

НЕДОСТАТКИ МЕТОДА ОБРАТНОГО РАСПРОСТРАНЕНИЯ ОШИБКИ

Акулин Евгений Валерьевич

аспирант, Сибирский государственный университет имени М. Ф. Решетнева РФ, г. Красноярск

DISADVANTAGES OF THE BACK PROPAGATION METHOD

Yevgeniy Akulin

Graduate student Siberian State University named after M. F. Reshetnev, Russia, Krasnoyarsk

Аннотация. Явные и неявные методы исследования – это подходы, которые используются в численном анализе для последующего получения численных приближений к решениям зависящих от дифференциальных уравнений в частных производных, применяющихся в современном компьютерном моделировании физических процессов. В свою очередь, явные методы исследования вычисляют состояние системы, неявные же методы изучения находят решение, с помощью решения уравнений, включающих текущее состояние системы.

Abstract. Explicit and implicit research methods are approaches that are used in numerical analysis to subsequently obtain numerical approximations to solutions of partial differential equations that depend on partial differential equations used in modern computer modeling of physical processes. In turn, explicit research methods calculate the state of the system, while implicit research methods find a solution by solving equations that include the current state of the system.

Ключевые слова: метод; соединение; стандартизация; функции; обучение; система; уравнение.

Keywords: method; connection; standardization; functions; training; system; equation.

В процессе обучения метод обратного распространения ошибок имеет ряд преимуществ над прямым методом, который использует распределение Коши, когда определенные шаги выполняются в неверном направлении из-за случайности выбора требующих достаточно много времени. Для большего ускорения обучения необходимо соединить прямой и обратный методы обучения. Коррекция весов, вычисленная алгоритмом обратного распространения, а так же случайный шаг, который задается алгоритмом Коши, приводит в итоге к системе, которая находит глобальный минимум гораздо быстрее, чем система, производящая обучение каждым из методов по отдельности.

Трудности метода обратного распространения ошибки. Несмотря на существенную мощь метода обратного распространения, в процессе его применения может возникнуть ряд определенных трудностей, некоторая часть из которых упрощается при использовании нового алгоритма.

Сходимость. Довольно хорошо известно, доказательство сходимости происходит на языке дифференциальных уравнений при частных производных, которые могут быть справедливы только в случае, когда сама коррекция весов будет выполнена при помощи бесконечно малых шагов в самом пространстве весов. В свою очередь, бесконечно малые шаги могут привести к бесконечному времени сходимости. В таком случае, метод будет терять свою силу в практических применениях, потому как нет доказательства, того что обратное распространение может сходиться при конечном количестве шагов. Проведенные эксперименты показали, что ИНС обычно обучаются, но при этом время обучения довольно велико и практически непредсказуемо.

Паралич. При определенных условиях ИНС, обучении может попасть в состояние, при котором модификация весов не приведет к действительным изменениям ИНС. Такой «паралич» является довольно серьезной проблемой. Он способен серьезно увеличить время обучения. Паралич возникает во время обучения ИНС, в процессе которого у большей части нейронов значения весов может становиться значительно большим, чтобы дать большее значение взвешенной сумме. Это способно привести к тому, что большинство или даже все нейроны смогут функционировать при гораздо больших значениях аргументов сжимающей функции активации в той области, при которой производная функция может приближаться к нулю.

Значительное количество эвристики применяется для предохранения от паралича, но на данный момент времени они могут рассматриваться только лишь как экспериментальные. Для преодоления же данной ситуации существуют модификации алгоритма, которые принимают участие лишь в направлении антиградиента и совсем не учитывающие саму его величину. Также существует адаптивный алгоритм самого выбора шага, который способен автоматически корректировать размер шага в процессе обучения.

Локальные минимумы. В методе исследования обратного распространения ошибки для получения коррекции значений весов связей ИНС применяется градиентный спуск по поверхности ошибки в самом пространстве весов связей в направлении локального минимума. Сам же локальный минимум может являться приемлемым или неприемлемым решением. Довольно часто на практике поверхность ошибки сложной ИНС очень сильно изрезана и может состоять из складок, долин и холмов в пространстве высокой размерности. ИНС способен попасть в локальный минимум и при этом не обнаружить глобальный минимум. В самой же точке локального минимума все направления ведут вверх, при котором ИНС не будет неспособна из него самостоятельно выбраться.

Для преодоления изложенной выше трудности обучения ИНС, может расширяться размерность пространства весов с помощью значительного увеличения скрытых слоев и самого числа нейронов скрытого слоя или же определенное число раз проводят обучение и уже затем выбирают вариант обученной ИНС, способный дать наиболее лучшие результаты. В некоторых случаях, к изменению значений весов связей можно добавить шум. В свою очередь, это позволяет ИНС выйти из локальных минимумов. Даже когда сеть обучена, не всегда возможно уверенно сказать, найден ли при помощи обратного распространения глобальный минимум. Если же решение будет неудовлетворительно, необходимо давать весам совершенно новые изначальные случайные значения и затем уже повторно обучать сеть без самой гарантии, того что обучение закончится на данной попытке или даже то, что глобальный минимум вообще будет найден.

Переобучение. Довольно большая точность на обучающей выборке, способна привести к определенной неустойчивости результатов на примере тестовой выборки. Чем наилучшим образом ИНС адаптирована к определенным условиям, тем значительно меньше она будет способна к экстраполяции и даже обобщению. В данном случае ИНС способна моделировать не функцию, а только шум, который присутствует в обучающей выборке. Данное явление имеет название переобучение. Значительное средство борьбы с данным недостатком – активное использование определенной выборки примеров, которое сможет использоваться для конкретного выявления переобучения ИНС. Определенное ухудшение характеристик ИНС при непосредственной работе с подтвержденной выборкой будет указывать на возможное переобучение.

Большим недостатком данного приема является определенное уменьшение числа примеров, которое можно в свою очередь использовать в обучающем множестве, потому что определенное уменьшение количества обучающей выборки может снижать качество самой работы ИНС. Так же может возникать проблема наиболее оптимального разбиения наших исходных данных на тестовую, обучающую и подтверждающую промерную выборку. В свою очередь, при случайной выборке довольно разные разбиения базы данных могут дать и довольно разные оценки.

Несмотря на значительное улучшение скорости обучения, которое предоставляется машиной Коши в сравнении с машиной Больцмана, при этом время сходимости может в сотни раз превышать время для алгоритма обратного распространения. Необходимо отметить то, что сетевой паралич особенно опасен для самого алгоритма обучения Коши, и в особенности для ИНС с значительной нелинейностью типа логистической функции. Сама же бесконечная дисперсия распределения Коши может приводить к значительным изменениям весов неограниченной величины. Набольшие изменения весов может приниматься даже тогда, когда они совсем неблагоприятны, при этом зачастую приводя к значительному насыщению сетевых нейронов с последующим вытекающим отсюда риском паралича.

В наше время, появление алгоритма стало огромным событием и положительно оказало влияние на развитие нейросетей, ведь с помощью его реализуется наиболее эффективный способ обучения многослойного персептрона. В это же время, алгоритм предлагает более оптимальное решение всех возможных проблем. С помощью его, был развеян возможный пессимизм относительно самого машинного обучения многослойных машин.

Список литературы:

1. Голубев Ю. Ф. Нейросетевые методы в мехатронике. – М: Изд-во Моск. унта, 2007. – 157 с.
2. Тарков М.С. Нейрокомпьютерные системы - Интернетуниверситет информационных технологий - ИНТУИТ.ру, БИНОМ. Лаборатория знаний, 2006. -144 с.
3. Хайкин С. Нейронные сети: полный курс, 2-е издание, исправленное: Пер. с англ. – М: ООО «И.Д. Вильямс», 2006. – 1104 с.