

УДК 004.032.26

А.А. КРОЩЕНКО

О СПОСОБАХ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ МЕТОДА ОБРАТНОГО РАСПРОСТРАНЕНИЯ ОШИБКИ

Для обучения нейронных сетей успешно применяется так называемый метод обратного распространения ошибки, представляющий собой следующую последовательность шагов [1]:

1. Задается шаг обучения α ($0 < \alpha < 1$) и желаемая среднеквадратичная ошибка нейронной сети E_m .

2. Случайным образом инициализируются весовые коэффициенты и пороговые значения нейронной сети.

3. Последовательно подаются образы из обучающей выборки на вход нейронной сети. При этом для каждого входного образа выполняются следующие действия:

а) производится фаза прямого распространения входного образа по нейронной сети. При этом вычисляется выходная активность всех нейронных элементов сети:

$$y_j = F\left(\sum_i \omega_{ij} y_i - T_j\right),$$

где индекс j характеризует нейроны следующего слоя по отношению к слою i .

б) производится фаза обратного распространения сигнала, в результате которой определяется ошибка γ_j , $j=1,2,\dots$ нейронных элементов для всех слоев сети. При этом соответственно для выходного и скрытого слоев:

$$\begin{aligned} \gamma_j &= y_j - t_j, \\ \gamma_j &= \sum_i \gamma_i F'(S_i) \omega_{ji}. \end{aligned}$$

В последнем выражении индекс i характеризует нейронные элементы следующего слоя по отношению к слою j .

в) Для каждого слоя нейронной сети происходит изменение весовых коэффициентов и порогов нейронных элементов:

$$\begin{aligned} \omega_{ij}(t+1) &= \omega_{ij}(t) - \alpha \gamma_j F'(S_j) y_i, \\ T_j(t+1) &= T_j(t) + \alpha \gamma_j F'(S_j). \end{aligned}$$

4. Вычисляется суммарная среднеквадратичная ошибка нейронной сети:

$$E = \frac{1}{2} \sum_{k=1}^L \sum_j (y_j^k - t_j^k)^2,$$

где L – размерность обучающей выборки.

5. Если $E > E_m$ то происходит переход к шагу 3 алгоритма. В противном случае алгоритм обратного распространения ошибки заканчивается.

Метод обратного распространения ошибки является эффективным методом, но, учитывая, что в его основе лежит градиентный метод, имеет медленную скорость сходимости.

К факторам, влияющим на скорость, относятся: используемый шаг обучения, наличие локальных минимумов функции среднеквадратичной ошибки, инициализация весов НС, архитектура самой сети.

Исходя из перечисленных факторов, нами применялись следующие подходы, улучшающие скорость сходимости:

1. Использование адаптивного шага обучения

Шаг в этом случае для каждого слоя НС будет определяться по формуле

$$\alpha(t) = \frac{\sum_j \gamma_j^2 F'(S_j)}{F'(0) \left(1 + \sum_i y_i^2 \right) \left(\sum_j \gamma_j^2 F'(S_j) \right)^2}$$

где γ_j – ошибка j-ого нейронного элемента.

Следует учитывать тот факт, что если постоянный шаг выбран удачно, то использование адаптивного шага не улучшит быстроту сходимости метода.

2. Модификация градиентного метода с моментом

Изменение синаптических связей при использовании такого подхода происходит в соответствии со следующим выражением:

$$\Delta \omega_{ij}(t+1) = -\alpha \gamma_j F'(S_j) y_i + \gamma \Delta \omega_{ij}(t)$$

где γ – величина постоянная, которая называется моментным параметром.

Значение моментного параметра выбирается из диапазона [0,1]. На практике часто используется значение $\gamma = 0.9$ [2].

Использование данной модификации позволяет на порядок сократить количество итерации при обучении.

3. Использование гетерогенных нейронных сетей

Гетерогенная нейронная сеть – сеть, слои которой имеют различные функции активации нейронных элементов. Нами использовалась нейронная сеть, имеющая на выходном слое отличную от скрытых слоев функцию активации.

Данный подход дает самый лучший результат для проведенного нами численного эксперимента, позволяя на несколько порядков сократить количество эпох обучения НС.

Вычислительный эксперимент. Нами проводился численный эксперимент, заключающийся в обучении нейронной сети распознаванию функции логического «И». Использовалась многослойная нейронная сеть с

фиксированной архитектурой, имеющей один входной слой с двумя нейронами и два обрабатывающих слоя с 4 и 1 нейронами соответственно. Во всех случаях использовался адаптивный шаг обучения и модификация алгоритма обратного распространения с моментом ($\gamma = 0.9$), желаемая среднеквадратичная ошибка – $E_m = 10^{-7}$.

В приведенной ниже таблице представлены результаты эксперимента. Число итераций было получено как среднее из 100 экспериментов с одинаковыми параметрами конфигурации сети.

<i>Функции активации</i>	<i>Число итераций</i>
sigm (all)	949511
sigm (hidden) – bisigm (output)	427852
sigm (hidden) – huptan (output)	208994
sigm (hidden) – lin (output)	514

Используемые функции активации: sigm – сигмоидная, bisigm – бисигмоидная, huptan – гиперболический тангенс, lin – линейная. Слои: all – все, hidden – скрытый, output – выходной.

Как видно из результатов, полученных для данной, достаточно простой задачи, наиболее приемлемым является использование гетерогенной нейронной сети с сигмоидной и линейной функциями активации для скрытого и выходного слоя соответственно. Стоит отметить, что использование подобной конфигурации сети позволило сократить число итерации приблизительно на 4 порядка от первоначально достигнутого результата (для сети без модификации с моментом).

Таким образом, использование гетерогенной нейронной сети наравне с применением модификации алгоритма обратного распространения ошибки с моментом способно для некоторых задач существенно ускорить процесс обучения.

СПИСОК ЛИТЕРАТУРЫ

1. Головки, В.А. Нейроинтеллект: теория и применение. Книга 1: Организация и обучение нейронных сетей с прямыми и обратными связями / Головки В.А. – Брест.: Изд. БПИ, 1999. – 264 с.
2. Hertz, J. Introduction to the theory of neural computation / J. Hertz, A. Knogh, R. Palmer. – Addison Wesley Publishing Company, 1991. – 327 p.