УДК 0048.032.26

АДАПТИВНЫЙ ШАГ ОБУЧЕНИЯ ДЛЯ НЕКОНТРОЛИРУЕМОГО ОБУЧЕНИЯ ГЛУБОКИХ НЕЙРОННЫХ СЕТЕЙ

В.А. Головко, А.А. Крощенко, Е.В. Михно, А.М. Соловчук

Брестский государственный технический университет, ул. Московская 267, 224017, г. Брест, Беларусь, vladimir.golovko@gmail.com
Corresponding author: vladimir.golovko@gmail.com

В статье получены выражения для адаптивного шага обучения на этапе предварительного обучения глубоких нейронных сетей. Для вывода выражений использовался метод наискорейшего спуска. Использование адаптивного шага обучения позволяет повысить скорость обучения глубокой нейронной сети на этапе предварительного обучения.

Ключевые слова: глубокое обучение; адаптивный шаг обучения; RBM машина.

Введение

Существуют два основных метода глубокого обучения.

1. *Метод с предварительным обучением*, который состоит из двух этапов [1–9]:

• предобучение нейронной сети, начиная с первого слоя. Данное обучение осуществляется без учителя и базируется на ограниченной машине Больцмана (RBM) или на автоэнкодерной нейронной сети ;

• настройка синаптических связей всей сети при помощи алгоритма обратного распространения ошибки.

2. *Метод стохастического градиента* (SGD) с ректификационной функцией активации (ReLU) нейронных элементов [4, 6].

В настоящее время принята следующая парадигма для обучения глубоких нейронных сетей. Если обучающая выборка большая, т. е. размерность обучающей выборки намного больше, чем количество настраиваемых параметров сети, то используется метод стохастического градиента с функцией активации ReLU нейронных элементов. Если размерность обучающей выборки сравнима или меньше количества настраиваемых параметров сети или используются сигмоидальные функции активации нейронных элементов, то применяется предварительное обучение нейронной сети и алгоритм обратного распространения ошибки на заключительном этапе обучения [4,6].

Основные недостатки существующих подходов:

1. метод стохастического градиента требует большого объема обучающей выборки и делает невозможным использование сигмоидальных функций активации в скрытых слоях нейронной сети [4,6].
2. методы предварительного обучения базируются на линейной ограниченной машине Больцмана с точки зрения минимизации суммарной квадратичной ошибки сети или классическом автоэнкодерном подходе, со всеми присущими их недостатками (выбор походящего шага обучения, достижение приемлемой начальной точки обучения с точки зрения хорошей обобщающей способности и редуцирования параметров нейронной сети) [6].

В данной работе получены выражения для адаптивного шага обучения на этапе предварительного обучения нейронной сети для ReLU функции активации, что позволяет повысить качество обучения.

1. Адаптивный шаг обучения

Для ускорения процедуры обучения градиентного спуска, вместо постоянного шага обучения можно использовать *адаптивный шаг* $\alpha(t)$.

Адаптивным называется шаг обучения, который выбирается на каждом этапе алгоритма таким образом, чтобы минимизировать квадратичную ошибку сети [6].

Пусть дана однослойная нейронная сеть, которая состоит из распределительного слоя нейронных элементов и выходного слоя (рисунок).

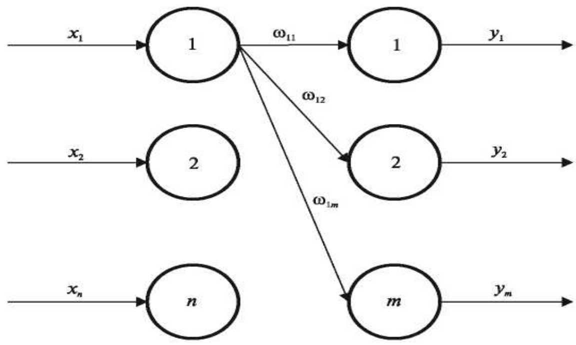

Рисунок – Однослойная нейронная сеть

В качестве нейронов выходного слоя используются нейронные элементы с ректификационной функцией активации:

$$y_j(t) = F(S_j(t)) = r_j(t)S_j(t), \tag{1}$$

$$r_j(t) = \begin{cases} r_1, \ S_j(t) \geq 0; \\ r_2, \ S_j(t) < 0. \end{cases} \tag{2}$$

Где $r_1 \neq r_2$. Каждый нейрон распределительного слоя имеет синаптические связи со всеми нейронами обрабатывающего слоя. Взвешенная сумма *j*-го нейрона сети определяется как

$$S_j(t) = \sum_i \omega_{ij}(t) x_i(t) + T_j(t). \tag{3}$$

Суммарная квадратичная ошибка сети для всей обучающей выборки вычисляется по формуле

$$E_s = \frac{1}{2}\sum_k \sum_j (y_j^k - e_j^k)^2 \tag{4}$$

Соответственно квадратичную ошибку для одного образа можно представить в виде:

$$E = \frac{1}{2}\sum_j (y_j - e_j)^2 \tag{5}$$

Для нахождения адаптивного шага обучения будем использовать метод наискорейшего спуска.

Теорема 1. Для однослойной нейронной сети с ReLU функцией активации при последовательном обучении значение адаптивного шага обучения вычисляется на основе следующего выражения:

$$\alpha(t) = \frac{\sum_{j=1}^{m} r_j(t+1) b_j (r_j(t+1) S_j(t) - e_j)}{\sum_{j=1}^{m} r_j^2(t+1) b_j^2},$$

$$b_j = r_j(t)(y_j - e_j)(1 + \sum_{i=1}^{n} x_i^2(t), \tag{6}$$

где $r_j(t+1) = \begin{cases} r_1, \; e_j(t) \geq 0; \\ r_2, \; e_j(t) < 0. \end{cases}$

Данная теорема справедлива, когда обучение происходит последовательно после подачи каждого образа на вход нейронной сети.

Весовые коэффициенты и пороговые значения изменяются следующим образом:

$$\omega_{ij}(t+1) = \omega_{ij}(t) - \alpha(t) r_j(t)(y_j - e_j) x_i(t),$$
$$T_j(t+1) = T_j(t) - \alpha(t) r_j(t)(y_j - e_j) \tag{7}$$

Адаптивный шаг обучения позволяет значительно повысить скорость обучения однослойной нейронной сети и достичь оптимального решения задачи.

2. Предобучение глубоких нейронных сетей с адаптивным шагом

Рассмотрим использование адаптивного шага обучения для ограниченной машины Больцмана (RBM).

В процессе обучения автоэнкодерной сети для каждого входного образа производится три цикла преобразования информации: прямое (сжатие), обратное (восстановление) и прямое (сжатие). Такое преобразование информации эквивалентно сэмплированию Гиббса с единичным шагом. После этого производится настройка весовых коэффициентов сети. Для наглядности процесса распространения информации введем обозначения. Пусть $x_i(0),\ i = 1, n$ входной вектор, поступающий на вход сети в начальный момент времени $t = 0$. Тогда формулы выходных значений сжимающего слоя можно записать как

$$y_j(0) = F(S_j) = r_j(0)S_j(0) = r_j(0)\left(\sum_{i=1}^{n} w_{ij}x_i(0) + T_j\right) \tag{8}$$

где $j = \overline{1, p}$.

Выходные значения восстанавливающего слоя в момент времени $t = 1$:

$$x_i(1) = F(S_i) = r_i(1)S_i(1) = r_i(1)\left(\sum_{j=1}^{P} w'_{ji}y_j(0) + T_i\right) \tag{9}$$

где $i = \overline{1, n}$.

На следующем этапе распространения информации выходной вектор восстанавливающего слоя подается на вход сети и опять определяются выходные значения сжимающего слоя:

$$x_i(1) = F(S_i) = r_i(1)S_i(1) = r_i(1)\left(\sum_{j=1}^{P} w'_{ji}y_j(0) + T_i\right)$$

$$y_j(1) = F(S_j(1) = r_j(1)S_j(1) = r_j(1)F\left(\sum_{i=1}^{n} w_{ij}x_i(1) + T_j\right) \tag{10}$$

где $j = \overline{1, p}$.

Для нелинейной ограниченной машины Больцмана правило модификации синаптических связей в случае CD-1 будет следующим:

$$w_{ij}(t+1) = w_{ij}(t) - \alpha_1\big(y_j(1) - y_j(0)\big)r_j(1)x_i(1) - \alpha_2(x_i(1) - x_i(0))r_i(1)y$$

$$T_i(t+1) = T_i(t) - \alpha_2\big(x_i(1) - x_i(0)\big)r_i(1)),$$
$$T_j(t+1) = T_j(t) - \alpha_1\big(y_j(1) - y_j(0)\big)r_j(1)) \qquad (11)$$

Адаптивные шаги для последовательного обучения вычисляются следующим образом:

$$\alpha_1(t) = \frac{\sum_{j=1}^{m} r_j(t+1)b_j(r_j(t+1)S_j(1) - y_j(0))}{\sum_{j=1}^{m} r_j^2(t+1)b_j^2},$$
$$b_j = r_j(1)(y_j(1) - y_j(0))(1 + \sum_{i=1}^{n} x_i^2(1), \qquad (12)$$

где $r_j(t+1) = \begin{cases} r_1,\ y_j(0) \geq 0; \\ r_2,\ y_j(0) < 0. \end{cases}$

$$\alpha_2(t) = \frac{\sum_{i=1}^{n} r_i(t+1)b_i(r_i(t+1)S_i(1) - x_i(0))}{\sum_{i=1}^{n} r_i^2(t+1)b_i^2},$$
$$b_i = r_i(1)(x_i(1) - x_i(0))(1 + \sum_{i=1}^{m} y_i^2(0), \qquad (13)$$

где $r_i(t+1) = \begin{cases} r_1,\ x_i(0) \geq 0; \\ r_2,\ x_i(0) < 0. \end{cases}$

Заключение

1. Получены выражения для вычисления адаптивного шага обучения в однослойном персептроне при использовании ReLU функции активации;

2. Произведено отображение полученных выражений для предварительного обучения глубоких нейронных сетей.

Данная работа выполнена при поддержке белорусского республиканского фонда фундаментальных исследований БРФФИ, проект Ф22КИ-046.

Библиографические ссылки

1. Hinton G. E., Osindero S., Yee-Whye Teh. A fast learning algorithm for deep belief nets // Neural computation. 2006. № 18. P. 1527−1554.
2. Hinton G., Salakhutdinov R. Reducing the dimensionality of data with neural networks // Science. 2006. № 313 (5786). P. 504−507.

3. Hinton G. E. A practical guide to training restricted Boltzmann machines. Toronto, 2010.
4. LeCun Y., Bengio Y., Hinton G. Deep learning Nature, 2015, 521 (7553). P. 436–444.
5. Головко В.А. От многослойных персептронов к нейронным сетям глубокого доверия: парадигмы обучения и применение: лекции по нейроинформатике. М.: 2015.
6. Головко В.А., Краснопрошин В.В. Нейросетевые технологии обработки данных. Минск: БГУ, 2017. 263 с.
7. V. Golovko, A. Kroshchanka, V. Turchenko, S. Jankowski, and D. Treadwell A new technique for restricted Boltzmann machine learning // IEEE 8th Int.Conf. on Intelligent Data Acquisition and Advanced Computing.
8. Systems: Technology and Applications (IDAACS), Warsaw: IEEE, 2015. P. 182–186.
9. V. Golovko, A. Kroshchanka, and D. Treadwell, “The nature of unsupervised learning in deep neural networks:A new understanding and novel approach // Opt. Memory Neural Networks, 2016. Vol. 25. P. 127–141.
10. V. Golovko . Deep neural networks: a theory, application and new trends / Proc. of the 13th intern, conf. on pattern recognition and inform, processing, Minsk, 3−5 oct. 2016. Minsk, 2016. P. 33−37.