Hyperparameter Selection for Multi-armed Bandit Problems

Aniket Deshmukh[1], Feng Wei [2], Clayton Scott [1]

1. Electrical and Computer Engineering, University of Michigan Ann Arbor, 2. Mathematics, University of Michigan Ann Arbor

Introduction

Hyperparameter selection in batch learning setup is a well studied problem in recent years. However, we know very little about how to select hyperparameters for online learning in a partial feedback setting. The problem in online learning with partial feedback is challenging because even a validation phase incurs the regret and one needs to be careful about how many resources are allotted to explore hyperparameter space. The need to control cumulative regret and limited resources lead to many difficulties. In this project, we present a bandit based algorithm for hyperparameter selection together with simulations and theoretical guarantees.

Motivation - Hyperparameters in Bandit Algorithms

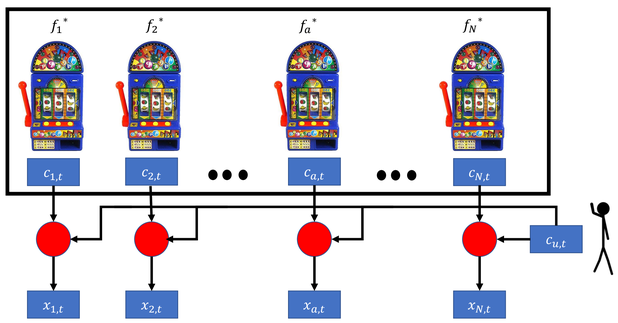

Figure 1: Contextual Bandits

Hyperparameter selection as MAB

Require: Budget B, M configurations and let $[M] = \{1, ..., M\}$
for $t = 0, ..., B-1$ **do**
Choose an arm $I_t \in [M]$ based on predefined selection strategy.
Play an bandit problem with hyperparameter configuration I_t.
Receive a loss $l_{I_t,t} \in \mathbb{R}$
end for
Output: Singleton set I_B at the end of budget B.

- Simple Regret: $SR_B = l_{I_B,B}$.
- Cumulative Regret: $CR_B = \Sigma_{i=0}^{B-1} l_{I_i,i}$.
- We aim to minize the weighted combination of the regret, $WR_B = \alpha CR_B + (1-\alpha)SR_B$
- For hyperparameter selection in batch learning, $\alpha = 0$.
- For hyperparameter selection when there is a partial feedback and we have a limited budget for the training phase, $\alpha \in [0, 1]$.

Cumulative Regret vs Simple Regret

- Bubeck et. al show that in case of stochastic multi arm bandit problems, upper bound on expected cumulative regret lead to upper bound on expected simple regret.
- Upper bound on expected cumulative regret also lead to lower bound on expected simple regert.
- For optimal simple regret one may need to do more exploration than in the case of cumulative regret [1].

Assumptions - Discussion on envelope

- Let M be the total number of hyperparameter configurations for a bandit algorithm.
- For each configuration $i \in [M]$, let $\mu_i f(T)$ be the expected cumulative regret.
- For simplicity, let $\mu_1 \leq \mu_2 \leq ... \leq \mu_M$.
- Let $g_i(t)$ be the realization of the cumulative regret and $\gamma_i(t)$ be the envelope of expected cumulative regret such that $|g_i(t) - \mu_i f(t)| \leq \mu_i \gamma_i(t) f(t)$.
- Let $\bar{\gamma}(t) = \max_{i \in [M]} \gamma_i(t)$, $\hat{\gamma} = \sup_t \bar{\gamma}(t)$ and $\bar{\mu} = \max_{i \in [M]} \mu_i(t)$.

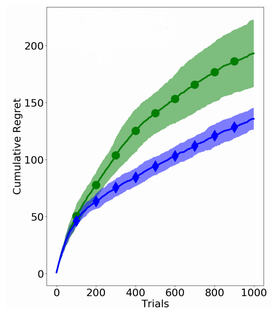

Figure 2: Example of an envelope in a contextual bandit experiment

Proposed Algorithm: $\beta-$aggressive Resource Allocation

Require: Budget B, M configurations and β. Let cumulative regret for the configuration i at the end of step k be CR_k^i. Define B_0 s.t. $B = B_0 \Sigma_{k=1}^{\log_2 M} \beta^{k-1}$.
Initialize: $S_1 = [M]$
for $k = 1, \cdots, \log_2(M)$ **do**
Pull each arm in S_k for $r_k = B_0 \frac{\beta^{k-1}}{|S_k|}$ resources.
Receive cumulative regret CR_k^i for $i \in S_k$.
$S_{k+1} =$ select $|S_k|/2$ arms which give the least CR_k^i.
end for
Output: Singleton set $S_{\log_2(M)}$

Theoretical Analysis

Theorem 1 (Upper Bound on Simple Regret) *Given the budget $B, \bar{\gamma}(t)$ and the expected cumulative regret $\mu_i f(t), i \in [M]$ of underlying base algorithms, simple regret is bounded by -*

$$SR_B \leq \left(\sum_{k=1}^{\log_2(M)} 2\bar{\gamma}(R_k)\bar{\mu} \right) f(B)$$

Theorem 2 (Upper Bound on Cumulative Regret) *Given the budget B, M number of configurations, let $f(B) = B^\eta$ then,*

$$CR_B \leq \frac{(1+\hat{\gamma})\bar{\mu}B^\eta M^{1-\eta}}{(\Sigma_{j=1}^{\log_2 M} \beta^{j-1})^\eta} \sum_{k=1}^{\log_2(M)} \frac{1}{2^{k-1}} \left(\sum_{i=1}^{k} (2\beta)^{i-1} \right)^\eta$$

and if $\beta > C_\eta$ then $CR_B \leq C(1+\hat{\gamma})\bar{\mu}B^\eta$.

Experimental Results

In the Fig. 3, we compare the different algorithms to choose exploration parameter of UCB algorithm on the synthetic data.

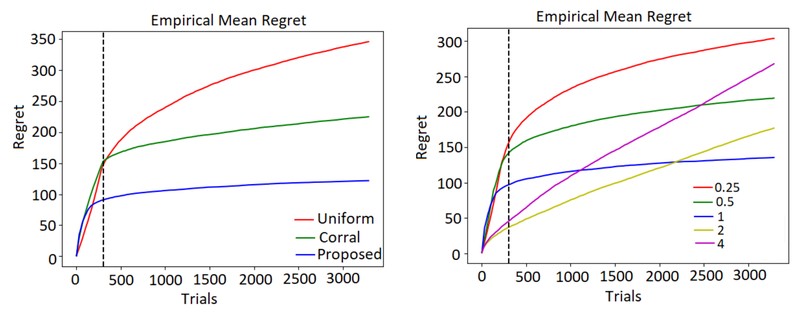

Figure 3: Mean Regret Comparison

Challenges and Future Work

- Choosing β based on the given budget B and α.
- Choosing more meaningful envelope $\bar{\gamma}(t)$
- Tighter simple and cumulative regret bound (which don't depend on $\bar{\mu}$)
- Conclusive experimental results on more bandit algorithms.

References

[1] Bubeck, Sebastien, Remi Munos, and Gilles Stoltz. "Pure exploration in finitely-armed and continuous-armed bandits." Theoretical Computer Science 412.19 (2011): 1832-1852.

[2] Agarwal, Alekh, Haipeng Luo, Behnam Neyshabur, and Robert E. Schapire. "Corralling a band of bandit algorithms." In Conference on Learning Theory, pp. 12-38. 2017.

[3] Jamieson, Kevin, and Ameet Talwalkar. "Non-stochastic best arm identification and hyperparameter optimization." In Artificial Intelligence and Statistics, pp. 240-248. 2016.