

## ОБ ОДНОМ АЛГОРИТМЕ ОБУЧЕНИЯ МНОГОСЛОЙНОЙ НЕЙРОННОЙ СЕТИ

РУДЕНКО О.Г., ШТЕФАН А.

Предлагается рекуррентный алгоритм настройки параметров многослойной нейронной сети, основанный на методе наименьших квадратов и использующий ограниченное число обучающих образов.

Искусственные нейронные сети (ИНС) находят все более широкое применение при решении задач классификации и распознавания образов [1-3], прогнозирования [4], оценивания и идентификации [5-8], управления сложными объектами [5, 9], обработки сигналов [10-12]. ИНС являются альтернативой классическим методам, использующим математические модели заданной структуры в виде различных полиномов. Основной особенностью ИНС является их способность к обучению, осуществляемому путем коррекции весовых параметров, используемых при описании ИНС, и основанному на сравнении выходных сигналов нейронной сети с обучающими образами, поступающими в последовательные моменты времени. Так как ИНС представляют собой многослойные структуры, при коррекции этих параметров используется информация о желаемых (оптимальных) сигналах скрытых слоев. Несмотря на то, что такая информация практически всегда отсутствует, обучение нейронной сети возможно.

Наиболее широкое распространение для коррекции параметров в ИНС получил back propagation algorithm [1-3]. Однако в последнее время все большее внимание исследователей привлекает метод наименьших квадратов (МНК) и его модификации, эффективность которых при решении данной задачи обучения подтверждается многочисленными работами [6, 8, 9, 11, 12]. В этих работах изучается МНК, рекуррентный МНК (РМНК) и взвешенный РМНК, причем оценки последнего более привлекательны, так как могут быть применены и для коррекции нестационарных параметров.

Рассмотрим еще одну модификацию МНК — РМНК со скользящим окном (с ограниченной или фиксированной памятью), который, как и взвешенный РМНК, удобен для коррекции изменяющихся во времени параметров.

Трехслойная нейронная структура, представленная на рис., содержит один скрытый слой и  $L$ ,  $M$  и  $N$  узлов во входном, скрытом и выходном слоях соответственно.

Обучающие образы поступают в последовательные моменты времени  $n=0, 1, 2, \dots$ . Для любого момента времени  $n$  данная структура может быть охарактеризована матрицами:  $X(n) \in R^{L \times n}$  — матрица входов, составленная из текущих векторов входного обучающего образа

$x(i) = (-1, x_1(i), x_2(i), \dots, x_L(i))^T$  размерности  $L \times 1$  ( $i = \overline{1, n}$ );

$Z(n), Z^*(n), \tilde{Z}(n), \tilde{Z}^*(n) \in R^{N \times n}$  — матрицы значений действительных и желаемых выходов, действительных и желаемых входов выходного слоя соответственно;

$Y(n), Y^*(n), \tilde{Y}(n), \tilde{Y}^*(n) \in R^{M \times n}$  — матрицы значений действительных и желаемых выходов, действительных и желаемых входов скрытого слоя соответственно;  $V \in R^{L \times M}$ ,  $W \in R^{M \times N}$  — матрицы весов скрытого и выходного слоев соответственно;  $f(\cdot)$ ,  $\sigma(\cdot)$  — функции активации (например,  $\{1 + \exp[-(\cdot)]\}^{-1}$  или  $\tanh(\cdot)$ ).

Задача настройки нейронной сети [1-3] сводится к минимизации некоторого наперед выбранного функционала (критерия качества). Так, искомыми матрицами весов скрытого  $V$  и выходного  $W$  слоев найдем, минимизируя функционалы:

$$I_1 = \text{tr} \left\{ \left( \tilde{Z}^*(n) - WY(n) \right)^T \left( \tilde{Z}^*(n) - WY(n) \right) \right\}; \quad (1)$$

$$I_2 = \text{tr} \left\{ \left( \tilde{Y}^*(n) - VX(n) \right)^T \left( \tilde{Y}^*(n) - VX(n) \right) \right\}. \quad (2)$$

Используя правила дифференцирования матричных выражений, из условий

$$\frac{\partial I_1}{\partial W} = 0 \quad \text{и} \quad \frac{\partial I_2}{\partial V} = 0$$

для случая, когда число указаний учителя превышает число неизвестных параметров, получаем следующие выражения для искомых оценок:

$$\bar{W}(n) = \tilde{Z}^*(n) Y^T(n) \left[ Y(n) Y^T(n) \right]^{-1}; \quad (3)$$

$$\bar{V}(n) = \tilde{Y}^*(n) X^T(n) \left[ X(n) X^T(n) \right]^{-1}. \quad (4)$$

Отметим, что в выражении (3) используются выходные сигналы скрытого слоя  $Y(n)$ , информация о которых обычно отсутствует. Из условия

$$\frac{\partial I_1}{\partial Y} = \tilde{Z}^*(n) - WY(n) = 0$$

можно получить следующее соотношение для определения  $Y(n)$ :

$$Y(n) = \left( W^T W \right)^{-1} W^T \tilde{Z}^*(n). \quad (5)$$

Здесь принято во внимание, что  $n \geq M$ . Тогда компоненты желаемого сетевого входа выходного слоя  $\tilde{y}_i^*(n)$ ,  $i = \overline{1, M}$  определяются как  $\tilde{y}_i^* = f^{-1}(y_i^*(n))$ , а действительный сетевой вход выходного слоя вычисляется так:

$$\tilde{Y}(n) = VX(n). \quad (6)$$

С другой стороны, при известном желаемом выходе  $Z^*(n)$  сети желаемый вход выходного слоя  $Y(n)$  может быть определен по формуле:

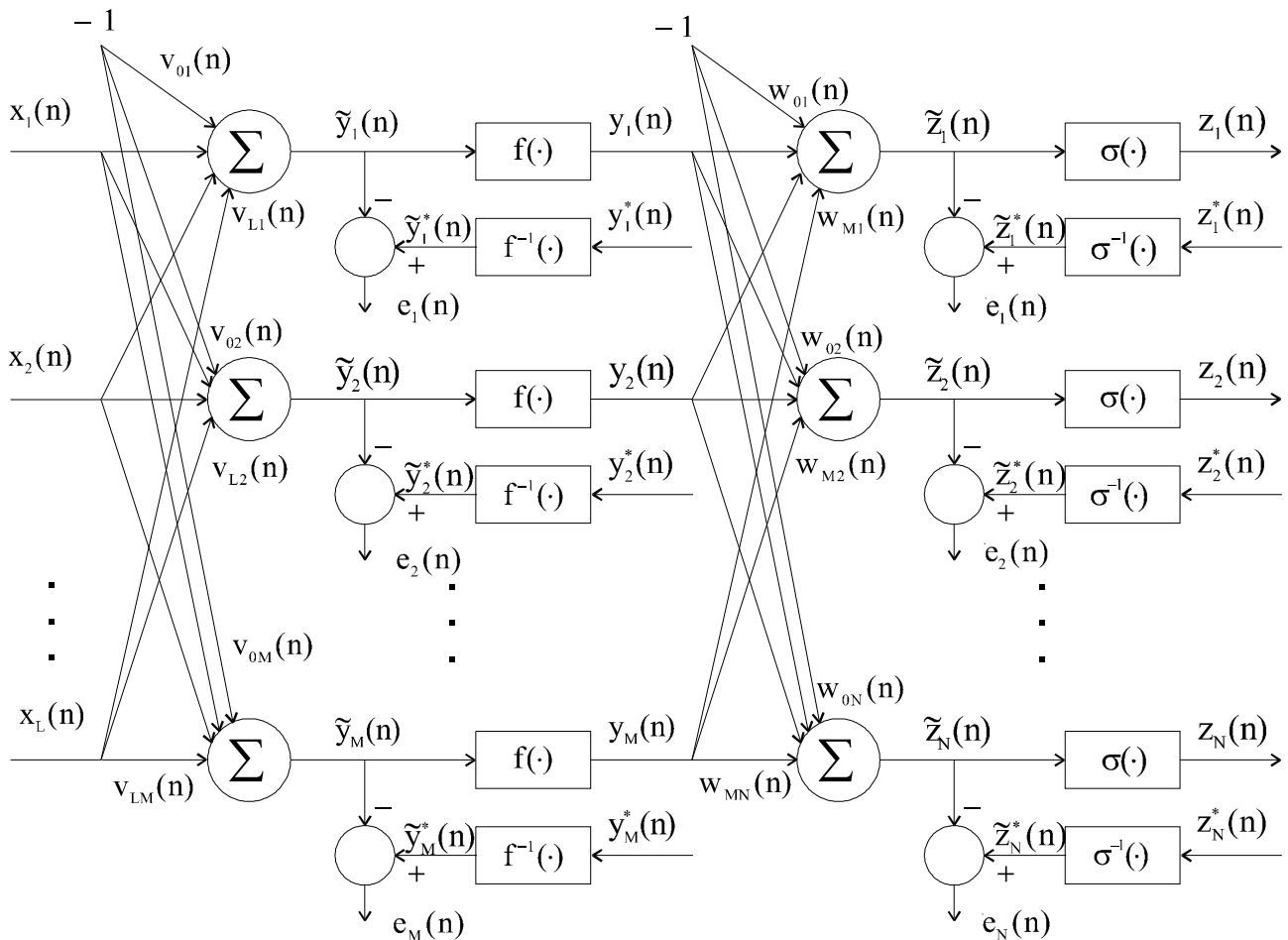


Рис. Трехслойная нейронная структура

$$Y(n) = [\bar{W}^T(n)\bar{W}(n)]^{-1} \bar{W}^T(n)\tilde{Z}^*(n), \quad (7)$$

где  $\bar{W}(n)$  – матрица оценок искоемых весов выходного слоя;  $\tilde{Z}^*(n)$  –  $N \times n$ -матрица, элементами которой являются  $\tilde{z}_i^*(j) = \sigma^{-1}(z_i^*(j))$ ,  $i = \overline{1, N}$ ;  $j = \overline{1, n}$ .

Выражения (3), (4), (5), (7) являются МНК-оценками, рекуррентные формы которых применительно к обучению нейронных структур приведены в [6, 8, 9, 11]. Как уже отмечалось, более привлекательными являются оценки взвешенного МНК и его рекуррентные аналоги [4, 11]. На наш взгляд, также более гибкими по сравнению с обычными МНК-оценками являются оценки, основанные на МНК, использующие однако ограниченное число обучающих образов, – оценки МНК с ограниченной (фиксированной) памятью (окном).

Обозначим буквой  $S$  фиксированное число используемых в алгоритме обучающих образов. Предположим, что  $S \geq M$  и  $S \geq L$ . Отметим, что схема вывода рекуррентной формы остается такой же, если при настройке матрицы весов  $\bar{W}$  используются  $S$  образов ( $S \geq M$ ), а при настройке весовой матрицы  $V$  –  $S'$  образов ( $S' \geq L$ ). Тогда соответствующие оценки (3), (4) примут вид:

$$\bar{W}_S(n) = \bar{Z}_S^*(n)Y_S^T(n)[Y_S(n)Y_S^T(n)]^{-1}; \quad (8)$$

$$\bar{V}_S(n) = \tilde{Y}_S^*(n)X_S^T(n)[X_S(n)X_S^T(n)]^{-1}, \quad (9)$$

где индекс  $S$  говорит о том, что в алгоритмах используется информация об  $S$  последних обучающих образах.

Особенностью алгоритмов с  $S = \text{const}$  является то, что используемые в них матрицы формируются следующим образом: в матрицы после поступления каждого образа включается информация о вновь поступившем  $n$ -м образе, а из нее исключается информация об  $(n-s)$ -м. В зависимости от того, как формируется новая матрица (добавляется ли сначала новая информация, а затем исключается устаревшая либо же сначала исключается устаревшая, а затем добавляется новая), возможны две рекуррентные формы МНК с окном. Остановимся на этом подробнее.

Так как рекуррентные формы для (8) и (9) получаются аналогично, рассмотрим рекуррентную форму оценки (8). Пусть на основе  $(n-1)$ -го образа получена оценка:

$$\bar{W}(n-1) = \tilde{Z}_S^*(n-1)Y_S^T(n-1)[Y_S(n-1)Y_S^T(n-1)]^{-1}. \quad (10)$$

Обозначим

$$R_S^{-1}(n-1) = Y_S(n-1)Y_S^T(n-1). \quad (11)$$

При поступлении нового  $(n)$ -го образа строим новую вспомогательную оценку с использованием  $(S+1)$ -образа:

$$\tilde{W}(n) = \tilde{Z}_{S+1}^*(n) Y_{S+1}^T(n) P_{S+1}(n), \quad (12)$$

где  $P_{S+1}(n) = \left[ Y_{S+1}(n) Y_{S+1}^T(n) \right]^{-1} =$

$$= \left[ R_S^{-1}(n-1) + y(n) y^T(n) \right]^{-1}. \quad (13)$$

Применение к (13) леммы об обращении матриц при условии, что матрица  $R_S^{-1}(n-1)$  – неособенная, дает

$$P_{S+1}(n) = R_S(n-1) - \frac{R_S(n-1) y(n) y^T(n) R_S(n-1)}{1 + y^T(n) R_S(n-1) y(n)}. \quad (14)$$

Подставив (14) в (12) с учетом (8), после несложных преобразований получим:

$$\begin{aligned} \tilde{W}(n) &= \bar{W}(n-1) + \\ &+ \left[ \tilde{z}^*(n) - \bar{W}_S(n-1) y(n) \right] y^T(n) P_{S+1}(n). \end{aligned} \quad (15)$$

А так как для получения МНК-оценки с окном  $S = \text{const}$  необходимо исключить  $(n-S+1)$ -й образ, можно записать

$$\bar{W}(n) = \tilde{Z}_S^*(n) Y_S^T(n) P_S(n), \quad (16)$$

где  $R_S(n) = \left[ Y_{S+1}(n) Y_{S+1}(n) - y(n-S+1) y(n-S+1)^T \right]^{-1};$

$$\tilde{Z}_S^*(n) Y_S^T(n) = \tilde{Z}_{S+1}^*(n) Y_{S+1}^T(n) - z(n-S+1) y^T(n-S+1).$$

С учетом введенных обозначений и при тех же условиях, что и выше, получаем

$$\begin{aligned} R_S(n) &= P_{S+1}(n) + \\ &+ \frac{P_{S+1}(n) y(n-S+1) y^T(n-S+1) P_{S+1}(n)}{1 - y^T(n-S+1) P_{S+1}(n) y(n-S+1)}. \end{aligned} \quad (17)$$

Подстановка данного выражения в (16) приводит к следующей рекуррентной процедуре:

$$\begin{aligned} \bar{W}(n) &= \tilde{W}(n) - \\ &- \left[ \tilde{z}^*(n-S+1) - \tilde{W}(n) y(n-S+1) \right] y^T(n-S+1) R_S(n). \end{aligned} \quad (18)$$

Таким образом, рекуррентный алгоритм настройки матрицы весов  $W$ , получаемый путем добавления нового  $(n)$ -го обучающего образа и последующего исключения старого  $(n-S+1)$ -го, описывается соотношениями (14), (15), (17), (18).

Если же при настройке весовой матрицы  $W$  сначала исключается самый старый,  $(n-S+1)$ -й, образ, а затем добавляется вновь поступивший,  $(n)$ -й, то, как нетрудно показать, рекуррентная процедура настройки будет иметь вид:

$$\begin{aligned} \bar{W}(n) &= \tilde{W}(n-1) + \\ &+ \left[ \tilde{z}^*(n) - \tilde{W}(n-1) y(n) \right] y^T(n) P_{S-1}(n-1); \end{aligned} \quad (19)$$

$$\begin{aligned} \tilde{W}(n) &= \bar{W}(n-1) - \left[ \tilde{z}^*(n-S+1) - \right. \\ &- \left. \bar{W}(n-1) y(n-S+1) \right] y^T(n-S+1) R_S(n-1); \end{aligned} \quad (20)$$

$$P_{S-1}(n-1) = R_S(n-1) +$$

$$+ \frac{R_S(n-1) y(n-S+1) y^T(n-S+1) R_S(n-1)}{1 - y^T(n-S+1) R_S(n-1) y(n-S+1)}; \quad (21)$$

$$\begin{aligned} R_S(n) &= P_{S-1}(n-1) - \\ &- \frac{P_{S-1}(n-1) y(n) y^T(n) P_{S-1}(n-1)}{1 + y^T(n) P_{S-1}(n-1) y(n)}. \end{aligned} \quad (22)$$

Начальные значения матриц  $P$  и  $R$  выбираются, как в обычном рекуррентном МНК. Как уже отмечалось, рекуррентные процедуры оценивания матрицы весов  $V$  легко могут быть получены аналогично. В соответствии с (4) в алгоритмах будут использоваться сетевые входы  $x(1), x(2), \dots, x(n)$  и желаемые выходы узлов скрытого слоя:

$$\tilde{y}^*(1), \dots, \tilde{y}^*(n).$$

Если ИНС содержат более одного скрытого слоя, процедуры коррекции матриц весов этих слоев будут иметь аналогичный вид и использовать как желаемые выходные сигналы данного слоя, так и выходные сигналы предыдущего скрытого слоя [12].

**Литература:** 1. *Rojas R.* Theorie der neuronalen Netze Springer. Verlag, Berlin: Heidelberg, New York. – 1997. – 446 s. 2. *Bishop C.M.* Neural networks for pattern recognition. Oxford: University Press. – 1995. – 482 p. 3. *Scherer A.* Neuronale Netze. Grundlagen und Anwendungen. Braunschweig/Wiesbaden: Vieweg. – 1997. – 249 s. 4. *Chen C. LP.* A rapid supervised learning neural network for function interpolation and approximation // IEEE Trans. Neural Networks. – 1996. – V.7. – №5. – P.1220-1229. 5. *Narendra K. S., Parthasarathy K.* Identification and control of dynamical systems using neural networks // IEEE Trans. Neural Networks. – 1990. – V.1. – №1. – P.4-27. 6. *Tiguni Y., Sakai H., Tokomura H.* A real-time learning algorithm for a multilayered neural network based on the extended Kalman filter // IEEE Trans. Signal Processing. – 1992. – V.40. – P.959-966. 7. *Nelles O., Ernst S., Isermann R.* Neuronale Netze zur Identifikation nichtlinearer, dynamischer Systeme: Ein Ueberblick // Automatisierungstechnik. – 1997. – V.45. – №6. – S.251-262. 8. *Chen S., Billings S. A.* Neural network for nonlinear dynamic system modelling and identification // Int. J. Control. – 1992. – V.56. – №2. – P.319-346. 9. *Jagannathan S., Lewis F. L.* Multilayer discrete-time neural-net controller with guaranteed performance // IEEE Trans. Neural Networks. – 1996. – V.7. – №1. – P.107-130. 10. *Cichocki A., Unbehauen R.* Neural networks for optimization and signal processing, John Wiley & Sons. – 1997. – 521 p. 11. *Chen S., Cowan C. F.N., Grant P. M.* Orthogonal least squares learning algorithm for radial basis function networks // IEEE Trans. Neural Networks. – 1991. – V.2. – №2. – P.302-309. 12. *Wang G.-J., Chen C.-Ch.* A fast multilayer neural-network training algorithm based on the layer-by-layer optimizing procedures. // IEEE Trans. Neural Networks. – 1996. – V.7. – №3. – P.768-775.

Поступила в редколлегия 28.12.97

**Руденко Олег Григорьевич**, д-р техн. наук, профессор, зав. кафедрой ЭВМ ХТУРЭ. Научные интересы: адаптивные системы, нейронные сети. Увлечения: изобразительное искусство, южноамериканская литература. Адрес: 310726, Украина, Харьков, пр. Ленина, 14, тел. (0572) 47-15-12.

**Штефан Андреас**, д-р-инженер, руководитель фирмы «Dr. Stephan & Parnter, System- und Softwarehaus», Ильменау, Германия. Научные интересы: адаптивные системы. Увлечения: путешествия. Тел. 84-10-67.