Stochastic k-Submodular Bandits with Full Bandit Feedback

Extended Abstract

Guanyu Nie
Iowa State University
Ames, IA, United States
nieg@iastate.edu

Vaneet Aggarwal
Purdue University
West Lafayette, IN, United States
vaneet@purdue.edu

Christopher John Quinn
Iowa State University
Ames, IA, United States
cjquinn@iastate.edu

ABSTRACT

In this paper, we present the first sublinear α-regret bounds for online k-submodular optimization problems with full-bandit feedback, where α is a corresponding offline approximation ratio. Specifically, we propose online algorithms for multiple k-submodular stochastic combinatorial multi-armed bandit problems, including (i) monotone functions and individual size constraints, (ii) monotone functions with matroid constraints, (iii) non-monotone functions with matroid constraints, (iv) non-monotone functions without constraints, and (v) monotone functions without constraints. We transform approximation algorithms for offline k-submodular maximization problems into online algorithms through the offline-to-online framework proposed by [9]. A key contribution of our work is analyzing the robustness of the offline algorithms.

KEYWORDS

k-submodular; multi-armed bandits; bandit feedback

ACM Reference Format:

Guanyu Nie, Vaneet Aggarwal, and Christopher John Quinn. 2025. Stochastic k-Submodular Bandits with Full Bandit Feedback: Extended Abstract. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 3 pages.

1 INTRODUCTION

In sequential decision-making problems like sensor placement and influence maximization, decisions involve selecting subsets of elements, making assignments, and observing outcomes. These problems often exhibit diminishing returns. For example, in a multi-agent social network content-spreading scenario, multiple companies cooperate to spread k types of content. The more influencers each company sponsors, the (marginal) increase in diffusion size due to any particular influencer will diminish.

The offline version of such problems can be modeled as k-submodular optimization problems [2]. However, maximizing a k-submodular function is NP-hard [16]. There has been progress in offline approximation algorithms [3, 10, 11]. The online version can be modeled as a stochastic combinatorial multi-armed bandit (CMAB) problem with k-submodular expected rewards, constraints, and bandit feedback. We address the CMAB problem with (only) bandit feedback.

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

Our Contributions: We propose and analyze the first CMAB algorithms with sub-linear α-regret for k-submodular expected rewards using full-bandit feedback. For different scenarios (non-monotone and monotone functions, with and without constraints), we achieve sub-linear regret bounds by analyzing the robustness of offline algorithms. The detailed results are summarized in Table 1, where the right side of the vertical line is obtained from our analysis.

Related Works: For k-submodular CMAB, [12] considered unconstrained problems under semi-bandit feedback in an adversarial setting. For submodular CMAB ($k = 1$), there are algorithms for full-bandit feedback and different constraints [1, 4, 7–9, 13]. Many works rely on additional "semi-bandit" feedback [6, 15, 17, 18], but we focus on full-bandit feedback.

Table 1: Summary of offline α-approximation algorithms for k-submodular maximization with our δ-robustness analysis and α-regret bounds for our proposed algorithms for k-submodular CMAB with full-bandit feedback. N is an upper bound on the query complexity of the offline algorithm. B is the total budget. M is the rank of the matroid.

Ref.	Mono.	Constraint	α	δ	N	Our α-regret
[3]	×	Unconstr.	1/2	$20n$	nk	$\tilde{O}(nk^{\frac{1}{3}}T^{\frac{2}{3}})$
[3]	✓	Unconstr.	$k/(2k-1)$	$(16-\frac{2}{k})n$	nk	$\tilde{O}(nk^{\frac{1}{3}}T^{\frac{2}{3}})$
[10]	✓	Total Size	1/2	$B+1$	nkB	$\tilde{O}(n^{\frac{1}{3}}k^{\frac{1}{3}}BT^{\frac{2}{3}})$
[14]	×	Total Size	1/3	$4/3(B+1)$	nkB	$\tilde{O}(n^{\frac{1}{3}}k^{\frac{1}{3}}BT^{\frac{2}{3}})$
[10]	✓	Indiv. Size	1/3	$4/3(B+1)$	nkB	$\tilde{O}(n^{\frac{1}{3}}k^{\frac{1}{3}}BT^{\frac{2}{3}})$
[11]	✓	Matroid	1/2	$M+1$	nkM	$\tilde{O}(n^{\frac{1}{3}}k^{\frac{1}{3}}MT^{\frac{2}{3}})$
[14]	×	Matroid	1/3	$4/3(M+1)$	nkM	$\tilde{O}(n^{\frac{1}{3}}k^{\frac{1}{3}}MT^{\frac{2}{3}})$

2 PRELIMINARIES

***k*-Submodular Functions.** Let k be a positive integer for the number of *types* (i.e., types of stories) and $V = [n]$ be the ground set of *elements* (i.e., users in a social network). Let $(k+1)^V := \{(X_1, \ldots, X_k) | X_i \subseteq V, i \in \{1, \ldots, k\}, X_i \cap X_j = \emptyset, \forall i \neq j\}$. A function $f : (k+1)^V \to \mathbb{R}$ is called *k-submodular* if, for any $\boldsymbol{x} = (X_1, \ldots, X_k)$ and $\boldsymbol{y} = (Y_1, \ldots, Y_k)$ in $(k+1)^V$, we have $f(\boldsymbol{x}) + f(\boldsymbol{y}) \geq f(\boldsymbol{x} \sqcup \boldsymbol{y}) + f(\boldsymbol{x} \sqcap \boldsymbol{y})$ where $\boldsymbol{x} \sqcap \boldsymbol{y} := (X_1 \cap Y_1, \ldots, X_k \cap Y_k)$, $\boldsymbol{x} \sqcup \boldsymbol{y} := (X_1 \cup Y_1 \setminus (\cup_{i \neq 1} X_i \cup Y_i), \ldots, X_k \cup Y_k \setminus (\cup_{i \neq k} X_i \cup Y_i))$. Define the *marginal gain* of assigning type $i \in [k]$ to element e given a current solution $\boldsymbol{x}$ (provided e has not been assigned any type in $\boldsymbol{x}$), $\Delta_{e,i} f(\boldsymbol{x}) = f(X_1, \ldots, X_{i-1}, X_i \cup \{e\}, X_{i+1}, \ldots, X_k) - f(X_1, \ldots, X_k)$. A k-submodular function satisfies *orthant submodularity* and *pairwise monotonicity* [16]. A function $f : (k+1)^V \to \mathbb{R}$ is *monotone* if $\Delta_{e,i} f(\boldsymbol{x}) \geq 0$ for any $\boldsymbol{x} \in (k+1)^V$, $e \notin \bigcup_{\ell \in [k]} X_\ell$, and $i \in [k]$.

CMAB. In the CMAB framework, a learner makes sequential decisions over time horizon T. At each step t, a feasible action (subset of the ground set) A_t is selected, and a stochastic reward $f_t(A_t)$ is received. The goal is to maximize the cumulative reward. When the analogous offline optimization problem is NP-hard, and there is a known α-approximation algorithm, performance is measured by expected cumulative α-regret,

$$\mathbb{E}[\mathcal{R}_T] = \alpha T f(\mathrm{OPT}) - \mathbb{E}\left[\sum_{t=1}^{T} f_t(A_t)\right]. \tag{1}$$

Problem Statement. We consider the stochastic CMAB problem where the expected reward function is k-submodular. Each arm is an item-type pair, and we only have full-bandit feedback. We aim to transform offline k-submodular optimization algorithms to online algorithms and use expected α-regret as the performance metric.

Offline-to-Online Framework. We adopt the offline-to-online transformation framework proposed in [9]. In [9], they introduced (α, δ, N)-robustness of an offline approximation algorithm (see Definition 1 below). They showed that this property alone is sufficient to guarantee that the offline algorithm can be adapted to solve CMAB problems in the corresponding online setting with just bandit feedback and achieve sub-linear regret. Specifically, they showed that the expected cumulative α-regret of C-ETC is at most $O\left(\delta^{\frac{2}{3}} N^{\frac{1}{3}} T^{\frac{2}{3}} \log(T)^{\frac{1}{3}}\right)$ with $T \geq \max\left\{N, \frac{2\sqrt{2}N}{\delta}\right\}$. More importantly, the CMAB adaptation will not rely on any special structure of the algorithm design, instead employing it as a black box. We restate the robustness definition in the following.

Definition 1 ((α, δ, N)-Robust Approximation [9])**.** Algorithm $\mathcal{A}$ is an (α, δ, N)-robust approximation algorithm for the combinatorial optimization problem of maximizing a function $f : 2^{\Omega} \to \mathbb{R}$ over a finite domain $D \subseteq 2^{\Omega}$ if its output S^* using a value oracle $\hat{f}$, provided that for any $\epsilon > 0$, $|f(S) - \hat{f}(S)| \leq \epsilon$ for all $S \in D$, satisfies $\mathbb{E}[f(S^*)] \geq \alpha f(\mathrm{OPT}) - \delta\epsilon$, where OPT is optimal under f, Ω is the ground set, the expectation is over the randomness of $\mathcal{A}$, and algorithm $\mathcal{A}$ uses at most N value oracle queries.

3 MAIN RESULTS

Non-monotone Functions without Constraints: We adopt the offline algorithm proposed in [3]. We first show that the Algorithm 2 in [3] is $(\frac{1}{2}, 20n, nk)$-robust. Then, by the C-ETC framework, we obtain the expected cumulative 1/2-regret bound of $O\left(nk^{\frac{1}{3}} T^{\frac{2}{3}} \log(T)^{\frac{1}{3}}\right)$ given $T \geq nk$.

Monotone Functions without Constraints: We use Algorithm 3 in [3]. In the original algorithm, it was stated as $\beta \leftarrow \sum_{i=1}^{k} y_i^t$. This particular step is not robust to noise and we showed it can be changed to $\beta \leftarrow \sum_{i=1}^{k} [y_i]_+^t$ to yield a $(\frac{k}{2k-1}, (16-\frac{2}{k})n, nk)$-robustness guarantee. By the C-ETC framework, we obtain the expected cumulative $\frac{k}{2k-1}$-regret bound of $O\left(nk^{\frac{1}{3}} T^{\frac{2}{3}} \log(T)^{\frac{1}{3}}\right)$ given $T \geq \max\{nk, \frac{2\sqrt{2}k}{16-\frac{2}{k}}\}$.

Monotone Functions with Individual Size (IS) Constraints: In IS, each type i has a limit B_i on the maximum number of pairs of that type i, with $B = \sum_i B_i$ as the total budget. We consider the offline greedy Algorithm 3 proposed in [10]. We first show that Algorithm 3 in [10] is $(\frac{1}{3}, \frac{4}{3}(B+1), nkB)$-robust. Then, by the C-ETC framework, we obtain the expected cumulative 1/3-regret bound of $O\left(n^{\frac{1}{3}} k^{\frac{1}{3}} B T^{\frac{2}{3}} \log(T)^{\frac{1}{3}}\right)$ given $T \geq nk \max\{1, \frac{3\sqrt{2}B}{2(B+1)}\}$.

Monotone Functions with Matroid Constraints: We adapt Algorithm 3.1 in [11]. We first show that Algorithm 3.1 in [11] is $(\frac{1}{2}, M+1, nkM)$-robust, where M is the rank of the matriod. Applying the C-ETC framework, we obtain the expected cumulative 1/2-regret bound of $O\left(n^{\frac{1}{3}} k^{\frac{1}{3}} M T^{\frac{2}{3}} \log(T)^{\frac{1}{3}}\right)$ given $T \geq nk \max\{1, \frac{3\sqrt{2}M}{2(M+1)}\}$. As a special case of the matroid constraint, we can obtain a similar regret bound for the Total Size (TS) constraint [10].

Non-monotone Functions with Matroid Constraints: The proposed algorithm in [14] is shown to achieve a 1/3 approximation ratio. We show that the algorithm is $(\frac{1}{3}, \frac{4}{3}(M+1), nkM)$-robust. Using C-ETC, the expected cumulative 1/3-regret bound is $O\left(n^{\frac{1}{3}} k^{\frac{1}{3}} M T^{\frac{2}{3}} \log(T)^{\frac{1}{3}}\right)$ given $T \geq nk \max\{1, \frac{3\sqrt{2}M}{2(M+1)}\}$.

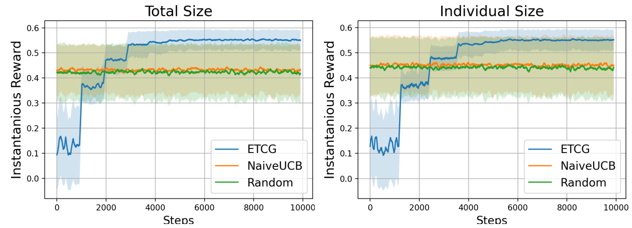

Figure 1: Instantaneous Rewards on Influence Maximization experiments.

4 EVALUATIONS

We evaluate our methods in the context of online influence maximization with $k = 3$ topics. We used the the *k-topic independent cascade* (k-IC) model from Ohsaka and Yoshida [10] on a subgraph with 350 users and 2,845 edges of the ego-Facebook network [5]. We evaluate our algorithms under both TS (budget $B = 6$) and IS (each topic has a budget of 2) constraints for a horizon of $T = 10^4$.

Instantaneous reward plots are shown in Figure 1. Means and standard deviations are calculated over 10 independent runs. We compare with NaiveUCB and random selection. The results show that our algorithm (ETCG) catches up in later stages and achieves lower cumulative regret, while NaiveUCB has a poor performance due to the large number of actions to explore.

5 CONCLUSION

We investigated online k-submodular maximization under bandit feedback. We proposed CMAB algorithms by adapting offline algorithms and analyzing their robustness, obtaining sublinear regret bounds in various settings. Numerical experiments verified the effectiveness of our methods. Future work could focus on further improving the regret bounds and more complex scenarios.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under grants CCF-2149588 and CCF-2149617.

REFERENCES

[1] Fares Fourati, Christopher John Quinn, Mohamed-Slim Alouini, and Vaneet Aggarwal. 2024. Combinatorial stochastic-greedy bandit. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 12052–12060.

[2] Anna Huber and Vladimir Kolmogorov. 2012. Towards Minimizing k-Submodular Functions. In *International Symposium on Combinatorial Optimization*.

[3] Satoru Iwata, Shin-ichi Tanigawa, and Yuichi Yoshida. 2015. Improved Approximation Algorithms for k-Submodular Function Maximization. In *ACM-SIAM Symposium on Discrete Algorithms*.

[4] Samir Khuller, Anna Moss, and Joseph Seffi Naor. 1999. The budgeted maximum coverage problem. *Inform. Process. Lett.* 70, 1 (1999), 39–45.

[5] Jure Leskovec and Julian Mcauley. 2012. Learning to discover social circles in ego networks. *Advances in neural information processing systems* 25.

[6] Tian Lin, Jian Li, and Wei Chen. 2015. Stochastic Online Greedy Learning with Semi-bandit Feedbacks.. In *Proceedings of the 29th International Conference on Neural Information Processing Systems*. 352–360.

[7] Rad Niazadeh, Negin Golrezaei, Joshua R Wang, Fransisca Susan, and Ashwinkumar Badanidiyuru. 2021. Online learning via offline greedy algorithms: Applications in market design and optimization. In *Proceedings of the 22nd ACM Conference on Economics and Computation*. 737–738.

[8] Guanyu Nie, Mridul Agarwal, Abhishek Kumar Umrawal, Vaneet Aggarwal, and Christopher John Quinn. 2022. An explore-then-commit algorithm for submodular maximization under full-bandit feedback. In *Uncertainty in Artificial Intelligence*. PMLR, 1541–1551.

[9] Guanyu Nie, Yididiya Y. Nadew, Yanhui Zhu, Vaneet Aggarwal, and Christopher John Quinn. 2023. A Framework for Adapting Offline Algorithms to Solve Combinatorial Multi-Armed Bandit Problems with Bandit Feedback. In *Proceedings of the 40th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 202)*. PMLR, 26166–26198.

[10] Naoto Ohsaka and Yuichi Yoshida. 2015. Monotone k-Submodular Function Maximization with Size Constraints. In *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett (Eds.), Vol. 28. Curran Associates, Inc.

[11] Shinsaku Sakaue. 2017. On maximizing a monotone k-submodular function subject to a matroid constraint. *Discrete Optimization* 23 (2017), 105–113.

[12] Tasuku Soma. 2019. No-regret algorithms for online k-submodular maximization. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics (Proceedings of Machine Learning Research, Vol. 89)*, Kamalika Chaudhuri and Masashi Sugiyama (Eds.). PMLR, 1205–1214.

[13] Matthew Streeter and Daniel Golovin. 2008. An Online Algorithm for Maximizing Submodular Functions. In *Proceedings of the 21st International Conference on Neural Information Processing Systems*. 1577–1584.

[14] Yunjing Sun, Yuezhu Liu, and Min Li. 2022. Maximization of k-Submodular Function with a Matroid Constraint. In *Theory and Applications of Models of Computation*, Ding-Zhu Du, Donglei Du, Chenchen Wu, and Dachuan Xu (Eds.). Springer International Publishing, Cham, 1–10.

[15] Sho Takemori, Masahiro Sato, Takashi Sonoda, Janmajay Singh, and Tomoko Ohkuma. 2020. Submodular Bandit Problem Under Multiple Constraints. In *Conference on Uncertainty in Artificial Intelligence*. PMLR, 191–200.

[16] Justin Ward and Stanislav Živnỳ. 2016. Maximizing k-submodular functions and beyond. *ACM Transactions on Algorithms* 12, 4 (2016), 1–26.

[17] Baosheng Yu, Meng Fang, and Dacheng Tao. 2016. Linear submodular bandits with a knapsack constraint. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 30.

[18] Yisong Yue and Carlos Guestrin. 2011. Linear submodular bandits and their application to diversified retrieval. *Advances in Neural Information Processing Systems* 24 (2011).