Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization

- 作者：Lisha Li, Kevin Jamieson, Giulia DeSalvo et.al.
- 机构：UCLA, UC Berkely, NYU, Google
- 会议：ICLR2017
- 地址：https://arxiv.org/abs/1603.06560
- 代码：https://github.com/automl/HpBandSter

论文主要内容

摘要

机器学习算法的performance很大程度依赖于超参的选择。贝叶斯优化方法可以自适应选择超参。本文提出HyperBand，简单灵活，理论上合理，能够自适应分配预定义的资源的一种early-stopping策略。

Motivation

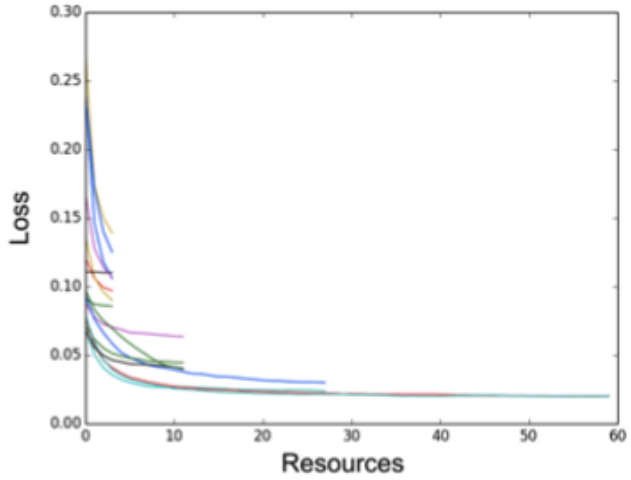
- SH：在算法早期停掉一些没有希望的config
- HB：解决SH中的trade-off问题

方法

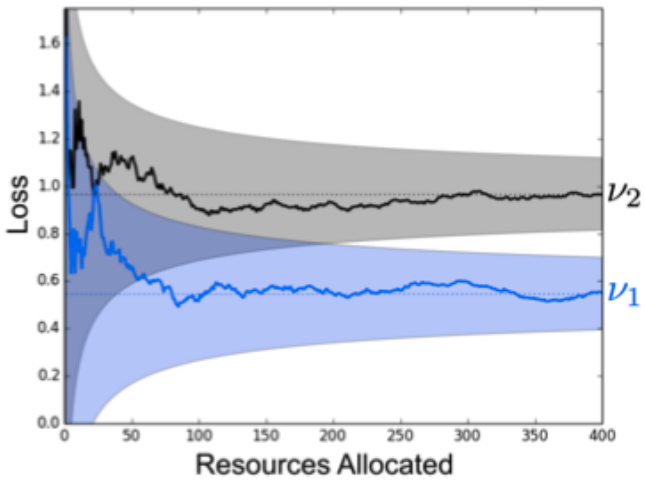
介绍

Successive Havling（SH）

- 给一个超参数配置集合分配一个统一的预算（budget），评估超参数集合中所有配置，扔掉结果最坏的那半配置，并重复过程。
- 需要输入一个 n 来表示共有多少个超参数配置
- 如果 $n = 33$ 则需要 $\log_2(33) + 1$ ，取整需要6次折半迭代(s=6)



(b) Configuration Evaluation



(c) Envelopes

// begin SUCCESSIVEHALVING with (n, r) inner loop
T =get_hyperparameter_configuration(n)
for $i \in \{0, \ldots, s\}$ **do**
  $n_i = \lfloor n\eta^{-i} \rfloor$
  $r_i = r\eta^i$
  $L = \{\text{run\_then\_return\_val\_loss}(t, r_i) : t \in T\}$
  T =top_k($T, L, \lfloor n_i/\eta \rfloor$)
end

- 符号说明
  - s ： 为总轮数； B ： 为找到最优config定下的总budget
  - η 决定SH中每轮丢弃的比例。例如 $\eta = 3$ 表示SH每轮后保留top 1/3的config
  - n_i ： SH中第 i 轮，config数量
  - r_i ： SH中第 i 轮，每个config被分配到的资源量（ $n_i \cdot r_i$ 为第 i 轮，configs集合的总资源）
  - $B = s \cdot n \cdot r$ 。每轮次中 $n_i \cdot r_i = \frac{B}{s} = R$ 固定，即每轮的config集合总budget
  - $n <= R$,因为第一轮中分配给每个config的资源至少为1
  - **最后一轮必只有一个config，独占资源** R
- 在总budget B固定下， n 越大，能比较的config就**越多**；但B/n越小，每个config的elarly stoping就越激进，comparison就越**不准**。反之亦然。

- 只能trade-off：无法找到每个config能被区分好坏的最低资源

方法：HyperBand（HB）

- 本质上就是对 SH 的grid search策略，外循环不断迭代尝试不同的n与r。避免trade-off
- 给定一次n与r，就执行一遍完整的SH（子程序）。HB中过一遍完整的SH称为一个**bracket**

Algorithm 1: HYPERBAND algorithm for hyperparameter optimization.

input : R, η (default $\eta = 3$)
initialization : $s_{\max} = \lfloor \log_\eta(R) \rfloor$, $B = (s_{\max} + 1)R$

for $s \in \{s_{\max}, s_{\max} - 1, \ldots, 0\}$ **do**
  $n = \lceil \frac{B}{R} \frac{\eta^s}{(s+1)} \rceil$, $r = R\eta^{-s}$
  // begin SUCCESSIVEHALVING with (n, r) inner loop
  T =get_hyperparameter_configuration(n)
  for $i \in \{0, \ldots, s\}$ **do**
    $n_i = \lfloor n\eta^{-i} \rfloor$
    $r_i = r\eta^i$
    $L = \{\text{run\_then\_return\_val\_loss}(t, r_i) : t \in T\}$
    T =top_k($T, L, \lfloor n_i/\eta \rfloor$)
  end
end
return *Configuration with the smallest intermediate loss seen so far.*

- 两个输入 R、η 决定了要进行多少次SH（**brackets数量，即** $s_{max}+1$ ）
  - R **是SH最后轮每config的资源（2，6行）**
- r 为SH第一轮次的每config上的资源，**为了确保** $r >= 1$ ：$s_{\max} = \lfloor \log_\eta(R) \rfloor$
- B 为一个bracket的budget

- 当s最大，即减半操作最多，能评估的config数量n就最多=>最激进的early-stopping
- 当s最小=0，即每config分配到的资源最多(R)=>退化为经典随机搜索
- 相比SH，总开销多一个**对数**乘子， $B_{HB} = log_\eta(R) \cdot B_{SH}$

	$s=4$		$s=3$		$s=2$		$s=1$		$s=0$	
i	n_i	r_i	n_i	r_i	n_i	r_i	n_i	r_i	n_i	r_i
0	81	1	27	3	9	9	6	27	5	81
1	27	3	9	9	3	27	2	81		
2	9	9	3	27	1	81				
3	3	27	1	81						
4	1	81								

Table 1: Values of n_i and r_i for the brackets of HYPERBAND when $R = 81$ and $\eta = 3$.

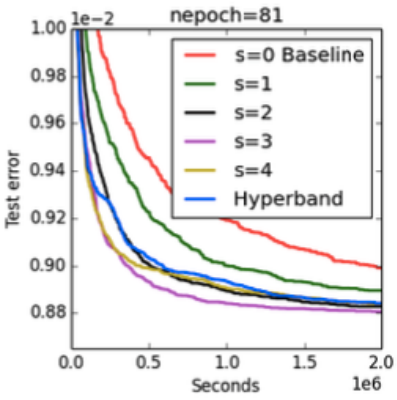


Figure 2: Performance of individual brackets s and HYPERBAND.

- s=3是最优的，hyperband表现接近直接选择最优的bracket

实验结果

Results

- model：卷积神经网络
- 搜索空间：SGD的6个超参、response normalization layer的2个超参
- Datasets：CIFAR-10、MRBI、SVHN
- HyperBand：一个资源单位对应100个mini-batch。
  - CIFAR-10设置R=300、MRBI设置R=300、SVHN设置600
  - $\eta = 4$

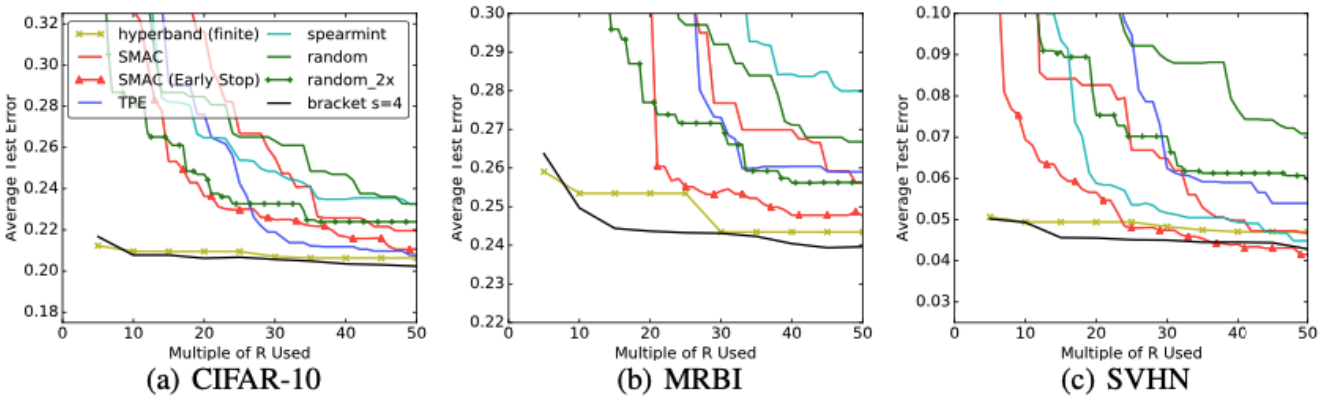


Figure 3: Average test error across 10 trials is shown in all plots. Label "SMAC_early" corresponds to SMAC with the early stopping criterion proposed in Domhan et al. (2015) and label "bracket $s = 4$" corresponds to repeating the most exploratory bracket of HYPERBAND.

数据下采样

- 数据下采样，数据量的大小作为资源budget

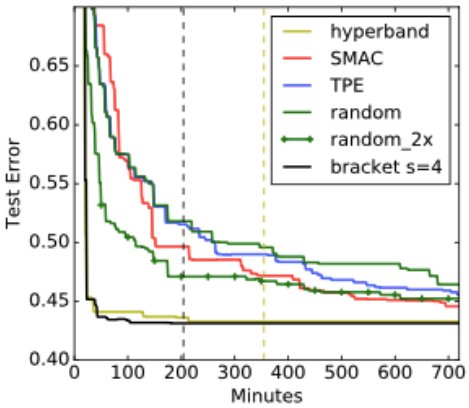


Figure 4: Average test error of the best kernel regularized least square classification model found by each searcher on CIFAR-10. The color coded dashed lines indicate when the last trial of a given searcher finished.

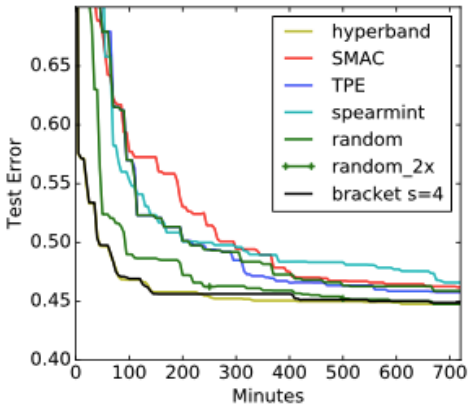


Figure 5: Average test error of the best random features model found by each searcher on CIFAR-10. The test error for HYPERBAND and bracket $s = 4$ are calculated in every evaluation instead of at the end of a bracket.

特征下采样

- 特征丰富程度作为资源budget（见figure5