

Автоматическое построение нейросети оптимальной сложности

Маркин В. О., Бахтеев О. Ю., Стрижов В. В.

markin1198@mail.ru

Московский физико-технический институт

В работе рассматривается задача построения оптимальной структуры нейронной сети и исследуется вопрос устойчивости построенной модели. Для оптимизации структурных параметров используется переход от выбора конкретной архитектуры к выбору комбинации различных архитектур сети и вариационный подход. Также исследуется влияние изменения данных на структуру сети. Для оценки качества и устойчивости моделей, построенных при помощи данного метода, проводятся эксперименты на выборке Boston, MNIST и синтетических данных. Проводится сравнение предложенного алгоритма с другими методами поиска оптимальных моделей нейронной сети.

Ключевые слова: *нейронные сети, оптимизация гиперпараметров, устойчивость нейросетевой модели.*

1 Введение

При использовании нейростетевых моделей в анализе данных часто встает вопрос о выборе архитектуры модели. Нейронная сеть имеет большое число гиперпараметров и долгое время - для выбора архитектуры использовался перебор и различные эвристические соображения. Такой подход вычислительно неэффективен и не дает гарантий оптимальности полученной модели.

В данной работе рассматривается задача построения оптимальной нейронной сети. Под оптимальной моделью понимается не та модель, которая дает максимально хорошее качество, а та, которая еще и имеет максимально простую структуру. Под структурой нейронной сети понимается набор гиперпараметров, таких как число слоев, размерность каждого слоя, функции активации и параметры регуляризации.

Один из подходов поиска оптимальной структуры — оптимальное прореживание [1], заключается в обучении максимально большой сети последующем удалении части связей. Другой подход заключается в предсказании структуры модели другой нейросетью [2]. В данной работе для выбора оптимального набора гиперпараметров проводится процедура релаксации [3] - переход от дискретного множества возможных значений гиперпараметров к непрерывному множеству их комбинаций. Эта процедура позволяет параметризовать структуру модели некоторым действительным вектором. Такой подход дает возможность применять различные методы оптимизации для нахождения наилучшего набора гиперпараметров. Оптимизация гиперпараметров проводится градиентными методами [4], [5], [6] либо с использованием Гауссовских процессов и Байесовской оптимизации.

Проводится вычислительный эксперимент на выборках Boston, MNIST [7] и синтетических данных. В ходе экспериментов оценивается не только качество, которое дает полученная модель но и её вычислительная сложность и устойчивость. Также в эксперименте проводится сравнение различных алгоритмов построения оптимальной модели

Литература

- [1] Yann Le Cun, John S. Denker, and Sara A. Solla. Advances in neural information processing systems 2. chapter Optimal Brain Damage, pages 598–605. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1990.

- [2] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, Cambridge, MA, USA, 2014. MIT Press.
- [3] Hanxiao Liu, Karen Simonyan, and Yiming Yang. Darts: Differentiable architecture search. *CoRR*, abs/1806.09055, 2018.
- [4] Dougal Maclaurin, David Duvenaud, and Ryan Adams. Gradient-based hyperparameter optimization through reversible learning. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 2113–2122, Lille, France, 07–09 Jul 2015. PMLR.
- [5] Luca Franceschi, Michele Donini, Paolo Frasconi, and Massimiliano Pontil. Forward and reverse gradient-based hyperparameter optimization. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1165–1173, International Convention Centre, Sydney, Australia, 06–11 Aug 2017. PMLR.
- [6] Fabian Pedregosa. Hyperparameter optimization with approximate gradient. pages 737–746, 2016.
- [7] Yann LeCun and Corinna Cortes. MNIST handwritten digit database. 2010.