исследованиях. — Вестник СевКавГТУ. Экономика, № 2(13), 2004, Эл. ресурс: [http://science.ncstu.ru/articles/econom/13/09.pdf/file\\_download](http://science.ncstu.ru/articles/econom/13/09.pdf/file_download).

- [10] Verma B., Blumenstein M., Kulkarni S. A New Compression Technique Using an Artificial Neural Network. — Journal of Intelligent Systems, № 8, 1999. — 39 — 53 с. (english)
- [11] Verma B., Blumenstein M., Kulkarni S. Image Compression using a Direct Solution Method based Neural Network. — Proceedings of the Tenth Australian Joint Conference on Artificial Intelligence, Perth, Australia, 1997. — 114 — 119 с. (english)
- [12] 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, 1997. — 12 — 16 с. (english)
- [13] Vilovic I. An Experience in Image Compression Using Neural Networks. — 48th International Symposium ELMAR-2006, 07-09 June 2006, Zadar, Croatia, 2006. — 95 — 98 с. (english)
- [14] Mahoney M. The PAQ1 Data Compression Program, 2002, Эл. ресурс: [http://www.compression.ru/download/articles/cm/mahoney\\_2002\\_paq1.pdf](http://www.compression.ru/download/articles/cm/mahoney_2002_paq1.pdf). (english)
- [15] Стремоухов В.Д. Practical usage of artificial neural networks for tasks of classification and image compression. — Научно-технический вестник Санкт-Петербургского государственного университета информационных технологий, механики и оптики. — Т. 27, 2007. — 122 — 128 с., Эл. ресурс: [http://vestnik.ifmo.ru/ntv/27/ntv\\_27.2.6.pdf](http://vestnik.ifmo.ru/ntv/27/ntv_27.2.6.pdf). (english)
- [16] Durai A.S., Saro A.E., Phil M. An Improved Image Compression approach with Self-Organizing Feature Maps using Cumulative Distribution Function. — GVIP Journal. — Т. 6, 2006. — 122 — 128 с. (english)
- [17] Durai A.S., Saro A.E., Phil M. Image Compression with Back-Propagation Neural Network using Cumulative Distribution Function. — Proceedings of World Academy of Science, Engineering and Tchnology. — Т. 17, 2006. — 60 — 64 с. (english)
- [18] Watta P., Desaie B., Dannug N., Hassoun M. Image Compression using Backprop. — Proceedings of the IEEE, 1996, Эл. ресурс: <http://neuron.eng.wayne.edu/bpImageCompression9PLUS/bp9PLUS.html>. (english)
- [19] Dony R., Haykin S. Neural Network Approaches to Image Compression. — Proceedings of the IEEE. — Т. 2, 1995. — 288 — 303 с., Эл. ресурс: <http://neuron.eng.wayne.edu/bpImageCompression9PLUS/bp9PLUS.html>. (english)

V. P. Fralenko. *Data prediction, compression and filtering using neural networks* // Proceedings of Program Systems institute scientific-practical conference “Program systems: Theory and applications”, devoted to the 15<sup>th</sup> anniversary of Pereslavl University named A. K. Ailamazyan. — Pereslavl-Zalesskij, 2008. — p. 235 — 252. — ISBN 978-5-901795-13-2 (*in Russian*).

ABSTRACT. The paper presents some outcomes of using neural network for time series prediction, lossy image compression and image filtering. It have been proved that a feedforward network may be used for time series prediction with paralleling techniques. The feedforward and Kohonen networks have been used for direct image compression. More correct technique of compression ratio accounting is executed. Some new schemes of image filtering have been suggested and tested and their efficiency while deleting patters on strongly noised images have been proved.

*Перевод проверен:* ст. преп. Н. А. Прохорова