

УДК 519.7

К ВОПРОСУ ОБ ОПТИМИЗАЦИИ КОНСТРУКТИВНОГО МЕТОДА ОБУЧЕНИЯ НЕЙРОННЫХ СЕТЕЙ*

М. А. Казаков

Институт прикладной математики и автоматизации – филиал ФГБНУ «Федеральный научный центр «Кабардино-Балкарский научный центр Российской академии наук» (ИПМА КБНЦ РАН), 360000, г. Нальчик, ул. Шортанова, 89А

E-mail: f_wolfgang@mail.ru

Предлагается конструктивный метод обучения нейронных сетей, для обеспечения устойчивости качества работы сети по отношению к обучению на новых классах данных, при котором нейроны, добавленные непосредственно перед текущей эпохой обучения, принимают на себя основную нагрузку обучения на новом классе. Приводятся результаты вычислительных экспериментов

Ключевые слова: нейронные сети, машинное обучение, конструктивные методы обучения

© Казаков М. А., 2018

MSC 68T27

ON THE OPTIMIZATION OF THE CONSTRUCTIVE METHOD OF TRAINING NEURAL NETWORKS

M. A. Kazakov

Institute of Applied Mathematics and Automation of Kabardin-Balkar Scientific Centre of RAS (IAMA KBSC RAS), 360000, Nalchik, Shortanova st., 89A

E-mail: f_wolfgang@mail.ru

The article suggests a constructive method for training neural networks in which neurons added just before the current epoch of training assume the main training load on the new class to ensure the stability of the network in relation to learning on new data classes. The results of computational experiments are presented

Key words: neural networks, machine learning, constructive training methods.

© Kazakov M. A., 2018

*Работа выполнена при поддержке гранта РФФИ № 18-01-00050-а.

Введение

Одной из важных задач при проектировании нейронных сетей является проектирование топологии и оценка оптимального количества нейронов в сети. Существуют методы, которые позволяют произвести расчёт оптимального количества нейронов, в зависимости от поставленной задачи и выбранной топологии, однако сложность такой оценки существенно возрастает с ростом глубины нейронной сети. Другая проблема связана с обучением нейронных сетей на выборке нового класса данных. В этом случае имеется риск ухудшения качества работы нейронной сети на данных из классов, она обучалась на предыдущих этапах. Одним из эффективных методов решения задачи оценки количества нейронов является конструктивный метод обучения, который заключается в добавлении в сеть новых нейронов в процессе обучения по мере необходимости. Преимуществом такого метода является адаптивность и гибкость, позволяющая менять структуру и масштаб нейронной сети непосредственно в процессе обучения. Однако оценка необходимого количества добавляемых нейронов на каждой эпохе обучения и их топологическое размещение вызывает свои трудности.

Постановка задачи

Для обеспечения устойчивости качества работы нейронной сети по отношению к обучению на новых классах данных, предлагается конструктивный метод обучения [1], при котором нейроны, добавленные непосредственно перед текущей эпохой обучения, принимают на себя основную нагрузку обучения на новом классе. Это достигается за счет коррекции алгоритма обучения при котором параметры интенсивности обучения для "старых" нейронов снижаются, в то время как новые нейроны обучаются интенсивно. Такой метод одновременно обеспечивает высокую пластичность обученной сети за счет новых неспециализированных нейронов и при этом сохраняет эффективность работы нейронной сети на данных предыдущих классов, на которых была обучена нейронная сеть.

Теоретические предпосылки

Метод обратного распространения ошибок является одним из самых эффективных методов обучения нейронных сетей. Кроме всего прочего, этот метод хорошо интегрируется с другими методами, что открывает широкое поле для исследования и совершенствования методов машинного обучения. Одним из характерных особенностей, который возникает в процессе градиентного спуска является застревание в локальном минимуме [2]. Существует ряд методов, помогающих бороться с этим явлением [3]. При этом, выход из локального минимума не страхует от возможности очередного попадания в другой локальный минимум. Конструктивный метод обучения позволяет бороться с этим явлением [4]. При разбиении обучения на этапы, при котором на начальном этапе присутствует лишь часть классов и относительно небольшое количество нейронов, предоставляет возможность проведения градиентного спуска на гораздо меньшем числе измерений (число переменных, от которых зависит функция оценки равно количеству параметров сети – число всех весов и смещений).

Предполагается, что конструктивный метод обучения, при котором нейроны, добавленные непосредственно перед текущей эпохой обучения, будет лучше спарвляться с проблемой впадения в локальные минимумы.

Построение программы для испытательных экспериментов

Для проведения численных экспериментов, позволяющих сравнивать рассматриваемые методы, была разработана программа на Python, с использованием библиотек `numpy` и `matplotlib`. Для экспериментов была выбрана классическая задача распознавания рукописных чисел.

В качестве тренировочной и тестовой выборки использовалась база данных MNIST. Исходные данные представлены в виде 20 изображений (10 для тренировочной выборки и 10 для тестовой). Каждое изображение тренировочной выборки, соответствующая определенной цифре, содержит около 6000 уникальных экземпляров, представляющих собой изображения 28x28 пикселей. Тестовая выборка содержит около 1000 уникальных экземпляров для каждой цифры, отличных от экземпляров тренировочной выборки. Для тестирования использовалась модель нейронной сети прямого распространения, где вход нейронов связаны со всеми выходами предыдущего слоя. Функция активации для всех нейронов представляет собой экспоненциальную сигмоиду:

$$f(s) = \frac{1}{1 + \exp(-\alpha s)}$$

Обучение производилось методом обратного распространения ошибок. В качестве функции оценки использовалась сумма квадратов расстояний между выходными сигналами y_i и эталонными значениями a_i :

$$E = \sum (y_i - a_i)^2$$

Нейрон выдает на выходе значение в интервале (0, 1). Каждый выходной нейрон соответствует одному из распознаваемых цифр. Результатом распознавания нейронной сети будет цифра, которая соответствует выходному нейрону с наибольшим значением выходного сигнала. Поэтому количество выходных нейронов в точности совпадает с количеством распознаваемых классов. Например, если на данном этапе обучения используются пять цифр, то выходных нейронов будет тоже пять. Количество нейронов первого слоя и скрытых слоев а так же количество слоев может быть произвольным.

На каждом шаге обучения на вход нейронной сети поступают данные изображения одного экземпляра 28x28 пикселей, стохастически выбираемого из всего множества тренировочных примеров. Предварительно множество изображений преобразуется в пригодный для нейронной сети тип данных. Множество значений пикселей одного экземпляра записывается в одномерный массив из 784 элементов, и преобразуется в набор данных в интервале (-0.3, 0.7). Первый слой нейрона получает данные в виде такого массива, значений которого характеризуют яркость соответствующего пикселя.

Первым делом создается нейронная сеть заданной топологии (количество слоев и нейронов в каждом слое). Веса и смещения устанавливаются случайным образом. Далее есть возможность загрузить значения весов из сохраненного ранее файла,

или же приступить непосредственно к процедуре обучения. При обучении можно осуществить произвольное количество эпох, на каждом из которых можно менять значения α и скорости обучения. В результате обучения визуализируется динамика изменения функции оценки. Обученную нейронную сеть можно протестировать или же сохранить его параметры в файл.

Вычислительные эксперименты

Во всех испытаниях строилась нейронная сеть, состоящая из трех слоев, $\alpha = 0.05$, общее число циклов обучения равно 200000. В качестве результата спользовалось значение эффективности работа сети – доля распознаваемых цифр из всего множества, и динамика изменения функции оценки в процессе обучения.

МЕТОД 1. Стандартный метод градиентного спуска. Первый слой нейронной сети содержит 30 нейронов, второй слой 20 нейронов и третий слой – 10 нейронов. Проводится 4 этапа обучения по всем выборкам:

1. 200 циклов обучения с коэффициентом скорости 4;
2. 1800 циклов со коэффициентом 3;
3. 18000 циклов с коэффициентом 2;
4. 180000 с коэффициентом 1.

МЕТОД 2. Конструктивный метод обучения. Создается сеть, содержащая 15 нейронов первого слоя, 10 нейронов второго слоя и 5 нейронов выходного слоя. Обучение проводится по выборкам соответствующим цифрам 0, 1, 2, 3, 4:

1. 100 циклов обучения с коэффициентом скорости 4;
2. 900 циклов с коэффициентом скорости 3;
3. количество нейронов каждого слоя удваивается;
4. 100 циклов обучения с коэффициентом скорости 4;
5. 900 циклов с коэффициентом скорости 3;
6. 18000 циклов с коэффициентом 2;
7. 180000 с коэффициентом 1.

МЕТОД 3. Конструктивный метод обучения с избирательной скоростью обучения. Алгоритм обучения аналогичен предыдущему, за исключением пунктов 4 и 5. В данном случае указанный коэффициент скорости применяется только к новым нейронам, в то время как старые нейроны обучаются с коэффициентом существенно ниже ($\times 0.01$).

Таким образом все три коэффициента приводятся некому единообразию, позволяющему сравнить их между собой. В графике, иллюстрирующему динамику спада функции оценки, приводятся результаты, соответствующие поледним двум пунктам обучения, который идентичен для каждого эксперимента. На рисунке приводятся соответствующие графики.

После проведения серии идентичных испытаний и усреднения результатов эффективности распознавания, были получены следующие результаты: первый метод дал 96.22%, второй метод – 96.41%, третий метод 96.57%.

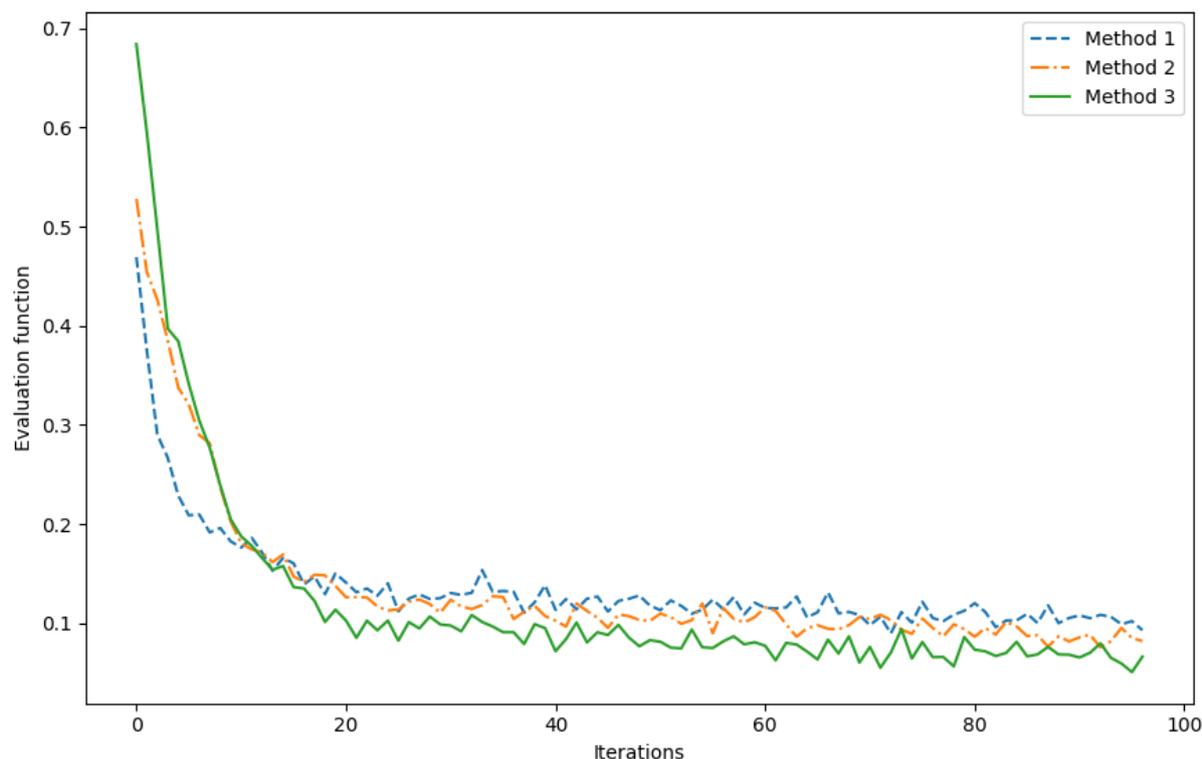


Рисунок. Динамика спада функции оценки

Выводы

На основе полученных результатов можно сделать вывод, что предложенный метод дает незначительное увеличение в эффективности распознавания нейронной сети. График динамики функции оценки показал, что скорость спада выше при стандартном методе градиентного спуска, однако при дальнейшем обучении, в предложенном методе значение функции ошибки становится меньше.

Список литературы

- [1] Шибзухов З. М., *Конструктивные методы обучения СП - нейронных сетей*, Наука, Москва, 2006, 159 с. [SHibzuhov Z. M., *Konstruktivnye metody obucheniya СП - nejronnyh setej*, Nauka, Moskva, 2006, 159 pp.]
- [2] Тимофеев А.В., Косовская Т.М., “Нейросетевые методы логического описания и распознавания сложных образов”, *Труды СПИИРАН.*, **27** (2013), 144-155. [Timofeev A.V., Kosovskaya T.M., “Nejrosetevye metody logicheskogo opisaniya i raspoznavaniya slozhnyh obrazov”, *Trudy SPIIRAN.*, **27** (2013), 144-155].
- [3] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton, “Deep learning”, *Nature* **521.7553**, 2015, 436-444.
- [4] Sutskever I., Martens J., Dahl G., Hinton G., “On the importance of initialization and momentum in deep learning”, *J. of Machine Learning Research*, **28** (2013), 1139-1147.

Список литературы (ГОСТ)

- [1] Шибзухов З. М. Конструктивные методы обучения $\Sigma\Pi$ - нейронных сетей. М.: Наука, 2006. 159 с.
- [2] Тимофеев А.В., Косовская Т.М. Нейросетевые методы логического описания и распознавания сложных образов // Труды СПИИРАН. 2013. Т. 27. С. 144-155.
- [3] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton Deep learning // Nature. 2015. vol. 521.7553. pp. 436-444.
- [4] Sutskever I., Martens J., Dahl G., Hinton G. On the importance of initialization and momentum in deep learning // J. of Machine Learning Research. 2013. vol. 28. pp. 1139-1147.

Для цитирования: Казаков М. А. К вопросу об оптимизации конструктивного метода обучения нейронных сетей // *Вестник КРАУНЦ. Физ.-мат. науки.* 2018. № 3(23). С. 184-189. DOI: 10.18454/2079-6641-2018-23-3-184-189

For citation: Kazakov M. A. On the optimization of the constructive method of training neural networks, *Vestnik KRAUNC. Fiz.-mat. nauki.* 2018, **23**: 3, 184-189. DOI: 10.18454/2079-6641-2018-23-3-184-189

Поступила в редакцию / Original article submitted: 08.06.2018