

АЛЬТЕРНАТИВНЫЙ ПОДХОД К ОРГАНИЗАЦИИ ОБУЧЕНИЯ СИГМОИДАЛЬНЫХ НЕЙРОНОВ В АРХИТЕКТУРЕ МНОГОСЛОЙНОЙ НЕЙРОННОЙ СЕТИ

Ю.В. Савицкий

Брестский государственный технический университет,
Брест, Беларусь, yury.savitsky@tut.by

The article formulates and mathematically substantiates a modification of the Back Propagation Error (BPE) algorithm for accurate training of neural elements with a sigmoid activation function in the architecture of a multilayer neural network. The proposed training rules for modifying the synaptic connections of neurons in the output and hidden layers can be used to construct efficient algorithms that reduce the time and computational complexity of the learning process for multilayer architectures.

Введение

В последнее время в мире активизировались исследования в области глубокого обучения многослойных нейронных сетей. Это связано с определенными успехами в данной области, достигнутыми рядом исследователей [1, 2], а также высокой практической значимостью сильно-многослойных нейронных сетей (СМНС). Так, ряд разработчиков интеллектуального программного обеспечения (корпорации Google, Microsoft и др.) с успехом применяют технологии глубоких нейронных сетей в различных своих приложениях. При этом перспективным считается подход к предобучению (pre-training) СМНС не только с помощью ограниченной машины Больцмана (RBM), но и с применением нейросетевых автоэнкодеров (Autoencoder). Каждый такой нейросетевой автоэнкодер представляет собой трехслойный персептрон архитектуры $N \rightarrow M \rightarrow N$, где параметр N соответствует количеству входов текущего предобучаемого слоя, M – количеству нейронов указанного слоя сильно-многослойной нейросетевой архитектуры. Последовательное (начиная с входного слоя СМНС) обучение совокупности таких автоэнкодеров на входной обучающей выборке позволяет получить наборы весовых коэффициентов для финальной настройки синаптических связей всей СМНС (fine-tuning). При этом, для обучения как нейросетевых автоэнкодеров, так и СМНС, как правило, применяется алгоритм обратного распространения ошибки (Back Propagation Error, BPE) [3-5]. Очевидно, что эффективность алгоритма BPE напрямую определяет эффективность (точность обучения, обобщающие свойства) результирующей модели СМНС в целом.

В данной работе предлагается методика точного обучения нейронных элементов (НЭ) сигмоидального типа в составе многослойной нейронной сети; выполняется анализ особенностей предложенных решений.

1. Обобщенная архитектура нейронной сети

На рисунке 1 приведена обобщенная архитектура многослойной нейронной сети, являющаяся объектом исследования, а также введены обозначения параметров сети.

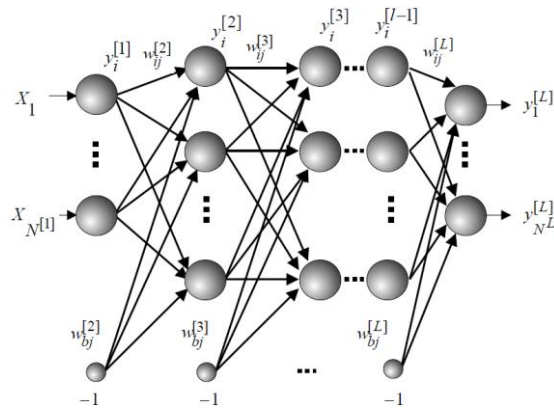


Рисунок 1 – Архитектура многослойной нейронной сети

Нейронный элемент слоя l осуществляет функцию преобразования некоторого вектора входных сигналов $Y^{[l-1]}$ в выходную активность $y_j^{[l]}$ по следующему правилу:

$$\begin{cases} S_j^{[l]} = \sum_{i=1}^{N^{[l-1]}} y_i^{[l-1]} w_{ij}^{[l]} - w_{bj}^{[l]}, \\ y_j^{[l]} = g^{[l]}(S_j^{[l]}), j = 1, \dots, N^{[l]}, \end{cases} \quad (1)$$

где $S_j^{[l]}$ - взвешенная сумма входных активностей НЭ j , находящегося в слое l ;

$w_{ij}^{[l]}$ - значение синаптического веса i -го входа НЭ;

$w_{bj}^{[l]}$ - значение порога активационной функции $g^{[l]}(S_j^{[l]})$ НЭ;

$N^{[l-1]}, N^{[l]}$ - соответственно количество входов НЭ слоя l и количество НЭ данного слоя.

2. Алгоритм ВРЕ

Как было сказано выше, точность решения практических задач в большой степени определяется параметрами обучающих процедур. Наиболее распространенным методом обучения базовых архитектур многослойных нейронных сетей является алгоритм ВРЕ. Он же является базисом для предлагаемого в работе альтернативного подхода к обучению НЭ. Поэтому рассмотрим детали применения алгоритма ВРЕ для обучения многослойной нейронной сети.

Согласно методу градиентного спуска, итерационное изменение весовых коэффициентов и порогов НЭ для каждого слоя нейронной сети происходит по следующим правилам:

$$\begin{cases} \Delta w_{ij}^{[l,p]}(t+1) = -\alpha \frac{\partial E^p(t)}{\partial w_{ij}^{[l,p]}(t)} = -\alpha \cdot \gamma_j^{[l,p]}(t) (g^{[l]})'(S_j^{[l,p]}(t)) y_i^{[l-1,p]}(t), \\ \Delta w_{bj}^{[l,p]}(t+1) = -\alpha \frac{\partial E^p(t)}{\partial w_{bj}^{[l,p]}(t)} = \alpha \cdot \gamma_j^{[l,p]}(t) (g^{[l]})'(S_j^{[l,p]}(t)), \end{cases} \quad (2)$$

где α - константа, определяющая шаг обучения;

$(g^{[l]})'(S_j^{[l,p]}(t))$ - производная активационной функции НЭ;

$\partial E^p(t)/\partial w_{ij}^{[l,p]}(t)$, $\partial E^p(t)/\partial w_{bj}^{[l,p]}(t)$ - частные производные функции ошибки нейронных связей, вычисляемые на каждой итерации обучения для каждого эталона p , $p = \{1, \dots, P\}$;

$\gamma_j^{[l,p]}(t)$ - ошибка НЭ j , определяемая как:

$$\gamma_j^{[L],p}(t) = y_j^{[L],p}(t) - D_j^p \quad (3)$$

для нейронов выходного слоя, либо:

$$\gamma_j^{[L],p}(t) = \sum_{j=1}^{N^{[L]}} y_j^{[L],p}(t) (g^{[L]})'(S_j^{[L],p}(t)) w_{ij}^{[L],p}(t) \quad (4)$$

для нейронов скрытых слоев сети;

$E^p(t)$ - среднеквадратичная ошибка нейронной сети для эталона p определяемая как:

$$E^p(t) = \sum_{j=1}^{N^{[L]}} E_j^p(t), \quad (5)$$

$$E_j^p(t) = \frac{1}{2} (y_j^{[L],p}(t) - D_j^p)^2 \quad (6)$$

где D_j^p — эталонное выходное значение j -го нейрона.

В результате выполнения каждой новой итерации обучения происходит минимизация общей ошибки сети, определяемой как:

$$E(t) = \sum_{p=1}^P E^p(t). \quad (7)$$

Таким образом, задача вычисления градиентов функции ошибки для элементов сети и модификации значений синаптических связей сводится к послойному определению ошибок нейронов в направлении от выходного слоя сети по правилам (3), (4) и использовании выражений (2).

3. Альтернативный подход к обучению НЭ с сигмоидной функцией активации

Следующая теорема, приводимая в рамках данной работы без доказательства, определяет альтернативный подход к обучению НЭ с сигмоидной функцией активации.

Теорема 1. Правила модификации синаптических связей НЭ j , находящегося в слое L , с сигмоидной функцией активации $g^{[L]}$, минимизирующие среднеквадратичную ошибку (6) данного НЭ для эталона p на итерации обучения t , определяются следующим образом:

$$\left\{ \begin{array}{l} w_{ij}^{[L],p}(t+1) = w_{ij}^{[L],p}(t) - \frac{S_j^{[L],p}(t) - \ln\left(\frac{D_j^p}{1-D_j^p}\right)}{1 + \sum_{k=1}^{N^{[L-1]}} (y_k^{[L-1],p})^2} y_i^{[L-1],p}(t), \\ w_{bj}^{[L],p}(t+1) = w_{bj}^{[L],p}(t) + \frac{S_j^{[L],p}(t) - \ln\left(\frac{D_j^p}{1-D_j^p}\right)}{1 + \sum_{k=1}^{N^{[L-1]}} (y_k^{[L-1],p})^2}. \end{array} \right. \quad (8)$$

Результаты Теоремы 1 (с учетом алгоритма ВРЕ) можно обобщить на НЭ последующих слоев многослойной нейронной сети.

Утверждение 1. Правила модификации синаптических связей НЭ j , находящегося в скрытом слое l , с сигмоидной функцией активации $g^{[L]}$, минимизирующие ошибку данного НЭ для эталона P на итерации обучения t , можно определить следующим образом:

$$\left\{ \begin{array}{l} w_{ij}^{[l],P}(t+1) = w_{ij}^{[l],P}(t) - \frac{S_j^{[L],P}(t) - \ln\left(\frac{y_j^{[l],P}(t) - \gamma_j^{[l],P}(t)}{1 - y_j^{[l],P}(t) + \gamma_j^{[l],P}(t)}\right)}{1 + \sum_{k=1}^{N^{[l-1]}} (y_k^{[l-1],P})^2} y_i^{[l-1],P}(t), \\ w_{bj}^{[l],P}(t+1) = w_{bj}^{[l],P}(t) + \frac{S_j^{[L],P}(t) - \ln\left(\frac{y_j^{[l],P}(t) - \gamma_j^{[l],P}(t)}{1 - y_j^{[l],P}(t) + \gamma_j^{[l],P}(t)}\right)}{1 + \sum_{k=1}^{N^{[l-1]}} (y_k^{[l-1],P})^2}, \end{array} \right. \quad (9)$$

где $\gamma_j^{[l],P}(t)$ — ошибка j -го НЭ.

Утверждение 1 вытекает из Теоремы 1, а также из допущения, что ошибку НЭ скрытого слоя $\gamma_j^{[l],P}(t)$ можно интерпретировать как разность между текущим значением выходной активности $y_j^{[l],P}(t)$ и желаемым значением выходной активности данного НЭ.

4. Анализ результатов

Анализ ряда вычислительных экспериментов показал, что наибольшую эффективность полученных результатов (Теоремы 1 и Утверждения 1) продемонстрировал алгоритм последовательного послойного (по ходу распространения активностей) обучения. В частности, для индивидуальных эталонов различного вида (в том числе из открытых баз MNIST и NIST), нормализованных в диапазон значений сигмоидной функции активации, данный алгоритм продемонстрировал сходимость процесса обучения многослойной НС до нулевой ошибки E^P за одну итерацию обучения.

Однако следует отметить, что более перспективным автор считает применение полученных результатов в алгоритме группового обучения (batch training); именно в этом направлении в настоящий момент проводятся экспериментальные исследования.

Список использованных источников

1. Hinton G. E., Osindero S., Teh Y. A fast learning algorithm for deep belief nets // Neural computation. – 2006. – № 18. – P. 1527–1554.

2. Головкин, В. А. Нейросетевые технологии обработки данных : учеб. пособие / В. А. Головкин, В. В. Краснопрошин. – Минск : БГУ, 2017. – 263 с. – (Классическое университетское издание).

3. Rumelhart D., Hinton G., Williams R. Learning representation by backpropagation errors // Nature. – 1986. – № 323. – P. 533–536.
4. Rumelhart D., Hinton G., Williams R. Learning internal representations by errors propagation // Parallel distributed processing. – Cambridge, 1986.
5. Golovko V., Savitsky Yu., Maniakov N. Neural networks for signal processing in measurement analysis and industrial applications: the case of chaotic signal processing // Neural networks for instrumentation, measurement analysis and related industrial applications. – Amsterdam, 2003. – P. 119–143.

УДК 339.137.22

ОСНОВНЫЕ ПОДХОДЫ К ПОВЫШЕНИЮ КОНКУРЕНТОСПОСОБНОСТИ ОРГАНИЗАЦИИ В УСЛОВИЯХ ЦИФРОВИЗАЦИИ ЭКОНОМИКИ

Г.Д. Какалыева

Туркменский государственный институт экономики и управления,
Ашхабад, Туркменистан, oguljan65@mail.ru

The processes for implementing a strategy to improve the competitiveness of an organization are complex, because are innovative, individual, cover all aspects from the problems of improving the quality of goods to personnel management. The main approaches to ensure the competitiveness of the organization are: benchmarking; reengineering of business processes; coordinating-stabilizing approach, etc.

In this paper, the benchmarking method is considered. The benchmarking method is based on the analysis of the best practices of the most innovatively active enterprises and copying this experience by inefficient enterprises. It is used to study the experience of leading competitors and apply it to the enterprise under study. In its simplest form, benchmarking can be viewed as a system of beliefs for solving problems of improving competitiveness, based on “comparison with a sample”.

Благодаря последовательным и целенаправленным усилиям уважаемого Президента Туркменистана в стране осуществляется инновационная индустриализация и цифровизация всех сфер жизни общества и государства. Одним из ярких примеров приверженности страны к дальнейшему инновационному развитию является Концепция развития цифровой экономики в Туркменистане на 2019-2025 г.г.

Процессы по реализации стратегии повышения конкурентоспособности организации в условиях цифровизации экономики сложны, т.к. носят инновационный характер, индивидуальны, охватывают все аспекты от проблем повышения качества товара до управления персоналом [1, с. 31].

Основными подходами для обеспечения конкурентоспособности организации являются:

- бенчмаркинг;
- реинжиниринг бизнес-процессов;
- координирующе-стабилизирующий подход и др. [1, с. 34];

В этой статье рассматривается метод бенчмаркинга.

Метод бенчмаркинга основан на анализе передового опыта наиболее иннова-