

Исследование эффективных трансформеров

М. Д. Ковалева

Московский физико-технический институт

14 декабря 2022 г.

Постановка задачи

Проблема

Рассматривается задача поиска нейронных архитектур (NAS) и оптимизации гиперпараметров. Основной проблемой алгоритмов оптимизации поиска нейронной архитектуры являются большие затраты по времени.

Цель работы —

Улучшение алгоритма оптимизации гиперпараметров Hyperband¹ с помощью ранней остановки, основанной на алгоритмах обнаружения точки изменения (алгоритм CUSUM),

Задачи работы

- 1) изучить статьи
- 2) добавить в алгоритм Hyperband раннюю остановку основанную на алгоритме CUSUM
- 3) протестировать на бенчмарках nas-bench nlp и cifar10 и сравнить с оригинальным Hyperband

¹Li, L. et al. (2018) Hyperband: A novel bandit-based approach to hyperparameter optimization, arXiv.org. Available at: <https://arxiv.org/abs/1603.06560> (Accessed: 18 May 2023).

Hyperband

Основная идея

Ускорении случайного поиска за счет адаптивного распределения ресурсов выделяемых на обучение моделей. Используется идея Successive Halving, но с различными конфигурациями.

Детали алгоритма

- ▶ R - максимальное количество ресурсов выделяемое для одной конфигурации
- ▶ η параметр контролирующий количество отсеиваемых конфигураций отсеиваемых на



Алгоритм

Algorithm An algorithm with caption

```
Require:  $R, \eta(\text{default} = 3)$ 
 $s_{\max} \leftarrow \lfloor \log_{\eta}(R) \rfloor$ 
 $B \leftarrow (s_{\max} + 1)R$ 
for  $s \in \{s_{\max}, s_{\max} - 1, \dots, 0\}$  do
     $n \leftarrow \lceil \frac{B}{R} \frac{\eta^s}{(s+1)} \rceil$ 
     $r \leftarrow R\eta^{-s}$ 
     $T \leftarrow \text{get\_hyperparameter\_configuration}(n)$ 
    for  $i \in \{0, \dots, s\}$  do
         $n_i \leftarrow \lfloor n\eta^{-i} \rfloor$ 
         $r_i \leftarrow r\eta^i$ 
         $L \leftarrow \{\text{return\_val\_loss}(t, r_i) : t \in T\}$ 
         $T \leftarrow \text{top\_k}(T, L, \lfloor n_i/\eta \rfloor)$ 
    end for
end for
return Configuration with the smallest intermediate loss seen so far
```

CUSUM

- ▶ CUSUM - широко используемый в задачах детектирования разладки (change point detection) алгоритм, который включает вычисление кумулятивной суммы какого-то процесса x_t :

$$S_0 = 0, S_{t+1} = \max\{0, S_t + x_t\}$$

Процесс сигнализирует о разладке когда кумулятивная сумма достигает заранее выбранного порогового значения.

- ▶ Так как мы хотим детектировать переобучение модели то в качестве процесса x_t у нас будет выступать уменьшение валидационной функции потерь на каждой эпохе: $x_t = L_t - L_{t-1}$. То есть когда валидационная ошибка начинает увеличиваться мы детектируем переобучение.
- ▶ Преимущество: онлайн алгоритм
- ▶ Недостаток: выбор порогового значения остается открытым вопросом

Результаты

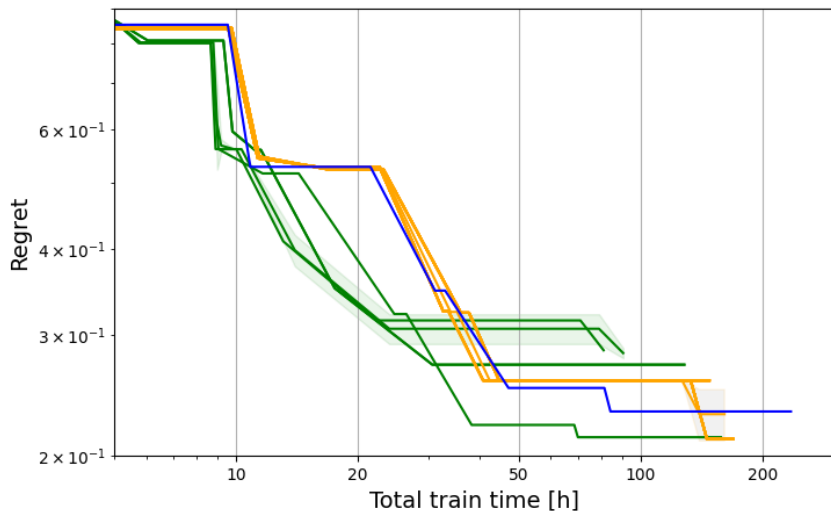


Рис.: зависимость лучшего найденного значения функции потерь от времени для разных подходов (синий - оригинальный Hyperband; зеленый - остановка по первой разладке (парето фронт), оранжевый - остановка по первому пику статистики в алгоритме CUSUM (парето фронт))

Список литературы

1. Li, L. et al. (2018) *Hyperband: A novel bandit-based approach to hyperparameter optimization*, arXiv.org. Available at: <https://arxiv.org/abs/1603.06560> (Accessed: 18 May 2023).
2. Shiryayev, A. *Stochastic Change-Point Detection Problems*. MCCME (in Russian), Moscow, 2017.