

ГИБРИДНЫЙ АЛГОРИТМ ОБУЧЕНИЯ НЕЙРОННЫХ СЕТЕЙ

В. В. Мацкевич

matskevich1997@gmail.com;

*Научный руководитель — В. В. Краснопрошин, доктор технических наук,
профессор*

На сегодняшний день для обучения нейронных сетей повсеместно применяются градиентные алгоритмы. Однако они не всегда обеспечивают необходимое качество решения. Алгоритмы случайного поиска, как правило, повышают качество решения, но медленно сходятся. В работе предлагается алгоритм обучения на основе объединения идей градиентного спуска и случайного поиска. На примере решения ряда задач показано, что данный алгоритм эффективнее современных алгоритмов обучения.

Ключевые слова: алгоритмы обучения; нейронные сети; случайный поиск; метод градиентного спуска.

ВВЕДЕНИЕ

На заре развития нейронных сетей, вычислительные мощности компьютеров были крайне малы. Для обучения нейронных сетей необходим был алгоритм не требующий большого объема вычислений.

Метод градиентного спуска обеспечивает высокую скорость сходимости за счет чего получил широкое распространение. Однако по мере развития цифровой техники, размерность и сложность входных данных выросла. Это привело к усложнению архитектур нейронных сетей и увеличения числа настраиваемых параметров. Более того, появились прикладные задачи, требующие высокого качества полученного решения. Для решения данной проблемы разрабатывались модификации градиентного метода с целью повышения качества полученного решения.

Однако эффективности модификаций оказалось недостаточно для решения ряда прикладных задач. Для повышения качества полученного решения были разработаны алгоритмы на основе случайного поиска. Однако они обладают крайне медленной скоростью сходимости и нечасто применяются на практике. Таким образом, проблема обучения нейронных сетей является актуальной.

АНАЛИЗ ПРОБЛЕМЫ

Проведем сравнение преимуществ и недостатков методов обучения.

Градиентные алгоритмы крайне быстро сходятся и зачастую на практике обеспечивают приемлемое качество решения. Однако к прикладной задаче предъявляется требование к дифференцируемости целевой функции, что существенно ограничивает класс решаемых прикладных задач.

Более того, для решения проблемы «взрывного» или «затухающего» градиента при обучении нейронных сетей на архитектуру накладывается ряд ограничений. Градиентные алгоритмы могут сходиться в точки, где значения градиента крайне близки к нулю. Это приводит к существенному снижению качества полученного решения.

Алгоритмы случайного поиска при решении задачи рассматривают все пространство решений. Это существенно повышает вероятность обнаружения оптимального решения. Единственным требованием является вычислимость оценки целевой функции для рассматриваемого решения.

Благодаря этому существенно расширяется класс решаемых прикладных задач. Однако, для рассмотрения всего пространства решений и точной оценки значения целевой функции, необходим огромный объем вычислений, что ограничивает его применение на практике.

ГИБРИДНЫЙ АЛГОРИТМ ОБУЧЕНИЯ

Идея гибридного алгоритма заключается в уменьшении рассматриваемого множества решений по сравнению со случайным поиском и его расширением по сравнению с градиентными алгоритмами.

Рассмотрим задачу минимизации некоторой целевой функции f . Будем предполагать, что оптимизируемые параметры можно разбить на наборы, в каждом из которых свой предполагаемый оптимальный отрезок значений. Алгоритм состоит из следующих шагов.

Предварительный шаг. Инициализация начального решения x_0 случайным образом и вычисление $f(x_0)$.

Шаг 1. Генерация случайного решения y на основе текущего решения x и вычисление $f(y)$. Этот шаг можно разбить на две части.

Шаг 1.1 Генерация случайных величин. Генерируется m равномерно распределенных на отрезке от нуля до количества параметров в наборе дискретных случайных величин $a_1, a_2, \dots, a_m$. Генерируется m случайных перестановок длиной, равной количеству в наборе параметров. Первые $a_1, a_2, \dots, a_m$ элементов каждой перестановки задают индексы изменяемых параметров в каждом наборе параметров соответственно.

Шаг 1.2 Генерация нового решения. Для каждого изменяемого параметра генерируется равномерно распределенная на отрезке $[-l/2; l/2]$ случайная величина b . Величина l зависит от того, какому набору принадлежит изменяемый параметр и равна $l_1, l_2, \dots, l_m$ соответственно. Значения l для каждого набора задаются как параметры алгоритма. Пусть x_i – изменяемый параметр, а его новое значение x_i' , тогда: $x_i' = x_i + b$.

Шаг 2. Замена текущего решения x на y , если $f(x) \geq f(y)$.

Шаг 3. Проверка критерия останова. Если за N подряд совершенных итераций не было переходов в новое решение, то алгоритм завершается. Иначе осуществляется переход на Шаг 1. N – параметр алгоритма.

С помощью настройки генерации случайного решения можно настраивать размер пространства поиска решений. При генерации решений в широкой окрестности от текущего, можно за несколько итераций сгенерировать практически любое решение из множества. Если для генерации нового решения задать ε -окрестность текущего решения, то алгоритм вырождается в простой градиентный метод. Данный алгоритм является компромиссом между высоким качеством полученного случайным поиском решения и быстрой сходимостью градиентных методов.

ЭКСПЕРИМЕНТЫ

Для сравнения эффективности разработанного алгоритма с градиентными методами рассмотрим решение задач сжатия цветных изображений и распознавания образов. Для сравнения с градиентными методами будет использован наилучший алгоритм обучения – following the moving leader (FTML) [1]. Для экспериментов использованы выборки CIFAR-10 [2] и STL-10 [3]. Выборка STL-10 использовалась для сжатия цветных изображений разрешением 96×96 , а CIFAR-10 – для классификации образов на цветных изображениях разрешением 32×32 . Для классификации образов использовались не очень сложно разделимые классы для упрощения сравнения эффективности алгоритмов обучения. Для оценки качества сжатия использовались наиболее распространенные функционалы качества: mean squared error (MSE), peak signal to noise ratio (PSNR), with human visual system (PSNR-HVS), structural similarity image measure (SSIM). Для сжатия цветных изображений была спроектирована глубокая доверительная сеть на основе ограниченных машин Больцмана. Для классификации образов – на основе автоэнкодеров с классификатором в конце в виде многослойного персептрона (см. Табл. 1, 2).

В решении задачи классификации образов (см. Табл. 1) гибрид не уступает ни по скорости, ни по качеству градиентам. Добиться преимущества в данной задаче достаточно сложно из-за сложности выборки.

Таблица 1

Результаты классификации образов (выборка CIFAR-10)

Пары классов	6-8	1-4	5-9	0-2
FTML	93,3	89,9	88,2	83,6
Гибридный	93,3	89,1	88,4	83,3

Таблица 2

Результаты сжатия изображений (выборка STL-10)

Алгоритм обучения	Степень сжатия, бит/пиксель	MSE	PSNR	PSNR-HVS	SSIM	Время обучения, ч
FTML	1,5	397	22,3	22,5	0,673	4,00
	0,75	756	19,4	19,5	0,509	6,00
Гибридный	1,5	415	22,1	22,3	0,661	10,0
	0,75	707	19,7	19,9	0,524	11,5

Из результатов сжатия цветных изображений видно, что по мере усложнения решаемой задачи (см. Табл. 2) разработанный алгоритм начинает существенно превосходить градиентные алгоритмы обучения. Он оказался в среднем вдвое медленнее градиента, но в данном случае это не критично, т.к. время обучения остается в разумных пределах.

ЗАКЛЮЧЕНИЕ

В работе предложен гибридный алгоритм обучения на основе совмещения идей случайного поиска и градиентного спуска. Экспериментально показано, что данный алгоритм превосходит по качеству полученного решения градиентные алгоритмы, однако немного уступает по скорости обучения, что не критично. При решении задачи классификации образов предложенный алгоритм показал равную с градиентами эффективность.

Предложенный в работе алгоритм не имеет существенных ограничений в применении на практике. Более того, по мере роста вычислительных мощностей увеличивается рассматриваемое алгоритмом множество решений за отведенное на обучение нейронной сети время, что приводит к росту качества полученного решения. Таким образом, можно сделать вывод, что предложенный в работе гибридный алгоритм обучения обладает определенной перспективой при решении прикладных задач.

Библиографические ссылки

1. Zheng Sh., Kwok J. T. Follow the moving leader in deep learning // Proc. of the 34-th International Conference on Machine Learning, 2017. Vol. 70. P. 4110–4119.
2. Выборка CIFAR-10 [Электронный ресурс]: Режим доступа: <https://www.cs.toronto.edu/~kriz/cifar.html> – Дата доступа: 04.03.2023.
3. Выборка STL-10 [Электронный ресурс]: – Режим доступа: web.archive.org/web/20110803194852/stanford.edu/~acoates/stl10/ Дата доступа: 24.04.2023.