Decentralized Competing Bandits in Many-to-One Matching Markets

Extended Abstract

Yirui Zhang
Tsinghua University
Shanghai Qi Zhi Institute
China
zhangyr22@mails.tsinghua.edu.cn

Zhixuan Fang
Tsinghua University
Shanghai Qi Zhi Institute
China
zfang@mail.tsinghua.edu.cn

ABSTRACT

Two-sided matching is a classic and well-studied problem. As the participants are usually not aware of the accurate preferences towards the other side, the model of competing bandits characterizes the process of learning uncertainty through interactions in one-to-one matching markets. However, it does not apply to many cases, such as the online labor market where employers may have multiple vacancies. Thus, in this paper, we study the generalized problem of competing bandits in many-to-one matching markets and focus on the fully decentralized setting. We propose an algorithm and show that it achieves $O(\log T)$ regret compared with the optimal stable matching, for the first time without restricted assumptions on preferences and observability in previous literature.

KEYWORDS

Multi-agent Reinforcement Learning; Bandits; Matching

ACM Reference Format:
Yirui Zhang and Zhixuan Fang. 2024. Decentralized Competing Bandits in Many-to-One Matching Markets: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

1 INTRODUCTION

Online matching markets (e.g., TaskRabbit, Thumbtack) that match employers with workers have become prevalent in the last decade. As the market grows rapidly, participants in the matching market face increasing uncertainties due to the lack of information on the other side. For example, consumers may not know the service qualities of service providers, and workers may not know the value brought by the provided positions. In these cases, agents do not have clear preferences towards the other side. They have to learn and form their preferences during repeated matches. To capture such processes, Liu et al. [5] introduce the framework of the one-to-one competing bandits by adopting the classical MAB model [3] into two-sided matching markets [1, 2, 7]. In this framework, employers or different kinds of works are considered as arms, and workers as agents. Both sides have preferences over the other side, but agents need to learn their preferences through repeated interactions. When multiple agents select the same arm, only the arm's most preferred agent wins (e.g., gets the work) and receives a non-zero stochastic reward.

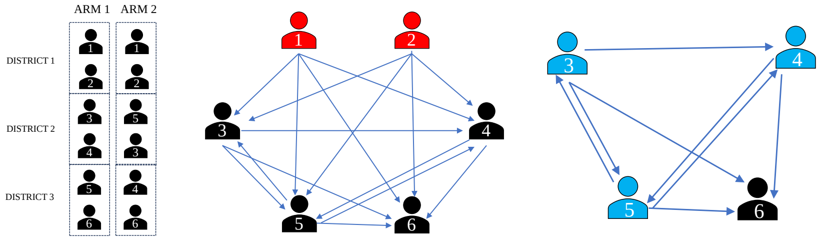
However, the one-to-one setting is somewhat limited since it is common for an employer or a kind of work to have multiple positions. The many-to-one setting where arms are able to match multiple agents is more general and practical. We focus on the many-to-one setting and want to design algorithms that achieve some commonly desired properties.

- **Fully decentralized**. Previous work has highlighted the importance and generality of the decentralized setting [5, 8]. In such case, there is no central authority and no explicit communication among agents. Beyond these properties, we argue that in the fully decentralized setting, agents do not have a predetermined identity (i.e., no predetermined index), as this usually requires a centralized identity assignment or mutual communication.
- **Arbitrary and private arm preferences**. This property requires that employers do not reveal their personal preferences to others, and they may prefer different kinds of workers without restrictions.
- **No observation of winner**. This is the most general assumption on observed information in reality. This reflects a common practice that a worker is only informed by the company of her own result (rejection or acceptance), without knowing the competition's winners or even the identities of the competitors. However, many works [4, 6, 9] focus on scenarios involving observation, where all winning agents on all arms are broadcast to all agents.
- **Regret on optimal stable matching**. While low regret is always the objective in bandit-based problems, regret on optimal stable matching rather than pessimal stable matching provides a much tighter metric on how well the agents are matched. This is because when there exist multiple stable matches, the gap between regret on optimal and pessimal stable matching could be up to a linear order.

Contribution. In our work, we propose the algorithm SUBMARket IteratioN Explore-then-commit (SUBMARINE), which is the first algorithm that achieves all the above desired properties and obtains the tight regret of $O(\log T)$ in the many-to-one setting. Moreover, we propose a general model to characterize the preference structure and new techniques to handle the challenges brought by the absence of communication channels and observation information in the fully decentralized scenario.

This work is supported by Tsinghua University Dushi Program. Corresponding author: Zhixuan Fang at Tsinghua University, Beijing, China (zfang@mail.tsinghua.edu.cn).

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).



(i) Districts (ii) Influence Graph (iii) Subgraph of (ii)

Fig. 1: An example of preference structure in the many-to-one setting. Figure (i) shows the districts, figure (ii) illustrates the influence graph, and figure (iii) is a subgraph of (ii) excluding agent 1 and agent 2. Agents with red color represent the favorite agents and agents with blue color form the outer closed circle. Note that if the favorite agents, i.e. agent 1 and agent 2, leave the influence graph, the set of agents $\{3, 4, 5\}$ becomes the outer closed circle of the subgraph.

2 PREFERENCE STRUCTURE

We introduce some new definitions in order to help figure out the complex preference structures in the many-to-one scenario.

Definition 1 (District and favorite agent). *An agent in district $d \in \{1, 2, ...\}$ on arm k means that the rank of the agent j on arm k is within the set $\{(d-1)S_k + 1, (d-1)S_k + 2, ..., dS_k\}$. An agent is a favorite agent if and only if she is in district 1 on every arm.*

The districts reflect part of the priority on each arm. The favorite agents are a set of agents who are most preferred by all arms. Figure 1 shows an example including 2 arms and 6 agents, and (i) shows the preferences on both arms and the corresponding districts.

Agents' preferences on arms also convey their influence on other agents. Specifically, if agent j_1 can potentially "squeeze" agent j_2 out, we say that agent j_1 can influence agent j_2.

Definition 2 (Influence graph in the many-to-one setting). *An influence graph $G(\mathcal{M}, \mathcal{K}, S)$, where $\mathcal{M}$ is the set of agents, $\mathcal{K}$ is the set of arms, and S denotes the vector of arm capacity, is a directed graph with every agent as one vertex. There is a directed edge from vertex i to vertex j in G if and only if there exists at least one arm $k \in \mathcal{K}$ such that $i \succ_k j$ and $j \notin \mathcal{D}_k^1$.*

If there exists a path from agent j_1 to agent j_2 in the influence graph, then it indicates that agent j_1 can influence agent j_2. Based on the preferences shown in (i), Figure (ii) depicts the influence graph with two favorite agents.

Definition 3 (Outer closed circle). *[10] An outer closed circle is a non-empty subset of vertexes in a directed graph G, i.e., $M \subseteq V(G)$, which satisfies that: i) M is connected; ii) there is no incoming edge from vertices in $V(G) \setminus M$ to any vertex in M.*

The outer closed circle in the influence graph represents the set of agents with the highest priority. They can influence each other and other agents but cannot be influenced by other agents. Figure (iii) is the subgraph of (ii) excluding the two favorite agents 1 and 2, and agents 3, 4, 5 form an outer closed circle.

Lemma 1. *If there exists no favorite agent, then there exists a set of agents that are the outer closed circle of the influence graph $G(\mathcal{M}, \mathcal{K}, S)$.*

3 ALGORITHM

In Section 2, we recognize that there always exists a set of agents with a high influence level, capable of influencing others but impervious to influence within any given preference structure. Based on this property, we can sequentially divide the matching market into several smaller sub-markets and assist agents in finding their optimal stable pairs through a recursion-based method. Specifically, in every sub-market, influential agents will match with their optimal stable pairs and subsequently leave the market. Other agents will then form a new sub-market and repeat the process.

The SUBMARINE algorithm consists of 3 phases: the initialization phase, the sub-market phase and the exploitation phase.

- **Initialization.** Every agent will receive an index and learn about her ranks or districts on all arms.
- **Submarket.** The algorithm will proceed in a round based way. After every round r, some agents will be settled on some arms satisfactorily and enter the exploitation phase. Agents and arms that are unsettled remain in the market, actively learning to find the stable match.
- **Exploitation.** Every agent will always choose her empirical optimal arm.

The following theorem shows that SUBMARINE achieves an optimal sublinear regret for every agent.

Theorem 1. *(Informal) If every agent runs SUBMARINE, then the optimal regret will be upper bounded by:*

$$Reg^*(T, j) = O(\frac{\max\{K, M\} \log T}{\Delta^2}).$$

4 CONCLUSION

In this work, we study competing bandits in many-to-one matching markets. We conduct a generalized analysis of the complex preference structure in the many-to-one setting. We propose an ETC-based algorithm, which is the first algorithm to relax restricted assumptions on winner observation, predetermined index, and special or public arm preferences, etc. The algorithm largely improves prior regret bounds and achieves $O(\log T)$ optimal regret, by utilizing new ideas and techniques such as sub-market, new communication protocols, and random pull.

REFERENCES

[1] Itai Ashlagi, Anilesh K Krishnaswamy, Rahul Makhijani, Daniela Saban, and Kirankumar Shiragur. 2020. Assortment Planning for Two-Sided Sequential Matching Markets. In *Web and Internet Economics: 16th International Conference, WINE 2020, Beijing, China, December 7–11, 2020, Proceedings*, Vol. 12495. Springer Nature, 476.

[2] David Gale and Lloyd S Shapley. 1962. College admissions and the stability of marriage. *The American Mathematical Monthly* 69, 1 (1962), 9–15.

[3] Michael N Katehakis and Arthur F Veinott Jr. 1987. The multi-armed bandit problem: decomposition and computation. *Mathematics of Operations Research* 12, 2 (1987), 262–268.

[4] Fang Kong and Shuai Li. 2023. Player-optimal Stable Regret for Bandit Learning in Matching Markets. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 1512–1522.

[5] Lydia T Liu, Horia Mania, and Michael Jordan. 2020. Competing bandits in matching markets. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 1618–1628.

[6] Lydia T Liu, Feng Ruan, Horia Mania, and Michael I Jordan. 2021. Bandit learning in decentralized matching markets. *Journal of Machine Learning Research* 22, 211 (2021), 1–34.

[7] Alvin E Roth and Xiaolin Xing. 1997. Turnaround time and bottlenecks in market clearing: Decentralized matching in the market for clinical psychologists. *Journal of political Economy* 105, 2 (1997), 284–329.

[8] Abishek Sankararaman, Soumya Basu, and Karthik Abinav Sankararaman. 2021. Dominate or delete: Decentralized competing bandits in serial dictatorship. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 1252–1260.

[9] Zilong Wang, Liya Guo, Junming Yin, and Shuai Li. 2022. Bandit Learning in Many-to-One Matching Markets. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 2088–2097.

[10] Yirui Zhang, Siwei Wang, and Zhixuan Fang. 2022. Matching in Multi-arm Bandit with Collision. *Advances in Neural Information Processing Systems* 35 (2022), 9552–9563.