

Автоматическое построение нейронной сети оптимальной сложности

Товкес А. А.¹, Бахтеев О. Ю.¹, Стрижов В. В.²

tovkes.aa@phystech.edu; bakhteev@phystech.edu; strijov@phystech.edu

Московский физико-технический институт¹;

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

Работа посвящена задаче выбора оптимальной по сложности модели нейросети. Нейросеть представляется в виде вычислительного графа, где ребрам соответствуют базовые функции, а вершинам — промежуточные представления выборки под действием этих функций. Параметры сети разделяются на непосредственно параметры модели, которые определяют итоговое качество классификации; гиперпараметры, определяющие процесс обучения и предотвращение переобучения; структурные параметры, определяющие непосредственно структуру модели. Для решения задачи оптимизации предлагается проводить релаксацию структуры нейросети. Рассмотрено изменение характеристик нейросети при возмущении структурных параметров. Для анализа качества представленного алгоритма проводятся эксперименты на выборках Boston, MNIST и CIFAR-10.

Ключевые слова: *нейронные сети, оптимизация гиперпараметров, оптимальная структура нейронной сети.*

1 Введение

В данной работе рассматривается задача построения оптимальной нейронной сети. В силу высокой вычислительной сложности, время оптимизации нейронных сетей может занимать большое количество времени. Поэтому построение и выбор оптимальной структуры нейронной сети также является вычислительно сложной процедурой, которая значительно влияет на итоговое качество модели.

Под оптимальной моделью понимается та модель, которая дает хорошее качество и имеет максимально простую структуру. Под структурой понимается набор гиперпараметров нейронной сети.

Существует ряд подходов к построению нейронной сети. В работе [1] предлагается использовать метод прореживания модели. Он заключается в построении переусложненной модели, с последующим удалением параметров, не влияющих на качество. В [2] используется предсказание структуры модели другой нейросетью. Кроме того в [3] используется метаобучение, которое по некоторой входной выборке, возвращает оптимальные гиперпараметры.

В данной работе исследуется метод построения нейросети, когда параметры и гиперпараметры оптимизируются одновременно. В основе разработанного метода лежит алгоритм DARTS, предложенный в работе [4]. Для выбора оптимального набора гиперпараметров предлагается параметризовать структуру модели некоторым действительным вектором, таким образом переходя от дискретного множества значений к непрерывному.

Проверка и анализ метода проводится на выборке Boston Housing [5], MNIST [6], CIFAR-10 [7] и синтетических данных. Результат сравнивается с моделью, полученной при помощи базовых алгоритмов. Оценивается не только качество, но и вычислительная сложность, устойчивость к возмущениям параметров.

Литература

- [1] Yann Le Cun, John S. Denker and Sara A. Solla. Optimal Brain Damage. 1989.

- [2] *Han Cai, Tianyao Chen, Weinan Zhang, Yong Yu, Jun Wang.* Efficient Architecture Search by Network Transformation.. 2017. URL: <https://arxiv.org/abs/1707.04873>.
- [3] *D. Maclaurin and D. Duvenaud and R. Adams.* Gradient-based Hyperparameter Optimization Through Reversible Learning 2015.
- [4] *Hanxiao L., Simonyan K., Yang .Y.* DARTS: Differentiable Architecture Search. 2018. URL: <https://arxiv.org/abs/1806.09055>.
- [5] *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/>.
- [6] *Yann LeCun, Corinna Cortes, Christopher J.C. Burges.* The MNIST Database of Handwritten Digits 1998. URL: <http://yann.lecun.com/exdb/mnist/>
- [7] *A. Krizhevsky, V. Nair, G. Hilton.* The CIFAR-10 dataset 2009. URL: <http://www.cs.toronto.edu/~kriz/cifar.html>