

Е. В. БОДЯНСКИЙ, Н. Е. КУЛИШОВА

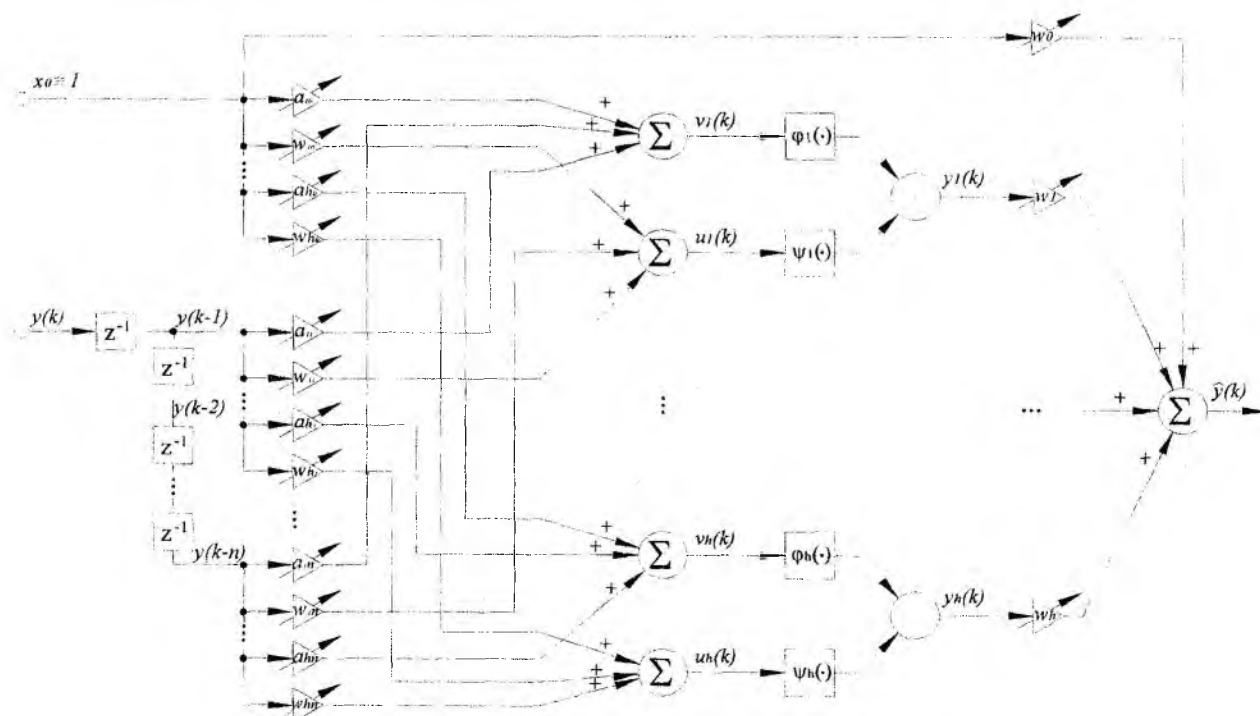
ОБОБЩЕННЫЙ АЛГОРИТМ ОБУЧЕНИЯ ПРОГНОЗИРУЮЩЕЙ ИСКУССТВЕННОЙ Σ -П НЕЙРОННОЙ СЕТИ

Искусственные нейронные сети (ИНС) в настоящее время получают все более широкое распространение при решении различных задач обработки информации и, в частности, прогнозирования нестационарных стохастических сигналов произвольной природы. Здесь наиболее популярными являются многослойные архитектуры как с прямой передачей информации [1–21], так и рекуррентные (с обратной связью) [22–26], а также прогнозирующие радиально-базисные нейронные сети [27, 28].

Эффективность многослойных сетей с сигмоидальными Σ -П функциями активации объясняется, прежде всего, их универсальными аппроксимирующими свойствами [29–30], однако низкая скорость обучения, основанного на концепции обратного распространения ошибок, может создавать определенные проблемы при прогнозировании существенно нестационарных процессов. Альтернативой многослойным могут служить радиально-базисные нейронные сети. Эти сети содержат только один скрытый слой, а функции активации здесь, в отличие от сигмOID многослойных ИНС – немонотонные, например, гауссианы. Поскольку выходной сигнал в этом случае формируется как линейная комбинация радиально-базисных функций, для настройки синаптических весов могут быть использованы оптимальные по быстродействию алгоритмы адаптивной идентификации, при этом, что очень важно, по своим аппроксимирующим свойствам радиально-базисные сети не уступают многослойным [31, 32]. Важнейшим преимуществом радиально-базисных сетей является то, что они позволяют сократить время обучения на один-три порядка по сравнению с сетями, использующими обратное распространение ошибок [33], обеспечивая при этом высокое качество аппроксимации. Вместе с тем следует помнить, что количество нейронов скрытого слоя должно быть весьма велико, причем оно растет в экспоненциальной зависимости от размерности аппроксимируемой функции.

Сочетание положительных свойств многослойных и радиально-базисных ИНС обеспечивается так называемыми Σ -П (сигма-пи) нейронными сетями [33, 34], получившими к настоящему времени весьма ограниченное распространение и применение.

Примером такой схемы может служить прогнозирующая искусственная Σ -П нейронная сеть, образованной элементами задержки, адаптивными линейными ассоциаторами (адалинами), блоками нелинейных функций активации и умножителями (рис. 1).

Рис. 1. Схема прогнозирующей искусственной Σ -П сети

Сигнал прогнозируемой нестационарной стохастической последовательности $y(k)$, $k=0, 1, 2, \dots$ подается на входной слой сети, образованной элементами задержки, в результате чего формируется вектор предыстории $x(k) = (1, y(k-1), y(k-2), \dots, y(k-n))^T = (x_0(k), x_1(k), x_2(k), \dots, x_n(k))^T$ размерности $(n+1) \times 1$, поступающий затем на скрытый слой. Этот слой образован h нейронами, каждый из которых содержит по два сумматора, с $n+1$ настраиваемыми входами каждый, два блока нелинейных функций активации, одна из них сигмоидальная (например, типа гиперболического тангенса), а другая – радиально-базисная (например, гауссовская), и умножитель. Выходной слой образован одним нейроном типа адаптивного линейного ассоциатора с $h+1$ входами. Всего рассмотренная сеть содержит $1+h(2n+3)$ настраиваемых синаптических весов.

Прогноз, формируемый Σ - Π сетью, имеет вид

$$\hat{y}(k) = w_0(k) + \sum_{i=1}^h w_i(k) \varphi_i(v_i(k)) \psi_i(u_i(k)), \quad (1)$$

где $v_i(k) = \sum_{j=0}^n a_{ij}(k) x_j(k)$, $u_i(k) = \sum_{j=0}^n w_{ij}(k) x_j(k)$; $i = 1, 2, \dots, h$; $j = 0, 1, \dots, n$,

$y_i(k) = \varphi_i(v_i(k)) \psi_i(u_i(k))$ – выходной сигнал i -го нейрона скрытого слоя;

$w_0(k), w_i(k), w_{ij}(k), a_{ij}(k)$ – настраиваемые синаптические веса в k -й момент времени;

$\varphi_i(\bullet)$ – радиально-базисная функция активации i -го нейрона;

$\psi_i(\bullet)$ – сигмоидальная функция активации того же нейрона.

С учетом введенных обозначений можно переписать выражение (1) в виде

$$\hat{y}(k) = w_0(k) + \sum_{i=1}^h w_i(k) y_i(k), \quad (2)$$

откуда видно, что сигнал выходного слоя формируется подобно выходу радиально-базисной сети, а скрытый слой обладает как персептронными, так и радиально-базисными свойствами.

Введя в рассмотрение локальный критерий обучения

$$E_k = \frac{1}{2} e^2(k), \quad (3)$$

где $e(k) = y(k) - \hat{y}(k)$ – ошибка прогнозирования,

и $1+h(2n+3) \times 1$ – вектор настраиваемых синаптических весов

$$W(k) = (w_0(k), w_1(k), \dots, w_h(k), \dots, w_{ij}(k), \dots, a_{ij}(k), \dots)^T,$$

можно записать алгоритм обучения сети в виде

$$W(k+1) = W(k) - \mu(k) \nabla_{W(k)} E_k \quad (4)$$

(здесь $\mu(k)$ – коэффициент усиления алгоритма) или покомпонентно:

$$\begin{cases} w_0(k+1) = w_0(k) + \mu_0(k) e(k), \\ w_i(k+1) = w_i(k) + \mu_i(k) e(k) y_i(k), \\ w_{ij}(k+1) = w_{ij}(k) + \mu_{ij}(k) e(k) w_i(k) \varphi_i(v_i(k)) \frac{d\psi_i(u_i(k))}{du_i(k)} x_j(k), \\ a_{ij}(k+1) = a_{ij}(k) + \mu_{ij}(k) e(k) w_i(k) \psi_i(u_i(k)) \frac{d\varphi_i(v_i(k))}{dv_i(k)} x_j(k). \end{cases} \quad (5)$$

Громоздкость алгоритма (5) и необходимость вычисления на каждом шаге производных функции активации заставляют искать пути упрощения процедуры обучения сети. Здесь весьма полезным оказывается тот факт, что сигмоидальная функция активации $\psi(\gamma q)$, заданная на квадрате $-1 \leq q \leq 1$, $-1 < \psi(\gamma q) < 1$, $\gamma > 0$, может быть представлена в виде степенного ряда [35]

$$\begin{aligned} \psi(\gamma q) &= \alpha_0 \gamma q + \alpha_1 \gamma^3 q^3 + \alpha_2 \gamma^5 q^5 + \alpha_3 \gamma^7 q^7 + \dots = \sum_{l=0}^L \alpha_l (\gamma q)^{2l+1}, \\ \alpha_{l+1} &= \beta \alpha_l, \quad -1 < \beta < 0 \end{aligned} \quad (6)$$

с высокой степенью точности при достаточно малом числе членов разложения L . Первая производная этого разложения

$$\begin{aligned} \frac{\partial \psi(\gamma q)}{\partial q} &= \alpha_0 \gamma + 3\alpha_1 \gamma^3 q^2 + 5\alpha_2 \gamma^5 q^4 + 7\alpha_3 \gamma^7 q^6 + \dots = \\ &= \sum_{l=0}^L \gamma (2l+1) \alpha_l (\gamma q)^{2l} = \sum_{l=0}^L \alpha'_l (\gamma q)^{2l}, \\ \alpha'_l &= \gamma (2l+1) \alpha_l \end{aligned} \quad (7)$$

является радиально-базисной функцией, а вторая производная

$$\begin{aligned} \frac{\partial^2 \psi(\gamma q)}{\partial q^2} &= 2 \cdot 3\alpha_1 \gamma^3 q + 4 \cdot 5\alpha_2 \gamma^5 q^3 + 6 \cdot 7\alpha_3 \gamma^7 q^5 + \dots = \\ &= \sum_{l=0}^L 2\gamma^2 (l+1)(2l+3) \alpha_{l+1} (\gamma q)^{2l+1} = \sum_{l=0}^L \alpha''_l (\gamma q)^{2l+1}, \\ \alpha''_l &= 2\gamma^2 (l+1)(2l+3) \alpha_{l+1} \end{aligned} \quad (8)$$

опять имеет структуру, подобную (6).

Разложения (6)–(8) позволяют придать компонентам, входящим в алгоритм (5), унифицированную форму

$$\left\{ \begin{aligned} \psi(\gamma u) &= \sum \alpha_l (\gamma u)^{2l+1}, \\ \varphi(\gamma v) &= \sum \alpha'_l (\gamma v)^{2l}, \\ \frac{d\psi(\gamma u)}{du} &= \sum \alpha'_l (\gamma u)^{2l}, \\ \frac{d\varphi(\gamma v)}{dv} &= \sum \alpha''_l (\gamma v)^{2l+1}, \\ \varphi(\gamma v) \frac{d\psi(\gamma u)}{du} &= \sum \alpha'_l (\gamma v)^{2l} \sum \alpha'_l (\gamma u)^{2l}, \\ \psi(\gamma u) \frac{d\varphi(\gamma v)}{dv} &= \sum \alpha_l (\gamma u)^{2l+1} \sum \alpha''_l (\gamma v)^{2l+1}, \\ \psi(\gamma u) \varphi(\gamma v) &= \sum \alpha_l (\gamma u)^{2l+1} \sum \alpha'_l (\gamma v)^{2l}, \end{aligned} \right. \quad (9)$$

после чего, полагая параметр крутизны $\gamma=1$, а функции активации всех нейронов однотипными, можно записать обобщенный алгоритм обучения сети:

$$\left\{ \begin{aligned} w_0(k+1) &= w_0(k) + \mu(k)e(k), \\ w_i(k+1) &= w_i(k) + \mu(k)e(k) \left(\sum_{l=0}^L \alpha'_l(v_i(k))^{2l} \sum_{l=0}^L \alpha_l(u_i(k))^{2l+1} \right), \\ w_{ij}(k+1) &= w_{ij}(k) + \mu(k)e(k)w_i(k) \left(\sum_{l=0}^L \alpha'_l(v_i(k))^{2l} \sum_{l=0}^L \alpha'_l(u_i(k))^{2l} \right) x_j(k), \\ a_{ij}(k+1) &= a_{ij}(k) + \mu(k)e(k)w_i(k) \left(\sum_{l=0}^L \alpha_l(u_i(k))^{2l+1} \sum_{l=0}^L \alpha'_l(v_i(k))^{2l+1} \right) x_j(k). \end{aligned} \right. \quad (10)$$

Несложно видеть, что алгоритм (10) достаточно прост с вычислительной точки зрения, не использует обратного распространения, что в свою очередь обеспечивает высокое быстродействие прогнозирующей Σ -П сети, являющейся обобщением многослойных и радиально-базисных ИНС.

Список литературы: 1. Schoeneburg E. Stock price prediction using neural networks a project report // Neurocomputing. 1990. Vol. 2. P. 17-27. 2. Wong F. S. Time series forecasting using backpropagation neural networks // Neurocomputing. 1990/1991. Vol. 2. P. 147-159. 3. Groot C. de, Wuertz D. Analysis of univariate time series with connectionist nets: A case study of two classical examples // Neurocomputing. 1991. Vol. 3. P. 177-192. 4. Pham D. T., Liu X. Neural Networks for Identification, Prediction and Control. London: Springer-Verlag, 1995. 238 p. 5. Masters T. Neural, Novel & Hybrid Algorithms for Time Series Prediction. N. Y.: John Wiley & Sons, Inc., 1995. 514 p. 6. Saxen H. Nonlinear time series analysis by neural network a case of study // Int. J. Neural Systems. 1996. Vol. 7. № 2. P. 195-201. 7. Kaastra I., Boyd M. Designing a neural network for forecasting financial and economic time series // Neurocomputing. 1996. Vol. 10. P. 215-236. 8. Chenoweth T., Obradovic Z. A multicomponent nonlinear prediction system for the S&P500 index // Neurocomputing. 1996. Vol. 10. P. 275-290. 9. Khotanzad A., Afkhami-Ronani R., Lu T.L., Abade A., Davis M., Maratukulam D. J. ANNSTLF A neural networks and robust time series prediction // IEEE Trans. on Neural Networks. 1994. Vol. 5. P. 240-254. 10. Hansen J. K., Nelson R. D. Neural networks and traditional time series methods: A synergistic combination in state economic forecasts // IEEE Trans. on Neural Networks. 1977. Vol. 8. P. 863-875. 11. Yu H.Y., Bang S.Ya. An improved time series prediction by applying the layer-by-layer learning method to FIR neural networks // Neural Networks. 1997. Vol. 10, № 9. P. 1717-1729. 12. Nie J. Nonlinear time-series forecasting: A fuzzy-neural approach // Neurocomputing. 1997. Vol. 16. P. 63-76. 13. Zirilli J. S. Financial Prediction Neural Networks. London: Int. Thomson Computer Press, 1997. 135 p. 14. Kingdom J. Intelligent Systems and Financial Forecasting. Berlin: Springer-Verlag, 1997. 227 p. 15. Bodayanskiy Ye., Pliss I., Stephan A. Multilayer neural networks for nonlinear dynamic system modelling and identification // «Теория и техника передачи, приема и обработки информации»: Тез. докл. 3-й междунар. конф. Харьков-Туапсе, 1997. С. 150. 16. Conway A. J., Macpherson K. P., Brown J. C. Delayed time series predictions with neural networks // Neurocomputing. 1998. Vol. 18. P. 81-89. 17. Бодянский Е. В., Воробьев С. А., Костюк О. В., Любчик Л. М. Фильтрация и прогнозирование тренд-сезонных временных рядов на основе искусственных нейронных сетей // Радиоэлектроника и информатика. 1998. № 3. С. 76-83. 18. Bodayanskiy Ye., Popov S., Stephan A. Adaptive prediction of stochastic polyharmonic processes using neural network technologies // 45. Internationales Wissenschaftliches Kolloquium. Tagungsband. TU Ilmenau (Thuer), 2000. S. 123-128. 19. Бодянский Е. В., Попов С. В., Штефан А. Нейросетевой упредитель многомерных стохастических последовательностей // «Автоматика 2000»: Праці міжнар. конф. з автоматичного управління. Т. 2. Львів: ДНДІ інформаційної інфраструктури, 2000. С. 40-42. 20. Бодянский Е. В., Попов С. В. Прогнозирующая нейронная сеть и алгоритмы ее обучения // Радиоелектроніка, інформатика, управління. 2000. № 1. С. 60-64. 21. Bodayanskiy Ye., Popov S. Adaptive forecasting of stochastic polyharmonic processes using neural network technologies // «Теория и техника передачи, приема и обработки информации»: Сб. науч. тр. 6-й междунар. конф. Харьков: ХТУРЭ, 2000. С. 324-325. 22. Connor J. T., Martin R. D., Atlas L. E. Recurrent neural networks robust time series prediction // IEEE Trans. on Neural Networks. 1994. Vol. 5. P. 240-254. 23. Aussem A., Murtagh F., Sarazin M. Dynamical recurrent neural networks towards environmental time series prediction // Int. J. Neural Systems. 1995. Vol. 6. P. 145-170. 24. Madhavan P. Y. A new recurrent neural network learning algorithm for time series prediction // J. Intelligent Systems. 1997. Vol. 7, № 1-2. P. 103-116. 25. Бодянский Е. В., Кулишова Н. Е., Руденко О. Г. Рекуррентна прогноуюча штучна нейронна мережа: архітектура та алгоритми навчання // Адаптивні системи автоматичного управління. Дніпропетровськ: Системні технології, 1999. Вип. 2 (22). С. 129-137. 26. Бодянский Е. В., Воробьев С. А. Рекуррентная нейронная сеть для обнаружения изменений свойств нелинейных стохастических последовательностей // Автоматика и телемеханика. 2000. № 7. С. 55-67. 27. Chang E. S., Chen S., Mulgrew B. Gradient radial basis function networks for nonlinear and nonstationary time series prediction // IEEE Trans. on Neural Networks. 1996. Vol. 7. P. 190-194. 28. Billings S. A., Hong X. Dual ortogonal radial basis function networks for nonlinear time series prediction // Neural Networks. 1998. Vol. 11. P. 479-493. 29. Hornik K., Stinchcombe M., White H. Multilayer feedforward networks are universal approximators // Neural Networks.

layer feedforward networks are universal approximators // *Neural Networks*. 1989. Vol. 2. P. 359-366. 30. *Hornik K.* Approximation capabilities of multilayer feedforward networks // *Neural Networks*. 1991. Vol. 4. P. 251-257. 31. *Park J., Sandberg I. W.* Universal approximation using radial basis function networks // *Neural Computation*. 1991. P. 246-257. 32. *Chen S., Billings S. A.* Neural networks for nonlinear dynamic system modelling and identification/ In: "Advances in Intelligent Control". Ed. by C. J. Harris. London: Taylor and Francis, 1994. P. 85-112. 33. *Cichocki A., Unbehauen R.* Neural Networks for Optimization and Signal Processing. Stuttgart: Teubner, 1993. 526 p. 34. *Bothe H. - H.* Neuro-Fuzzy-Methoden. Einfuehrung in Theorie und Anwendungen. Berlin: Springer-Verlag, 1998. 305 p. 35. *Бодянский Е. В., Кулишова Н. Е., Руденко О. Г.* Об одной модели формального нейрона // *Доповіді НАН України*. 2001. № 4. С. 69-73.

Поступила в редколлегию 22.02.2001