Categorized Bandits

Matthieu Jedor
Centre Borelli, ENS Paris-Saclay & Cdiscount
matthieu.jedor@ens-paris-saclay.fr

Jonathan Louëdec
Cdiscount
jonathan.louedec@cdiscount.com

Vianney Perchet
ENSAE & Criteo AI Lab
vianney.perchet@normalesup.org

ABSTRACT

With e-commerce as the motivating example, we introduce a new multi-armed bandit setting where arms are grouped inside "ordered" categories. We conduct an analysis on real data to highlight that those ordered categories actually exist in practice. Finally, we provide algorithms that fully leverage the structure of the model and experimental results show the superiority of our proposed policies.[1]

CCS CONCEPTS

• **Computing methodologies** → **Online learning settings**; **Sequential decision making**; • **Applied computing** → **Online shopping**; • **Information systems** → *Recommender systems.*

KEYWORDS

multi-armed bandits, recommender systems, e-commerce

1 INTRODUCTION

In the multi-armed bandit problem, an agent faces several possible decisions (or arms) and chooses sequentially one of them at each time step. This generates a sequence of rewards and the objective is to maximize their cumulative sum. The performance of an algorithm is evaluated through the "regret", which is the difference between the cumulative reward of an oracle (that knows the best arm) and the one of the algorithm. There is a clear trade-off arising between gathering information on uncertain arms and using the information already available. The traditional bandit model must however be adapted to specific applications to unleash its full power.

Consider for instance e-commerce. One of the core optimization problem is to decide which products to recommend to a user, in the objective of maximizing the click-through-rate. Arms of recommender systems are the different products that can be displayed. The number of products, even if finite, is prohibitively huge as the regret typically scale linearly with the number of arms. So agnostic bandit algorithms take too much time to complete their learning phase. Thankfully, there is an inherent structure behind a typical catalogue: products are gathered into well defined categories. As customers are generally interested in a few of them, it seems beneficial to gather information across products to speed up the learning phase and, ultimately, to make more refined recommendations.

2 MODEL

As motivated in the introduction, the total number of arms can be prohibitively large, but we assume that arms are grouped in a small number M of categories and each category has the same number of arms K.[2] At time step $t \in [T]$, the agent selects a category C_t and an arm $A_t \in C_t$ in this category. This generates a reward $X_{A_t}^{C_t} = \mu_{A_t}^{C_t} + \eta_t$ where μ_k^m is the unknown expected reward of the arm k of category m and η_t is some independent 1 sub-Gaussian white noise. We assume the existence of an optimal category with respect to a partial order defined below. Some categories might not be pairwise comparable, but we assume that the optimal category is comparable to, and dominates, all the others. As in any multi-armed bandit problem, the overall objective of an agent is to maximize her expected cumulative reward until time horizon T or identically, to minimize her expected cumulative regret.

We consider three dominances that are gradually weaker so that the setting is more and more general. Let $\mathcal{A} = \{\mu_1^{\mathcal{A}}, \dots, \mu_K^{\mathcal{A}}\} \subset \mathbb{R}$ and $\mathcal{B} = \{\mu_1^{\mathcal{B}}, \dots, \mu_K^{\mathcal{B}}\} \subset \mathbb{R}$ be a pair of categories,

Group-sparse dominance $\mathcal{A}$ group-sparsely dominates $\mathcal{B}$, denoted by $\mathcal{A} \succeq_s \mathcal{B}$, if each element of $\mathcal{A}$ is non-negative and at least one is positive, and each element of $\mathcal{B}$ are non-positive, i.e., $\max_{k\in[K]} \mu_k^{\mathcal{A}} > \min_{k\in[K]} \mu_k^{\mathcal{A}} \geq 0 \geq \max_{k\in[K]} \mu_k^{\mathcal{B}}$.

Strong dominance $\mathcal{A}$ strongly dominates $\mathcal{B}$, denoted by $\mathcal{A} \succeq_0 \mathcal{B}$, if each element of $\mathcal{A}$ is bigger than any element of $\mathcal{B}$, i.e., $\min_{k\in[K]} \mu_k^{\mathcal{A}} \geq \max_{k\in[K]} \mu_k^{\mathcal{B}}$.

First-order dominance $\mathcal{A}$ first-order dominates $\mathcal{B}$, denoted by $\mathcal{A} \succeq_1 \mathcal{B}$, if $\sup_{x\in\mathbb{R}} F_{\mathcal{A}}(x) - F_{\mathcal{B}}(x) \leq 0$, where $F_{\mathcal{A}}(x) = \frac{1}{K}\sum_{k=1}^{K} \mathbf{1}\{\mu_k^{\mathcal{A}} \leq x\}$ is the cumulative distribution function of a uniform random variable over $\mathcal{A}$.

3 EMPIRICAL EVIDENCE OF DOMINANCE

We illustrate these assumptions on a real dataset. We have collected the CTR of products in four different categories over one month on the e-commerce website Cdiscount, one of the leading e-commerce companies in France. *CAT* 1 to 3 are three of the largest categories in terms of revenue while *CAT 4* is a smaller category. We have represented the cumulative distribution function of the four categories in Figure 1. The following dominances can be highlighted. For the strong dominance, *CAT 1* $\succeq_0$ *CAT 4* and *CAT 2* $\succeq_0$ *CAT 4*. For the first-order dominance, *CAT 2* $\succeq_1$ *CAT 3* and *CAT 3* $\succeq_1$ *CAT 4*, the last assertion is not verified in the strong case. *CAT 1* and *CAT 2* are not comparable with respect to any partial order. Notice that, had the first item of *CAT 2* performed only 5% worse than observed,[3] then *CAT 1* would have been optimal in the first-order sense.

[1] This extended abstract describes work published at NeurIPS 2019 [1].

[2] All of our results immediately generalize to categories with different number of arms.

[3] The CTR of the best item of *CAT 2* is so higher than the second one, we could expect it is actually an outlier, i.e., an artefact of the choice of that specific month and category.

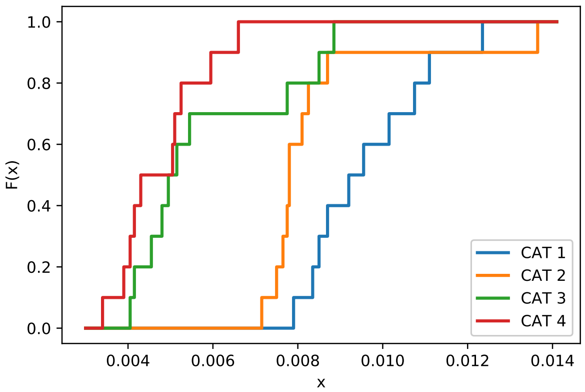

Figure 1: Cdf of the 4 categories

4 ALGORITHMS

4.1 Category Successive Elimination (CatSE)

The main concept of CatSE is to successively eliminate suboptimal categories. It behaves in three different ways depending on the number of categories that are called "active". The definition of an active category will depend on the assumption of dominance. Formally, let $\delta \in (0,1)$ be a confidence level. At time step t, it computes the set of active categories, denoted $\mathcal{A}(t,\delta)$. Then it behaves according to the assertion that is verified:

(1) $|\mathcal{A}(t,\delta)| = 0$: pulls all arms.
(2) $|\mathcal{A}(t,\delta)| = 1$: performs UCB in the active category.
(3) $|\mathcal{A}(t,\delta)| > 1$: pulls all arms inside active categories.

4.2 Murphy Sampling (MS)

The MS algorithm [2] is derived from Thompson Sampling (TS), the difference being that the sampling respects some inherent structure of the problem. To define MS, we denote by $\mathcal{F}(t) = \sigma(A_1, X_1, \ldots, A_t, X_t)$ the information available after t steps and $\mathcal{H}_d$ the assumption of dominance considered. Let $\Pi_t = \mathbb{P}(\cdot|\mathcal{F}_t)$ be the posterior distribution of the means parameters after t rounds. The algorithm samples, at each time step, from the posterior distribution $\Pi_{t-1}(\cdot|\mathcal{H}_d)$ and then pulls the best arm, which, by definition, is in the best category sampled at this time step. In comparison, TS would sample from Π_{t-1} without taking into account any structure.

5 EXPERIMENTS

Numerical experiments illustrating the performance of our algorithms are presented for two scenarios: in the strong dominance on Figure 2 and in the first-order dominance on Figure 3. In both experiments, rewards are drawn from Gaussian distribution with unit variance and we report the average regret as a function of time. Both CatSE and MS outperform baseline algorithms, with a clear advantage for the latter due to it being a fully sequential strategy.[4]

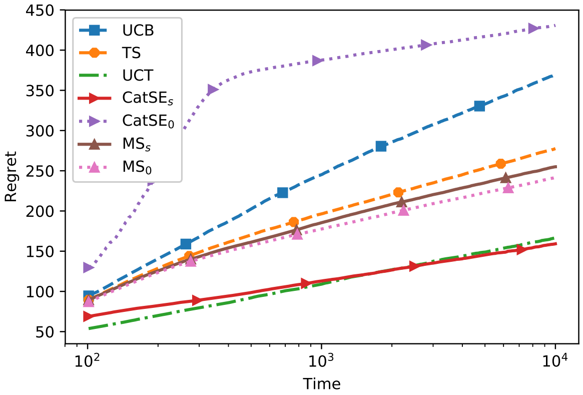

Figure 2: Regrets in the strong dominance scenario

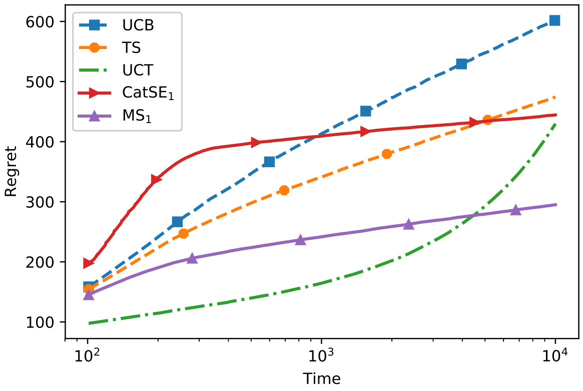

Figure 3: Regrets in the first-order dominance scenario

6 CONCLUSION

We introduced a novel bandit framework inspired by e-commerce applications where arms are assumed to belong to ordered categories. We confirmed the veracity of our model on real data and presented two generic algorithms for this setting.

Two problems remain open: the first one is a better exploration phase in CatSE since it heavily impacts the regret; and the second is an upper bound on the regret of the MS algorithm since it is highly competitive in practice. We believe that it is asymptotically optimal and that it can be applied to other model of structured bandits.

ACKNOWLEDGMENTS

This work was supported by a public grant as part of the Investissement d'avenir project, reference ANR-11-LABX-0056-LMH, LabEx LMH, in a joint call with Gaspard Monge Program for optimization, operations research and their interactions with data sciences.

REFERENCES

[1] Matthieu Jedor, Vianney Perchet, and Jonathan Louedec. 2019. Categorized Bandits. In *Advances in Neural Information Processing Systems*. 14399–14409.

[2] Emilie Kaufmann, Wouter Koolen, and Aurelien Garivier. 2018. Sequential Test for the Lowest Mean: From Thompson to Murphy Sampling. *arXiv preprint arXiv:1806.00973* (2018).

[4] The performance of CatSE$_s$ is due to a more efficient sampling.