

Автоматическое построение нейросети оптимальной сложности.

Забазнов А. Г.¹, Бахтеев О. Ю.¹, Стрижов В. В.^{1,2}

antoniozabaznov@yandex.ru; bakhteev@phystech.edu; strijov@phystech.edu

Московский физико-технический институт¹;

Вычислительный центр им. А. А. Дородницына ФИЦ ИУ РАН^{1,2}

В данной работе рассматривается задача выбора оптимальной модели нейросети и оптимизация её параметров. В общем случае нейросеть представляется графом, ребрами которого являются нелинейные операции, а вершины – промежуточные представления выборки, полученные под действием этих операций. Параметры сети можно разделить на три типа: параметры, отвечающие за итоговое качество классификации; гиперпараметры, отвечающие за процесс обучения и предотвращение переобучения; структурные параметры, отвечающие непосредственно за структуру сети, такие как количество слоев и тип нелинейных операций. Предлагается подход выбора структуры нейросети на основе вариационного вывода и алгоритма выбора оптимальных значений гиперпараметров с использованием релаксации, учитывающий неточности при оптимизации параметров и позволяющий находить наиболее устойчивые модели.

Ключевые слова: *нейронные сети, автоматическое построение нейронных сетей, оптимальная структура нейронной сети*

1 Введение

Выбор оптимальной модели в машинном обучении является одной из ключевых задач. Под оптимальной моделью понимается структура обучаемой сети и совокупность её гиперпараметров, которая даёт приемлемое качество классификации или регрессии при небольшом количестве параметров. В данной работе в качестве критерия выбора модели предлагается сложность модели, то есть величина, учитывающая сложность описания совокупности выборки и модели. Под описанием выборки понимается приближенная оценка сложности модели, основанная на связи с её правдоподобием [1]

Существует несколько подходов выбора модели оптимальной сложности. В работе [2] используется метод прореживания модели. Он заключается в построении заведомо переусложнённой модели с дальнейшим удалением параметров, не влияющих на качество классификации, таким образом получается сеть наименьшего размера. Ещё одним способом, предложенным в работе [3], являются байесовские методы оптимизации параметров нейронных сетей. В работе [4] для оптимизации модели предлагается использовать метод градиентного спуска.

Одна из проблем оптимизации моделей глубокого обучения – большое количество параметров и гиперпараметров, которое может достигать миллионов. Кроме того, сам процесс оптимизации становится ресурсоёмким. Задача выбора модели глубокого включает в себя выбор стратегии построения модели, эффективной по вычислительным ресурсам. Существуют методы градиентной оптимизации совокупности параметров и гиперпараметров.

В данной работе построение модели оптимальной сложности происходит в процессе самого обучения. В основе разработанного метода лежит алгоритм DARTS, предложенный в работе [5]. Для выбора оптимального набора гиперпараметров предлагается параметризовать структуру модели некоторым действительным вектором, путём перехода от дискретного множества возможных значений гиперпараметров к непрерывному множеству их комбинаций.

Проверка и анализ метода проводится на выборке Boston Housing [6], MNIST [7] и синтетических данных. Проводится сравнение представленного метода с эвристическими алгоритмами выбора модели.

Литература

- [1] *Grunwald P.* A Tutorial Introduction to the Minimum Description Length Principle. 2005.
- [2] *Yann Le Cun, John S. Denker and Sara A. Solla.* Optimal Brain Damage. 1989.
- [3] *A. Neal and M. Radford* Bayesian Learning for Neural Networks.. 1995.
- [4] *J. Luketina, M. Berglund, T. Raiko, and K. Gref* Scalable gradient-based tuning of continuous regularization hyperparameters. 2016.
- [5] *Hanxiao L., Simonyan K., Yang .Y* DARTS: Differentiable Architecture Search. 2018. URL: <https://arxiv.org/abs/1806.09055>.
- [6] *Harrison Jr. , Rubinfeld D., Daniel L.* Hedonic housing prices and the demand for clean air. 1978. URL: <https://archive.ics.uci.edu/ml/machine-learning-databases/housing/>.
- [7] *Yann LeCun, Corinna Cortes, Christopher J.C. Burges,* The MNIST Database of Handwritten Digits 1998. URL: <http://yann.lecun.com/exdb/mnist/>

Поступила в редакцию