

нет и достаточное количество времени для проверки учебных баз каждого студента. К сожалению, типовыми учебными планами такая «облачная» нагрузка преподавателей не предусмотрена – эффективность учебного процесса полностью зависит от энтузиазма преподавателя.

*В.А. ГОЛОВКО, А.А. КРОЩЕНКО*

БрГТУ (г. Брест, Беларусь)

## **ОБ ОДНОМ МЕТОДЕ ОБУЧЕНИЯ НЕЙРОННЫХ СЕТЕЙ ГЛУБОКОГО ДОВЕРИЯ**

Д. Хинтоном был предложен подход обучения нейронных сетей глубокого доверия (DBNN – Deep Belief Neural Network), состоящий из двух этапов [1, 2]:

1. Послойная предварительная тренировка DBNN без учителя. Обучение производится с использованием «жадного» алгоритма: последовательно тренируется каждый слой сети как RBM с последующей передачей преобразованных входных данных на следующие слои. В качестве метода обучения на данном этапе используется Contrastive Divergence (CD).

2. «Тонкая» настройка всей нейронной сети методом обратного распространения ошибки.

Нами был предложен альтернативный подход к обучению RBM [3]. Целью данного подхода является минимизация среднеквадратичной ошибки восстановления:

$$E_s = \frac{1}{2} \sum_{k=1}^L \sum_{j=1}^m (y_j^k(1) - y_j^k(0))^2 + \frac{1}{2} \sum_{k=1}^L \sum_{i=1}^n (x_i^k(1) - x_i^k(0))^2.$$

Веса и пороги сети изменяются итеративно в соответствии со следующими правилами:

$$w_{ij}(t+1) = w_{ij}(t) - \alpha \frac{\partial E}{\partial w_{ij}(t)},$$

$$T_j(t+1) = T_j(t) - \alpha \frac{\partial E}{\partial T_j(t)},$$

$$T_j(t+1) = T_j(t) - \alpha \frac{\partial E}{\partial T_j(t)}.$$

Функция ошибки для одного образца определяется следующим образом:

$$E = \frac{1}{2} \sum_{j=1}^m (y_j(1) - y_j(0))^2 + \frac{1}{2} \sum_{i=1}^n (x_i(1) - x_i(0))^2. \quad (1)$$

Далее дифференцируя (1) по  $w_{ij}$ ,  $T_i$  и  $T_j$ , мы получаем следующие правила обучения RBM:

$$\begin{aligned}w_{ij}(t+1) &= w_{ij}(t) - \alpha(y_j(1) - y_j(0))F'(S_j(1))x_i(1) + (x_i(1) - x_i(0))F'(S_i(1))y_j(0), \\T_j(t+1) &= T_j(t) - \alpha(y_j(1) - y_j(0))F'(S_j(1)), \\T_i(t+1) &= T_i(t) - \alpha(x_i(1) - x_i(0))F'(S_i(1)).\end{aligned}$$

Нами была выполнена сравнительная характеристика производительности классического и альтернативного подхода на примере обучения глубокого семислойного автоэнкодера. Данные, которые использовались при обучении, генерировались моделью, известной как система Лоренца:

$$\begin{cases} \dot{x}_1 = -a(x_1 - x_2), \\ \dot{x}_2 = -x_1x_3 + bx_1 - x_2, \\ \dot{x}_3 = x_1x_2 - cx_3, \end{cases}$$

где  $a = 10$ ,  $b = 28$ ,  $c = 8.0/3.0$  – параметры системы.

Размер обучающей и тестовой выборки – 1000 образцов. Скорость обучения  $\alpha = 0.9$  (0.05 для классического метода обучения), моментный параметр – 0.5, ограничение на количество эпох – 50 (CD) / 200 (BP).

Результаты эксперимента представлены в следующей таблице. Здесь *Stepscount* – количество шагов метода CD, *MSE* – среднеквадратичная ошибка на тренировочной выборке, *MS* – среднеквадратичная ошибка на тестовом наборе данных для проверки обобщающей способности.

Таблица

Подход	Steps count	MSE	MS
Классический	1	20,234	15,647
	5	19,915	15,302
	10	20,850	16,434
REBA	1	17,037	13,322
	5	53,954	56,160
	10	49,965	51,042
BP	–	70,953	76,823

#### СПИСОК ЛИТЕРАТУРЫ

1. Bengio, Y. Learning deep architectures for AI / Y. Bengio // Foundations and Trends in Machine Learning. – 2009. – Vol. 2, № 1. – P. 1–127.
2. Hinton, G. A practical guide to training restricted Boltzmann machines / G. Hinton // Tech. Rep. 2010–000. – Toronto : Machine Learning Group, University of Toronto, 2010.
3. Golovko, V. learning technique for deep belief neural networks / V. Golovko [et al.] // Neural Networks and Artificial Intelligence. – Springer, 2014. – Vol. 440 : Communication in Computer and Information Science. – P. 136–146.