Introduction to Bandits in Recommender Systems

Andrea Barraza-Urbina
Insight Centre for Data Analytics, Data Science Institute, NUI Galway
Galway, Ireland
andrea.barraza@insight-centre.org

Dorota Glowacka
Department of Computer Science, University of Helsinki
Helsinki, Finland
glowacka@cs.helsinki.fi

ABSTRACT

The multi-armed bandit problem models an agent that simultaneously attempts to acquire new knowledge (exploration) and optimize his decisions based on existing knowledge (exploitation). The agent attempts to balance these competing tasks in order to maximize his total value over the period of time considered. There are many practical applications of bandit algorithms, including clinical trials, adaptive routing or portfolio design. Over the last decade there has been an increased interest in developing bandit algorithms to address specific issues in recommender systems, such as improved product recommendation, the cold start problem, or personalization. The aim of this tutorial is to provide a brief introduction to the bandit problem with an overview of the various applications of bandit algorithms in recommendation.

CCS CONCEPTS

• **Information systems → Personalization**; *Information retrieval diversity*; • **Computing methodologies → Sequential decision making**; **Online learning settings**; • **Mathematics of computing →** *Probability and statistics.*

KEYWORDS

Recommender Systems, Recommendation Systems, Bandit algorithms, Multi-Armed Bandits, Exploration-Exploitation Trade-off, Reinforcement Learning, Information Retrieval, Interactive Search, Evaluation Framework.

ACM Reference Format:
Andrea Barraza-Urbina and Dorota Glowacka. 2020. Introduction to Bandits in Recommender Systems. In *Fourteenth ACM Conference on Recommender Systems (RecSys '20), September 22–26, 2020, Virtual Event, Brazil.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3383313.3411547

1 INTRODUCTION

The multi-armed bandit (MAB) problem models an agent that simultaneously attempts to acquire new knowledge (exploration) and optimize his decisions based on existing knowledge (exploitation). The agent attempts to balance these competing tasks in order to maximize his total value over the period of time considered. There are many practical applications of the bandit model, such as clinical trials, adaptive routing or portfolio design. Over the last decade there has been an increased interest in developing bandit algorithms for specific problems in recommender systems (RS), such as news and ad recommendation, the cold start problem in recommendation, personalization, collaborative filtering with bandits, or combining social networks with bandits to improve product recommendation.

The aim of this tutorial is to provide participants with the basic knowledge of the following concepts: (a) the exploration-exploitation dilemma and its connection to learning through interaction; (b) framing of the RS problem as an interactive sequential decision–making task that needs to balance exploration and exploitation; (c) basic fundamentals behind bandit approaches that address the exploration-exploitation dilemma; and (d) a general picture of the state-of-the-art of bandit-based RS. With this tutorial we hope to enable participants to start working on bandit-based RS and to provide a framework that would empower them to develop more advanced approaches. This tutorial follows a series of tutorials on similar topics that took place in recent years [17, 19].

This introductory tutorial is aimed at an audience with background in computer science, information retrieval or RS who have a general interest in the application of machine learning techniques in RS. The prerequisite knowledge is basic familiarity with machine learning and basic knowledge of statistics and probability theory. The tutorial will provide practical examples based on Python code and Jupyter Notebooks.

2 OVERVIEW OF THE TUTORIAL

The tutorial is divided into three sections focused on: (1) general motivation and introduction to classic bandit approaches; (2) hands-on session where a simple synthetic recommendation task representing a bandit problem with linear rewards will be used; and (3) overview of a variety of applications of bandit algorithms in recommendation systems summarizing the current state and an outline of challenges applying bandit algorithms in recommendation systems.

The following sections describe in more detail the topics covered in the tutorial.

2.1 Introduction to Classic Bandit Approaches

This section provides an introduction to the core concepts needed to understand basic/classic bandit approaches. The underlying assumptions and intuitions behind classic approaches serve as an essential foundation to understanding how bandit ideas are applied to RS.

- *Motivation*: Introduce the exploration-exploitation dilemma and its relevance to recommendation systems [6, 18]. Discuss example bandit-based recommendation use cases and real-world applications [1, 20, 29, 30, 37, 41].

RecSys '20, September 22–26, 2020, Virtual Event, Brazil

ACM ISBN 978-1-4503-7583-2/20/09.
https://doi.org/10.1145/3383313.3411547

- *Introduction to Classic Multi-Armed Bandits (MAB)*: Describe the stochastic MAB problem and its assumptions [36]. Introduce classic bandit approaches, including e-greedy [5, 36], Upper Confidence Bound (UCB) [5, 27] and Thompson Sampling [11]. The main goals of this part of the tutorial are to:
  - Introduce the concept of regret and reward.
  - Discuss the impact of using different notions of uncertainty to define a bandit approach, such as unguided/naive exploration (e.g., e-greedy) or guided exploration (e.g., UCB).
  - Highlight enhancements to classic approaches that address the stochastic MAB problem (e.g., e-first [5], e-decreasing [5, 36], UCB variations [5, 27]).
- *Bandits and Reinforcement Learning:* Introduction to Reinforcement Learning (RL) [36] and its connection to bandit algorithms [8]. Introduction to Environments as a representation of the task (in this case, stochastic MAB problem) that is to be solved by the Agent (aka bandit algorithm or recommender system) [7, 36].
- *Bandit Variations:* Highlight different variations of the stochastic MAB problem and why they exist, e.g., multiple-play bandits [33], adversarial bandits [10] and contextual bandits [30].

2.2 Hands-on Session

Participants will have the opportunity to try out in a practical setting different configurations of bandit algorithms and observe results. During the hands-on section, the BEARS framework [8][1] will be used and Jupyter Notebooks will be made available. BEARS is an open-source python framework that aims to provide a clean and well documented code to be used in an academic/research setting to enable reproducible evaluations of bandit-based recommendation systems.

- *Introduction to Contextual Bandits:* Context-free stochastic (classic) bandits cannot fully represent the complexities of the RS problem. An important MAB variation was the introduction of contextual bandits and the intuition of linear rewards with respect to contextual data[2]. The tutorial offers an emphasis on contextual bandits and discusses its application in recommender systems as exemplified by Yahoo News recommendation with LinUCB [30].
- *Hands-on Exercise:*
  - Description of a synthetic recommendation systems environment with linear rewards.
  - Introduction to BEARS [8] and setting up experiments (Environment, Agent, Evaluator and Experiment components). A Jupyter Notebook will be provided for this exercise.
  - Allow participants to engage on their own with the Jupyter Notebook. BEARS will enable participants to experiment with different bandit algorithms (and different parameter values), different experiment configurations (e.g. by changing the number of iterations/horizon and episodes/runs) and different metric set-ups (e.g., reward, cumulative reward, regret).
  - Discussion on any remarks/findings participants had with regards to the exercise. For instance, how setting exploration values too high or too low had an impact on the results.
  - Highlight the importance of sharing all the details of experiments towards reproducible evaluations (e.g., random seeds).
- *Evaluation Challenges:* For the exercise, a synthetic environment was developed. Nevertheless, creating environments for RS (e.g., based on a dataset) to achieve unbiased evaluations is a challenge [7]. The presenters aim to briefly review the importance of topics related to bias when dealing with missing values and review solutions that have been proposed (e.g., rejection sampling [30, 31] and counterfactual reasoning [9, 22]).

2.3 Bandits in Recommender Systems

The goal is to provide an overview of existing representative solutions to show how MAB are used in recommendation system. Throughout the presentation, we will provide explanations based on the general components previously introduced, related to the RL framework and BEARS. In this way, participants will have a general framework to understand the variety of solutions. This part of the tutorial will focus on the application of bandit algorithms in various areas of recommender systems as well as issues related to implementation, scalability, training and dealing with specific types of data (e.g., ads, newspaper articles, multimedia, etc.) [18]. Topics covered include the following application of bandit algorithms in the context of recommender systems:

- The cold start problem [38]
- Social networks and recommender systems [15, 16]
- Collaborative filtering and matrix factorization in recommender systems [21, 32, 40, 42]
- Recommendation with a limited lifespan [28, 39]
- News item recommendation [30, 31]
- Online advertising [11]
- Multimedia recommendation and retrieval: images, music, video [23, 26, 41]
- Ranked Bandits and recommending lists [25, 28, 33]
- Examples of interactive retrieval and recommendation systems based on bandit algorithms [3, 14, 20, 34]
- Personalization and system optimization [2, 4, 12, 13, 24, 35]

Further references to bandit algorithms in and recommender systems can be found in [18].

3 TUTORIAL MATERIALS

All materials, including slides and code, will be available after the tutorial in a public repository[3].

ACKNOWLEDGMENTS

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant

[1]The framework is open-source and can be found at: https://gitlab.insight-centre.org/andbar/bears.

[2]Note that the definition of context within the bandit field is different to the definition of context in the field of recommender systems.

[3]https://gitlab.insight-centre.org/andbar/bears/tree/master/tutorials/RECSYS2020

Number SFI/12/RC/2289_P2, cofunded by the European Regional Development Fund.

REFERENCES

[1] Himan Abdollahpouri and Steve Essinger. 2018. Towards effective exploration/exploitation in sequential music recommendation. *arXiv preprint arXiv:1812.03226* (2018).

[2] Kumaripaba Ahukorala, Alan Medlar, Kalle Ilves, and Dorota Glowacka. 2015. Balancing exploration and exploitation: Empirical parameterization of exploratory search systems. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 1703–1706.

[3] Salvatore Andolina, Khalil Klouche, Jaakko Peltonen, Mohammad Hoque, Tuukka Ruotsalo, Diogo Cabral, Arto Klami, Dorota Głowacka, Patrik Floréen, and Giulio Jacucci. 2015. Intentstreams: smart parallel search streams for branching exploratory search. In *Proceedings of the 20th international conference on intelligent user interfaces*. ACM, 300–305.

[4] Kumaripaba Athukorala, Alan Medlar, Antti Oulasvirta, Giulio Jacucci, and Dorota Glowacka. 2016. Beyond relevance: Adapting exploration/exploitation in information retrieval. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*. ACM, 359–369.

[5] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. 2002. Finite-time analysis of the multiarmed bandit problem. *Machine learning* 47, 2-3 (2002), 235–256.

[6] Andrea Barraza-Urbina. 2017. The exploration-exploitation trade-off in interactive recommender systems. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*. 431–435.

[7] Andrea Barraza-Urbina and Mathieu d'Aquin. 2019. Towards sharing task environments to support reproducible evaluations of interactive recommender systems. *arXiv preprint arXiv:1909.06133* (2019).

[8] Andrea Barraza-Urbina, Georgia Koutrika, Mathieu D'Aquin, and Conor Hayes. 2018. BEARS: Towards an Evaluation Framework for Bandit-based Interactive Recommender Systems. In *Proceedings of the Workshop on Offline Evaluation for Recommender Systems (REVEAL'18). Workshop Programme of the 12th ACM Conference on Recommender Systems (RecSys)*.

[9] Léon Bottou, Jonas Peters, Joaquin Quiñonero-Candela, Denis X Charles, D Max Chickering, Elon Portugaly, Dipankar Ray, Patrice Simard, and Ed Snelson. 2013. Counterfactual reasoning and learning systems: The example of computational advertising. *The Journal of Machine Learning Research* 14, 1 (2013), 3207–3260.

[10] Giuseppe Burtini, Jason Loeppky, and Ramon Lawrence. 2015. A survey of online experiment design with the stochastic multi-armed bandit. *arXiv preprint arXiv:1510.00757* (2015).

[11] Olivier Chapelle and Lihong Li. 2011. An empirical evaluation of thompson sampling. In *Advances in neural information processing systems*. 2249–2257.

[12] Pedram Daee, Joel Pyykkö, Dorota Glowacka, and Samuel Kaski. 2016. Interactive intent modeling from multiple feedback domains. In *Proceedings of the 21st international conference on intelligent user interfaces*. ACM, 71–75.

[13] Louis Dorard, Dorota Glowacka, and John Shawe-Taylor. 2009. Gaussian process modelling of dependencies in multi-armed bandit problems. In *Int. Symp. Op. Res.* 77–84.

[14] Yuan Gao, Kalle Ilves, and Dorota Głowacka. 2015. Officehours: A system for student supervisor matching through reinforcement learning. In *Proceedings of the 20th International Conference on Intelligent User Interfaces Companion*. ACM, 29–32.

[15] Claudio Gentile, Shuai Li, Purushottam Kar, Alexandros Karatzoglou, Giovanni Zappella, and Evans Etrue. 2017. On context-dependent clustering of bandits. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 1253–1262.

[16] Claudio Gentile, Shuai Li, and Giovanni Zappella. 2014. Online clustering of bandits. In *International Conference on Machine Learning*. 757–765.

[17] Dorota Glowacka. 2017. Bandit algorithms in interactive information retrieval. In *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*. 327–328.

[18] Dorota Głowacka. 2019. *Bandit algorithms in information retrieval.* Now Publishers.

[19] Dorota Glowacka. 2019. Bandit algorithms in recommender systems. In *Proceedings of the 13th ACM Conference on Recommender Systems*. 574–575.

[20] Dorota Glowacka, Tuukka Ruotsalo, Ksenia Konuyshkova, Samuel Kaski, Giulio Jacucci, et al. 2013. Directing exploratory search: Reinforcement learning from user interactions with keywords. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. ACM, 117–128.

[21] Frédéric Guillou, Romaric Gaudel, and Philippe Preux. 2016. Scalable explore-exploit collaborative filtering. In *Pacific Asia Conference On Information Systems (PACIS)*. Association For Information System.

[22] Katja Hofmann, Lihong Li, and Filip Radlinski. 2016. Online evaluation for information retrieval. *Foundations and trends in information retrieval* 10, 1 (2016), 1–117.

[23] Sayantan Hore, Lasse Tyrvainen, Joel Pyykko, and Dorota Glowacka. 2015. A reinforcement learning approach to query-less image retrieval. In *International Workshop on Symbiotic Interaction*. Springer, 121–126.

[24] Antti Kangasrääsiö, Dorota Glowacka, and Samuel Kaski. 2015. Improving controllability and predictability of interactive recommendation interfaces for exploratory search. In *Proceedings of the 20th international conference on intelligent user interfaces*. ACM, 247–251.

[25] Pushmeet Kohli, Mahyar Salek, and Greg Stoddard. 2013. A fast bandit algorithm for recommendation to users with heterogenous tastes. In *Twenty-Seventh AAAI Conference on Artificial Intelligence*.

[26] Ksenia Konyushkova and Dorota Glowacka. 2013. Content-based image retrieval with hierarchical Gaussian Process bandits with self-organizing maps.. In *ESANN*.

[27] Volodymyr Kuleshov and Doina Precup. 2014. Algorithms for multi-armed bandit problems. *arXiv preprint arXiv:1402.6028* (2014).

[28] Anisio Lacerda. 2017. Multi-objective ranked bandits for recommender systems. *Neurocomputing* 246 (2017), 12–24.

[29] Anisio Lacerda, Rodrygo LT Santos, Adriano Veloso, and Nivio Ziviani. 2015. Improving daily deals recommendation using explore-then-exploit strategies. *Information Retrieval Journal* 18, 2 (2015), 95–122.

[30] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. 2010. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*. 661–670.

[31] Lihong Li, Wei Chu, John Langford, and Xuanhui Wang. 2011. Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms. In *Proceedings of the fourth ACM international conference on Web search and data mining*. 297–306.

[32] Shuai Li, Alexandros Karatzoglou, and Claudio Gentile. 2016. Collaborative filtering bandits. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 539–548.

[33] Jonathan Louëdec, Max Chevalier, Josiane Mothe, Aurélien Garivier, and Sébastien Gerchinovitz. 2015. A multiple-play bandit algorithm applied to recommender systems. In *The Twenty-Eighth International Flairs Conference*.

[34] Alan Medlar, Kalle Ilves, Ping Wang, Wray Buntine, and Dorota Glowacka. 2016. Pulp: A system for exploratory search of scientific literature. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 1133–1136.

[35] Alan Medlar, Joel Pyykkö, and Dorota Glowacka. 2017. Towards fine-grained adaptation of exploration/exploitation in information retrieval. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*. ACM, 623–627.

[36] R. Sutton and Barto A. 2018. *Reinforcement learning: An introduction.* MIT press.

[37] Choon Hui Teo, Houssam Nassif, Daniel Hill, Sriram Srinivasan, Mitchell Goodman, Vijai Mohan, and SVN Vishwanathan. 2016. Adaptive, personalized diversity for visual discovery. In *Proceedings of the 10th ACM conference on recommender systems*. 35–38.

[38] Hastagiri P Vanchinathan, Isidor Nikolic, Fabio De Bona, and Andreas Krause. 2014. Explore-exploit in top-n recommender systems via gaussian processes. In *Proceedings of the 8th ACM Conference on Recommender systems*. 225–232.

[39] Huazheng Wang, Qingyun Wu, and Hongning Wang. 2016. Learning hidden features for contextual bandits. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. 1633–1642.

[40] Huazheng Wang, Qingyun Wu, and Hongning Wang. 2017. Factorization bandits for interactive recommendation. In *Thirty-First AAAI Conference on Artificial Intelligence*.

[41] Xinxi Wang, Yi Wang, David Hsu, and Ye Wang. 2014. Exploration in interactive personalized music recommendation: a reinforcement learning approach. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 11, 1 (2014), 1–22.

[42] Qingyun Wu, Huazheng Wang, Quanquan Gu, and Hongning Wang. 2016. Contextual bandits in a collaborative environment. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 529–538.