

# Автоматическое построение нейросети оптимальной сложности

Губанов<sup>1</sup> С.Е.

sergey.gubanov@phystech.edu

<sup>1</sup>Московский физико-технический институт

Работа посвящена оптимизации структуры нейронной сети. Оптимизация нейронной сети предполагает заданную структуру и значения гиперпараметров. Подобная оптимизация приводит к чрезмерному количеству параметров и неоптимальности структуры, что приводит к невысокой скорости оптимизации и переобучению. В данной работе предлагается новый метод оптимизации, который позволяет учитывать особенности задачи, подстраивая структуру и гиперпараметры в процессе оптимизации. Результатом работы предложенного метода является устойчивая модель, дающая приемлемое качество результатов при меньшей вычислительной сложности.

**Ключевые слова:** *нейронные сети, оптимизация гиперпараметров, вычислительный граф, прореживание нейронной сети, устойчивость.*

## Введение

Современные глубокие нейронные сети являются вычислительно емкими моделями и содержат сотни миллионов параметров [1]. Это обуславливает не только длительное время оптимизации, но и ресурсоемкость эксплуатации. Переусложненная модель требует много ресурсов и затрудняет использование в переносимых устройствах и микроконтроллерах. Также существует риск переобучения из-за чрезмерного числа параметров [2]. Целью данной работы является алгоритм построения нейросети, чтобы эти проблемы, а также проблема устойчивости модели, были учтены.

Идея автоматического поиска архитектуры нейросети (NAS) известна давно [3], а в современных работах такие алгоритмы показывают сравнимые со state-of-the-art архитектурами результаты [4]. Однако, используемая обычно методология оптимизации дискретной структуры нейросети [5], значительно ограничивает эффективность оптимизаций, не позволяя использовать методы градиентной оптимизации.

Альтернативный подход подразумевает переход от дискретной параметризации структуры нейросети к непрерывной. В работе [6], такой переход производится над функциями активации. Затем используется градиентная оптимизация [7], и выбирается функция с наибольшим весом в каждом отдельном случае.

В данной работе развивается идея релаксации. Оптимизируются не только функции активации, но и остальные структурные параметры нейросети. Предлагается ввести регуляризацию структуры, позволяющую калибровать дискретность параметризации структуры нейросети [8]. При снижении температуры распределение значений структурных параметров приближается к дискретному, что упрощает итоговый выбор структуры нейросети.

Для оценки полученной системы используются выборки MNIST [9], CIFAR-10. Предметом оценки является не только точность ответов на тестовой подвыборке, но и устойчивость результатов.

## Литература

- [1] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [2] Gaurang Panchal, Amit Ganatra, Parth Shah, and Devyani Panchal. Determination of over-learning and over-fitting problem in back propagation neural network. *International Journal on Soft Computing*, 2(2):40–51, 2011.
- [3] Geoffrey F Miller, Peter M Todd, and Shailesh U Hegde. Designing neural networks using genetic algorithms. In *ICGA*, volume 89, pages 379–384, 1989.
- [4] Barret Zoph and Quoc V Le. Neural architecture search with reinforcement learning. *arXiv preprint arXiv:1611.01578*, 2016.
- [5] Renato Negrinho and Geoff Gordon. Deeparchitect: Automatically designing and training deep architectures. *arXiv preprint arXiv:1704.08792*, 2017.
- [6] Hanxiao Liu, Karen Simonyan, and Yiming Yang. Darts: Differentiable architecture search. *arXiv preprint arXiv:1806.09055*, 2018.
- [7] Joel Grus. *Data science from scratch: first principles with python*. "O'Reilly Media, Inc. 2015.
- [8] Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144*, 2016.
- [9] Yann LeCun and Corinna Cortes. MNIST handwritten digit database. 2010.