

Тархов Д.А., Радченко Д.С. Распределенное обучение нейронных сетей. // Проблемы информатики в образовании, управлении, экономике и технике: Сб. статей XV Междунар. научно-техн. конф. – Пенза: ПДЗ, 2015. – С. 70-73.

УДК 004.75, 004.85, 519.683, 519.688

РАСПРЕДЕЛЕННОЕ ОБУЧЕНИЕ НЕЙРОННЫХ СЕТЕЙ

Д.А. Тархов, Д.С. Радченко

THE DISTRIBUTED TRAINING OF ARTIFICIAL NEURAL NETWORKS

D.A. Tarkhov, D.S. Radchenko

Аннотация. Рассмотрены методы обучения нейронных сетей на нескольких компьютерах, сообщающихся между собой через Интернет. Рассмотрена ситуация, когда обучение нейронных сетей производится на нескольких компьютерах, при этом вычислительные процессы на разных компьютерах взаимодействуют друг с другом.

Ключевые слова: нейронные сети, распределенные вычисления, глобальный экстремум, кластеризация, генетический алгоритм.

Abstract. This article discusses methods of neural networks training on several computers, communicating with each other via the Internet. It contains the situation where the training of neural networks is performed on several computers, at the same time computational processes on different computers communicate with each other.

Keywords: neural networks, distributed computing, global extremum, clustering, genetic algorithm.

Ввиду большой ресурсоёмкости процессов обучения нейронных сетей большой размерности возникают проблемы в задачах, требующих их использования. В таком случае актуальным является увеличение вычислительной производительности за счет их распараллеливания и реализации в распределённых системах.

На данный момент существует несколько подходов к распределённому обучению нейронных сетей, реализация которых имеет ряд недостатков, связанных с отсутствием эффективных распределённых алгоритмов многомерной оптимизации.

Проблема обучения нейронной сети с вычислительной точки зрения – это проблема нахождения глобального экстремума сложной многоэкстремальной функции с большим числом аргументов. Самый простой способ её решения состоит в том, чтобы использовать метод рестартов, т.е. на каждом из компьютеров запустить процесс поиска локального экстремума из своей точки, после чего выбрать лучший результат. Даже в этом простом случае возникает ряд интересных алгоритмических особенностей, если рассматривать не каждый локальный алгоритм в отдельности, а механизм взаимодействия локальных процессов.

Распределенный вариант метода рестартов

Если в каждый момент времени мы можем в каждом узле получить информацию обо всех процессах сразу, то не возникает задачи распределённого обучения, так как отсутствуют описанные выше специфические трудности. Если

мы такой информации получить не можем, то можно предложить следующие варианты организации вычислений:

Шаг 1. На каждом из компьютеров запускаем процесс поиска локального экстремума из своей точки.

Шаг 2. Каждый узел через определённые промежутки времени посылает в другой узел, выбранный случайно или по определённому закону, свой идентификатор и значение функционала ошибки.

Шаг 3. Узел, получивший такую информацию, сравнивает присланное значение функционала с тем, которое получено локальным процессом и отправляет далее информацию о лучшем результате. Является ли адрес посылки сообщения случайным или подчиняется определённому закону, надо чтобы информация поступала в узлы максимально равномерно.

Шаг 4. Если какой-либо узел долго не получает информацию из других узлов, то он начинает посылать свою более часто – таким образом система будет адаптироваться к замедлению работы сети и возможным сбоям в ней.

Шаг 5. Полученная информация сразу или через заданный промежуток времени сказывается на ходе работы локального алгоритма. Самый простой способ работы с поступившей информацией состоит в том, чтобы после истечения заданного времени работы сравнить свой результат с лучшим из присланных результатов.

Шаг 6. Когда критерием остановки вычислительного процесса является достижение некоторого фиксированного заранее значения минимизируемого функционала, тогда возможны следующие варианты: если это значение достигнуто локально, то процесс оптимизации останавливается и происходит рассылка результатов, если в другом узле – то этот узел должен прислать свои результаты – веса и т.д., если соответствующий узел их ещё не прислал, то их следует запросить.

Аналогично рассматриваются распределенные варианты других методов – например, многогранника (Нелдера-Мида):

Шаг 1. На каждом из компьютеров запускаем процесс поиска локального экстремума из своей точки.

Шаг 2. Каждый компьютер посылает на какой-либо другой компьютер из этой совокупности результаты своей работы – веса сети и величину ошибки.

Шаг 3. Получивший эту информацию компьютер пересчитывает среднее значение весов и пересылает эту информацию дальше, присовокупив величину ошибки, достигнутой его локальным методом.

Шаг 4. Повторяем шаг 3, при этом каждый новый узел в цепочке получает среднее значение весов сети и совокупность значений функционалов ошибки, достигнутых предыдущими компьютерами. Принявший информацию компьютер может продолжить эту цепочку, т.е. пересчитать среднее значение весов по известной рекуррентной формуле.

Шаг 5. Если выполнены заданные условия (например, значение ошибки в этом узле больше, чем заданное число ошибок в предыдущих узлах), то совершается отражение весов относительно среднего в соответствии с методом многогранника и с этой точки продолжается работа локального метода минимизации. При этом цепочка пересылки данных может обрываться, продолжаться дальше или начинаться новая.

Шаг 6. Если сеть ненадёжна и компьютер долго не получает таких «посланий» от других узлов, он сам инициирует новую цепочку.

Работа поддержана грантом РФФИ №14-01-00660А

Библиографический список

1. Калинин А.В., Подвальный С.Л. Технология нейросетевых распределенных вычислений. – Воронеж: ВГУ, 2004. –121 с.
2. Царегородцев В.Г. Перспективы распараллеливания программ нейросетевого анализа и обработки данных // Материалы III Всерос. конф. «Математика, информатика, управление». – Иркутск, 2004.
3. Haykin S. Neural Networks. A comprehensive foundation. – 1999. – 823 с.
4. Тархов Д.А. Нейросетевые модели и алгоритмы. – М.: Радиотехника, 2014. – 352 с.
5. Vasilyev A.N., Tarkhov D.A. Mathematical Models of Complex Systems on the Basis of Artificial Neural Networks // Nonlinear Phenomena in Complex Systems. 2014. vol. 17, no. 3. – P. 327–335.
6. Nelder, J.A. and Mead, R., “A simplex method for function minimization”, Comput. J., (1965) 7. – P. 308–313.
7. Lagarias, J.C., Reeds, J.A., Wright, M.H., and Wright, P.E. “Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions”, SIAM J. Optim. (1998), 9. – P. 112–147.

Тархов Дмитрий Альбертович
Санкт-Петербургский
государственный политехнический
университет Петра Великого,
г. Санкт-Петербург, Россия
E-mail: dtarkhov@gmail.com

Tarkhov D.A.
Peter the Great Saint-Petersburg
Polytechnical University,
Saint-Petersburg, Russia

Радченко Дарья Сергеевна
Санкт-Петербургский
государственный политехнический
университет Петра Великого,
г. Санкт-Петербург, Россия
E-mail: darya.darya.radchenko@mail.ru

Radchenko D.S.
Peter the Great Saint-Petersburg
Polytechnical University,
Saint-Petersburg, Russia