

Градиентные методы оптимизации гиперпараметров моделей глубокого обучения

*Бахтеев Олег Юрьевич*¹

bakhteev@phystech.edu

¹Московский физико-технический институт

Решается задача оптимизации гиперпараметров моделей глубокого обучения. *Модель* — суперпозиция функций, решающая задачу классификации или регрессии. *Гиперпараметры* модели — параметры распределения параметров модели.

Нахождение оптимальных гиперпараметров сильно влияет на правдоподобие модели. Рассматриваются методы, основанные на градиентной оптимизации. Основным их преимуществом перед другими методами оптимизации гиперпараметров является возможность эффективно оптимизировать большое количество гиперпараметров одновременно. Другой особенностью данных алгоритмов является возможность их применения для произвольной дифференцируемой функции потерь. В данной работе сравниваются градиентные методы оптимизации гиперпараметров: жадный алгоритм оптимизации, DrMAD и HOAG. Сложность рассматриваемых алгоритмов сопоставима со сложностью оптимизации параметров модели.

В экспериментальной части в качестве критерия выбора модели выступают вариационная нижняя оценка правдоподобия модели и ошибка на валидационной части выборки. Исследуется поведение алгоритмов на выборках большой мощности, таких как WISDM и MNIST. Рассматриваемые алгоритмы показывают высокое качество в случае, когда количество гиперпараметров велико. Экспериментально проиллюстрировано, что в данном случае оптимизация гиперпараметров методом случайного поиска дает значительно худшие результаты за то же количество итераций.

Исследование поддержано РФФИ, проект 16-07-01160.

- [1] Бахтеев О.Ю., Попова М.С., Стрижов В.В. Системы и средства глубокого обучения в задачах классификации // Системы и средства информатики Москва: Институт Проблем Информатики, РАН, 2016. — С. 4–22. http://www.ipiran.ru/journal_system/article/08696527160201.html.

Gradient-based hyperparameter optimization for deep learning models

*Bakhteev Oleg*¹

bakhteev@phystech.edu

¹Moscow Institute of Physics and Technology

The paper describes methods of hyperparameter optimization. A model is a function superposition, hyperparameters are parameters of parametrical distribution.

The paper describes gradient-based hyperparameter optimization methods, which allow to optimize a large amount of hyperparameters effectively. Gradient-based methods can be used with arbitrary differential loss function. The reviewed algorithms are HOAG, DrMad and greedy gradient-based algorithm.

The experiments were conducted on WISDM accelerometer records dataset, MNIST handwritten digits dataset and simulation data. The experiments show that the reviewed methods handle hyperparameter optimization effectively in case when the amount of the hyperparameters is large. We compare the reviewed methods with random search using the same amount of optimization iterations. The experiments show that the gradient-based methods can optimize hyperparameters significantly better.

This research is supported by RFBR project 16-07-01160.

- [1] *O. Yu. Bakhteev, M. S. Popova, and V. V. Strijov.* Systems and means of deep learning for classification problems // Systems and Means of Informatics , Moscow: Institute of Informatics Problems, Russian Academy of Sciences, 2016, 2. — p. 4–22. http://www.ipiran.ru/journal_system/article/08696527160201.html.