

Автоматическое построение нейросети оптимальной сложности

Губанов¹ С.Е.

sergey.gubanov@phystech.edu

¹Московский 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

Глубокие нейронные сети являются большими вычислительными комплексами. Это обуславливает не только долгое время обучения, но и долгое время применения, а также риск переобучения из-за чрезмерного числа параметров. Целью данной работы является создание алгоритма построения нейросети, чтобы эти проблемы, а также проблема устойчивости модели, были учтены.

Работы по задаче автоматического поиска архитектуры (NAS) показывают сравнимые со state-of-the-art архитектурами результаты [1]. Однако, используемая обычно методология оптимизации дискретных параметров нейросети, значительно ограничивает эффективность оптимизаций, не позволяя использовать методы градиентной оптимизации.

Альтернативный подход подразумевает переход от дискретных параметров нейросети к непрерывным. Например, в работе [2], такой переход производится над функциями активации. Затем используется градиентная оптимизация, и выбирается функция с наибольшим весом в каждом отдельном случае.

Суть данной работы состоит в развитии описанной идеи и релаксации не только функции активации, но и остальных гиперпараметров нейросети. Таким образом строится многослойный персептрон с оптимальной структурой, функциями активации и связями между узлами.

Для оценки полученной системы используются стандартные выборки, такие как MNIST[3], CIFAR-10 и другие. Это дает возможность сравнить результаты с полученными во многих других работах. Предметом оценки является не только точность ответов на тестовой подвыборке, но и устойчивость результатов.

Литература

- [1] Barret Zoph and Quoc V Le. Neural architecture search with reinforcement learning. *arXiv preprint arXiv:1611.01578*, 2016.
- [2] Hanxiao Liu, Karen Simonyan, and Yiming Yang. Darts: Differentiable architecture search. *arXiv preprint arXiv:1806.09055*, 2018.
- [3] Yann LeCun and Corinna Cortes. MNIST handwritten digit database. 2010.