

Автоматическое построение нейросети оптимальной сложности

Улитин А.Ю. , Бахтеев О.Ю. , Стрижов В.В.

ulitin.ayu@phystech.edu

Московский физико-технический институт

Работа посвящена поиску оптимальной модели нейросети. Нейросеть представляется как граф, где ребрам соответствуют нелинейные операции, а вершины - промежуточные представления. Параметры сети разделим на три типа: Параметры, отвечающие за итоговое качество классификации, гиперпараметры, отвечающие за процесс переобучения и предотвращение переобучения, а также структурные параметры, которые отвечают за структуру модели. Структура нейросети определяется вершинами симплекса. Будем проводить релаксацию структуры для решения задачи оптимизации.

Ключевые слова: *нейросети, оптимизация гиперпараметров, робастность модели.*

Введение

В данной работе рассматривается метод построения оптимальной нейронной сети. Под оптимальной сетью понимается ее максимальная простая структура. Под структурой понимается набор гиперпараметров: количество слоев, нейронов в каждом слое, а также функции активации в каждом нейроне. Рассмотрим наиболее распространенный композиция нейронов — многослойный персептрон. Нейроны располагаются слоями, при этом каждый нейрон из следующего слоя связан со всеми нейронами предыдущего. А построив задачу на простой композиции нейронов, мы будем переходить к другим видам.

Существует несколько способов построения оптимальной нейронной сети. Один из основных - оптимальное прореживание [1]. Он заключается в том, что из максимально сложной модели мы убираем связи и получаем упрощенную сеть. В работе [2] предложен байесовский метод оптимизации сети, а в работе [3] рассмотрен метод градиентного спуска.

Ввиду того, что у моделей огромное количество параметров и гиперпараметров, процесс оптимизации может быть затратным. В данной работе используется эффективный по ресурсам метод, в основе которого лежит алгоритм DARTS [4], где на вход мы получаем некоторый набор входных данных, а также функции активации. Оптимизируя параметры и гиперпараметры параллельно, мы на выходе получим оптимальную нейронную сеть.

Проверка и анализ метода проводится на выборках [5], [6] и синтетических данных. В эксперименте проводится сравнение полученного результата с моделями, полученными другими базовыми алгоритмами.

Литература

- [1] *Yann Le Cun, John S. Denker and Sara A. Solla. Optimal Brain Damage. 1989.*
- [2] *A. Neal and M. Radford Bayesian Learning for Neural Networks.. 1995.*
- [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] *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-datab...>*

- [6] *Yann LeCun, Corinna Cortes, Christopher J.C. Burges*, The MNIST Database of Handwritten Digits 1998. URL:<http://yann.lecun.com/exdb/mnist/>