

УДК 004.032.26

Ю.Л. ИВАСЬКИВ, В.В. ЛЕВЧЕНКО, О.Л. ЛЕЩИНСКИЙ

ФОРМИРОВАНИЕ НЕЧЕТКИХ ОБУЧАЮЩИХ МНОЖЕСТВ ДЛЯ НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧАХ СЖАТИЯ ДАННЫХ БЕЗ ПОТЕРЬ

Abstract: The features of training and functioning neural networks in problem of lossless data compression in the stream mode are considered. The method of forming of training sets using fuzzy presentation of output values is offered. The experimental analysis of influence of kind of membership function of fuzzy output value on the training process of network is executed.

Key words: training set, fuzzy data presentation, membership-function, alpha-cut.

Анотація: Розглянуті особливості навчання та функціонування нейронних мереж при розв'язанні задачі стиснення даних без втрат у потоковому режимі. Запропоновано метод формування навчальних множин з використанням нечіткого представлення вихідних значень. Виконано експериментальний аналіз впливу виду функції приналежності нечіткого вихідного значення на процес навчання мережі.

Ключові слова: навчальна множина, нечітке представлення даних, функція приналежності, альфа-зріз.

Аннотация: Рассмотрены особенности обучения и функционирования нейронных сетей при решении задачи сжатия данных без потерь в потоковом режиме. Предложен метод формирования обучающих множеств с использованием нечеткого представления выходных значений. Выполнен экспериментальный анализ влияния вида функции принадлежности нечеткого выходного значения на процесс обучения сети.

Ключевые слова: обучающее множество, нечеткое представление данных, функция принадлежности, альфа-срез.

1. Введение

Специфические свойства искусственных нейронных сетей (ИНС) позволяют использовать их при решении различных задач обработки данных в цифровых системах [1–3]. К числу таких свойств, в частности, относится возможность адаптации сети к обрабатываемым данным. Свойство адаптации позволяет использовать ИНС для обработки данных различной природы без изменения архитектуры сети в так называемом потоковом режиме. Такой режим характеризуется тем, что данные поступают на входы ИНС последовательно во времени и их дальнейшее использование после обработки оказывается невозможным.

Потоковый режим применяется в цифровых системах, реализующих различные функции обработки в режиме реального времени, например, фильтрацию, модуляцию, шифрование, предиктивное кодирование, дискретные преобразования, сжатие данных [3–7]. Нейросетевые методы реализации перечисленных функций обработки отличаются высокой степенью параллелизма вычислительного процесса, надежностью и устойчивостью [4–10]. Известные методы предполагают использование одного из следующих подходов к построению ИНС, выполняющей целевую функцию обработки:

1) структура ИНС разрабатывается применительно к конкретному виду реализуемой функции обработки, а численные значения параметров ИНС задаются априорно, исходя из условия решаемой задачи;

2) разрабатываемая структура ИНС не предусматривает ориентацию на конкретную функцию обработки, при этом численные значения параметров ИНС подбираются в процессе настройки (обучения) системы.

Подход, ориентированный на разработку структуры ИНС применительно к конкретному виду реализуемой функции обработки, позволяет строить наиболее экономные и эффективные с точки зрения используемых вычислительных ресурсов системы, однако, такие системы лишены свойства настраиваемости на решаемую задачу.

Подход, не предусматривающий ориентацию разрабатываемой структуры ИНС на конкретную функцию обработки, открывает перспективы построения систем, способных настраиваться на решение различных классов задач. Однако, несмотря на возможность получения определенных преимуществ, при его практическом применении возникает ряд трудностей, связанных в первую очередь со сложностью учета свойств обрабатываемых данных различной природы. Для учета таких свойств при использовании ИНС в задачах сжатия данных без потерь требуется разработка специальных методов и подходов к формированию обучающего множества.

2. Постановка задачи

Применение ИНС для сжатия данных без потерь в цифровых системах основывается на использовании предиктивных свойств нейронных сетей [14]. Предиктивные свойства ИНС широко используются при прогнозировании временных рядов, когда нейронная сеть обучается на множестве, составленном из значений временного ряда, погруженного в лаговое пространство [14–16].

В работах, посвященных нейросетевой реализации цифровых систем сжатия данных без потерь, описывается также использование ИНС для статистического моделирования источника сжимаемых данных [13, 17]. В процессе моделирования ИНС функционирует в режиме предиктора, когда на основе предсказаний строится распределение вероятностей появления различных значений в потоке сжимаемых данных.

Несмотря на сходство нейросетевых методов прогнозирования и моделирования, между ними существуют и принципиальные отличия, требующие применения разных подходов при их реализации. Сходство проявляется, в частности, в том, что как при прогнозировании, так и при моделировании используются одна архитектура и алгоритм обучения ИНС. Вместе с тем, при прогнозировании временных рядов существует возможность предварительного обучения ИНС на всем множестве значений ряда с требуемой точностью, тогда как при моделировании обучение ИНС должно быть совмещено с функционированием и выполняться на потоке обрабатываемых данных. Кроме того, при прогнозировании временных рядов входные и выходные значения ИНС являются номинальными, что допускает наличие в них некоторой погрешности, в то время как при моделировании дискретные выходные значения являются, в общем случае, категориальными, что порождает специальные требования к ошибке предсказания [18].

Таким образом, при использовании ИНС в цифровых системах сжатия данных без потерь возникает необходимость в разработке таких методов и способов формирования обучающего множества, которые учитывают свойства обрабатываемых данных и обеспечивают адаптацию к ним обрабатывающей системы в процессе ее функционирования.

3. Основные результаты

При решении задачи прогнозирования временных рядов адаптация ИНС к обрабатываемым

данным может быть достигнута в результате применения предварительной обработки обучающего множества (устранение противоречивых примеров, нормирование входных и выходных значений, использование специальных способов представления данных и т.п.) либо за счет подбора параметров обучающего процесса (например, количества нейронов, типа функции активации, коэффициента скорости обучения, момента обучения).

Предварительная обработка обучающего множества, как правило, предполагает возможность доступа к любому его элементу в произвольный момент времени, что противоречит принципу потоковой обработки данных. Поэтому в цифровых системах сжатия данных предлагается ограничить предварительную обработку данных знаковым представлением входных значений [17, 18], что, в конечном счете, приводит к упрощению процесса формирования входных векторов ИНС.

Однако решение задачи адаптации сети к обрабатываемым данным не ограничивается их предварительной обработкой. Дело в том, что в процессе сжатия данных без потерь обучающее множество в каждый момент времени состоит из пары “входной вектор – выходной вектор”. При этом входным вектором является двоичный контекст, а целевым выходным является значение, рассматриваемое как прогнозируемое (рис. 1).

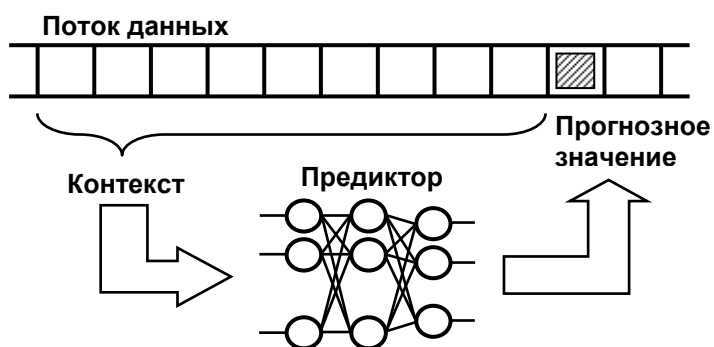


Рис. 1. Схема функционирования ИНС в предиктивном режиме

Такое обучающее множество не может считаться достаточно информативным, поскольку не дает представления о виде многомерной функции, интерполируемой нейронной сетью в процессе обучения и требует соответствующего расширения. Расширение можно было бы получить за счет, например, применения оконного режима обучения [18]. В оконном режиме обучающее множество формируется из значений, которые предшествуют обрабатываемому с некоторой задержкой во времени (рис. 2). Подтверждением эффективности применения оконного режима могут служить выполненные в ходе проводимых исследований эксперименты, которые показали, что такой режим позволяет увеличить количество правильных прогнозов в зависимости от типа обрабатываемых данных более чем на 15 %.

Оконный режим, однако, не позволяет учитывать тот факт, что на категориальных двоичных данных ИНС фактически обучается классификации обрабатываемых контекстов. Известно, что для успешной классификации обучающее множество должно включать как положительные, так и отрицательные примеры для каждого класса [19]. Поэтому предлагается принципиально иной подход к формированию обучающего множества: входной вектор ИНС разделяется на статическую и динамическую части, причем статическая часть представляет собой фиксированный контекст,

предшествующий прогнозируемому элементу данных, а динамическая часть – сам прогнозируемый элемент (рис. 3).

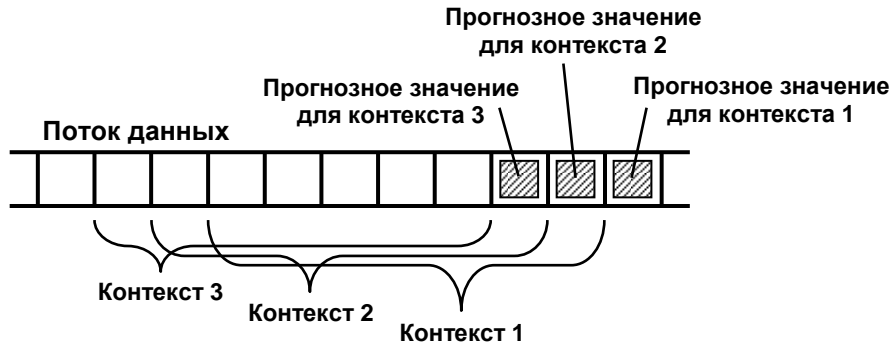


Рис. 2. Расположение контекстов в потоке данных

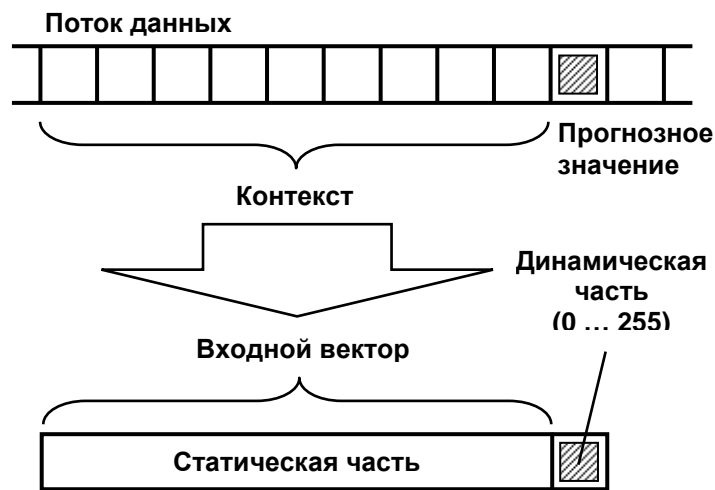


Рис. 3. Схема формирования нечеткого обучающего множества

При байтовом представлении элемент данных, поступающий на вход ИНС, может принимать $2^8 = 256$ значений, из которых лишь одно соответствует прогнозируемому. Значение этого элемента данных предлагается рассматривать как нечеткое число (НЧ). Для оценки меры сходства каждого из 256 входных значений с прогнозируемым вводится функция принадлежности $\mu(x)$, где x – некоторая мера удаленности рассматриваемого значения от прогнозируемого. Задачей обучения ИНС является аппроксимация функции принадлежности $\mu(x)$ для каждого обрабатываемого элемента, поступающего в потоке данных.

В процессе компьютерного моделирования были использованы треугольная, экспоненциальная, параболическая и вырожденная толерантная плоская функции принадлежности [20]. Наиболее простой из перечисленных является треугольная функция принадлежности, применяемая в задачах прогнозирования финансовых временных рядов [21]:

$$\mu(h) = 1 - \frac{h}{8}, \quad (1)$$

где h – расстояние Хемминга между рассматриваемым и правильным векторами, бит.

Особенностью треугольной функции (1) является то, что ее мода совпадает с левой

границей. График такой функции представлен на рис. 4.

Экспоненциальная функция задается выражением

$$\mu(h) = e^{-kh}, \quad (2)$$

где h – расстояние Хемминга между рассматриваемым и правильным векторами, бит;

k – коэффициент крутизны, который выбирается из диапазона (0; 1].

Такая функция характеризуется наличием экспоненциальной кривизны, выделяющей модальное значение (рис. 5).

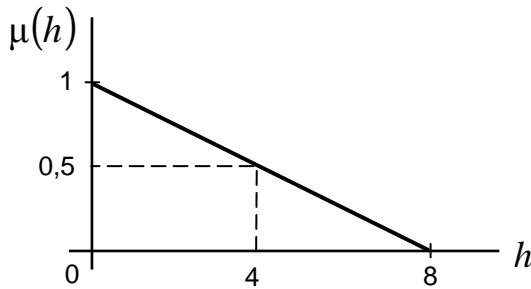


Рис. 4. График треугольной функции принадлежности (1)

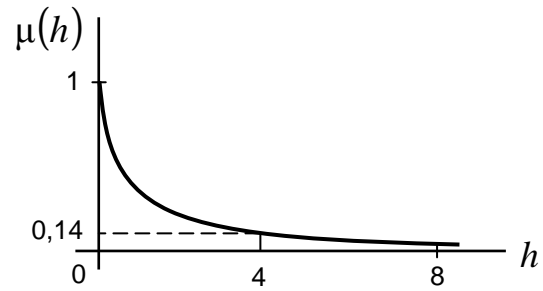


Рис. 5. График экспоненциальной функции принадлежности (2) с коэффициентом крутизны $k = 0,5$

В качестве аргумента для функций принадлежности (1) и (2) было использовано расстояние Хемминга между рассматриваемым и правильным значениями. Пример обучающего множества, составленного для байта со значением 65_{16} с использованием нелинейной функции принадлежности при $k = 0,5$, представлен в табл. 1.

Таблица 1. Зависимость значения экспоненциальной функции принадлежности от расстояния Хемминга

Расстояние Хемминга – h , бит	Количество различных байтовых значений – C_8^h	Значение функции принадлежности
0	1	1,000
1	8	0,607
2	28	0,368
3	56	0,223
4	70	0,135
5	56	0,082
6	28	0,050
7	8	0,030
8	1	0,018

Вырожденная толерантная плоская и параболическая функции принадлежности задаются выражениями (3) и (4) соответственно. Эти функции в качестве аргумента используют модуль разности между рассматриваемым и правильным векторами. Графики функций (3) и (4) представлены на рис. 6, 7 соответственно.

$$\mu(x) = 1 - \sqrt{\frac{|x-b|}{256}}, \quad (3)$$

$$\mu(x) = 1 - \left(\frac{|x - b|}{256} \right)^2, \quad (4)$$

где x – динамическая часть рассматриваемого вектора;

b – динамическая часть правильного вектора.

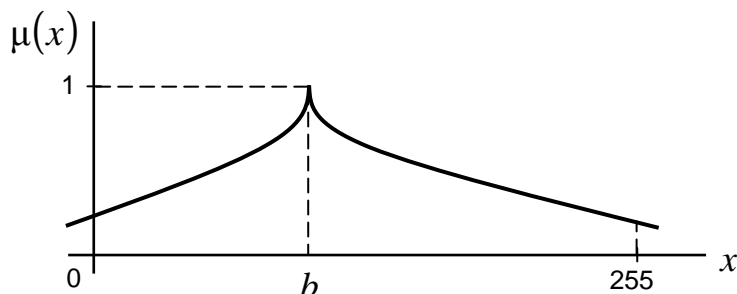


Рис. 6. График вырожденной толерантной плоской функции принадлежности (3)

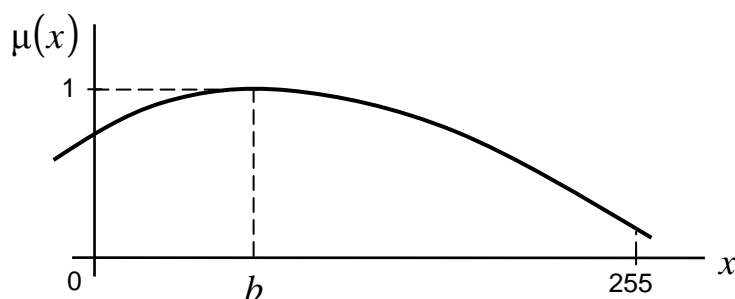


Рис. 7. График параболической функции принадлежности (4)

Оценка предиктивных возможностей ИНС производилась на файлах из набора Calgary Corpus Test, доступного в сети Internet и предназначенного для тестирования архиваторов. Для моделирования использовалась ИНС со следующими экспериментально определенными характеристиками:

- количество слоев – 2;
- количество нейронов в скрытом слое – 24 ... 32;
- функция принадлежности нейронов скрытого слоя – сигмоидная.

Быстрая сходимость обучающего процесса обеспечивалась моментом обучения, равным 0,9. Наилучшие результаты с точки зрения относительного количества правильных прогнозов были получены при обработке цифровых данных категориального типа (тексты на естественных и искусственных языках) с использованием экспоненциальной функции принадлежности. При такой функции принадлежности на протяжении всего процесса обучения наблюдалось стабильное увеличение процента правильных прогнозов, выполненных нейросетевым предиктором. Типичный график зависимости процента правильных прогнозов от обрабатываемого контекста представлен на рис. 8. В процессе обучения и функционирования сети среднее расстояние Хемминга между прогнозными и реальными значениями на выборке данных объемом до 10 % составило около 3,5 бит, а средняя скорость его снижения – около 0,5 бит на килобайт обрабатываемых данных, что позволяет достигнуть приемлемого соотношения верных и неверных прогнозов при решении широкого круга прикладных задач.

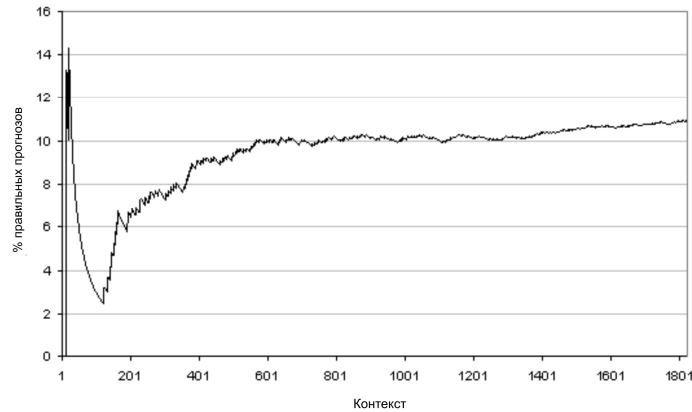


Рис. 8. Зависимость количества правильных прогнозов от номера обрабатываемого контекста

При использовании функций (3) и (4) в процессе обработки данных различного вида было установлено, что более 75 % прогнозных значений, сформированных нейросетевым предиктором, попадают в область альфа-среза на уровне 0,7 (рис. 9). Такой результат позволяет значительно уменьшить диапазон рассматриваемых прогнозных значений при построении вероятностной модели источника данных в процессе сжатия данных без потерь.

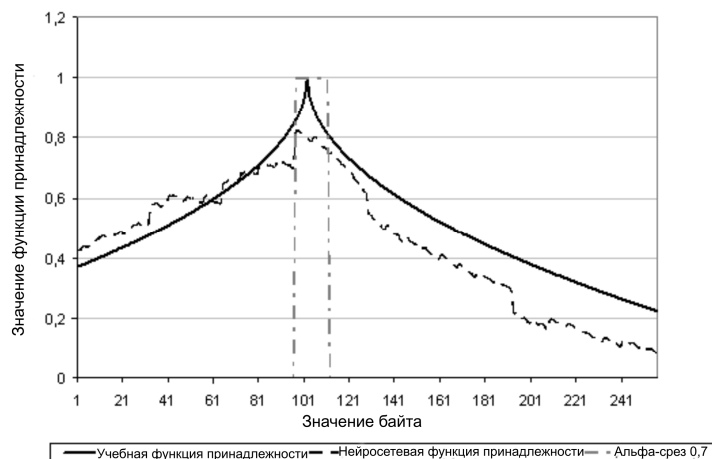


Рис. 9. График функции принадлежности, смоделированной на ИНС

4. Выводы

Для обработки данных различного типа в потоковом режиме при решении задачи сжатия данных без потерь предлагается использовать многослойную нейронную сеть, обучаемую по модифицированному (с учетом момента) алгоритму обратного распространения ошибки.

Возникающую необходимость предобработки входных данных предлагается выполнять за счет знакового представления двоичных значений, поступающих на входы сети.

Обучающее множество для ИНС в задачах сжатия данных без потерь предлагается формировать с использованием нечеткого представления прогнозных значений, а целью обучения ИНС считать аппроксимацию функции принадлежности нечеткого числа, характеризующего прогнозируемое значение.

Наибольшее количество правильных прогнозов, полученных на ИНС, достигается при использовании экспоненциальной функции принадлежности прогнозного значения.

При использовании параболической и вырожденной толерантной плоской функций принадлежности наибольшее количество прогнозов, полученных на ИНС, попадают в альфа-срез на уровне 0,7, что позволяет с достаточной для практического использования точностью оценивать вероятности появления байтовых значений в потоке данных.

Роботу виконано за підтримки Державного фонду фундаментальних досліджень Міністерства освіти і науки України. Договір Ф25/129-2008 від 14.05.2008.

СПИСОК ЛИТЕРАТУРЫ

1. Нейроматематика. – Кн. 6: Учеб. пособие для вузов / А.Д. Агеев, А.В. Балухто и др.; Под общ. ред. А.И. Галушкина. – М.: ИПРЖР, 2002. – 448 с.
2. Осовский С. Нейронные сети для обработки информации / Пер. с польск. Н.Д. Рудинского. – М.: Финансы и статистика, 2004. – 344 с.
3. Милов В.Р. Адаптивная обработка сигналов на основе нейронных RBF-сетей // Материалы 5-й Междунар. конф. “Цифровая обработка сигналов и ее применение”. – 2003. – № 2. – С. 125 – 128.
4. Хрящев В.В. Линейная цифровая фильтрация на клеточных нейронных сетях // Материалы 4-й Междунар. конф. “Цифровая обработка сигналов и ее применение”. – 2002. – № 2. – С. 43 – 46.
5. Хрящев В.В. и др. Нейросетевое восстановление амплитуды дискретного сигнала по его фазовому спектру / В.В. Хрящев, Е.А. Соколенко, А.Л. Приоров // Материалы 5-й Междунар. конф. “Цифровая обработка сигналов и ее применение”. – 2003. – № 2. – С. 622 – 624.
6. Ланнэ А.А. Нейронные цепи и синтез нелинейных операторов обработки сигналов // Материалы 4-й Междунар. конф. “Цифровая обработка сигналов и ее применение”. – 2002. – № 1. – С. 34 – 42.
7. Daniel A. Jimenez. Fast Path-Based Neural Branch Prediction // Proc. of the 36th Annual IEEE/ACM International Symposium on Microarchitecture. – 2003 – P. 243 – 252.
8. Verma B. et al. A New Compression Technique Using an Artificial Neural Network / B. Verma, M. Blumenstein, S. Kulkarni // Journal of Intelligent Systems. – 1999. – Vol. 9. – P. 39 – 53.
9. Балухто А.Н., Галушкин А.Н. Нейрокомпьютерная реализация сжатия изображений на основе преобразований // Материалы 3-й Междунар. конф. “Цифровая обработка сигналов и ее применение”. – 2000. – № 3. – С. 14 – 15.
10. Скрибцов П. Методы оптимизации программ, эмулирующие нейронные сети на Intel-совместимых процессорах с использованием технологии SIMD (Single Instruction Multiple Data) // Материалы 5-й Междунар. конф. “Цифровая обработка сигналов и ее применение”. – 2003. – № 2. – С. 320 – 324.
11. Cottrell G.W. et al. Image compression by back propagation: An example of extentional programming / G.W. Cottrell, P. Munro, D. Zipser // Proc. 9th Annual Confer., Cognitive Soc. – 1987. – P. 461 – 473.
12. Иваськів Ю.Л., Левченко В.В. Вопросы использования репликативных нейронных сетей в задачах сжатия изображений // Математичні машини і системи. – 2006. – № 4. – С. 42 – 50.
13. Schmidhuber J. Sequential neural text compression // IEEE transaction on Neural Networks. – 1996. – Vol. 7. – P. 142 – 146.
14. Ежов А.А., Шумский С.А. Нейрокомпьютеринг и его применение в экономике и бизнесе (серия “Учебники экономико-аналитического института МИФИ”) / Под ред. проф. В.В. Харитонова. – М.: МИФИ, 1998. – 244 с.
15. Малинецкий Г.Г., Потапов А.Б. Современные проблемы нелинейной динамики. – М.: Эдиториал УРСС, 2000. – 336 с.
16. Забалдіна Ю.Б. и др. Дві адаптивні моделі короткострокового прогнозування туристичних потоків / Ю.Б. Забалдіна, В.В. Левченко, О.Л. Лещинський // Стратегія розвитку України (економіка, соціологія, право): Наук. журнал. – Вип. 1 / Гол. ред. О.П. Степанов. – К.: Книжкове вид-во НАУ, 2006. – С. 404 – 410.
17. Иваськів Ю.Л. и др. Обучение нейронных сетей в задачах сжатия данных без потерь / Ю.Л. Иваськів, О.Л. Лещинський, В.В. Левченко // Матеріали VIII-ї Міжнар. науково-технічної конф. “Авіа-2007”. – 2007. – 25 – 27 квітня. – Т. 1. – С. 13.57 – 13.60.
18. Ivaskiv Yu., Levchenko V. Forming of learning set for neural networks in problems of losless data compression // XIII-th International Conf. “Knowledge – Dialogue – Solution”. – Varna, 2007. – Vol. 1. – P. 55 – 59.
19. Царегородцев В.Г. Оптимизация предобработки данных для обучаемой нейросети: критерии оптимальности предобработки // XIV Междунар. конф. по нейрокибернетике. – Ростов-на-Дону, 2005. – Т. 2. – С. 64 – 67.
20. Нечеткие множества в моделях управления и искусственного интеллекта / Под ред. Д.А. Поспелова – М.: Наука, 1986. – 312 с.
21. Горбачова О.М. и др. Прогнозування фінансових часових рядів, представлених послідовностями трикутних нечітких чисел на нейронних мережах / О.М. Горбачова, О.Л. Лещинський, В.В. Левченко // Стратегія розвитку України (економіка, соціологія, право): Наук. журнал. – Вип. 1 / Гол. ред. О.П. Степанов. – К.: Книжкове вид-во НАУ, 2007. – С. 487 – 492.

Стаття надійшла до редакції 15.10.2008