ближенно определить как

$$\alpha = \frac{t_s + t_2}{4} \qquad (25)$$

Тогда общее время вычислений

$$t_0 = \frac{(4+L)(t_s + t_2)}{4} \qquad (26)$$

Для эффективности вычислений в конвейерном режиме необходимо выполнение следующего неравенства:

$$\frac{(4+L)(t_s + t_2)}{4} < t_s L \qquad (27)$$

Выражая из последнего выражения L, получим

$$L > \frac{4(t_s + t_2)}{3t_s - t_2} \qquad (28)$$

Проведем анализ трудоемкости вычислений каждой стадии конвейера. Используя результаты предыдущего раздела, можно получить количество операций, вычисляемых каждым процессором

$$V(P_1) = m(2n+1) \qquad (29)$$

$$V(P_2) = (2m+1) \qquad (30)$$

$$V(P_3) = (4m+4) \qquad (31)$$

$$V(P_4) = m(4n+4) \qquad (32)$$

Отсюда следует, что различные ступени конвейера характеризуются различной сложностью вычислений. При этом время такта конвейера будет определяться временем прохождения самой медленной ступени P_4. Для нейтрализации этого недостатка можно предложить конвейерно-параллельную схему. Она заключается в том, что самая медленная ступень разбивается, например, на m параллельно работающих процессоров, как это показано на рис. 6 для четвертой ступени.

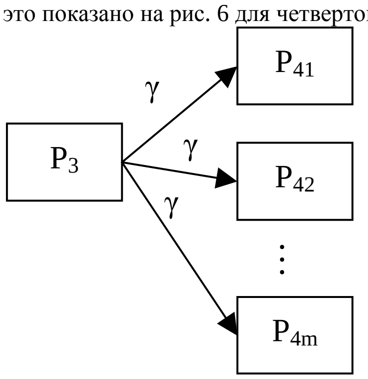

Рисунок 6 – Схема конвейерно-параллельной параллелизации.

В результате этого количество операций четвертой ступени конвейера можно сократить в среднем в m раз. Аналогичные действия можно применить для других ступеней конвейера, что позволяет повысить производительность вычислений.

Следующий вариант распараллеливания алгоритма обучения – групповая параллелизация. Такая схема предполагает разбиение обучающего множества на K групп и использования для каждой группы своей нейронной сети (рис. 7).

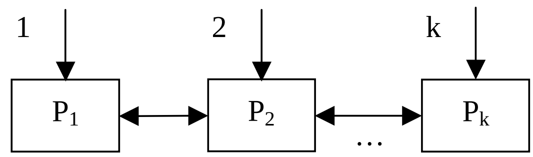

Рисунок 7 – Групповая параллелизация.

Количество образов в группе равняется L/K. Фаза межмодульного обмена в такой схеме происходит после подачи L/K образов в каждый процессор и заключается в изменении соответствующих синаптических связей. В качестве процессора P_i в такой схеме можно использовать последовательный процессор, параллельный процессор, рассмотренный в предыдущем разделе, и одну из конвейерных схем. Если применяется последовательный процессор, то производительность групповой параллелизации в K раз больше по сравнению с одним процессором. Аналогичное соотношение наблюдается для других схем параллелизации.

ЗАКЛЮЧЕНИЕ

Таким образом, в данной работе рассмотрены различные схемы реализации нейронных сетей на многопроцессорных системах. Приведены аналитические соотношения, позволяющие оценивать эффективность различных схем параллелизации. Окончательный выбор варианта пареллелизации зависит от архитектуры многопроцессорной системы и допустимых аппаратных издержек, затрачиваемых на реализацию соответствующего алгоритма параллелизации.

СПИСОК ИСПОЛЬЗОВАНЫХ ИСТОЧНИКОВ

1. Frank M. Thiesing, Ulrich Middelberg, Oliver Vornberger. Parallel Back-Propagation for Sales Prediction on Transputer System. Harrogate. UK. Proc. Of World Transputer Congress. 1995.
2. Головко В.А. Нейроинтеллект: теория и применение. Книга 1: Организация и обучение нейронных сетей с прямыми и обратными связями. Брест. Изд. БПИ, 1999 – 264 с.
3. Weigend A., Gershenfeld N. Time series prediction: forecasting the future and understanding past // Proceedings of the Santa Fe Institute. New Mexico: Addison–Wesley. – 1992. – 336 p.

УДК 681.324: 519.711.7

Махнист Л.П.

ОБУЧЕНИЕ НЕЙРОННЫХ СЕТЕЙ С ИСПОЛЬЗОВАНИЕМ МЕТОДА СОПРЯЖЕННЫХ ГРАДИЕНТОВ

ВВЕДЕНИЕ

Рассмотрим нейронную сеть, состоящую из *n* нейронных элементов распределительного слоя и m - выходного слоя (рис. 1).

Для данной сети каждый нейрон распределительного слоя имеет синаптические связи w_{ij}, $\left(i = \overline{1,n},\ j = \overline{1,m}\right)$ со всеми нейронами обрабатывающего слоя. В качестве нейронов выходного слоя используются элементы с некоторой функцией активации F [1, 2]. На вход сети подаются входные

Махнист Леонид Петрович. *К.т.н., доцент каф. высшей математики Брестского государственного технического университета.*
Беларусь, БГТУ, 224017, г. Брест, ул. Московская, 267.

образы – векторы $\overline{x^k} = \left(x_1^k, \ldots, x_n^k\right) \left(k = \overline{1, L}\right)$.

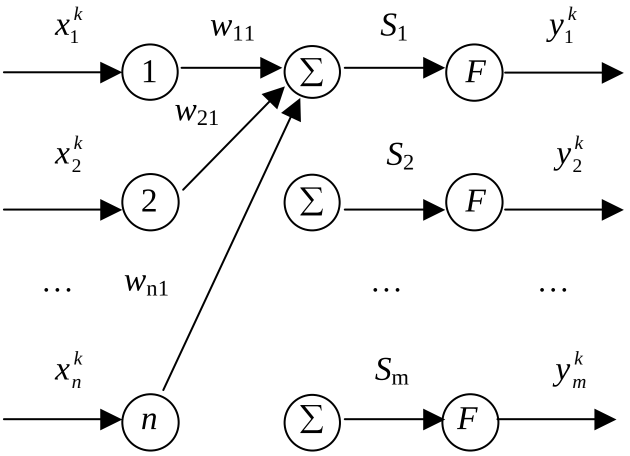

Рисунок 1 – Схема функционирования нейронной сети.

Выходное значение j-ого нейрона сети для k-ого образа определяется выражением:

$$y_j^k = F\left(S_j^k\right),$$

где

$$S_j^k = \sum_{i=1}^{n} w_{ij} x_i^k - T_j, \quad j = \overline{1, m}, \quad k = \overline{1, L}.$$

Задача обучения нейронной сети с фиксированной функцией активации F состоит в нахождении весовых коэффициентов w_{ij} $\left(i = \overline{1, n}, \; j = \overline{1, m}\right)$ и порогов нейронных элементов T_j $\left(j = \overline{1, m}\right)$, которые минимизируют некоторую ошибку сети E_S, как отклонение выходных значений y_j^k от эталонных значений t_j^k – j-ого нейрона сети для k-ого образа. В качестве ошибки сети можно рассмотреть "квадратичное отклонение" $E_S = \frac{1}{2}\sum_{k=1}^{L}\sum_{j=1}^{m}\left(y_j^k - t_j^k\right)^2$, которое будем называть квадратичной ошибкой сети.

Столбец

$$\overline{W} = \left(w_{11}, w_{21}, \ldots, w_{n1}, T_1, w_{12}, w_{22}, \ldots, w_{n2}, T_2, \ldots, w_{1m}, w_{2m}, \ldots, w_{nm}, T_m\right)^T$$

будем называть приближенным решением или просто решением системы (по методу наименьших квадратов):

$$F\left(\sum_{i=1}^{n} w_{ij} x_i^k - T_j\right) = t_j^k, \quad j = \overline{1, m}, \quad k = \overline{1, L},$$

если "квадратичное отклонение"

$$E_S = \frac{1}{2}\sum_{k=1}^{L}\sum_{j=1}^{m}\left(F\left(\sum_{i=1}^{n} w_{ij} x_i^k - T_j\right) - t_j^k\right)^2$$

достигает своего наименьшего значения. Для нахождения такого решения можно применять различные градиентные методы [1, 2], например, метод сопряженных градиентов и его модификации, которые будут рассмотрены ниже.

ВЫРАЖЕНИЯ КВАЗИОПТИМАЛЬНЫХ ПАРАМЕТРОВ ШАГА ОБУЧЕНИЯ НЕЙРОННОЙ СЕТИ С ИСПОЛЬЗОВАНИЕМ МЕТОДА СОПРЯЖЕННЫХ ГРАДИЕНТОВ

Получим выражения для квазиоптимальных параметров шага обучения нейронной сети с использованием метода сопряженных градиентов после подачи на вход сети нескольких образов $\overline{x^k} = \left(x_1^k, \ldots, x_n^k\right) \left(k = \overline{1, L}\right)$.

Рассмотрим дважды непрерывно дифференцируемую функцию $E_S(t)$ – ошибку сети, как функцию нескольких переменных:

$$E_S\left(w_{11}, w_{21}, \ldots, w_{n1}, T_1, w_{12}, w_{22}, \ldots, w_{n2}, T_2, \ldots, w_{1m}, w_{2m}, \ldots, w_{nm}, T_m\right)$$

Разложим функцию в ряд Тейлора, ограничиваясь частными производными второго порядка включительно:

$$E_S(t+1) = E_S(t) + \left(\nabla E_S(t), \; \overline{W}(t+1) - \overline{W}(t)\right) + \\ + \frac{1}{2}\left(\nabla^2 E_S(t) \cdot \left(\overline{W}(t+1) - \overline{W}(t)\right), \; \overline{W}(t+1) - \overline{W}(t)\right).$$

Учитывая, что в соответствии с идеей метода сопряженных градиентов

$$\overline{W}(t+1) = \overline{W}(t) - \alpha(t) \cdot \nabla E_S(t) + \beta(t) \cdot \left(\overline{W}(t) - \overline{W}(t-1)\right) \quad (1)$$

и введя обозначение $\overline{\Delta W}(t) = \overline{W}(t) - \overline{W}(t-1)$, получим

$$E_S(t+1) = E_S(t) + \left(\nabla E_S(t), \; -\alpha(t) \cdot \nabla E_S(t) + \beta(t) \cdot \overline{\Delta W}(t)\right) + \\ + \frac{1}{2}\left(\nabla^2 E_S(t) \cdot \left(-\alpha(t) \cdot \nabla E_S(t) + \beta(t) \cdot \overline{\Delta W}(t)\right), \right. \\ \left. -\alpha(t) \cdot \nabla E_S(t) + \beta(t) \cdot \overline{\Delta W}(t)\right) = \\ = E_S(t) - \alpha(t) \cdot \left(\nabla E_S(t), \; \nabla E_S(t)\right) + \\ + \beta(t) \cdot \left(\nabla E_S(t), \; \overline{\Delta W}(t)\right) + \\ + \frac{1}{2}\alpha^2(t)\left(\nabla^2 E_S(t) \cdot \nabla E_S(t), \; \nabla E_S(t)\right) - \\ - \alpha(t)\beta(t) \cdot \left(\nabla^2 E_S(t) \cdot \nabla E_S(t), \; \overline{\Delta W}(t)\right) + \\ + \frac{1}{2}\beta^2(t) \cdot \left(\nabla^2 E_S(t) \cdot \overline{\Delta W}(t), \; \overline{\Delta W}(t)\right).$$

Найдем критические точки функции $E_S(t+1) = E_S(\alpha; \beta)$, как функции двух переменных. Для этого найдем частные производные функции $E_S(\alpha; \beta)$ и решим систему уравнений:

$$\begin{cases} \dfrac{\partial E_S}{\partial \alpha} = -\left(\nabla E_S, \; \nabla E_S\right) + \alpha \cdot \left(\nabla^2 E_S \cdot \nabla E_S, \; \nabla E_S\right) - \\ - \beta \cdot \left(\nabla^2 E_S \cdot \nabla E_S, \; \overline{\Delta W}\right) = 0 \\ \dfrac{\partial E_S}{\partial \beta} = \left(\nabla E_S, \; \overline{\Delta W}\right) - \alpha \cdot \left(\nabla^2 E_S \cdot \nabla E_S, \; \overline{\Delta W}\right) + \\ + \beta \cdot \left(\nabla^2 E_S \cdot \overline{\Delta W}, \; \overline{\Delta W}\right) = 0 \end{cases}$$

Тогда приходим к системе:

$$\begin{cases} \alpha \cdot \left(\nabla^2 E_S \cdot \nabla E_S, \; \nabla E_S\right) - \beta \cdot \left(\nabla^2 E_S \cdot \nabla E_S, \; \overline{\Delta W}\right) = \\ = \left(\nabla E_S, \; \nabla E_S\right) \\ \alpha \cdot \left(\nabla^2 E_S \cdot \nabla E_S, \; \overline{\Delta W}\right) - \beta \cdot \left(\nabla^2 E_S \cdot \overline{\Delta W}, \; \overline{\Delta W}\right) = \\ = \left(\nabla E_S, \; \overline{\Delta W}\right) \end{cases}$$

Следовательно,

$$\begin{cases} \alpha = \dfrac{\begin{vmatrix} \left(\nabla E_S,\ \nabla E_S\right) & -\left(\nabla^2 E_S \cdot \nabla E_S,\ \overline{\Delta W}\right) \\ \left(\nabla E_S,\ \overline{\Delta W}\right) & -\left(\nabla^2 E_S \cdot \overline{\Delta W},\ \overline{\Delta W}\right) \end{vmatrix}}{\begin{vmatrix} \left(\nabla^2 E_S \cdot \nabla E_S,\ \nabla E_S\right) & -\left(\nabla^2 E_S \cdot \nabla E_S,\ \overline{\Delta W}\right) \\ \left(\nabla^2 E_S \cdot \nabla E_S,\ \overline{\Delta W}\right) & -\left(\nabla^2 E_S \cdot \overline{\Delta W},\ \overline{\Delta W}\right) \end{vmatrix}} \\ \beta = \dfrac{\begin{vmatrix} \left(\nabla^2 E_S \cdot \nabla E_S,\ \nabla E_S\right) & \left(\nabla E_S,\ \nabla E_S\right) \\ \left(\nabla^2 E_S \cdot \nabla E_S,\ \overline{\Delta W}\right) & \left(\nabla E_S,\ \overline{\Delta W}\right) \end{vmatrix}}{\begin{vmatrix} \left(\nabla^2 E_S \cdot \nabla E_S,\ \nabla E_S\right) & -\left(\nabla^2 E_S \cdot \nabla E_S,\ \overline{\Delta W}\right) \\ \left(\nabla^2 E_S \cdot \nabla E_S,\ \overline{\Delta W}\right) & -\left(\nabla^2 E_S \cdot \overline{\Delta W},\ \overline{\Delta W}\right) \end{vmatrix}} \end{cases}$$

или

$$\begin{cases} \alpha = \dfrac{-\left(\nabla E_S,\ \nabla E_S\right)\cdot\left(\nabla^2 E_S \cdot \overline{\Delta W},\ \overline{\Delta W}\right) + \left(\nabla E_S,\ \overline{\Delta W}\right)\cdot\left(\nabla^2 E_S \cdot \nabla E_S,\ \overline{\Delta W}\right)}{-\left(\nabla^2 E_S \cdot \nabla E_S,\ \nabla E_S\right)\cdot\left(\nabla^2 E_S \cdot \overline{\Delta W},\ \overline{\Delta W}\right) + \left(\nabla^2 E_S \cdot \nabla E_S,\ \overline{\Delta W}\right)^2} \\ \beta = \dfrac{\left(\nabla E_S,\ \overline{\Delta W}\right)\cdot\left(\nabla^2 E_S \cdot \nabla E_S,\ \nabla E_S\right) - \left(\nabla E_S,\ \nabla E_S\right)\cdot\left(\nabla^2 E_S \cdot \nabla E_S,\ \overline{\Delta W}\right)}{-\left(\nabla^2 E_S \cdot \nabla E_S,\ \nabla E_S\right)\cdot\left(\nabla^2 E_S \cdot \overline{\Delta W},\ \overline{\Delta W}\right) + \left(\nabla^2 E_S \cdot \nabla E_S,\ \overline{\Delta W}\right)^2} \end{cases}$$

Таким образом (2).

Вычислим частные производные второго порядка функции $E_S(\alpha;\beta)$:

$$\frac{\partial^2 E_S}{\partial \alpha^2} = \left(\nabla^2 E_S \cdot \nabla E_S,\ \nabla E_S\right),$$

$$\frac{\partial^2 E_S}{\partial \alpha \partial \beta} = \frac{\partial^2 E_S}{\partial \beta \partial \alpha} = -\left(\nabla^2 E_S \cdot \nabla E_S,\ \overline{\Delta W}\right) \text{ и}$$

$$\frac{\partial^2 E_S}{\partial \beta^2} = \left(\nabla^2 E_S(t) \cdot \overline{\Delta W},\ \overline{\Delta W}\right).$$

Тогда гессиан $\left|\nabla^2 E_S(\alpha;\beta)\right|$ функции $E_S(\alpha;\beta)$ равен

$$\left|\nabla^2 E_S(\alpha;\beta)\right| = \begin{vmatrix} \left(\nabla^2 E_S \cdot \nabla E_S,\ \nabla E_S\right) & -\left(\nabla^2 E_S \cdot \nabla E_S,\ \overline{\Delta W}\right) \\ -\left(\nabla^2 E_S \cdot \nabla E_S,\ \overline{\Delta W}\right) & \left(\nabla^2 E_S \cdot \overline{\Delta W},\ \overline{\Delta W}\right) \end{vmatrix} =$$

$$= \left(\nabla^2 E_S \cdot \nabla E_S,\ \nabla E_S\right)\cdot\left(\nabla^2 E_S \cdot \overline{\Delta W},\ \overline{\Delta W}\right) - \left(\nabla^2 E_S \cdot \nabla E_S,\ \overline{\Delta W}\right)^2.$$

В случае если $\left|\nabla^2 E_S(\alpha;\beta)\right| > 0$ и $\left(\nabla^2 E_S \cdot \nabla E_S,\ \nabla E_S\right) > 0$, то функция $E_S(\alpha;\beta)$ достигает минимального значения при $\alpha(t)$ и $\beta(t)$, определяемыми соотношениями (2).

Таким образом, доказано следующее утверждение.

Теорема. Величины квазиоптимальных параметров $\alpha(t)$ и $\beta(t)$ шага обучения нейронной сети с использованием метода сопряженных градиентов в момент времени t определяется соотношениями [3] (3, 4), где функция квадратичной ошибки сети

$$E_S(t) = E_S\left(w_{11}, w_{21}, ..., w_{n1}, T_1, w_{12}, w_{22}, ..., w_{n2}, T_2, -\right.$$
$$\left., ..., w_{1m}, w_{2m}, ..., w_{nm}, T_m\right)$$

дважды непрерывно дифференцируемая функция нескольких переменных,

$$\overline{W}(t) = \left(w_{11}, w_{21}, ..., w_{n1}, T_1, w_{12}, w_{22}, ..., w_{n2}, T_2, -\right.$$
$$\left., ..., w_{1m}, w_{2m}, ..., w_{nm}, T_m\right)^T$$

вектор переменных,

$$\nabla E_S(t) = \left(\frac{\partial E_S}{\partial w_{11}}, \frac{\partial E_S}{\partial w_{21}}, ..., \frac{\partial E_S}{\partial w_{n1}}, \frac{\partial E_S}{\partial T_1}, \frac{\partial E_S}{\partial w_{12}}, \frac{\partial E_S}{\partial w_{22}}, -\right.$$
$$\left., ..., \frac{\partial E_S}{\partial w_{n2}}, \frac{\partial E_S}{\partial T_2}, ..., \frac{\partial E_S}{\partial w_{1m}}, \frac{\partial E_S}{\partial w_{2m}}, ..., \frac{\partial E_S}{\partial w_{nm}}, \frac{\partial E_S}{\partial T_m}\right)^T$$

вектор градиента функции $E_S(t)$, $\nabla^2 E_S(t)$ – матрица Гессе вторых производных функции $E_S(t)$ в момент времени t.

Так как

$$\frac{\partial E_S}{\partial w_{ij}(t)} = \sum_p \left(y_j^p - t_j^p\right)\cdot F'\left(S_j^p\right)x_i^p,$$
$$\left(i = \overline{1,n},\ j = \overline{1,m}\right),$$

и

$$\frac{\partial E_S}{\partial T_j(t)} = -\sum_p \left(y_j^p - t_j^p\right)F'\left(S_j^p\right),\ \left(j = \overline{1,m}\right),$$

то, подставляя эти соотношения в (1), получим, что модификация синаптических связей с использованием квазиоптимальных параметров шага обучения определяется выражениями:

$$\begin{aligned} w_{ij}(t+1) &= w_{ij}(t) - \alpha(t)\sum_{k=1}^{L}\left(y_j^k(t) - t_j^k\right)\cdot F'\left(S_j^k(t)\right)x_i^k + \\ &+ \beta(t)\left(w_{ij}(t) - w_{ij}(t-1)\right),\ i = \overline{1,n},\ j = \overline{1,m} \end{aligned} \tag{5}$$

$$\begin{cases} \alpha(t) = \dfrac{\left\|\nabla E_S(t)\right\|^2\cdot\left(\nabla^2 E_S(t)\cdot\overline{\Delta W}(t),\ \overline{\Delta W}(t)\right) - \left(\nabla E_S(t),\ \overline{\Delta W}(t)\right)\cdot\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \overline{\Delta W}(t)\right)}{\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \nabla E_S(t)\right)\cdot\left(\nabla^2 E_S(t)\cdot\overline{\Delta W}(t),\ \overline{\Delta W}(t)\right) - \left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \overline{\Delta W}(t)\right)^2} \\ \beta(t) = \dfrac{\left\|\nabla E_S(t)\right\|^2\cdot\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \overline{\Delta W}(t)\right) - \left(\nabla E_S(t),\ \overline{\Delta W}(t)\right)\cdot\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \nabla E_S(t)\right)}{\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \nabla E_S(t)\right)\cdot\left(\nabla^2 E_S(t)\cdot\overline{\Delta W}(t),\ \overline{\Delta W}(t)\right) - \left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \overline{\Delta W}(t)\right)^2} \end{cases} \tag{2}$$

$$\alpha(t) = \frac{\left\|\nabla E_S(t)\right\|^2\cdot\left(\nabla^2 E_S(t)\cdot\overline{\Delta W}(t),\ \overline{\Delta W}(t)\right) - \left(\nabla E_S(t),\ \overline{\Delta W}(t)\right)\cdot\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \overline{\Delta W}(t)\right)}{\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \nabla E_S(t)\right)\cdot\left(\nabla^2 E_S(t)\cdot\overline{\Delta W}(t),\ \overline{\Delta W}(t)\right) - \left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \overline{\Delta W}(t)\right)^2} \tag{3}$$

$$\beta(t) = \frac{\left\|\nabla E_S(t)\right\|^2\cdot\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \overline{\Delta W}(t)\right) - \left(\nabla E_S(t),\ \overline{\Delta W}(t)\right)\cdot\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \nabla E_S(t)\right)}{\left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \nabla E_S(t)\right)\cdot\left(\nabla^2 E_S(t)\cdot\overline{\Delta W}(t),\ \overline{\Delta W}(t)\right) - \left(\nabla^2 E_S(t)\cdot\nabla E_S(t),\ \overline{\Delta W}(t)\right)^2} \tag{4}$$

$$T_j(t+1)=T_j(t)+\alpha(t)\sum_{k=1}^{L}\left(y_j^k(t)-t_j^k\right)F'\left(S_j^k(t)\right)+ \\ +\beta(t)\left(T_j(t)-T_j(t-1)\right),\ \ j=\overline{1,m} \qquad .(6)$$

Замечание. Для использования соотношений (3)-(6) необходимо знать начальные значения $\overline{W}(0)$ и $\overline{W}(1)$. Как правило, вектор $\overline{W}(0)$ получают случайной инициализацией, а вектор $\overline{W}(1)$ может быть вычислен на основании соотношений для шага обучения нейронной сети с использованием метода наискорейшего спуска [2]:

$$\alpha(t)=\frac{\sum_{k=1}^{L}\sum_{j=1}^{m}\left(y_j^k(t)-t_j^k\right)F'\left(S_j^k(t)\right)a_j^k(t)}{\sum_{k=1}^{L}\sum_{j=1}^{m}\left(\left(F'\left(S_j^k(t)\right)\right)^2+\left(y_j^k(t)-t_j^k\right)F''\left(S_j^k(t)\right)\right)\left(a_j^k(t)\right)^2} \qquad (7)$$

где $a_j^k(t)=\sum_{p=1}^{L}\left(y_j^p(t)-t_j^p\right)F'\left(S_j^p(t)\right)\left(\sum_{i=1}^{n}x_i^p x_i^k+1\right)$, и

$j=\overline{1,m},\ k=\overline{1,L}$

$$w_{ij}(t+1)=w_{ij}(t)-\alpha(t)\cdot\sum_{k=1}^{L}\left(y_j^k(t)-t_j^k\right)\cdot F'\left(S_j^k(t)\right)x_i^k,$$

$i=\overline{1,n},\ \ j=\overline{1,m}$

$$T_j(t+1)=T_j(t)+\alpha(t)\cdot\sum_{k=1}^{L}\left(y_j^k(t)-t_j^k\right)F'\left(S_j^k(t)\right),\ \ j=\overline{1,m}.$$

Легко показать, что такой же вектор $\overline{W}(1)$ может быть получен с использованием метода сопряженных градиентов, определяемого выражениями (5), (6), где $\beta(t)=0$, а $\alpha(t)$ будет вычислено в соответствии с (7).

АЛГОРИТМ ОБУЧЕНИЯ НЕЙРОННОЙ СЕТИ С ИСПОЛЬЗОВАНИЕМ МЕТОДА СОПРЯЖЕННЫХ ГРАДИЕНТОВ

Приведем алгоритм обучения нейронной сети с использованием метода сопряженных градиентов, использующий соотношения (3)-(6):

1. Задается минимальная квадратичная ошибка сети ε_m, которой необходимо достичь в процессе обучения.

2. Записывается число t=0 в счетчик числа итераций алгоритма.

3. Случайным образом инициализируются весовые коэффициенты сети $w_{ij}(t)$ $\left(i=\overline{1,n},\ j=\overline{1,m}\right)$, и пороговые значения нейронных элементов $T_j(t)$ $\left(j=\overline{1,m}\right)$.

4. Подаются входные образы $\overline{x^k}=\left(x_1^k,\dots,x_n^k\right)$ $\left(k=\overline{1,L}\right)$ на нейронную сеть и вычисляются векторы $\overline{y^k(t)}=\left(y_1^k(t),\dots,y_m^k(t)\right)$ $\left(k=\overline{1,L}\right)$ выходной активности сети, определяемые соотношениями (4).

5. Если $t\neq0$, то величины квазиоптимальных параметров $\alpha(t)$, $\beta(t)$ шага обучения с использованием метода сопряженных градиентов вычисляются в соответствии с соотношениями (3)-(4), в противном случае параметр $\alpha(t)$ определяется выражением (7), а $\beta(t)$ полагается равным нулю.

6. Производится изменение весовых коэффициентов $w_{ij}(t+1)$ $\left(i=\overline{1,n},\ j=\overline{1,m}\right)$ и порогов нейронной сети $T_j(t+1)$ $\left(j=\overline{1,m}\right)$ согласно выражениям (5) и (6), соответственно.

7. Полагается $t=t+1$.

8. Алгоритм завершает свою работу, если суммарная квадратичная ошибка сети $E_S(t)=\frac{1}{2}\sum_{j=1}^{m}\sum_{k=1}^{L}\left(y_j^k(t)-t_j^k\right)^2$ или норма вектора $\overline{\Delta W}(t)=\overline{W}(t)-\overline{W}(t-1)$ не превосходят заданной величины ε_m, т. е. $E_S(t)\le\varepsilon_m$ или

$$\left\|\overline{\Delta W}(t)\right\|= \\ =\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(w_{ij}(t)-w_{ij}(t-1)\right)^2+\sum_{j=1}^{m}\left(T_j(t)-T_j(t-1)\right)^2}<\varepsilon_m$$

в противном случае выполняется п. 4.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Головко В.А. Нейронные сети: обучение, организация и применение. Кн. 4: Учебное пособие для вузов / Общая ред. А.И. Галушкина. – М.: ИПРЖР, 2001. – 256 с.: ил. (Нейрокомпьютеры и их применение).
2. Гладкий И.И., Головко В.А., Махнист Л.П. Обучение нейронных сетей с использованием метода наискорейшего спуска // Вестник Брестского государственного технического университета. – Брест: БГТУ, 2001. – № 5: Физика, математика, химия.– С. 47-55.
3. Махнист Л.П. О решении одной системы разностно-дифференциальных уравнений и ее применении // Тезисы докладов международной математической конференции “Еругинские чтения – VIII”. – Брест: БрГУ, 2002.– С. 88-89.