Learning Augmented Online Learning Algorithms - The Adversarial Bandit with Knapsacks framework

Davide Drago[3], Andrea Celli[3] and Marek Eliáš[3]

[3]Bocconi University, Via Guglielmo Röntgen 1, Milan, 20136, Italy

Abstract

We delve into the Bandit with Knapsacks framework with the aim of creating a learning-augmented online algorithm with better competitive guarantees than the state-of-the-art classical worst-case algorithms. In particular, we obtain better competitive ratios when the input predictions are accurate, while also upholding worst-case scenario guarantees for imprecise predictions. Two unique algorithms are introduced — the first working in a full feedback environment and the other tailored for a bandit setting. Both algorithms integrate a static prediction in a worst-case α-competitive algorithm. This results in an optimized competitive ratio of $1/[\pi + \frac{1}{\alpha}(1-\pi)]$ in scenarios where the prediction is perfect, and a marginally compromised constant competitive ratio of $\alpha/(1-\pi)$ when the prediction is highly imprecise, with $\pi \in (0,1)$ parameter chosen by the decision-makers.

1. Introduction

The large availability of data in the application settings of the bandit with knapsacks framework [2] (e.g., online advertising) brings us to the central research question of this paper: Can machine learning predictions enhance the performance of traditional algorithms in the Bandit with Knapsacks framework? Our contribution yields two novel algorithms, one for the full feedback setting and the other for the bandit case. Both the algorithms have enhanced performances when equipped with a good prediction, but maintain worst-case guarantees for imprecise ones.

1.1. Related Work

Bandit with Knapsacks. In the case of adversarial bandits with knapsacks, Immorlica et al. [3] provide a competitive ratio of $O(m \log T)$. This was improved by Kesselheim and Singla [4] to $O(\log m \log T)$. Subsequently, Castiglioni et al. [5] provided the first constant-factor competitive ratio for the case in which $B = \Omega(T)$. Such competitive ratio is $1/\rho = T/B$.

Learning Augmented online algorithms. The framework of Learning Augmented online algorithms was formally established by Lykouris and Vassilvtiskii [6]. Applications of this framework are wide-ranging and include scheduling [7, 8], caching or paging algorithms [9, 10]. In addition, recently a general framework for integrating predictions into online primal-dual algorithms was introduced in [11].

IPS-RCRA-SPIRIT 2023: Italian Workshop on Planning and Scheduling, RCRA Workshop on Experimental evaluation of algorithms for solving problems with combinatorial explosion, and SPIRIT Workshop on Strategies, Prediction, Interaction, and Reasoning in Italy. November 7-9th, 2023, Rome, Italy [1]

davide.drago@studbocconi.it (D. Drago); andrea.celli2@unibocconi.it (A. Celli); marek.elias@unibocconi.it (M. Eliáš)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

2. Setting

The decision maker makes a sequence of T decisions, drawing actions from a finite set A. A randomized strategy is defined as $\xi_t \in \Delta(A)$. We denote by ξ^A is the best predicted mixed strategy, while ξ^* is the best-fixed distribution. The decision maker has m available resources and a budget B for each of them. A sequence of items γ is selected by an adversary. In our setting, γ_t will be composed of a reward $f_t \in [0,1]^n$ and a cost vector $c_t \in [0,1]^{n\times m}$. We focus on the case in which $B = \Omega(T)$. We denote as $\rho = B/T$ the ratio of budget to time horizon.
Benchmark. The benchmark used in the paper is the Fixed Distribution benchmark, defined in [3] and denoted as OPT^{FD}. Such a quantity is defined as the expected total reward of the distribution over actions ξ^*, maximizing $\mathbb{E}[\text{REW}]$.
Regret. To evaluate the algorithm we use the notion of pseudo-regret, expressed ac

$$\mathbb{E}[\text{REW}_{\text{ALG}}] \geq c\,\text{OPT}^{FD} + \texttt{reg}$$

where $1/c$ is the competitive ratio, OPT^{FD} is the profit of the fixed distribution benchmark, and $\texttt{reg}$ a sublinear regret term.

3. Algorithms

Full-feedback. In the full-feedback algorithm, at each iteration, with probability p the prediction is played, with probability ν the iteration is skipped, and with the remaining probability the worst-case algorithm is played. Both the prediction and the worst-case algorithm are assigned the full budget B and are stopped when the budget assigned would be depleted, had they been played for the full sequence. The worst-case algorithm is updated at each iteration.
Bandit. The difference in the bandit feedback algorithm lies in the update rules. The worst-case algorithm is updated in the iterations in which it is played, otherwise, we set the feedback at $(\mathbf{0}, \mathbf{0})$. Moreover, since the calculation of the expected stopping times is not possible with the bandit feedback, the budget must be divided preemptively between the two algorithms, proportionally to the probability of being played.
Results. Both algorithms have the same competitive ratio guarantees in their respective settings.

Theorem 3.1. *The algorithms with $\xi^A = \xi^*$ and $\nu = \frac{2\sqrt{2\log(1/\delta)}}{\rho T^{1/2}}$, for a sequence of inputs γ achieve w.h.p. a competitive ratio of $1/[p + \rho(1-p)]$. When $\sum f_t(\xi^A) = 0$, the competitive ratio degrades to $1/[\rho(1-p)]$.*

4. Conclusions

Our findings, although encouraging, have some limitations. Specifically, our algorithms do not ensure sublinear regret under stochastic inputs, and are not designed to adjust the probability parameter p in response to real-time performance. Future research could focus on adapting our framework to stochastic environments and creating algorithms capable of dynamically modifying the parameter p as system dynamics change. Moreover, enhancing our model to provide *best-of-both-worlds* guarantees may be useful in diverse applications.

References

[1] R. D. Benedictis, M. Castiglioni, D. Ferraioli, V. Malvone, M. Maratea, E. Scala, L. Serafini, I. Serina, E. Tosello, A. Umbrico, M. Vallati, Preface to the Italian Workshop on Planning and Scheduling, RCRA Workshop on Experimental evaluation of algorithms for solving problems with combinatorial explosion, and SPIRIT Workshop on Strategies, Prediction, Interaction, and Reasoning in Italy (IPS-RCRA-SPIRIT 2023), in: Proceedings of the Italian Workshop on Planning and Scheduling, RCRA Workshop on Experimental evaluation of algorithms for solving problems with combinatorial explosion, and SPIRIT Workshop on Strategies, Prediction, Interaction, and Reasoning in Italy (IPS-RCRA-SPIRIT 2023) co-located with 22th International Conference of the Italian Association for Artificial Intelligence (AI* IA 2023), 2023.

[2] A. Badanidiyuru, R. Kleinberg, A. Slivkins, Bandits with knapsacks, in: 2013 IEEE 54th Annual Symposium on Foundations of Computer Science, 2013, pp. 207–216. doi:10.1109/FOCS.2013.30.

[3] N. Immorlica, K. Sankararaman, R. Schapire, A. Slivkins, Adversarial bandits with knapsacks, Journal of the ACM 69 (2022). doi:10.1145/3557045.

[4] T. Kesselheim, S. Singla, Online learning with vector costs and bandits with knapsacks, in: J. Abernethy, S. Agarwal (Eds.), Proceedings of Thirty Third Conference on Learning Theory, volume 125 of *Proceedings of Machine Learning Research*, PMLR, 2020, pp. 2286–2305.

[5] M. Castiglioni, A. Celli, C. Kroer, Online learning with knapsacks: the best of both worlds, in: K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, S. Sabato (Eds.), Proceedings of the 39th International Conference on Machine Learning, volume 162 of *Proceedings of Machine Learning Research*, PMLR, 2022, pp. 2767–2783.

[6] T. Lykouris, S. Vassilvtiskii, Competitive caching with machine learned advice, in: J. Dy, A. Krause (Eds.), Proceedings of the 35th International Conference on Machine Learning, volume 80 of *Proceedings of Machine Learning Research*, PMLR, 2018, pp. 3296–3305.

[7] S. Lattanzi, T. Lavastida, B. Moseley, S. Vassilvitskii, Online Scheduling via Learned Weights, 2020, pp. 1859–1877. doi:10.1137/1.9781611975994.114.

[8] M. Mitzenmacher, Scheduling with predictions and the price of misprediction, 2019. arXiv:1902.00732.

[9] D. Rohatgi, Near-optimal bounds for online caching with machine learned advice, 2019. arXiv:1910.12172.

[10] A. Antoniadis, C. Coester, M. Elias, A. Polak, B. Simon, Online metric algorithms with untrusted predictions, in: H. D. III, A. Singh (Eds.), Proceedings of the 37th International Conference on Machine Learning, volume 119 of *Proceedings of Machine Learning Research*, PMLR, 2020, pp. 345–355.

[11] Étienne Bamas, A. Maggiori, O. Svensson, The primal-dual method for learning augmented algorithms, 2020. arXiv:2010.11632.