Exploring Bandit Algorithms in Sparse Environments

Does increasing the level of sparsity increase the advantage of sparsity-adapted Multi-Armed Bandit algorithms?

Author:
Rafał Owczarski
R.R.Owczarski@student.tudelft.nl

Affiliations:
EEMCS, Delft University of Technology, The Netherlands

Supervisor:
Assistant Professor Julia Olkhovskaya

Examiner:
Dr. Ranga Rao Venkatesha Prasad

References:
[1] Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multiarmed bandit problem. SIAM journal on computing, 32(1):48–77, 2002.
[2] Tor Lattimore and Csaba Szepesvári. Bandit algorithms. Cambridge University Press,2020.
[3] Wei Chu, Lihong Li, Lev Reyzin, and Robert Schapire. Contextual bandits with linear payoff functions. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, pages 208–214. JMLR Workshop and Conference Proceedings, 2011.
[4] Min-hwan Oh, Garud Iyengar, and Assaf Zeevi. Sparsity-agnostic lasso bandit. International Conference on Machine Learning, pages 8271–8280. PMLR, 2021.

01 Introduction

- Multi-Armed bandits (MAB) are a class of decision-making problems that aim to optimize the exploration-exploitation dilemma.
- The algorithms designed to model the problem aim to minimize the regret of not picking the optimal option at a given iteration.
- The concept is widely used in reinforcement learning, optimization, and economics. In these applications, contextual information provided can be highly dimensional, but only a small subset of the features influence the reward observed.

02 Contribution

- The research addresses the gap in the existing literature, where the algorithms are not compared against each other in environments different from those for which they were designed.
- This research uses an experimental approach to compare 4 algorithms in an environment with a fixed sparse reward function, noisy observations, and an i.i.d. context.
- The challenge is to identify the optimal algorithm to model a given scenario.

03 Methodology

The main research question to be answered is: *Does increasing the level of sparsity enhance the advantage of sparsity-adapted Multi-Armed Bandit algorithms?*

We will consider the following algorithms:

- Stochastic bandits - Upper Confidence Bound Algorithm (UCB) [1]
- Adversarial bandits - Exponential-weight algorithm for Exploration and Exploitation (EXP3) [2]
- Linear contextual bandits - LinUCB [3]
- Sparse bandits - Sparsity-Agnostic Lasso Algorithm [4]

- Synthetic data was created to represent the experimental environment.
- The reward was chosen based on some i.i.d. context and unknown sparse vector.
- The MAB algorithms were run in the same environment with 1000 iterations with 10 repetitions.
- The results obtained for each repetition were combined and averaged, and the standard deviation of each measurement was derived.
- The sparsity was fixed and measurements were made with varying context dimensions to measure the impact on algorithms' performance
- The algorithms were compared, with the main focus on obtaining cumulative regret.
- Additional comparisons were made regarding complexity, memory efficiency, and hyperparameter sensitivity.

04 Experiment Results

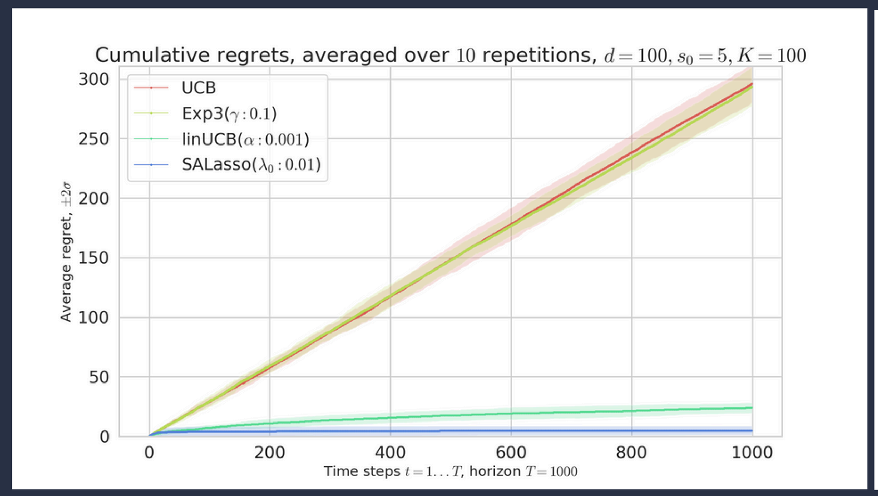

Figure 1. Cumulative regret obtained by all algorithms in an environment with context with 100 dimensions, generated using Exponential distribution with rate 3; 100 arms and sparsity 5.

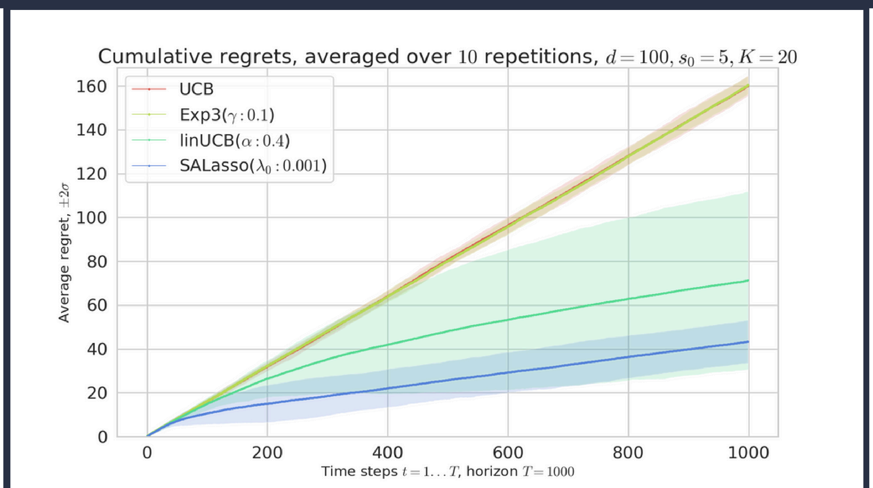

Figure 2. Cumulative regret obtained by all algorithms in an environment with context with 100 dimensions, generated using Multivariate Normal distribution with mean 0; 20 arms and sparsity 5.

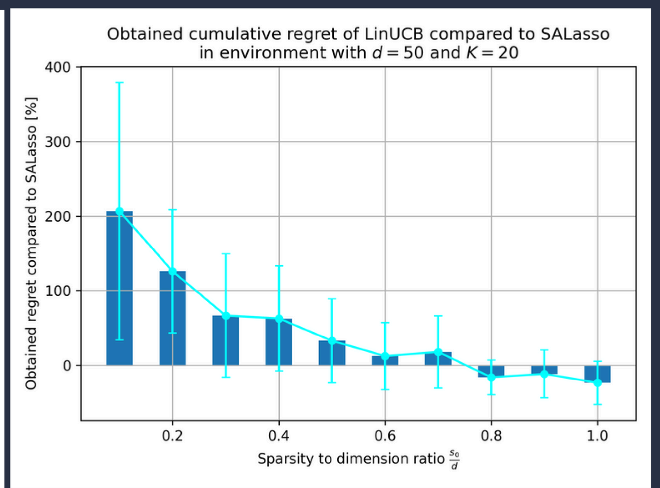

Figure 3. Cumulative regret of LinUCB compared to SALasso in an environment with context with 50 dimensions generated using Multivariate Normal distribution with mean 1; 20 arms and varying level of sparsity.

05 Limitations

- While the study highlights the influence of sparsity in controlled artificial environments with simulated noisy observations, the performance of the algorithms in real-life scenarios was not explored in the scope of this research.
- Other sparsity-adapted bandit algorithms could be compared to explore whether any algorithm of that class achieves more optimal regret than the traditional approaches. Similarly, the study can benefit from expanding the number of non-sparsity-adapted algorithms.
- The study focused on a fixed sparsity measure equal to 5. By experimenting with this value and the dimension and generation strategy of the context more results can be obtained.

06 Conclusions

- Increasing the sparsity indeed enhances the performance of sparsity-adapted Multi-Armed Bandit algorithms.
- In environments with high variability and common negative context coefficients, SALasso and LinUCB behave statistically identically.
- SALasso is highly sensitive to its hyperparameter, and not using its optimal value can cause traditional approaches to outperform the algorithm.
- As expected, stochastic bandits fail to converge to logarithmic regret bounds in sparse contextual environments.