

ОРТОСИНАПС, ОРТОНЕЙРОНЫ И НЕЙРОПРЕДИКТОР НА ИХ ОСНОВЕ

Введена специализированная конструкция нелинейного синапса с активационными функциями в виде ортогональных полиномов и на его основе архитектуры ортонейрона и двойного ортонейрона. Предложены алгоритмы обучения на основе процедур оптимизации первого и второго порядков. Рассмотрена и решена задача прогнозирования с помощью развиваемого подхода.

искусственные нейронные сети, ортогональные полиномы, ортонормированные полиномы, ортосинапс, ортонейрон, двойной ортонейрон

Введение

В настоящее время при решении проблем, связанных с обработкой и анализом нестационарных нелинейных сигналов произвольной природы в условиях априорной и текущей неопределённости, широко применяются методы вычислительного интеллекта и, прежде всего, искусственные нейронные сети и их гибриды типа нейро-фаззи-, взвешенно-нейро, нео-фаззи-систем. Как правило, это многослойные архитектуры (рекуррентные или с прямой передачей информации), узлы которых в качестве активационных используют либо сигмоидальные, либо колоколообразные функции. Одной из важных задач, связанных с обработкой сигналов и процессов, является прогнозирование и эмуляция состояния динамических систем в будущие моменты времени, с математической точки зрения сводящаяся к аппроксимации функций произвольного вида по экспериментальным «зашумлённым» данным.

Среди множества функциональных структур, используемых для аппроксимации нелинейных процессов, особого внимания заслуживают ортогональные полиномы [1], обладающие рядом привлекательных свойств, как с вычислительной точки зрения, так и с позиции обеспечиваемой ими точности. При этом одномерная система, описываемая в пространстве «вход-выход» некоторой неизвестной функциональной зависимости $y(x)$, сколь угодно точно может быть представлена суммой [1]

$$f(x) = w_0\phi_0(x) + w_1\phi_1(x) + \dots + w_h\phi_h(x) = \sum_{j=0}^h w_j\phi_j(x), \tag{1}$$

где x и $f(x)$ – входная и оцениваемая выходная переменные нелинейного процесса соответственно; $\phi_j(x)$ – ортогональный полином порядка j ($j = 0, 1, 2, \dots, h$), обладающий свойством ортогональности

$$\sum_{k=1}^N \phi_j(x(k))\phi_q(x(k)) = 0, \forall j \neq q; \tag{2}$$

j, k – неотрицательные целые числа; $k = 1, 2, \dots, N$ – текущее дискретное время или номер наблюдения в выборке.

Уравнение (1) может быть реализовано с помощью элементарной схемы, приведённой на рис. 1 и названной нами ортосинапсом OS_i .

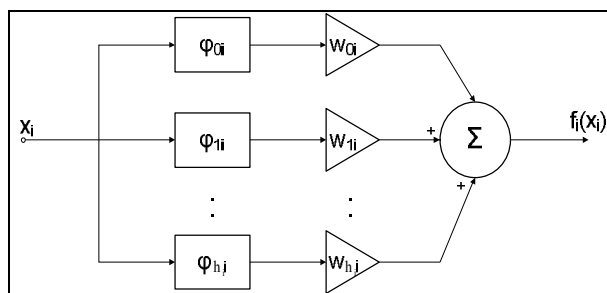


Рис. 1. Ортосинапс – OS_i

Здесь x_i представляет собой i -ю ($i = 1, 2, \dots, n$) компоненту многомерного сигнала $x = (x_1, x_2, \dots, x_n)^T$, w_{ji} ($j = 0, 1, 2, \dots, h_i$) синаптические веса, подлежащие определению, а выходной сигнал ортосинапса может быть записан в виде

$$f_i(x) = \sum_{j=0}^{h_i} w_{ji}\phi_{ji}(x_i). \tag{3}$$

Несложно заметить, что OS_i структурно совпадает с нелинейным синапсом нео-фаззи-нейрона [2 – 4], однако, вместо кусочно-линейной аппроксимации, обеспечиваемой треугольными функциями принадлежности нео-фаззи-нейрона, реализует более гладкое полиномиальное приближение на основе ортогональных функций.

Ортосинапс (3) является «строительным» блоком конструкции, названной нами ортонейроном ON и представленной на рис. 2.

Ортонейрон, структурно совпадающий с нео-фаззи-нейроном [2 – 4], реализует отображение вида

$$u = \sum_{i=1}^n f_i(x_i) = \sum_{i=1}^n \sum_{j=0}^{h_i} w_{ji}\phi_{ji}(x_i)$$

и обеспечивает высокое качество аппроксимации существенно нелинейных нестационарных сигналов и процессов [5 – 8]. В задаче прогнозирования как стохастических, так и хаотических процессов ортонейрон показал высокую точность предсказания и скорость сходимости процесса обучения [9].

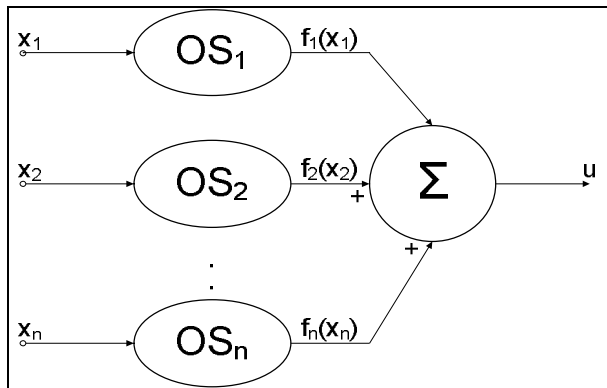


Рис. 2. Ортонейрон – ON

Постановка задачи. Целью настоящей работы является улучшение аппроксимирующих свойств ортонейрона и повышение точности прогнозирования в условиях неопределённости о свойствах предсказываемого процесса и действующих на него возмущений.

Введём в рассмотрение архитектуру двойного ортонейрона DON, представленную на рис. 3 и совпадающую по сути со структурой двойного взвешенного нейрона [10], однако отличающуюся от него используемыми функциями активации, что обеспечивает ему, как будет показано далее, ряд новых полезных свойств.

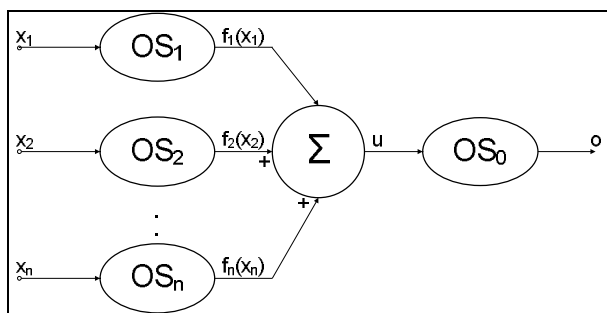


Рис. 3. Двойной ортонейрон – DON

При подаче на вход DON векторного сигнала $x = (x_1, x_2, \dots, x_n)^T$ на его выходе появляется скалярный сигнал

$$o = f_0(u) = \sum_{l=0}^{h_0} w_{l0} \varphi_{l0}(u) = \sum_{l=0}^{h_0} w_{l0} \varphi_{l0} \left(\sum_{i=1}^n \sum_{j=0}^{h_i} w_{ji} \varphi_{ji}(x_i) \right),$$

определяемый как синаптическими весами w_{ji} , w_{l0} , так и значениями ортогональных функций активации $\varphi_{ji}(x_i)$, $\varphi_{l0}(u)$. Как видно, двойной ортонейрон состоит из двух слоёв: скрытого слоя, содержа-

щего n ортосинапсов OS_i по $h_i + 1$ активационных функций в каждом, и выходного слоя, состоящего из одного ортосинапса OS_0 и содержащего $h_0 + 1$ активационную функцию.

В каждом ортосинапсе могут быть реализованы различные ортогональные функции [11] такие, как полиномы Чебышева, Эрмита, Лагерра, Лежандра и др. При этом в разных ортосинапсах могут быть использованы разные активационные функции. Так, например, если входные данные предварительно нормированы так, что $-1 \leq x_i \leq 1$, то в скрытом слое целесообразно использовать систему полиномов Лежандра, ортогональных на отрезке $[-1, 1]$ с весом равным единице,

$$P_j(x_i) = 2^{-j} \sum_{p=0}^{[j/2]} (-1)^p \frac{(2j-2p)!}{p!(j-p)!(j-2p)!} x_i^{j-2p}$$

(здесь $[.]$ – целая часть числа), при этом для упрощения расчетов можно воспользоваться рекуррентной формулой

$$P_{j+1}(x_i) = \frac{2j+1}{j+1} x_i P_j(x_i) - \frac{j}{j+1} P_{j-1}(x_i).$$

В выходном же слое, поскольку интервал варьирования переменной заранее определить невозможно, можно воспользоваться системой полиномов Эрмита, ортогональных на интервале $[-\infty, \infty]$ с весом e^{-u^2} :

$$H_l(u) = l! \sum_{p=1}^{[l/2]} (-1)^p \frac{(2u)^{l-2p}}{p!(l-2p)!}.$$

Такой прием исключает необходимость нормирования сигналов в скрытом слое по ходу процесса обучения и обеспечивает защиту от паралича, достаточно часто возникающего при настройке многослойных нейронных сетей.

Процесс обучения двойного ортонейрона сводится к настройке его синаптических весов в скрытом и выходном слоях путём минимизации применяемого критерия (целевой функции).

Обучение двойного ортонейрона

В качестве критерия обучения будем использовать стандартную локальную квадратичную функцию, принятую в теории искусственных нейронных сетей

$$E(k) = \frac{1}{2} (d(k) - o(k))^2 = \frac{1}{2} e(k)^2 \quad (4)$$

(здесь $d(k)$ – внешний обучающий сигнал), а само обучение проводить на основе обратного распространения ошибки.

Градиентный алгоритм обучения синаптических весов выходного слоя может быть записан в виде

$$w_{l0}(k+1) = w_{l0}(k) + \eta_0(k) e(k) \varphi_{l0}(u(k)),$$

или в векторной форме

$$w_0(k+1) = w_0(k) + \eta_0(k)e(k)\varphi_0(u(k)),$$

где $\eta_0(k)$ – параметр шага поиска, определяющий характер сходимости процесса обучения;

$w_0(k) = (w_{00}(k), w_{10}(k), \dots, w_{h_0 0}(k))^T$, $\varphi_0(k) = (\varphi_{00}(u(k)), \varphi_{10}(u(k)), \dots, \varphi_{h_0 0}(u(k)))^T$ – $(h_0 + 1) \times 1$ – вектор синаптических весов и активационных функций выходного слоя соответственно. Поскольку синаптические веса $w_0(k)$ входят в описание DON линейно, для их настройки может быть использован адаптивный алгоритм Качмажа-Уидроу-Хоффа [12-15] в виде

$$w_0(k+1) = w_0(k) + \frac{d(k) - o(k)}{\|\varphi_0(k)\|^2} \varphi_0(k), \quad (5)$$

при этом ортогональность сигналов на входах синапсов w_{10} обеспечивает алгоритму (5) максимально возможное быстроедействие [16].

При обработке зашумлённых сигналов в алгоритм обучения целесообразно ввести дополнительные сглаживания, для чего можно использовать процедуру [17]

$$\begin{cases} w_0(k+1) = w_0(k) + r_0^{-1}(k)(d(k) - o(k))\varphi_0(k); \\ r_0(k+1) = \alpha r_0(k) + \|\varphi_0(k)\|^2, \end{cases} \quad (6)$$

где $0 \leq \alpha \leq 1$ – параметр сглаживания. Заметим, что при $\alpha = 0$, процедура (6) совпадает с алгоритмом (5), а при $\alpha = 1$ – превращается в правило обучения типа стохастической аппроксимации.

Немаловажным фактом является то, что выходной ортосинапс может настраиваться с помощью стандартного метода наименьших квадратов, минимизирующего критерий обучения, заданный на всей обучающей выборке.

$$E^N = \sum_{k=1}^N E(k) = \frac{1}{2} \sum_{k=1}^N e^2(k), \quad (7)$$

и имеющего вид

$$\begin{aligned} w_0(N) &= \left(\sum_{k=1}^N \varphi_0(k)\varphi_0^T(k) \right)^{-1} \sum_{k=1}^N \varphi_0(k)d(k) = \\ &= P(N) \sum_{k=1}^N \varphi_0(k)d(k). \end{aligned} \quad (8)$$

Выражение (8) может также быть записано в рекуррентной форме

$$\begin{cases} w_0(k+1) = w_0(k) + P(k)(d(k) - o(k))\varphi_0(k); \\ P(k+1) = P(k) - \frac{P(k)\varphi_0(k+1)\varphi_0^T(k+1)P(k)}{1 + \varphi_0^T(k+1)P(k)\varphi_0(k+1)}, \end{cases}$$

представляющий собой, по сути, процедуру минимизации второго порядка целевой функции (7). Необходимо отметить, что в силу условия (2) матрицы $P(N)$, $P(k+1)$, $P(k)$ являются диагональными (или стремятся к таковым), что резко упрощает числен-

ную реализацию процесса обучения, обеспечивает ей вычислительную устойчивость и высокую скорость сходимости.

Для обучения ортосинапсов OS_i скрытого слоя перепишем целевую функцию (4) в виде

$$E(k) = \frac{1}{2} \left(d(k) - f_0 \left(\sum_{i=0}^n \sum_{j=0}^{h_i} w_{ji} \varphi_{ji}(x_i(k)) \right) \right)^2,$$

минимизация которой с помощью градиентного алгоритма приводит к процедуре обучения

$$\begin{aligned} w_{ji}(k+1) &= w_{ji}(k) + \eta_i(k)e(k)f'_0(k)\varphi_{ji}(x_i(k)) = \\ &= w_{ji}(k) + \eta_i(k)\delta(k)\varphi_{ji}(x_i(k)), \end{aligned}$$

или в векторной форме

$$w_i(k+1) = w_i(k) + \eta_i(k)\delta(k)\varphi_i(k),$$

где $\delta(k) = e(k)f'_0(k)$ – стандартная δ -ошибка, принятая в теории обучения многослойных нейронных сетей;

$$w_i(k) = (w_{0i}(k), w_{1i}(k), \dots, w_{h_i i}(k))^T;$$

$$\varphi_i(k) = (\varphi_{0i}(x_i(k)), \varphi_{1i}(x_i(k)), \dots, \varphi_{h_i i}(x_i(k)))^T.$$

Используя одношаговую процедуру Левенберга-Марквардта со сглаживанием [18, 19], можно записать алгоритм обучения ортосинапсов скрытого слоя в виде

$$\begin{cases} w_i(k+1) = w_i(k) + r_i^{-1}(k)\delta(k)\varphi_i(k); \\ r_i(k+1) = \alpha r_i(k) + \|\varphi_i(k+1)\|^2, 0 \leq \alpha \leq 1, \end{cases}$$

обладающий как фильтрующими, так и следящими свойствами.

Нейропредиктор на основе двойного ортонейрона

На основе двойного ортонейрона синтезирована схема нейропредиктора стохастических и хаотических процессов, приведённая на рис. 4.

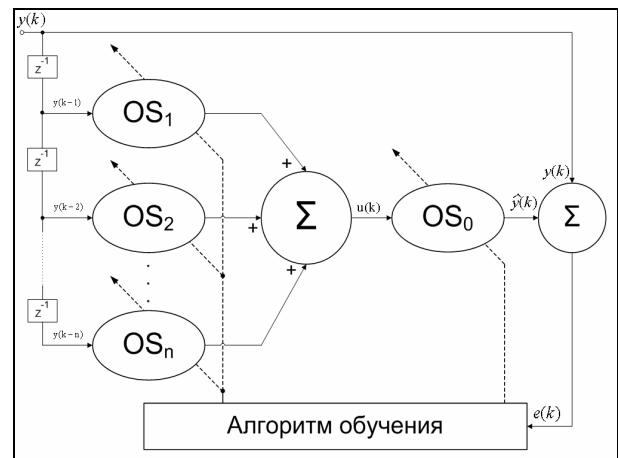


Рис. 4. Нейропредиктор на основе DON

В качестве тестовой решалась задача прогнозирования хаотического процесса, описываемая диф-

ференциальным уравнением Мэки-Гласса [20]

$$y'(t) = \frac{0,2t(t-\tau)}{1+y^{10}(t-\tau)} - 0,1y(t),$$

а при этом в процессе квантования с шагом 0,1 значения временного ряда в шагах $k = 1, 2, \dots$ вычислялись с помощью метода Рунге-Кутты четвёртого порядка. На основе обучающей выборки объёмом $N = 500$ элементов строился прогноз на 6 шагов вперёд по предыстории, сформированной k -м, $(k - 6)$ -м, $(k - 12)$ -м, $(k - 18)$ -м элементами ряда. В качестве тестовой выборки использовались 9500 элементов последовательности от 501-го до 10000-го.

Для оценки точности получаемых прогнозов принята стандартная нормированная среднеквадратическая ошибка

$$NRMSE(k, N) = \frac{\sum_{q=1}^N e^2(k+q)}{N\sigma},$$

где σ – среднеквадратичное отклонение прогнозируемого процесса на тестовой выборке.

В качестве активационных функций были использованы ортогональные полиномы Чебышева первого рода (Т1), ортогональные полиномы Чебышева второго рода (Т2), ортогональные полиномы Эрмита (Н1), ортонормированные полиномы Эрмита (Н2) и ортогональные полиномы Лагерра (L). Их количество в ортосинапсах $h_i = h_0 = h$ менялось от одного до восьми, с учётом нулевого, то есть фактически от двух до девяти. Ошибки, полученные в результате прогнозирования, приведены в табл. 1.

Таблица 1
Ошибки прогнозирования двойного ортонейрона

h	Базисная система многочленов				
	T1	T2	H1	H2	L
1	0,425	0,425	0,425	0,086	0,425
2	0,084	0,078	0,127	0,05	0,126
3	0,064	0,045	0,031	0,032	0,063
4	0,109	0,050	0,039	0,044	0,049
5	0,085	0,19	0,209	0,051	0,159
6	0,314	1,027	0,199	0,052	0,138
7	13,62	2,29	0,246	0,056	0,202
8	80,93	1131	0,26	1,583	0,252

Динамика процесса изменения ошибки с увеличением размерности ортогонального или ортонормированного базиса показана на рис. 5.

В данном случае, для решения тестовой задачи оптимальная размерность ортосинапсов составляет три-четыре полинома. Следовательно, количество настраиваемых параметров значительно меньше, чем в существующих распространённых конструкциях, что обеспечивает повышенную скорость настройки весовых коэффициентов сети.

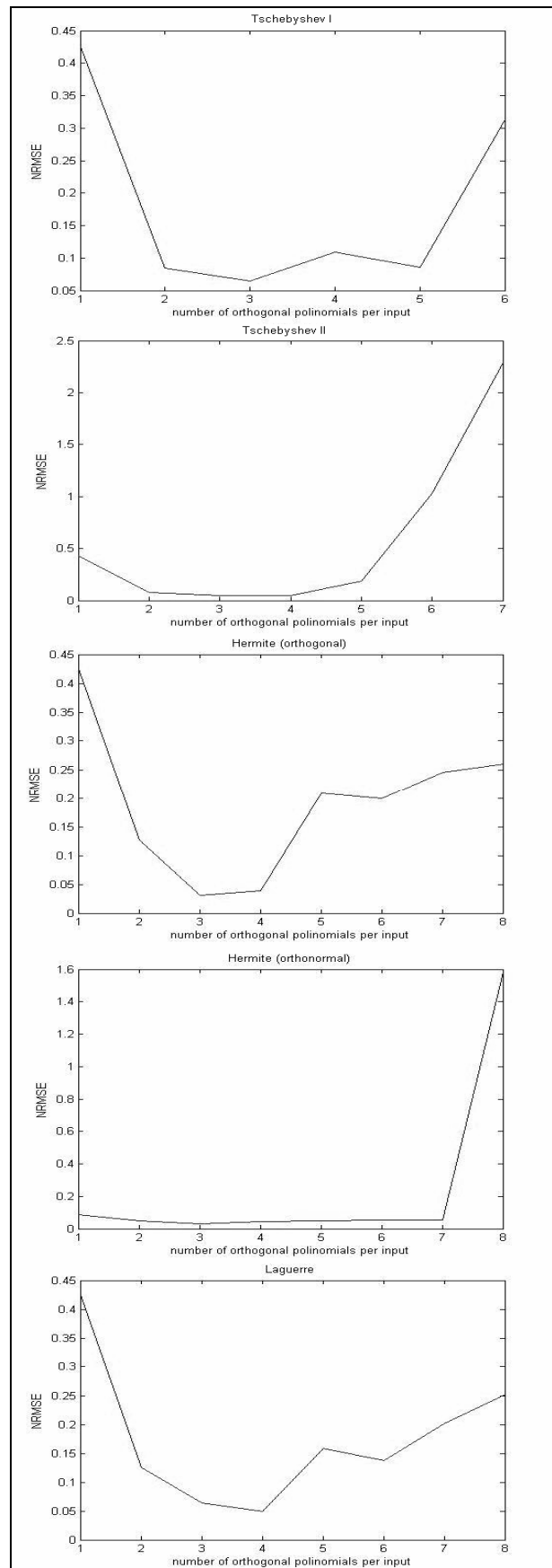


Рис. 5. Изменение ошибки прогнозирования двойного ортонейрона

С увеличением количества полиномов разложения в каждом ортосинапсе конечный результат

улучшался лишь до определённого момента. Это объясняется общеизвестным фактом, устанавливающим, что с возрастанием сложности модели рано или поздно её аппроксимирующие способности начнут снижаться.

Как видно из результатов, наилучшее качество прогнозирования было получено при использовании полиномов Эрмита, как ортогональных, так и ортонормированных. Это объясняется тем, что эта система полиномов ортогональна на отрезке $[-\infty, \infty]$, а интервал варьирования переменных, поступающих из скрытого слоя в выходной, не определён. В случае использования в качестве базисных систем ортогональных полиномов Чебышева первого рода, Чебышева второго рода или Лагерра вполне возможна ситуация, когда в ортосинапсе выходного слоя происходит разложение входного сигнала, который не попадает в интервал ортогональности выбранной системы многочленов, что приводит к снижению качества прогнозирования.

Аналогичный эксперимент был проведён и для ортонейрона. Полученные ошибки прогнозирования оказались в среднем больше в 1,6 раза. Однако следует заметить, что ортонейрон обучается быстрее, чем двойной ортонейрон, поскольку является однослойной архитектурой, синаптические веса которой линейно входят в её описание.

Заключение

В работе введены ортосинапс и ортонейроны, отличающиеся от известных систем вычислительно-интеллекта используемыми функциями активации в виде ортогональных полиномов, что позволяет обеспечить ряд полезных свойств, связанных, прежде всего, с улучшением качества процессов обучения. Решена тестовая задача прогнозирования хаотического процесса с помощью нейропредиктора, построенного из ортосинапсов. Предложенные конструкции просты в численной реализации, имеют повышенную скорость обучения и обеспечивают высокую точность прогнозирования.

Список литературы

1. Грон Л. Методы идентификации систем. – М.: Мир, 1979. – 302 с.
2. Yamakawa T., Uchino E., Miki T., Kusanagi H. A new fuzzy neuron and its applications to system identification and prediction of the system behavior // Proc. 2-nd Int. Conf. on Fuzzy Logic and Neural Networks "FIZUKA-92". – Iizuka, Japan, 1992. – P. 477-483.
3. Uchino E., Yamakawa T. Soft computing based signal prediction, restoration and filtering / Ed. by Da Ruan "Intelligent Hybrid Systems: Fuzzy Logic, Neural Networks and Genetic Algorithms". – Boston: Kluwer Academic Publisher, 1997. – P. 331-349.
4. Miki T., Yamakawa T. Analog implementation of neuro-fuzzy neuron and its on-board learning / Ed. by N.E. Mastorakis "Computational Intelligence and Applications". – Piraeus: WSES Press, 1999. – P. 144-149.

5. Bodyanskiy Ye., Kolodyazhniy V., Slipchenko O. Structural and synaptic adaptation in the artificial neural networks with orthogonal activation functions // Sci. Proc. of Riga Technical University. Comp. Sci., Int Technology and Management Sci. – 2004. – № 20. – P. 69-76.

6. Bodyanskiy Ye., Kolodyazhniy V., Slipchenko O. Artificial neural network with orthogonal activation functions for dynamic system identification / Ed. by O. Sawodny, P. Scharff "Synergies between Information Processing and Automation". – Aachen: Shaker Verlag, 2004. – P. 122-127.

7. Bodyanskiy Ye., Pliss I., Slipchenko O. Growing neural networks based on orthogonal activation functions // Proc. XII-th Int. Conf. "Knowledge – Dialog – Solution". – Varna, 2006. – P. 84-89.

8. Bodyanskiy Ye., Slipchenko O. Ontogenic neural networks using orthogonal activation functions // Prace. Naukowe Akademii Ekonomicznej we Wroclawiu. – 2006. – № 21. – P. 13-20.

9. Бодянский Е.В., Колодяжный В.В., Слипченко А.Н. Нейропредиктор с ортогональными функциями активации // Автоматизация виробничих процесів. – 2005. – № 1 (20). – С. 123-127.

10. Бодянский Е.В., Винокурова Е.А., Ламонова Н.С. Двойной взвешенный нейрон и алгоритм его обучения // Радиоэлектроника. Информатика. Управление. – 2006. – № 2. – С. 85-91.

11. Бейтмен Г., Эрдейи А. Высшие трансцендентные функции. – М.: Наука, 1974. – 12. – 295 с.

12. Kaczmarz S. Angenaherte Ausloesung von Systemen Linearer Gleichungen // Bull. Int. Acad. Polon. Sci. – 1937. – Let. A. – S. 355-357.

13. Kaczmarz S. Approximate solution of systems of linear equations // Int. J. Control. 1993. – 53. – P. 1269-1271.

14. Widrow B., Hoff Jr. M.E. Adaptive switching circuits // 1960 URE WESCON Convention Record. – N.Y.: IRE, 1960. – Part 4. – P. 96-104.

15. Widrow B., Lee M. 30 years of adaptive neural networks: perception, adaline and backpropagation // Proc. IEEE – 1990. – 78. – P. 1415-1442.

16. Райман Н.С., Чадаев В.М. Построение моделей процессов производства. – М.: Энергия, 1975. – 376 с.

17. Бодянский Е.В., Плисс И.П., Соловьева Т.В. Многошаговые оптимальные упреждители многомерных нестационарных стохастических процессов // Доклады АН УССР. – 1986. – Сер. А. – № 12. – С. 47-49.

18. Bodyanskiy Ye., Kolodyazhniy V., Stephan A. An adaptive learning algorithm for a neuro-fuzzy network / Ed. by B. Reusch. "Computational Intelligence. Theory and Applications." – Berlin – Heidelberg – New York: Springer, 2001. – P. 68-75.

19. Otto P. Bodyanskiy Ye, Kolodyazhniy V. A new learning algorithm for a forecasting neuro-fuzzy network // Integrated Computer-Aided Engineering. – Amsterdam: IOS Press, 2003 – 10. – № 4. – P. 399-409.

20. Mackey M.C., Glass L. Oscillation and chaos in physiological control systems // Science. – 1977. – 197. – P. 238-289.

Поступила в редколлегию 4.04.2007

Рецензент: д-р техн. наук, проф. В.А. Филатов, Харьковский национальный университет радиоэлектроники, Харьков.