A Comparative Study of Bandit Algorithms for Financial Asset Recommendation Using the FAR-Trans Dataset

Abstract

Financial asset recommendation (FAR) systems must balance exploration and exploitation in dynamic markets while personalizing recommendations to diverse investor profiles. This paper introduces the application of multi-armed bandit algorithms to financial asset recommendation and establishes cumulative regret as an evaluation metric for this domain. Building upon the FAR-Trans benchmark dataset, we conduct the first systematic evaluation of bandit algorithms for FAR, comparing LinUCB (contextual bandit), UCB1 (non-contextual bandit), Random selection, and Collaborative Filtering across four experimental scenarios with varying number of assets and customers (130K to 750K recommendation rounds). Our regret-based evaluation framework reveals algorithm learning dynamics and quantifies the opportunity cost of exploration, providing insights not captured by traditional accuracy or profitability metrics. Results demonstrate that LinUCB consistently achieves 2-4% lower cumulative regret compared to UCB1, with particularly strong performance in long-horizon scenarios. These findings establish bandit algorithms as effective approaches for FAR and demonstrate how regret analysis provides a principled framework for evaluating sequential recommendation systems in financial applications.

I. INTRODUCTION

The financial services industry has witnessed a paradigm shift toward algorithmic investment advice and robo-advisory platforms, creating unprecedented demand for intelligent recommendation systems. Financial asset recommendation (FAR), as defined by Sanz-Cruzado et al., is a specialized subdomain of recommender systems that identifies suitable financial securities for investors based on multiple data sources including pricing time series, customer profiles, and historical transactions. Unlike conventional recommendation domains such as e-commerce or content streaming, FAR faces unique challenges: severe consequences of suboptimal recommendations, highly dynamic reward structures driven by market volatility, regulatory compliance requirements, and the need to accommodate diverse investor risk profiles and time horizons.

Traditional approaches to FAR have been constrained by the lack of publicly available datasets containing realistic customer transaction data. Most research has focused on profitability prediction using freely available pricing data, but these approaches fail to capture the personalization aspects critical to real-world financial advisory. The recent introduction of FAR-Trans by Sanz-Cruzado et al. addresses this gap by providing the first public benchmark dataset containing pricing information, asset characteristics, and anonymized retail investor transactions from a large European financial institution spanning January 2018 to November 2022.

Multi-armed bandit (MAB) algorithms offer a principled framework for sequential decision-making under uncertainty, naturally addressing the exploration-exploitation dilemma inherent in recommendation systems. Contextual bandits can leverage user-specific and market-specific features to provide personalized recommendations while continuously learning from user interactions. Despite their theoretical appeal and success in other domains such as news recommendation, their application to financial asset recommendation using real transaction data remains underexplored.

This paper makes the following contributions:

1. **Applying Bandit Algorithms to Financial Asset Recommendation**: We introduce the application of multi-armed bandit algorithms (LinUCB, UCB1) to the financial asset recommendation problem, providing the first systematic evaluation of bandit approaches on the FAR-Trans benchmark dataset. While Sanz-Cruzado et al. evaluated eleven algorithms spanning profitability-based and transaction-based methods, their work did not include bandit algorithms.
2. **Regret-Based Evaluation Framework**: We establish cumulative regret as an evaluation metric for comparing financial recommendation algorithms. Unlike traditional information retrieval metrics (precision, recall, NDCG) or purely financial metrics (returns, Sharpe ratio), regret directly quantifies the opportunity cost of exploration and provides a principled framework for evaluating the exploration-exploitation trade-off inherent in sequential recommendation tasks. We demonstrate how regret analysis across multiple time horizons (130K to 750K rounds) reveals algorithm learning dynamics and long-term performance characteristics not captured by static metrics.

II. RELATED WORK

A. Financial Asset Recommendation

Financial asset recommendation has evolved from rule-based portfolio theory approaches to sophisticated machine learning models. Sanz-Cruzado et al. categorize FAR approaches into three groups based on their primary data sources: price-based methods that analyze time series patterns, transaction-based methods leveraging collaborative signals from investor behavior, and hybrid approaches combining multiple information sources. Early work focused on profitability prediction using technical indicators and price patterns, but lacked the personalization capabilities needed for diverse investor populations.

The FAR-Trans benchmark introduced by Sanz-Cruzado et al. evaluated eleven algorithms spanning these categories, finding that profitability-based algorithms effectively identified profitable assets while collaborative filtering excelled at predicting future purchases based on historical transaction patterns. However, their benchmark did not include bandit algorithms, which are particularly well-suited for online learning scenarios where recommendations must adapt continuously to changing market conditions and evolving user preferences.

B. Multi-Armed Bandit Algorithms

The multi-armed bandit problem, formalized by Robbins in 1952, provides a mathematical framework for sequential decision-making under uncertainty. Classical algorithms such as UCB1, proposed by Auer et al., balance exploration and exploitation through upper confidence bounds on action rewards. Under stochastic reward assumptions, UCB1 achieves logarithmic regret bounds.

Contextual bandits extend this framework by incorporating side information (context) to enable personalized decision-making. LinUCB, proposed by Li et al. for news article recommendation, maintains linear models of rewards as functions of context features and uses ridge regression with confidence ellipsoids for exploration. The algorithm achieves problem-dependent logarithmic regret bounds when rewards are linear in features. Variants include disjoint LinUCB (separate models per arm) and hybrid LinUCB (shared and arm-specific features).

C. Bandits in Financial Applications

Recent work has begun exploring bandit algorithms for financial applications including algorithmic trading, portfolio optimization, and dynamic pricing. However, most studies use synthetic data or focus on single-asset trading rather than multi-asset recommendation. The application of contextual bandits to realistic financial recommendation scenarios with real customer transaction data remains an open area. This paper addresses this gap by providing the first systematic evaluation of bandit algorithms on the FAR-Trans benchmark.

D. Evaluation Metrics in Recommendation Systems

Traditional recommender systems employ information retrieval metrics such as precision, recall, NDCG, and mean average precision to evaluate ranking quality. Financial recommendation systems additionally use domain-specific metrics including portfolio returns, Sharpe ratios, and profitability measures. The FAR-Trans benchmark evaluates algorithms using metrics such as nDCG@k for ranking quality and investment return metrics for financial performance.

However, these metrics have limitations for sequential recommendation scenarios. Static accuracy metrics evaluate recommendations in isolation without considering the temporal learning process, while return-based metrics do not explicitly quantify the cost of exploration needed to discover optimal recommendations. Cumulative regret, borrowed from the bandit literature, provides a complementary evaluation framework that explicitly measures the opportunity cost of suboptimal recommendations during the learning phase. This metric is particularly relevant for financial advisory where the system must balance exploring new assets to improve future recommendations against exploiting current knowledge to maximize immediate returns.

This paper introduces regret-based evaluation to the FAR domain, demonstrating how this metric reveals algorithm learning dynamics and provides insights into the exploration-exploitation trade-off that are not captured by traditional metrics.

E. Non-Stationarity and Adversarial Bandits

Financial markets exhibit non-stationarity where reward distributions change over time due to economic cycles, market regimes, and exogenous shocks. While classical bandit algorithms assume stationary rewards, extensions such as sliding-window UCB, discounted UCB, and adversarial bandits (e.g., EXP3) have been developed for non-stationary environments. Understanding how standard algorithms perform under real-world financial non-stationarity is critical for practical deployment.

III. METHODOLOGY

A. The FAR-Trans Dataset

B. Problem Formulation

C. Algorithms

D. Experimental Setup

IV. RESULTS

V. DISCUSSION

VI. CONCLUSION

ACKNOWLEDGMENT

We thank Sanz-Cruzado et al. for making the FAR-Trans dataset publicly available, enabling reproducible research in financial asset recommendation.

REFERENCES

[1] J. Sanz-Cruzado, N. Droukas, and R. McCreadie, "FAR-Trans: An investment dataset for financial asset recommendation," arXiv preprint arXiv:2407.08692, 2024.

[2] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," Machine Learning, vol. 47, no. 2-3, pp. 235-256, 2002.

[3] L. Li, W. Chu, J. Langford, and R. E. Schapire, "A contextual-bandit approach to personalized news article recommendation," in Proc. 19th International Conference on World Wide Web, 2010, pp. 661-670.

[4] W. Chu, L. Li, L. Reyzin, and R. Schapire, "Contextual bandits with linear payoff functions," in Proc. Fourteenth International Conference on Artificial Intelligence and Statistics, 2011, pp. 208-216.

[5] H. Markowitz, "Portfolio selection," The Journal of Finance, vol. 7, no. 1, pp. 77-91, 1952.

[6] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA: MIT Press, 2018.

[7] A. Slivkins, "Introduction to multi-armed bandits," Foundations and Trends in Machine Learning, vol. 12, no. 1-2, pp. 1-286, 2019.

[8] S. Bubeck and N. Cesa-Bianchi, "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," Foundations and Trends in Machine Learning, vol. 5, no. 1, pp. 1-122, 2012.

[9] O. Chapelle and L. Li, "An empirical evaluation of Thompson sampling," in Advances in Neural Information Processing Systems, 2011, pp. 2249-2257.

[10] A. Garivier and E. Moulines, "On upper-confidence bound policies for switching bandit problems," in Proc. International Conference on Algorithmic Learning Theory, 2011, pp. 174-188.

[11] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "The nonstochastic multiarmed bandit problem," SIAM Journal on Computing, vol. 32, no. 1, pp. 48-77, 2002.

[12] R. McCreadie, J. Sanz-Cruzado, and C. Macdonald, "Galago Meets RankSys: Reproducible financial recommendation," in Proc. ACM RecSys Workshop on Recommender Systems in Finance, 2022.

[13] D. Zibriczky, "Recommender systems meet finance: A literature review," in Proc. 2nd Workshop on Personalization & Recommender Systems in Financial Services, 2016.

[14] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," Computer, vol. 42, no. 8, pp. 30-37, 2009.

[15] J. Ramos and A. Ledezma, "An experimental comparison of bandit algorithms for recommender systems," Expert Systems with Applications, vol. 193, p. 116443, 2022.