Bandit Algorithms in Recommender Systems

Dorota Glowacka
Department of Computer Science
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 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, 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 an overview of the various applications of bandit algorithms in recommendation as well as issues related to their practical deployment and performance in real-life systems/applications.

1 TUTORIAL LENGTH AND TARGETED AUDIENCE

This introductory 90-minute tutorial is aimed at an audience with some background in computer science, information retrieval or recommender system who have a general interest in the application of machine learning techniques in recommender systems. The tutorial would ideally suit first or second year PhD students. The prerequisite knowledge is basic familiarity with machine learning, basic knowledge in statistics and probability theory.

2 THE ORGANIZER

Dorota Glowacka is an Assistant Professor in Machine Learning and Artificial Intelligence in the Department of Computer Science, University of Helsinki. Prior to that she was an Assistant Professor in Machine Learning in the School of Informatics, University of Edinburgh. In 2012, she completed a PhD in Machine Learning focusing on reinforcement learning and interactive systems in the Department of Computer Science, University College London (UCL). Her research interests are in the area of user modelling, interactive systems, interactive information retrieval, and reinforcement learning. She regularly publishes in top tier IR and HCI venues, such as CIKM, SIGIR and IUI, where she obtained the best paper award in 2013.

Over the past seven years, she co-organised a number of workshops related to interactive search and personalisation co-located with IUI, NIPS and ICML. Over the past three years she also designed and delivered two graduate courses at the University of Helsinki: Reinforcement Learning and its Applications and Reinforcement Learning in Information Retrieval. Both courses are related to the proposed tutorial. In 2017, she delivered a half-day tutorial entitled "Bandit algorithms in Interactive Information Retrieval" during the International Conference on the Theory of Information Retrieval (ICTIR). In May 2019, her book entitled "Bandit Algorithms in Information retrieval" will be published [?]. A substantial part of the book is devoted to online recommendation with bandits and the tutorial will be largely based on these parts of the book.

3 MOTIVATION AND RELEVANCE TO THE RECSYS COMMUNITY

With ever increasing amount and type of data available on the web, search engines and recommender systems have gradually developed into complex systems that combine many criteria with the aim of producing the optimal recommendation list in response to users' queries. Traditional approaches, where a recommender systems are trained off-line face a number of issues. For example, manually annotated data used for training can be very expensive due to the involvement of human experts. Further, users' preferences and product availability frequently change without sufficient time to allow the recommender system to be trained off-line in a timely fashion. Lastly, it may not always be possible for experts to annotate data as might be the case in personalised search and recommendation, where the perceived preference for a given item tend to be very subjective and cannot be captured by off-line product annotation.

A response to these issues has been emergence of new online approaches to user modelling and recommendation. These approaches often employ various reinforcement learning methods, of which bandit algorithms are the most popular. The 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. This aspect of bandit algorithms allow the recommender system to gradually build the user model without getting stuck in a local search space, while engaging the user in the search loop. The simplicity and ease of implementation of bandit algorithms also adds to their recent popularity in both the information retrieval and recommender system communities.

The goal of this tutorial is to bring together current efforts in the area, summarize the research performed so far and give a holistic view on the challenges of applying bandit algorithms in the information retrieval domain. After the tutorial, the attendees should be familiar with the following concepts: basic theory behind bandits, exploration-exploitation trade-off, types of bandit algorithms, what

types of recommender problems bandit algorithms are best suited to, what type of data different types of bandits are best suited to, scalability issues, deployment of bandits in interactive systems.

A website will be created for the workshop, where the slides used during the tutorial will be place. The website will also contain links to all the relevant literature referenced during the tutorial. The attendees will be given a course-pack that will include a reference list, an annotated bibliography of seminal works in the field.

4 FORMAT AND DETAILED SCHEDULE

The tutorial will consist of three parts: 1) overview of bandit algorithms; 2) application of bandits in recommender systems; 3) Optimization of interactive recommender/information retrieval systems based on bandit algorithms. The first part will provide an overview of bandit algorithms and how they develop overtime starting from Gittins indices and then gradually follow the algorithmic development of bandits: Upper Confidence Bound (UCB) algorithms, multi-armed bandits, dependent arm bandits, contextual bandits, dueling bandits, collaborative bandits, etc. The basic aspects of bandit algorithms, such as the reward function and the exploration-exploitation trade-off will be introduced as well. The first part of the tutorial aims to not only familiarize the audience with the mathematical and statistical foundations of of bandits but also provide them with an intuition how they can be applied to real-life problems through examples from various areas, such as clinical trials, economics or information retrieval. Topics covered in the first part of the tutorial will include: Gittins indices [4], UCB [2], multi-armed bandits [3], dependent arm bandits [7], contextual bandits [6], exploration-exploitation trade-off [2].

The second part of the tutorial will focus on the application of bandit algorithms in various areas of recommender systems with an emphasis on the limitations of different types of bandits in various applications as well as issues related to implementation, scalability, training and dealing with specific types of data (ads, newspaper articles, multimedia, etc.). This part of the tutorial will be largely based on selected chapters (with references within) from my recent book that focus on recommendation with bandit algorithms [?]. Topics covered in the second part of the tutorial:

- the cold start problem
- social networks and recommender systems
- collaborative filtering and recommender systems
- recommendation with a limited lifespan
- feature learning
- news item recommendation
- online advertising
- multimedia recommendation and retrieval: images, music, video [5, 10?]

In the third part of the tutorial I will present example information retrieval systems based on bandits and discuss how such systems can be optimized and personalized through user involvement. This part will be largely based on our recent research into this area. I will discuss user study methods that we specifically developed to optimize and test recommender/retrieval systems based on bandit algorithms. Topics covered in the third part of the tutorial:

- examples of information retrieval and recommender systems based on bandits [8? ? ? ? ?]

- issues related to deploying bandits in online systems [9?]
- personalization [1?]
- system optimization [? ?]

REFERENCES

[1] 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.

[2] P. Auer. 2002. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research* 3, Nov (2002), 397–422.

[3] P. Auer, N. Cesa-Bianchi, and P. Fischer. 2002. Finite-time analysis of the multi-armed bandit problem. *Machine learning* 47, 2-3 (2002), 235–256.

[4] John C Gittins. 1979. Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society. Series B (Methodological)* (1979), 148–177.

[5] K. Konyushkova and D. Glowacka. 2013. Content-based image retrieval with hierarchical Gaussian Process bandits with self-organizing maps. In *ESANN*.

[6] L. Li, W. Chu, J. Langford, and R. 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.

[7] S. Pandey, D. Chakrabarti, and D. Agarwal. 2007. Multi-armed bandit problems with dependent arms. In *Proceedings of the 24th international conference on Machine learning*. 721–728.

[8] T. Ruotsalo, J. Peltonen, M. Eugster, D. Glowacka, K. Konyushkova, K. Athukorala, I. Kosunen, A. Reijonen, P. Myllymäki, G. Jacucci, and S. Kaski. 2013. Directing Exploratory Search with Interactive Intent Modeling. In *Proceedings of the 22Nd ACM International Conference on Information & Knowledge Management*.

[9] A. Vorobev, D. Lefortier, G. Gusev, and P. Serdyukov. 2015. Gathering additional feedback on search results by multi-armed bandits with respect to production ranking. In *Proceedings of the 24th international conference on World wide web*.

[10] X. Wang, Y. Wang, D. Hsu, and Y. Wang. 2014. Exploration in interactive personalized music recommendation: a reinforcement learning approach. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 11, 1 (2014), 7.