

ДИНАМИЧЕСКАЯ ИНИЦИАЛИЗАЦИЯ ШАГА ОБУЧЕНИЯ В АЛГОРИТМЕ BACK PROPAGATION ERROR

Савицкий Ю. В., Савицкий А. Ю.

Кафедра интеллектуальных информационных технологий, Брестский государственный технический университет

Брест, Республика Беларусь

E-mail: yury.savitsky@tut.by, ant.sav95@tut.by

В работе предлагается и обосновывается методика динамической инициализации индивидуального шага обучения в алгоритме Back Propagation Error (обратного распространения ошибки), по настоящее время остающемся наиболее популярным для обучения многослойных нейронных сетей. Результаты вычислительных экспериментов демонстрируют практическую значимость и перспективность метода.

ВВЕДЕНИЕ

В настоящее время в мире накоплено значительное количество теоретических результатов и большой практический опыт в применении аппарата искусственных нейронных сетей для обработки информации в различных сферах. Предложен ряд высокоэффективных алгоритмов обучения (в том числе Back Propagation Error, ВРЕ), позволяющих улучшить адаптивные свойства нейросетевых моделей.

Ключевым моментом задачи проектирования нейронных систем по настоящее время остается проблема эффективного обучения нейронных сетей [1]. При этом используемые для обучения алгоритм ВРЕ и его известные модификации обладают принципиальными недостатками, связанными с необходимостью выбора констант обучения, следствием чего являются проблемы низкой сходимости и стабильности обучения, а также проблемы локальных минимумов [1,2,3,8,9]. Применяемые в настоящее время для обучения нейронных сетей более совершенные методы оптимизации (сопряженных градиентов [4, 5], Ньютона [5, 7], Левенберга-Марквардта [5, 10]), хотя и позволяют значительно улучшить процесс сходимости алгоритма, также обладают существенными недостатками, сильно ограничивающими их эффективное применение в практических задачах. Наиболее характерными из этих недостатков являются большая вычислительная сложность итерации обучения, высокая чувствительность алгоритма к погрешностям вычислений, низкая сходимость при большом удалении целевой функции обучения от точки минимума [5]. Все это является весьма серьезным ограничением для использования нейросетевых моделей в практических задачах.

Данная работа является продолжением комплекса научных исследований по оптимизации работы алгоритма ВРЕ в контексте решения задачи выбора значения шага обучения. Главной целью своей работы авторы считают снижение степени неопределенности значения шага обуче-

ния в алгоритме ВРЕ при решении практических задач организации нейросетевых моделей.

I. МЕТОДИКА ИНИЦИАЛИЗАЦИИ ШАГА ОБУЧЕНИЯ

Главная идея данной работы заключается в применении в алгоритме ВРЕ метода индивидуальной случайной инициализации шага обучения на каждой обучающей итерации. Предпосылки и суть идеи сводятся к следующему:

1. В алгоритме ВРЕ и во всех его наиболее распространенных модификациях традиционно используется стратегия инициализации каждого весового коэффициента нейронов случайным числом, равномерно распределенным на некотором диапазоне, границы которого задаются эмпирически. Необходимость такого подхода обусловлена неопределенностью информации о начальных значениях весовых коэффициентов. При этом результат обучения существенно зависит от удачной инициализации сети; в ряде случаев, для достижения приемлемой ошибки обучения приходится неоднократно повторять данный процесс.

2. Проблему неопределенности шага обучения в алгоритме ВРЕ предлагается решить на базе аналогичного подхода. Это означает, что на каждой итерации алгоритма для каждого весового коэффициента значение шага определяется методом вызова функции, генерирующей случайное равномерно распределенное число. Нижняя и верхняя границы инициализации определяются эмпирически как минимальное и максимальное приемлемые значения шага. Предпосылками подхода является то, что при неудачном выборе значения глобального шага существует вероятность попадания целевой функции обучения в локальный минимум (что часто и имеет место в реальных задачах обучения). С другой стороны, случайное варьирование шага для каждого весового коэффициента в допустимом диапазоне значений на каждой итерации алгоритма уменьшает такую вероятность, что должно обеспечивать

более высокую сходимость алгоритма к приемлемой ошибке обучения.

II. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

Были проведены две группы экспериментов. Многослойная нейронная сеть архитектуры 7-5-1 – 7 входных нейронов, 5 скрытых нейроэлементов с сигмоидной функцией активации и 1 нейрон линейного типа – обучена прогнозированию хаотического процесса Энона [6] по методу скользящего окна (взято 510 элементов ряда) [6]. Количество выполненных итераций обучения $NI_T=10000$.

1. Шаг обучения – случайный, выбираемый по вышеуказанной методике, с диапазоном инициализации $[0,01; 0,001]$. Результаты для 20 попыток обучения: средняя ошибка обучения (средний квадрат ошибки) $MSE=7,53E-06$; среднеквадратичное отклонение ошибки обучения $SD=5,08E-04$.

2. Шаг обучения – глобальный детерминированный, равный 0,045 (середина диапазона инициализации шага для предыдущего эксперимента). Результаты для 20 попыток обучения: $MSE=2,34E-05$; $SD=7,97E-03$. Таким образом, за счет применения методики случайной инициализации шага в алгоритме ВРЕ ошибку обучения удалось сократить в 3,1 раза, при этом еще более существенно уменьшается параметр SD . Последнее говорит о значительном повышении стабильности процесса сходимости алгоритма ВРЕ для вышеуказанного числа экспериментов обучения. Аналогичные результаты были получены и в других задачах обучения. В частности, при обучении нейронной сети архитектуры 5-4-1 на прогнозирование временных рядов, синтезированных на базе суперпозиции периодических функций, ошибка обучения сократилась в среднем в 1,74 раза, при одновременном уменьшении параметра SD почти в 2 раза (диапазон инициализации шага $[0; 0,1]$, размер глобального шага 0,05, количество попыток обучения 30).

III. ЗАКЛЮЧЕНИЕ

По мнению авторов, полученные результаты имеют практическую значимость, поскольку, во-первых, увеличивают скорость и стабильность процесса сходимости алгоритма ВРЕ, во-вторых, создают предпосылки для эффективной параллелизации процесса обучения на многокомпонентных вычислительных архитектурах

– предложенная методика динамической инициализации шага обучения не зависит от обучающего множества, промежуточных параметров обучения нейронной сети и других факторов, влияющих на эффективность процесса параллелизации.

В качестве объективных недостатков метода можно отметить некоторые незначительные осцилляции ошибки при переходе от итерации к итерации, которые в целом несущественно влияют на конечный результат обучения нейронной сети.

IV. СПИСОК ЛИТЕРАТУРЫ

1. Hertz J., Krogh A., Palmer R. Introduction to the Theory of Neural Computation. – Addison Wesley Publishing Company. – 1991. – 327 p.
2. Vladimir Golovko, Yury Savitsky, "New Approach of the Recurrent Neural Network Training", Proc. of the Int. Conf. on Neural Networks and Artificial Intelligence ICNNAI'99, 12-15 October 1999, Brest, Belarus. – pp. 32-35.
3. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Representations by Backpropagating Errors," Nature, vol. 323, 1986. – pp. 533-536.
4. E. M. Johansson, F. U. Dowla, and D. M. Goodman, "Backpropagation Learning for Multilayer Feedforward Neural Networks Using the Conjugate Gradient Method," Int. J. Neural Systems, vol. 2, no. 4, pp. 291-302, 1992. R. Battiti, "First- and Second-Order Methods for Learning: Between Steepest Descent and Newton Methods," Neural Comput., vol. 4, pp. 141-166. – 1992.
5. Поляк Б. Т. Введение в оптимизацию. – М: Наука, 1983. – 384 с.
6. V. Golovko, Y. Savitsky, N. Maniakov. Neural Networks for Signal Processing in Measurement Analysis and Industrial Applications: the Case of Chaotic Signal Processing // chapter of NATO book "Neural networks for instrumentation, measurement and related industrial applications". - Amsterdam: IOS Press, 2003. – pp.119-143.
7. S. Osowski, P. Bojarczak, and M. Stodolski, "Fast Second-Order Learning Algorithm for Feedforward Multilayer Neural Networks and its Applications," Neural Networks, vol. 9, no. 9, pp. 1583-1596. – 1996.
8. Осовский С. Нейронные сети для обработки информации / Пер. с польского И.Д. Рудинского. – М.: Финансы и статистика, 2002. – 334 с.
9. V. Golovko, Yu. Savitsky, Th. Laopoulos, A. Sachenko, L. Grandinetti. Technique of Learning Rate Estimation for Efficient Training of MLP // Proc. of Int. Joint Conf. on Neural Networks IJCNN'2000, Como, Italy. Vol. 1. – 2000. – pp. 323-329.
10. T. H. Martin and B. M. Mohammad, "Training Feedforward Network with Marquardt Algorithm," IEEE Trans. Neural Networks, vol. 5, Nov. 1996. – pp. 959-96.