

Автоматическое построение нейросети оптимальной сложности

Криницкий К. Д. , Бахтеев О. Ю. , Стрижов В. В.

krinitskiy.kd@phystech.edu; bakhteev@phystech.edu; strijov@phystech.edu

Аннотация: В этой статье рассматривается задача поиска оптимальной структуры нейронной сети. Нейросеть рассматривается как вычислительный граф, реализуемый с помощью библиотеки Pytorch. Предусматривается, что число структурных параметров можно уменьшить без существенной потери качества классификации или регрессии. Будут изучены изменения характеристик нейронной сети при колебании структурных параметров. Итоговым результатом будет являться модель, дающая приемлемое качество классификации либо регрессии, и не являющаяся избыточной по параметрам.

Ключевые слова: *нейронные сети, графовые вычисления, оптимизация гиперпараметров, вариационный вывод*

Введение

В данной работе решается вопрос построения нейронной сети оптимальной сложности. Под оптимальностью имеется ввиду та модель, которая является не избыточной по своим параметрам, но при этом дающая относительно приемлемый результат классификации либо регрессии. В данной статье, конкретно, рассматривается оптимизация гиперпараметров, таких как: размерность слоев и их количество, функция активации.

Как известно, существует немало способов выбора модели оптимальной сложности. Гауссовская модель [1], в которой поясняется как нужно оптимизировать структуру, в случае недостатка информации о входных данных. В [2] (глава 28) применяется байесовская модель, а также говорится о принципе "Бритва Оккама", который гласит, что из моделей одинаковой точности выбирается наиболее простая. В работе [3] рассматривается градиент спуск, также являющийся одним из способов оптимизации.

Построение оптимальной нейронной сети - задача ресурсоемкая и вычислительно трудная. Из-за огромного количества гиперпараметров время обучение сети сильно возрастает. В данной работе используется эффективный, с точки зрения ресурсозатратности, алгоритм, основанный на методе DARTS [4]. Выбор наилучших гиперпараметров происходит благодаря процедуре релаксации: переход от дискретного набора параметров к непрерывному.

Проверка полученного алгоритма произведена на данных MNIST [5], CIFAR-10 [6], Boston Housing [7] также на синтетических данных. Результат сравнивается с моделями, построенными на базовых алгоритмах.

Литература

- [1] *Carl E. Gaussian Processes in Machine Learning.* 2005.
- [2] *David J.C. MacKay Information Theory, Inference, and Learning Algorithms.* 2005.
- [3] *J. Luketina, M. Berglund, T. Raiko, and K. Gref Scalable gradient-based tuning of continuous regularization hyperparameters.* 2016.
- [4] *Hanxiao L., Simonyan K., Yang .Y DARTS: Differentiable Architecture Search.* 2018. URL: <https://arxiv.org/abs/1806.09055>.

- [5] *Yann LeCun, Corinna Cortes, Christopher J.C. Burges*, The MNIST Database of Handwritten Digits 1998. URL:<http://yann.lecun.com/exdb/mnist/>
- [6] *A. Krizhevsky, V. Nair, G. Hilton*. The CIFAR-10 dataset 2009. URL: <http://www.cs.toronto.edu/~kriz/cifar.html>
- [7] *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/>.