

Увеличение эффективности подбора гиперпараметров

A Preprint

Валентин А. Абрамов
Факультет вычислительной математики и кибернетики
МГУ имени М. В. Ломоносова

Abstract

Гиперпараметры являются важными параметрами алгоритмов машинного обучения, которые не могут быть определены в процессе обучения модели. Традиционные методы выбора гиперпараметров, такие как сеточный поиск или случайный поиск, могут быть неэффективными или требовать значительных вычислительных ресурсов, поэтому вместо них используют более сложные алгоритмы, например, эволюционные. Одним из таких алгоритмов является DEHB.

В данной статье предложено улучшение метода дифференциальной эволюции, лежащего в основе DEHB. Алгоритм DEHB жадный - постоянно переиспользует старые значения. Предложено добавлять шум во время мутации для увеличения покрытия пространства гиперпараметров.

Были представлены экспериментальные результаты, демонстрирующие эффективность этого метода в сравнении с базовым DEHB и традиционными методами выбора гиперпараметров. Результаты показывают, что изменение мутации может обеспечить более высокую точность модели при неизменном использовании вычислительных ресурсов.

Keywords First keyword · Second keyword · More

1 Введение

В последние годы машинное обучение стало одной из самых активно развивающихся областей в информационных технологиях. Чтобы достичь высокого качества моделей машинного обучения, необходимо правильно настроить гиперпараметры алгоритмов. Для каждой задачи значения параметров одной модели будут отличаться, и они могут сильно повлиять на целевую метрику. Гиперпараметрическая оптимизация - большая и перспективная область исследований. В статье Bischl et al. [2021] приводится подробный обзор существующих методов, все алгоритмы формально описаны, поднимаются вопросы переобучения гиперпараметров на валидационную выборку и параллелизации алгоритмов. Для глубоких нейронных сетей также необходимо подбирать гиперпараметры, такие как число слоёв, learning rate и даже связи между слоями. О задачах гиперпараметрической оптимизации в нейронных сетях подробно написано в статье Talbi [2020].

Задача автоматического подбора гиперпараметров активно исследуется. Ранее использовались методы перебора по сетке или случайного поиска значений, но эти методы имеют свои недостатки, такие как случайность, проклятие размерности, сложность подбора сетки, медленная скорость работы, невозможность учесть взаимосвязи между гиперпараметрами и зависимость целевой метрики от них. В 2011 году был создан подход SMBO - последовательная оптимизация, основанная на суррогатной функции (Hutter et al. [2011]). В качестве суррогатной функции изначально брали гауссовские процессы, но они подходят не для всех задач, так как предназначены только для непрерывных гиперпараметров. Также был придуман алгоритм Successive Halving (Jamieson and Talwalkar [2015]), представляющий из себя жадный алгоритм, выделяющий больше ресурсов для оценки гиперпараметров с хорошим потенциалом.

по формуле
тексту
должны
быть ---
вместо -

слишком
много текста
без ссылки

ссылки

ссылки

ссылки

На данный момент одним из лучших методов поиска является TPE (Bergstra et al. [2011]), основанный на байесовском подходе. Также популярны методы, основанные на подходе SMBO, такие как SMAC, использующий случайный лес в качестве суррогата, или модели, использующие суррогат на основе гауссовских процессов. Лучшие всего себя показывают методы, основанные на алгоритме Succesive Halving. На основе SH созданы алгоритмы HyperBand (Li et al. [2018]), автоматически выделяющий ресурсы для нескольких запусков SH, BOHB (Falkner et al. [2018]), добавляющий в HyperBand суррогат на основе TPE и DEHB Awad et al. [2021], добавляющий в HB идею дифференциальной эволюции. По текущим бенчмаркам DEHB является SOTA-алгоритмом подбора гиперпараметров.

В данный момент DEHB является одним из лучших алгоритмов, но у него есть недостатки. Из-за жадного переиспользования оптимальных значений при дифференциальной эволюции, модель застревает на плато в локальном минимуме. Чтобы этого избежать, предлагается добавлять шумовые векторы к мутантам при эволюции. Таким образом, благодаря шумовым векторам увеличится покрытие поверхности функции ошибки от гиперпараметров.

про твоя метод + 1.2 абзаца + про эксперименты.

2 Постановка задачи

Задача гиперпараметрической оптимизации состоит в том, чтобы имея набор данных $X \in \mathbb{R}^{N \times d}$, $y \in \mathbb{R}^N$, пространство гиперпараметров $\Theta \in \mathbb{R}^k$, модель $f(\theta) : \mathbb{R}^d \rightarrow \mathbb{R}$, $\theta \in \Theta$, $f \in \mathbb{H}$ и целевую метрику $c(f, X, y) : \mathbb{H} \times \mathbb{R}^{N \times d} \times \mathbb{R}^N \rightarrow \mathbb{R}$, найти θ^* такую, что выполнено

$$\theta^* = \arg \min_{\theta} c(f(\theta), X, y)$$

Таким образом, зная модель f , пространство гиперпараметров Θ , набор данных X, y и целевую метрику c , необходимо найти набор гиперпараметров θ^* , при котором целевая метрика c наименьшая при обучении модели f на данных X, y . В нашей работе будет модифицирован метод подбора оптимальных гиперпараметров DEHB - Differential Evolution HyperBand. В нём один раз сэмпелируются кандидаты-наборы гиперпараметров из равномерного распределения, далее запускается процесс эволюции, в котором лучшие кандидаты передаются горизонтально в Succesive Halving сетках для одинаковых значений бюджета. Для мутации сэмпелируются лучшие кандидаты из предыдущего значения бюджета в текущей SH сетке.

Базовый эксперимент! Теория? Выводы? Анализ ошибок?

Список литературы

- Bernd Bischl, Martin Binder, Michel Lang, Tobias Pielok, Jakob Richter, Stefan Coors, Janek Thomas, Theresa Ullmann, Marc Becker, Anne-Laure Boulesteix, Difan Deng, and Marius Lindauer. Hyperparameter optimization: Foundations, algorithms, best practices and open challenges, 2021.
- El-Ghazali Talbi. Optimization of deep neural networks: a survey and unified taxonomy. 2020. URL <https://api.semanticscholar.org/CorpusID:219427357>.
- Frank Hutter, Holger H. Hoos, and Kevin Leyton-Brown. Sequential model-based optimization for general algorithm configuration. In Carlos A. Coello Coello, editor, Learning and Intelligent Optimization, pages 507–523, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. ISBN 978-3-642-25566-3.
- Kevin Jamieson and Ameet Talwalkar. Non-stochastic best arm identification and hyperparameter optimization, 2015.
- James Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. Algorithms for hyperparameter optimization. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K.Q. Weinberger, editors, Advances in Neural Information Processing Systems, volume 24. Curran Associates, Inc., 2011. URL https://proceedings.neurips.cc/paper_files/paper/2011/file/86e8f7ab32cfd12577bc2619bc635690-Paper.pdf.
- Lisha Li, Kevin Jamieson, Giulia DeSalvo, Afshin Rostamizadeh, and Ameet Talwalkar. Hyperband: A novel bandit-based approach to hyperparameter optimization, 2018.
- Stefan Falkner, Aaron Klein, and Frank Hutter. Bohb: Robust and efficient hyperparameter optimization at scale, 2018.
- Noor Awad, Neeratyoy Mallik, and Frank Hutter. Dehb: Evolutionary hyperband for scalable, robust and efficient hyperparameter optimization, 2021.