УДК 681.513

Достаточные условия сходимости одного алгоритма обучения нейронной сети при идентификации нелинейно параметризуемых систем

С.А. Николаенко[1]

Аннотация – **The asymptotical behavior of neural networks used for the identification of nonlinearly parameterized systems is addressed in this paper. Sufficient conditions for the convergence of a learning algorithm are derived.**

Ключевые слова – **Идентификация, Нейронная сеть, Последовательное обучение, Градиентный алгоритм, Сходимость.**

I. Введение

Для управления сложными техническими и технологическими системами на базе настраиваемых моделей в последнее десятилетия в качестве таких моделей начинают использовать нейронные сети, ориентированные на обучение в реальном времени [1], [2]. Такие модели способны реализовать зависимости, нелинейные не только по переменным, которые входят в уравнение системы, но и по параметрам, представляющим собой неизвестные коэффициенты этого уравнения. А это существенно затрудняет обоснование сходимости процесса обучения.

Несмотря на обилие публикаций в области нейросетевого моделирования имеется не так уж много работ, посвященных анализу сходимости алгоритмов последовательного обучения нейронных сетей. Наиболее существенный результат получен в работах [3]-[5]. К сожалению, авторам этих работ удалось доказать сходимость таких алгоритмов, ограничившись случаем, когда обучающая последовательность формируется из конечной выборки; при этом сигналы на входе нейросети периодически повторяются, тогда как в задачах адаптивной идентификации такое ограничение оказывается неприемлемым.

Данная работа продолжит исследования в области анализа сходимости стандартных градиентных алгоритмов обучения нейросетей на бесконечной нестохастической обучающей последовательности, начатые в [6], [7]. Цель работы – установление условий, гарантирующих сходимость алгоритма с постоянным шаговым коэффициентом, как и в [4].

II. Постановка задачи

Пусть имеется некоторая динамическая система, описываемая разностным уравнением

$$y_n = F(Y_{n-1}, U_{n-1}) + v_n. \tag{1}$$

В этом уравнении $Y_n \in R$ – изеряемый выход в n-й момент времени,

$$Y_{n\text{-}1} = [y_{n-1}, ..., y_{n-N_y}]^T \tag{2}$$

– N_y-мерный вектор предыдущих выходов,

$$U_{n\text{-}1} = [u_{n-1}, ..., u_{n-N_u}]^T \tag{3}$$

– N_u-мерный вектор управлений. Далее, v_n – возмущение (помеха) в момент n, а $F(\cdot,\cdot): R^{N_y} \times R^{N_u} \to R$ – неизвестная нелинейная функция.

В качестве модели нелинейности $F(\cdot,\cdot)$ возьмем нейронную сеть, уравнение которой запишем, как и в [2], в таком общем виде:

$$y_n^{mod} = NN(X_{n-1}, w). \tag{4}$$

Здесь y_n^{mod} – выход нейросетевой модели; $X_{n\text{-}1} = [Y_{n\text{-}1}^T, U_{n-1}^T]^T - (N_y + N_u)$ –мерный вектор сигналов, поступающих на вход этой модели; w – вектор весов всех слоев нейронной сети; $NN(\cdot,\cdot)$ – нелинейная функция, реализуемая данной моделью.

Ошибка аппроксимации

$$e_n(w) = y_n - NN(X_{n-1}, w) \tag{5}$$

при каждом $X_{n-1} \in R^{N_y+N_u}$ зависит от структуры и параметров нейросети (от выбранной функции $NN(\cdot,\cdot)$ и вектора w).

Предположим, что при некотором пока неизвестном $w = w^*$ ошибка аппроксимации $e(w^*) = y - NN(X, w^*)$ будет отсутствовать:

$$F(X) \equiv NN(X, w^*). \tag{6}$$

Условие (6) можно рассматривать как условие полной адекватности нейросетевой модели (4) идентифицируемому объекту (1).

Для экспериментальной проверки асимптотических свойств алгоритма (7) проводилось моделирование процесса обучения простейшей нейросети, содержащей по одному нейрону во входном и скрытом слоях. Рассматривался один специальный случай, когда неизвестными полагались только величина смещения в скрытом слое и коэффициент синаптической связи между первым и скрытым слоями. Именно в этом случае выполнение условия (13) сходимости алгоритма обучения (7) гарантируется.

Характер поведения функции V_n и ошибки e_n, который наблюдался в процессе моделирования,

[1] Международный научно-учебный центр информационных технологий и систем НАН Украины и МОНМС Украины, пр. Академика Глушкова, 40, Киев-207, 03680, УКРАИНА, E-mail: s_nikolaenko@ukr.net

показал, что $V_n \to 0$ при $n \to \infty$, как и должно быть; при этом $\lim_{n\to\infty} w_n = w^*$.

Для обучения нейросети в реальном времени используем стандартный алгоритм градиентного типа

$$w_{n+1} = w_n - \eta e_n(w_n) \operatorname{grad}_w e_n(w_n). \tag{7}$$

в котором $\eta \equiv \text{const} > 0$ – шаговый коэффициент.

Задача состоит в том, чтобы в условии предположения (6) установить условие сходимости алгоритма (7) в смысле существования предела

$$\lim_{n\to\infty} w_n = w_\infty. \tag{8}$$

III. Основные результаты

Для исследования сходимости алгоритма (7) вводится функция

$$V_n = \left\| w^* - w_n \right\|^2. \tag{9}$$

Согласно [6], [7] при условии, что $\{V_n\}$– невозрастающая последовательность, т.е.

$$V_n \le V_{n-1} \quad \text{для всех } n, \tag{10}$$

при выполнении (8) и $U_n \equiv 0$ ошибка $e_n(w_n)$ в асимптоте стремится к нулю,

$$\lim_{n\to\infty} e_n(w_n) = 0; \tag{11}$$

при этом алгоритм (7) сходится. Однако такая сходимость вовсе еще не означает, что $w_\infty = w^*$.

Модельные эксперименты показали, однако, что требование (10) невозрастания V_n в принципе совершенно не обязательно для обеспечения асимптотического свойства (8) алгоритма (7) последовательного обучения нейросетевой модели (4). Для того, чтобы этот алгоритм сходился, достаточно существование предела

$$\lim_{n\to\infty} V_n < V_0. \tag{12}$$

Оказывается, что на определенных последовательностях $\{X_n\}$ алгоритм (7) может сходиться даже тогда, когда $V_n \to V_\infty > V_0$ при $n \to \infty$. В то же время не исключается (это подтверждают результаты моделирования), что на отдельных последовательностях $\{X_n\}$ предел $\{V_n\}$ при $n \to \infty$ может вовсе не существовать.

Основной результат, устанавливающий достаточные условия сходимости алгоритма обучения (7), формулируется следующим образом.

У т в е р ж д е н и е. Пусть $v_n \equiv 0$ и выполнено требование (6). Тогда если существует число $\Lambda > 0$ такое, что

$$\begin{aligned}&[NN(X, w^*) - NN(X, w)](w^* - w)^T \operatorname{grad}_w NN(X, w) \ge \\ &\ge \Lambda [NN(X, w^*) - NN(X, w)]^2 \left\| \operatorname{grad}_w NN(X, w) \right\|^2 \ge 0\end{aligned} \tag{13}$$

для всех X, w, то при любом η, удовлетворяющем ограни-чению

$$0 < \eta < \frac{1}{2\Lambda} \tag{14}$$

алгоритм (7) будет сходиться.

Условие (13), фигурирующее в утверждении, налагает довольно жесткие ограничения на $NN(X, w)$. Это условие уместно интерпретировать как своеобразное требование сильной монотонности функции $NN(X, w)$ в направлении вектора w. По существу, это означает, что

$$\inf \frac{[NN(X, w^*) - NN(X, w)](w^* - w)^T \operatorname{grad}_w NN(X, w)}{[NN(X, w^*) - NN(X, w)]^2 \left\| \operatorname{grad}_w NN(X, w) \right\|^2} > $$

$$> 0$$

(15)

для всех X и $w \ne w^*$.

IV. Вывод

При достаточно малом $\eta > 0$ сходимость градиентного алгоритма последовательного обучения нейросетевой модели для идентификации нелинейно параметризуемой системы при отсутствии помех и полной адекватности этой модели гарантируется, если функция, реализующая преобразование сигналов в нейросетях, обладает свойством так называемой сильной монотонности по направлению ее весового вектора.

Список ссылок

[1] Е. В. Бодянский, и О. В. Запорожец, "Адаптивный регулятор для нелинейного динамического объекта," *Изв. РАН. Теория и системы управления*, сс. 92-96, №2. 2002.

[2] Y. Z. Tsypkin, J. D. Mason, E. D. Avedyan, K. Warwick, and I.K. Levin, "Neural Networks for Identification of Nonlinear Systems Under Random Piecewise Poynomial Disturbances," *IEEE Trans. Neural Networks*, vol. 10, pp. 303-311, Mar. 1999.

[3] W. Wu, G. Feng, and X. Li, "Training multilayer perceptrons via minimization of ridge functions," *Advances in Comput. Mathematics*, vol. 17, pp. 331-347, 2002.

[4] W. Wu, G. Feng, X. Li, and Y. Xu, "Deterministic convergence of an online gradient method for BP neural networks," *IEEE Trans. Neural Networks*, vol. 16, pp. 1-9, 2005.

[5] H. Zhang, W. Wu, F. Liu, and M. Yao, "Boundedness and convergence of online gradient method with penalty for feedforward neural networks," *IEEE Trans. Neural Networks*, vol. 20, pp. 1050-1054, 2009.

[6] С. А. Николаенко, "Асимптотические свойства алгоритмов последовательного обучения нейросетевых моделей в нестохастической среде," в *Тез. доп. 17-ї Міжнародної конференції з автоматичного управління «Автоматика-2010»*, ХНУРЭ, Харьков, Сен. 2010, сс. 142-144.

[7] В. Н. Азарсков, Л. С. Житецкий, С. А. Николаенко, "О сходимости стандартных алгоритмов последовательного обучения нейросетевых моделей в нестохастической среде," в *Сб. научн. трудов 13-й Всеросийской научно-технической конференции «Нейроуправление-2011»*, НИЯУ МИФИ, Москва, Янв. 2011, сс. 258-260.