Таблица 2

Полезный эффект от применения предложенного корректирующего кода заключается в повышении верности считываемой информации и снижении требований к аппаратуре обеспечения.

Литература: 1. *Старичено О., Голуб В.* Коды и кодирование информации, штриховое кодирование выбор и применение штриховых кодов//Руководящий нормативный документ по стандартизации КНД 50-051-95. К.: Госстандарт Украины. 1995. 32 с. 2. *Али К. Абуд Аль-Амери.* Способы и средства помехоустойчивого штрихового кодирования – декодирования алфавитно-цифровой информации. К.1994. 24 с. 3. *Шувалов В. П., Захарченко Н. В. И др.* / Под ред. Шувалова В. П. Передача дискретных сообщений. М.: Радио и связь, 1990. 142 с. 4. *Зюко А. Г., Кловский Д. Д., Назаров М. В., Финк Л. М.* Теория передачи сигналов. М.: Радио и связь, 1986. 242с.

Поступила в редколлегию 12.03.98

Голуб Владимир Иванович, начальник инженерного центр украинского объединения почтовой связи “УКРПОЧТА”. Адрес: 252001, Украина, Киев-1, Хрещатик, 22, тел. (044) 228-37-12.

Жамхарьян Александр Сергеевич, аспирант кафедры сети связи, НИЛ-20. Адрес: 270021, Украина, Одесса, ул. Кузнечная, 1, тел. (8-0482) 23-61-80.

Фомина Ольга Владимировна, аспирант кафедры сети связи, НИЛ-20. Адрес: 270021, Украина, Одесса, ул. Кузнечная, 1, тел. (8-0482) 23-61-80.

УДК 681.51.015:519.7

ОБУЧЕНИЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ В СТОХАСТИЧЕСКИХ УСЛОВИЯХ

ВОРОБЬЕВ С.А.

Рассматривается задача обучения искусственных нейронных сетей в стохастических условиях для прямонаправленных сетей типа радиально-симметричной сети и многослойного перцептрона. При этом, при синтезе алгоритмов настройки синаптических весов сети, важным является требование наличия фильтрующих свойств у разрабатываемых алгоритмов. Исследуется сходимость предлагаемых алгоритмов в стохастических условиях обучения.

В последние годы для решения задач идентификации, моделирования и управления нелинейными системами широко применяются искусственные нейронные сети [1-4]. Основными задачами при проектировании нейронной сети являются выбор ее архитектуры и типа активационных функций нейронов, разработка алгоритмов настройки синаптических весов нейронов (алгоритмов обучения). Подробнее остановимся на задаче обучения нейронной сети в стохастических условиях.

Градиентная процедура настройки синаптических весов имеет вид

$$\begin{aligned}\omega(k+1) &= \omega(k)+\eta(k)\varepsilon(k,\omega)\nabla_{\omega} f(\varphi(k),\omega) = \\ &= \omega(k)+\eta(k)\varepsilon(k,\omega)G(\varphi(k),\omega).\end{aligned} \qquad (1)$$

где $\omega(k)=(\omega_1(k),\omega_2(k),...,\omega_{n_\omega}(k))^T$ – вектор синаптических весов нейронной сети; $\eta(k)$ – параметр шага поиска, принимаемый чаще всего постоянным; $\nabla_{\omega} f(\varphi(k),\omega)=G(\varphi(k),\omega)$ – градиент функции $f(\varphi(k),\omega)$ по синаптическим весам; $\varphi(k)=(y(k-1), ..., y(k-n), u(k-1),...,u(k-p))^T$; $f(\bullet)$ – некоторая функция;

$$\varepsilon(k,\omega)=y(k)-\hat{y}(k,\omega) \qquad (2)$$

– ошибка идентификации; $y(k)$ и $u(k)$ – соответственно выход и вход системы; $\hat{y}(k,\omega)=f(\varphi(k),\omega)$.

Далее остановимся на более конкретных задачах настройки нейронных сетей в стохастических условиях. Рассмотрим радиально-симметричную искусственную нейронную сеть (RBFN-сеть), которая состоит из трех слоев, называемых: входным, скрытым и выходным [5]. Функцией первого слоя является простое прохождение входного сигнала на уровень скрытого слоя, который выполняет нелинейное преобразование пространства входных сигналов в новое пространство. При этом преобразование строится с помощью активационной функции гауссовского типа

$$H_l = \exp[-\|x-c_l\|^2/\delta_l^2], \quad l=1,2,...,L,$$

где $\|\bullet\|$ – евклидова норма вектора; c_l и δ_l – соответственно центр и ширина гауссовской функции. Выходной слой сети представляет собой линейную комбинацию выходов нейронов скрытого слоя

$$y_m(k)=\sum_{l=1}^{L}\omega_{lm}(k)H_l(k), \quad m=1,2,...,s.$$

Здесь $\omega_{lm}(k)$ – настраиваемые весовые множители между скрытым и выходным слоем сети (синаптические веса). Таким образом, вектор настраиваемых синаптических весов m-го нейрона выходного слоя имеет вид

$$\hat{\omega}_m(k)=(\hat{\omega}_{1m}(k),\hat{\omega}_{2m}(k),...,\hat{\omega}_{Lm}(k))^T.$$

Для идентификации детерминированных объектов в теории искусственных нейронных сетей применяется алгоритм Уидроу-Хоффа [6]:

$$\hat{\omega}_m(k+1)=\hat{\omega}_m(k)+\frac{H(k)\varepsilon(k,\hat{\omega}_m)}{\|H(k)\|^2}H(k), \qquad (3)$$

где $H(k)=(H_1(k),H_2(k),...,H_L(k))^T$, $\varepsilon(k,\hat{\omega}_m)=$
$=y_m(k)-\hat{\omega}_m^T(k)H(k)$.

Однако при работе в стохастических условиях алгоритм (3) должен также обладать фильтрующими свойствами, что можно достичь, введя в (3) параметр сглаживания α:

$$\begin{cases}\hat{\omega}_m(k+1)=\hat{\omega}_m(k)+ar^{-1}(k)\varepsilon(k,\hat{\omega}_m)H(k),\\ r(k)=\alpha r(k-1)+\|H(k)\|^2,\\ 0\le\alpha\le 1,\quad 0<a<2,\quad r(0)=1.\end{cases} \tag{4}$$

Алгоритм (4) совпадает с (3) при $\alpha=0$ и с процедурой Гудвина-Рэмеджа-Кейнеса [7] при $\alpha=1$.

Проведем анализ сходимости алгоритма (4). Запишем (4) относительно ошибки настройки синаптических весов $\tilde{\omega}_m(k)=\omega_m(k)-\hat{\omega}_m(k)$:

$$\begin{aligned}&\tilde{\omega}_m(k)=\tilde{\omega}_m(k-1)+ar^{-1}(k)H(k)\times\\&\times\varepsilon(k,\hat{\omega}_m).\end{aligned} \tag{5}$$

Тогда функция Ляпунова примет вид

$$\begin{aligned}&V(k)=\tilde{\omega}_m^T(k)\tilde{\omega}_m(k)=V(k-1)+\\&+2ar^{-1}(k)\tilde{\omega}_m^T(k-1)H(k)(\varepsilon(k,\hat{\omega}_m)-\\&-\xi_m(k))+2ar^{-1}(k)\tilde{\omega}_m^T(k-1)H(k)\xi_m(k)+\\&+a^2r^{-2}(k)\|H(k)\|^2(\varepsilon(k,\hat{\omega}_m)-\xi_m(k))^2+\\&+2\xi_m(k)(\varepsilon(k,\hat{\omega}_m)-\xi_m(k))+\xi_m^2(k),\end{aligned} \tag{6}$$

где $\xi_m(k)$ – помеха с нулевым средним и ограниченной дисперсией $\sigma_{\xi_m}^2$. Вводя обозначения $\tilde{\xi}_m(k)=\varepsilon(k,\hat{\omega}_m)-\xi_m(k)$ и $b(k)=-H^T(k)\tilde{\omega}_m(k-1)$ и усредняя (6) по $\xi_m(k)$, получаем

$$\begin{aligned}&M\{V(k)/\Im_k\}=V(k-1)-2ar^{-1}(k)\times\\&\times b(k)\tilde{\xi}_m(k)+a^2r^{-2}(k)\|H(k)\|^2\tilde{\xi}_m^2(k)+\\&a^2r^{-2}(k)\|H(k)\|^2\sigma_{\xi_m}^2.\end{aligned} \tag{7}$$

Здесь $M\{\bullet/\bullet\}$ – символ условного математического ожидания; $\Im_k$ – σ-алгебра, порожденная величинами $\{y(0),y(1),...,y(k)\}$, $\Im_0\subset\Im_1\subset...\subset\Im_k$.

Далее, как и в [7], будем рассматривать последовательность (7) как супермартингал. При этом принципиальным моментом является выполнение на каждом шаге настройки сети условия $r^2(k)>r(k)r(k-1)$ или, $(\alpha(k)r(k-1)+\|H(k)\|^2)^2>(\alpha(k)r(k-1)+\|H(k)\|^2)r(k-1)$.

Следовательно, нахождение требуемого $\alpha(k)$ связано с решением неравенства

$$\begin{aligned}&\alpha^2(k)+\alpha(k)(2r(k-1)\|H(k)\|^2-\\&-r^2(k-1))r^{-2}(k-1)+(\|H(k)\|^4-\\&-\|H(k)\|^2r(k-1))r^{-2}(k-1)>0,\end{aligned}$$

которое принимает вид

$$|\beta(k)|>\sqrt{m(k)},\tag{8}$$

где

$$\beta(k)=\alpha(k)+\frac{2\|H(k)\|^2-r(k-1)}{2r(k-1)},\tag{9}$$

$$\begin{aligned}&m(k)=\left(\frac{2\|H(k)\|^2-r^2(k-1)}{2r(k-1)}\right)^2-\\&-\frac{\|H(k)\|^4-\|H(k)\|^2r(k-1)}{r^2(k-1)}=\frac{1}{4}.\end{aligned}\tag{10}$$

Из (9), (10) получаем

$$\alpha(k)+\frac{2\|H(k)\|^2-r(k-1)}{2r(k-1)}>\frac{1}{2},$$

откуда имеем

$$1-\frac{\|H(k)\|^2}{r(k-1)}<\alpha(k)\le 1.\tag{11}$$

Из (11) следует очевидное соотношение

$$(1-\alpha(k))\sum_{j=1}^{k}\alpha^{k-j-1}\|H(k)\|^2\le\|H(k)\|^2,$$

являющееся решением уравнения
$\alpha(k)=(\alpha_0\alpha(k-1)+(1-\alpha_0),\ 0\le\alpha_0\le 1.$

Следовательно, сходимость алгоритма (4) может обеспечиваться либо постоянным ростом величины $\|H(k)\|^2$, что можно сделать путем настройки центра и ширины гауссовской функции, либо соответствующим изменением параметра сглаживания $\alpha(k)$, который должен увеличиваться от 0 до 1.

Далее, используя в качестве характеристики скорости сходимости изменение функции Ляпунова на каждом шаге, можно оценить влияние $\alpha(k)$ на скорость сходимости алгоритма (9). Значение, обеспечивающее максимальную скорость сходимости, равно $r(k)=a\|H(k)\|^2$.

Кроме того, $r(k)$ определяется вторым соотношением из (4), откуда видно, что максимальное быстродействие алгоритма достигается при $a=1$ и $\alpha(k)=0$, т. е. обеспечивается процедурой (3).

Следовательно, в стохастических условиях для RBFN-сети с алгоритмом настройки синаптических весов (4) следящие свойства алгоритма (4), которые определяют скорость обучения сети, вступают в противоречие с его фильтрующими свойствами. Поэтому в процессе настройки синаптических весов целесообразно начинать работу с малых значений параметра $\alpha(k)$, обеспечивая тем самым высокую скорость обучения сети, далее увеличивать его до уровня, обеспечивающего компромисс между фильтрующими и следящими свойствами алгоритма (4).

Рассмотрим теперь задачу настройки многослойного перцептрона вида [8]:

$$y(k+1) = \sum_{i=1}^{H_q} \omega_i^{(q)} f\left[\sum_{t=1}^{H_q-1} \omega_{it}^{(q-1)} f\left[\ldots\right.\right.$$
$$\left.\left.f\left[\sum_{j=1}^{n} \omega_{sj}^{(1)} y(k-j+1) + \sum_{j=1}^{p} \omega_{s,n+j}^{(1)} u(k-j+1)\right]\ldots\right]\right], \quad (12)$$

где $\omega_{sj}^{(k)}$ – синаптические веса k-го слоя; $f(\bullet) = \{1+\exp[-(\bullet)]\}^{-1}$ или $f(\bullet) = \tanh(\bullet)$ – сигмоидальная функция активации или функция гиперболического тангенса соответственно. Запишем нелинейный одношаговый вариант алгоритма Марквардта [9]:

$$\omega(k+1) = \omega(k) + (G(\varphi(k),\omega)G^T(\varphi(k),\omega) + \\ + \rho(k)E)^{-1} G(\varphi(k),\omega)\varepsilon(k,\omega) \ . \quad (13)$$

Здесь $\omega(k) = (\omega_1(k), \omega_2(k), \ldots, \omega_{n_\omega}(k))^T$ $-(1\times n_\omega)$ -вектор всех синаптических весов; $n_\omega = \sum_{i=0}^{q}(n_i+1)n_{i+1} + n_{q-1}n_q$, q – количество слоев сети; $n_i, 1 \le q \le n_i$ – количество нейронов i-го слоя; $\varepsilon(k,\omega)$ определяется соотношением (2); $\rho(k) > 0$, E – единичная матрица, $G(\varphi(k),\omega) = \nabla_\omega f(\varphi(k),\omega)$. Используя соотношения

$$\begin{cases} \lim\limits_{\rho(k)\to 0} (G(\varphi(k),\omega)G^T(\varphi(k),\omega) + \\ + \rho(k)E)^{-1} = (G(\varphi(k),\omega)G^T(\varphi(k),\omega))^+, \\ (G(\varphi(k),\omega)G^T(\varphi(k),\omega))^+ G(\varphi(k),\omega) = \\ = (G(\varphi(k),\omega))^+ = G(\varphi(k),\omega)\|G(\varphi(k),\omega)\|^{-2}, \end{cases}$$

запишем оптимальный по быстродействию вариант алгоритма (13) в виде

$$\omega(k+1) = \omega(k) + \frac{y(k) - \hat{y}(k,\omega)}{\|G(\varphi(k),\omega)\|^2} G(\varphi(k),\omega), \quad (14)$$

который в линейном случае совпадает с алгоритмом настройки синаптических весов Уидроу-Хоффа [7]. Как и в предыдущем случае, для того, чтобы придать дополнительные сглаживающие свойства алгоритму (14), необходимые при его использовании для настройки MLP-сети в задаче идентификации нелинейного нестационарного стохастического объекта, введем следующую экспоненциально-взвешенную модификацию:

$$\begin{cases} \omega(k+1) = \omega(k) + r^{-1}(k)(y(k) - \hat{y}(k,\omega))G(\varphi(k),\omega), \\ r(k) = \alpha r(k-1) + \|G(\varphi(k),\omega)\|^2, \\ 0 \le \alpha \le 1, \quad r(0) = 1. \end{cases} \quad (15)$$

Анализ сходимости процедуры (15) можно провести с использованием той же техники, что была применена ранее. Однако следует учесть, что $G(\varphi(k),\omega)$ – это градиент активационной функции, а алгоритм (15) в целом является нелинейным.

Литература: 1. *Narendra K.S., Parthasarathy K.* Identification and control of dynamical systems using neural networks // IEEE Trans. on Neural Networks. 1990. Vol.1, N1. P.4-26. 2. *Sudharsanan S.I., Sudareshan M.K.* Supervised training of dynamical neural networks for associative memory design and identification of nonlinear maps // Int. J. Neural Systems. 1994. Vol.5, N3. P.165-180. 3. *Pham D.T., Liu X.* Neural Networks for Identification, Prediction and Control. London: Springer-Verlag, 1995. 238p. 4. *Chen S., Billings S.A.* Neural networks for nonlinear dynamic system modeling and identification // Int. J. Control. 1992. Vol.56, N2. P.319-346. 6. *Moody J., Darken C.J.* Fast learning in networks of locally-tuned processing units // Neural Computation. 1989. N1. P.281-294. 6. *Rojas R. Neural* Networks. A Systematic Introduction. Berlin: Springer-Verlag, 1996. 502p. 7. *Goodwin G.C., Ramadge P.J., Caines P.E.* A globally convergent adaptive predictor // Automatica. 1981. Vol.17, N1. P.135-140. 8. *Tan Y., van Cauwenberghe A.* Nonlinear one-step-ahead control using neural networks: control strategy and stability design // Automatica. 1996. Vol.32, N12. P.1701-1706. 9. *Marquardt D.* An algorithm for least-squares estimation on nonlinear parameters // SIAM J. Appl. Math. 1963. N11. P.431-441.

Поступила в редколлегию 05.03.98

Воробьев Сергей Анатольевич, канд. техн. наук, старший научный сотрудник проблемной научно-исследовательской лаборатории автоматизированных систем управления ХТУРЭ. Научные интересы: искусственные нейронные сети, фильтрация и прогнозирование нестационарных процессов, фракталы и фрактальная размерность, иррационализм. Хобби: психология, иностранные языки, музыка. Служебный адрес: 310726, Украина, Харьков, пр. Ленина, 14; тел. (0572)409890, (0572)434278, e-mail: svor@kture.kharkov.ua.