Bandit Algorithms in Information Retrieval

Other titles in Foundations and Trends® in Information Retrieval

Web Forum Retrieval and Text Analytics: A Survey
Doris Hoogeveen, Li Wang, Timothy Baldwin and Karin M. Verspoor
ISBN: 978-1-68083-350-8

Display Advertising with Real-Time Bidding (RTB) and Behavioural Targeting
Jun Wang, Weinan Zhang and Shuai Yuan
ISBN: 978-1-68083-310-2

Applications of Topic Models
Jordan Boyd-Graber, Yuening Hu and David Mimno
ISBN: 978-1-68083-308-9

Searching the Enterprise
Udo Kruschwitz and Charlie Hull
ISBN: 978-1-68083-304-1

Bandit Algorithms in Information Retrieval

Dorota Głowacka
University of Helsinki
glowacka@cs.helsinki.fi

the essence of knowledge
Boston — Delft

Foundations and Trends® in Information Retrieval

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

D Głowacka. *Bandit Algorithms in Information Retrieval*. Foundations and Trends® in Information Retrieval, vol. 13, no. 4, pp. 299–424, 2019.

ISBN: 978-1-68083-575-5

Foundations and Trends® in Information Retrieval

Volume 13, Issue 4, 2019

Editorial Scope

Topics

Foundations and Trends® in Information Retrieval publishes survey and tutorial articles in the following topics:

- Applications of IR
- Architectures for IR
- Collaborative filtering and recommender systems
- Cross-lingual and multilingual IR
- Distributed IR and federated search
- Evaluation issues and test collections for IR
- Formal models and language models for IR
- IR on mobile platforms
- Indexing and retrieval of structured documents
- Information categorization and clustering
- Information extraction
- Information filtering and routing
- Metasearch, rank aggregation and data fusion
- Natural language processing for IR
- Performance issues for IR systems, including algorithms, data structures, optimization techniques, and scalability
- Question answering
- Summarization of single documents, multiple documents, and corpora
- Text mining
- Topic detection and tracking
- Usability, interactivity, and visualization issues in IR
- User modelling and user studies for IR
- Web search

Information for Librarians

Foundations and Trends® in Information Retrieval, 2019, Volume 13, 5 issues. ISSN paper version 1554-0669. ISSN online version 1554-0677. Also available as a combined paper and online subscription.

Contents

Bandit Algorithms in Information Retrieval

Dorota Głowacka

University of Helsinki; glowacka@cs.helsinki.fi

ABSTRACT

Bandit algorithms, named after casino slot machines sometimes known as "one-armed bandits", fall into a broad category of stochastic scheduling problems. In the setting with multiple arms, each arm generates a reward with a given probability. The gambler's aim is to find the arm producing the highest payoff and then continue playing in order to accumulate the maximum reward possible. However, having only a limited number of plays, the gambler is faced with a dilemma: should he play the arm currently known to produce the highest reward or should he keep on trying other arms in the hope of finding a better paying one? This problem formulation is easily applicable to many real-life scenarios, hence in recent years there has been an increased interest in developing bandit algorithms for a range of applications. In information retrieval and recommender systems, bandit algorithms, which are simple to implement and do not require any training data, have been particularly popular in online personalization, online ranker evaluation and search engine optimization. This survey provides a brief overview of bandit algorithms designed to tackle specific issues in information retrieval and recommendation and, where applicable, it describes how they were applied in practice.

Dorota Głowacka (2019), "Bandit Algorithms in Information Retrieval", Foundations and Trends® in Information Retrieval: Vol. 13, No. 4, pp 299–424. DOI: 10.1561/1500000067.

1

Introduction

Over the last decade there has been an increased interest in application of bandit algorithms in information retrieval (IR) and recommender systems. The aim of this survey is to provide an overview of bandit algorithms inspired by various aspects of IR, such as click models, online ranker evaluation, personalization or the cold-start problem. Each section of the survey focuses on a specific IR problem and aims to explain how it was addressed with various bandit approaches. Within each section, all the algorithms are presented in chronological order. The goal is to show how specific concepts related to bandit algorithms, e.g. graph clustering with bandits, or a specific family of bandit algorithms, e.g. dueling bandits developed over time. Gathering all this information in one place allows us to explain the impact of IR on the development of new bandit algorithms as well as the impact of bandit algorithms on the development of new methods in IR. The survey covers papers published up to the end of 2017.

Why Bandits?

Bandit algorithms derive their name from casino slot machines, sometimes referred to as one-armed bandits. In this scenario, a gambler is faced with a row of such machines. The gambler has to make a number of decisions, such as which machines to play or how many times to play each machine. The problem is that each machine provides a random reward from a probability distribution specific to that machine. The gambler aims to maximize the sum of the rewards by playing different machines. Thus, the gambler needs to make a trade-off between exploiting the machine with the highest expected payoff so far and exploring other machines to get more information about their expected payoffs.

In the 1950's Herbert Robbins realized the importance of the problem and constructed convergent population selection strategies for sequential design of experiments (Robbins, 1985). A couple of decades later John Gittens constructed a theorem, called the Gittins index, that gave an optimal policy for maximizing the expected discounted reward (Gittins, 1979). Later on, some approximate solutions based on *epsilon* strategies (Sutton and Barto, 1998) as well as Bayesian methods, such as Thompson sampling (Thompson, 1933), were developed to solve the bandit problem. The last two decades has seen an immense interest in the study of bandit algorithms, starting with the development of Upper Confidence Bound (UCB) (Agrawal, 1995) strategies. In UCB algorithms, however, every bandit arm is independent and does not pass any information about its payoff generating distribution to other bandit arms. This led to the development of linear and contextual bandits (Auer, 2002; Li *et al.*, 2010b), where a linear dependency between the expected payoff of an arm and its context is assumed. In comparison to independent bandit strategies, linear bandits can lead to elimination of arms with low payoff earlier during the exploration phase thus allowing the player to focus on trying arms with a potentially higher payoff.

There are a number of reasons why bandit algorithms have gained a high level of popularity in many applications. They are quick and easy to implement, they do not require any training data, and they allow for continuous testing/learning, which makes them highly applicable to any online application with a continuous stream of data. Thus, over the years bandits have been applied in many areas: clinical trials (Villar *et al.*, 2015; Williamson *et al.*, 2017), adaptive routing (Awerbuch and Kleinberg, 2008), auctions (Nazerzadeh *et al.*, 2016), financial portfolio design (Shen *et al.*, 2015), cognitive modelling (Głowacka *et al.*, 2009), games (Kocsis and Szepesvári, 2006), and, as this survey shows, in information retrieval.

Organization of the Survey

The survey is organized as follows. Chapter 2 introduces bandit algorithms and gives a brief overview of four broad classes of bandit algorithms: *epsilon* strategies, independent arms bandits based on upper confidence bound, linear bandits with dependent arms, and Thompson sampling. These broad categories of bandit strategies form the basis of more specialized algorithms discussed in the remaining chapters. Other types of bandit algorithms with specific applications are introduced in relevant chapters rather than being briefly introduced in Chapter 2. Chapter 3 summarizes bandit algorithms inspired by three click models: the *Cascade Model* (Section 3.1), the *Dependent Click Model* (Section 3.2) and the *Position Based Model* 3.3. The following two chapters discuss bandit based approaches to ranking (Chapter 4) and ranker evaluation (Chapter 5). Of particular interest to the reader might be Section 4.1, where the first bandit algorithms applied to ranking are described. Chapter 5 focuses mostly on dueling bandits algorithms and their application to ranking. In Chapter 6, various bandit approaches used in recommender systems are described. The chapter talks about personalization (Section 6.1), social network based bandits (Section 6.2), collaborative filtering with bandits (Section 6.3), optimization through feature learning (Section 6.4) and multiple arms evaluation (Section 6.6). Section 6.1.1 talks in more detail about contextual bandits in the context of advertising and recommender systems by introducing some of the classic algorithms in this area, such as LinUCB (Li *et al.*, 2010a). Finally, Chapter 7 briefly touches on other areas of information retrieval where bandits are gradually introduced, such as short text recommendation (Section 7.1), multimedia retrieval (Section 7.2), and web-page layout optimization (Section 7.3). The appendices contain explanations of the abbreviations and mathematical symbols used throughout the survey.

Who is this Survey Intended for?

The survey is primarily intended for two groups of readers: (1) IR researchers interested in bandit algorithms or more broadly in reinforcement learning, and who would like to know how and where bandits algorithms have been applied in IR; (2) machine learning researchers generally interested in practical applications of machine learning techniques and challenges posed by such practical applications; (3) data scientists interested in algorithmic solutions to issues regularly encountered in information retrieval and recommender systems.

The survey provides a general overview of the bandits methods discussed and as such it should be accessible to anyone who completed introductory to intermediate level courses in machine learning and/or statistics. The reader is advised to consult specific papers referenced throughout the text to learn more about theoretical analysis or implementation details of specific algorithms. All the chapters are self-contained and can be read in isolation, although references to related concepts in other sections are provided throughout the survey. Each section provides a chronological development of a specific approach or family of algorithms, where most of the later developments build upon or improve earlier findings. Chapter 2 is primarily aimed at readers with little knowledge of reinforcement learning and bandits. Readers not familiar with these topics are encouraged to start with this chapter before proceeding to the rest of the survey.

References

Agarwal, D., B.-C. Chen, P. Elango, N. Motgi, S.-T. Park, R. Ramakrishnan, S. Roy, and J. Zachariah. 2009. "Online models for content optimization". In: *Advances in Neural Information Processing Systems*. 17–24.

Agrawal, R. 1995. "Sample mean based index policies by o (log n) regret for the multi-armed bandit problem". *Advances in Applied Probability*. 27(4): 1054–1078.

Agrawal, S. and N. Goyal. 2012. "Analysis of thompson sampling for the multi-armed bandit problem". In: *Conference on Learning Theory*. 39–1.

Ailon, N., Z. Karnin, and T. Joachims. 2014. "Reducing dueling bandits to cardinal bandits". In: *International Conference on Machine Learning*. 856–864.

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

Auer, P., N. Cesa-Bianchi, and P. Fischer. 2002a. "Finite-time analysis of the multiarmed bandit problem". *Machine Learning*. 47(2-3): 235–256.

Auer, P., N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. 2002b. "The nonstochastic multiarmed bandit problem". *SIAM Journal on Computing*. 32(1): 48–77.

Auer, P., Z. Hussain, S. Kaski, A. Klami, J. Kujala, J. Laaksonen, A. P. Leung, K. Pasupa, and J. Shawe-Taylor. 2010. "Pinview: Implicit Feedback in Content-Based Image Retrieval". In: *WAPA*. 51–57.

Awerbuch, B. and R. Kleinberg. 2008. "Online linear optimization and adaptive routing". *Journal of Computer and System Sciences*. 74(1): 97–114.

Bouneffouf, D., A. Bouzeghoub, and A. L. Gançarski. 2012. "A contextual-bandit algorithm for mobile context-aware recommender system". In: *International Conference on Neural Information Processing*. Springer. 324–331.

Bresler, G., G. H. Chen, and D. Shah. 2014. "A latent source model for online collaborative filtering". In: *Advances in Neural Information Processing Systems*. 3347–3355.

Brost, B., I. J. Cox, Y. Seldin, and C. Lioma. 2016a. "An improved multileaving algorithm for online ranker evaluation". In: *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM. 745–748.

Brost, B., Y. Seldin, I. J. Cox, and C. Lioma. 2016b. "Multi-dueling bandits and their application to online ranker evaluation". In: *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM. 2161–2166.

Busa-Fekete, R., E. Hüllermeier, and A. E. Mesaoudi-Paul. 2018. "Preference-based online learning with dueling bandits: A survey". *arXiv preprint arXiv:1807.11398*.

Busa-Fekete, R., B. Szörényi, and E. Hüllermeier. 2014. "PAC Rank Elicitation through Adaptive Sampling of Stochastic Pairwise Preferences". In: *AAAI*. 1701–1707.

Cai, F. and M. De Rijke. 2016. "A survey of query auto completion in information retrieval". *Foundations and Trends® in Information Retrieval*. 10(4): 273–363.

Caron, S. and S. Bhagat. 2013. "Mixing Bandits: A Recipe for Improved Cold-start Recommendations in a Social Network". In: *Proceedings of the 7th Workshop on Social Network Mining and Analysis*. *SNAKDD '13*. Chicago, Illinois: ACM. 11:1–11:9. DOI: 10.1145/2501025.2501029.

Cesa-Bianchi, N., C. Gentile, and G. Zappella. 2013. "A gang of bandits". In: *Advances in Neural Information Processing Systems*. 737–745.

Chakrabarti, D., R. Kumar, F. Radlinski, and E. Upfal. 2009. "Mortal multi-armed bandits". In: *Advances in Neural Information Processing Systems*. 273–280.

Chapelle, O. and L. Li. 2011. "An empirical evaluation of thompson sampling". In: *Advances in Neural Information Processing Systems*. 2249–2257.

Chen, L., A. Krause, and A. Karbasi. 2017. "Interactive Submodular Bandit". In: *Advances in Neural Information Processing Systems*. 140–151.

Chuklin, A., I. Markov, and M. d. Rijke. 2015. "Click models for web search". *Synthesis Lectures on Information Concepts, Retrieval, and Services*. 7(3): 1–115.

Combes, R., S. Magureanu, A. Proutiere, and C. Laroche. 2015. "Learning to Rank: Regret Lower Bounds and Efficient Algorithms". *SIGMETRICS Perform. Eval. Rev.* 43(1): 231–244. DOI: 10.1145/2796314.2745852.

Cormack, G. V., C. R. Palmer, and C. L. Clarke. 1998. "Efficient construction of large test collections". In: *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM. 282–289.

Craswell, N., O. Zoeter, M. Taylor, and B. Ramsey. 2008. "An experimental comparison of click position-bias models". In: *Proceedings of the 2008 international conference on web search and data mining*. ACM. 87–94.

Daee, P., J. Pyykkö, D. Glowacka, and S. Kaski. 2016. "Interactive Intent Modeling from Multiple Feedback Domains". In: *Proceedings of the 21st International Conference on Intelligent User Interfaces. IUI '16*. Sonoma, California, USA: ACM. 71–75. DOI: 10.1145/2856767.2856803.

Deshpande, Y. and A. Montanari. 2012. "Linear bandits in high dimension and recommendation systems". In: *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*. IEEE. 1750–1754.

Dorard, L., D. Glowacka, and J. Shawe-Taylor. 2009. "Gaussian process modelling of dependencies in multi-armed bandit problems". In: *Int. Symp. Op. Res.* 77–84.

Dudık, M., K. Hofmann, R. E. Schapire, A. Slivkins, and M. Zoghi. 2015. "Contextual dueling bandits". In: *Proceedings of The 28th Conference on Learning Theory*. 563–587.

Durand, A., J.-A. Beaumont, C. Gagné, M. Lemay, and S. Paquet. 2017. "Query Completion Using Bandits for Engines Aggregation". *arXiv preprint arXiv:1709.04095*.

Garivier, A. and O. Cappé. 2011. "The KL-UCB algorithm for bounded stochastic bandits and beyond". In: *Proceedings of the 24th annual Conference On Learning Theory*. 359–376.

Garivier, A. and E. Moulines. 2011. "On upper-confidence bound policies for switching bandit problems". In: *International Conference on Algorithmic Learning Theory*. Springer. 174–188.

Gentile, C., S. Li, P. Kar, A. Karatzoglou, G. Zappella, and E. Etrue. 2017. "On context-dependent clustering of bandits". In: *International Conference on Machine Learning*. 1253–1262.

Gentile, C., S. Li, and G. Zappella. 2014. "Online Clustering of Bandits". In: *ICML*. 757–765.

Gittins, J. C. 1979. "Bandit processes and dynamic allocation indices". *Journal of the Royal Statistical Society. Series B (Methodological)*: 148–177.

Głowacka, D., L. Dorard, A. Medlar, and J. Shawe-Taylor. 2009. "Prior Knowledge in Learning Finite Parameter Spaces". In: *International Conference on Formal Grammar*. Springer. 199–213.

Guo, F., C. Liu, and Y. M. Wang. 2009. "Efficient multiple-click models in web search". In: *Proceedings of the second ACM international conference on web search and data mining*. ACM. 124–131.

He, J., C. Zhai, and X. Li. 2009. "Evaluation of methods for relative comparison of retrieval systems based on clickthroughs". In: *Proceedings of the 18th ACM conference on Information and knowledge management*. ACM. 2029–2032.

Hill, D. N., H. Nassif, Y. Liu, A. Iyer, and S. Vishwanathan. 2017. "An Efficient Bandit Algorithm for Realtime Multivariate Optimization". In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '17*. Halifax, NS, Canada: ACM. 1813–1821. DOI: 10.1145/3097983.3098184.

Hofmann, K., S. Whiteson, and M. De Rijke. 2011a. "A probabilistic method for inferring preferences from clicks". In: *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM. 249–258.

Hofmann, K., S. Whiteson, and M. D. Rijke. 2013a. "Fidelity, soundness, and efficiency of interleaved comparison methods". *ACM Transactions on Information Systems (TOIS)*. 31(4): 17.

Hofmann, K., S. Whiteson, and M. de Rijke. 2011b. "Balancing exploration and exploitation in learning to rank online". In: *European Conference on Information Retrieval*. Springer. 251–263.

Hofmann, K., S. Whiteson, and M. de Rijke. 2013b. "Balancing exploration and exploitation in listwise and pairwise online learning to rank for information retrieval". *Information Retrieval*. 16(1): 63–90.

Hore, S., L. Tyrvainen, J. Pyykko, and D. Glowacka. 2015. "A reinforcement learning approach to query-less image retrieval". In: *International Workshop on Symbiotic Interaction*. Springer. 121–126.

Horvitz, D. G. and D. J. Thompson. 1952. "A generalization of sampling without replacement from a finite universe". *Journal of the American statistical Association*. 47(260): 663–685.

Hsieh, C.-C., J. Neufeld, T. King, and J. Cho. 2015. "Efficient Approximate Thompson Sampling for Search Query Recommendation". In: *Proceedings of the 30th Annual ACM Symposium on Applied Computing. SAC '15*. Salamanca, Spain: ACM. 740–746. DOI: 10.1145/2695664.2695748.

Jie, L., S. Lamkhede, R. Sapra, E. Hsu, H. Song, and Y. Chang. 2013. "A Unified Search Federation System Based on Online User Feedback". In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '13*. Chicago, Illinois, USA: ACM. 1195–1203. DOI: 10.1145/2487575.2488198.

Joachims, T. 2003. "Evaluating Retrieval Performance Using Clickthrough Data". *Text Mining*.

Kale, S., L. Reyzin, and R. E. Schapire. 2010. "Non-stochastic bandit slate problems". In: *Advances in Neural Information Processing Systems*. 1054–1062.

Katariya, S., B. Kveton, C. Szepesvari, C. Vernade, and Z. Wen. 2016a. "Stochastic rank-1 bandits". *arXiv preprint arXiv:1608.03023*.

Katariya, S., B. Kveton, C. Szepesvári, C. Vernade, and Z. Wen. 2017. "Bernoulli Rank-1 Bandits for Click Feedback". *arXiv preprint arXiv:1703.06513*.

Katariya, S., B. Kveton, C. Szepesvári, and Z. Wen. 2016b. "DCM Bandits: Learning to Rank with Multiple Clicks". In: *Proc. of ICML*.

Kaufmann, E., O. Cappé, and A. Garivier. 2012. "On Bayesian upper confidence bounds for bandit problems". In: *Artificial Intelligence and Statistics*. 592–600.

Kawale, J., H. H. Bui, B. Kveton, L. Tran-Thanh, and S. Chawla. 2015. "Efficient Thompson Sampling for Online Matrix-Factorization Recommendation". In: *Advances in Neural Information Processing Systems*. 1297–1305.

Kleinberg, R., A. Slivkins, and E. Upfal. 2008. "Multi-armed bandits in metric spaces". In: *Proceedings of the fortieth annual ACM symposium on Theory of computing*. ACM. 681–690.

Kocák, T., M. Valko, R. Munos, and S. Agrawal. 2014. "Spectral Thompson Sampling". In: *AAAI*. 1911–1917.

Kocsis, L. and C. Szepesvári. 2006. "Bandit based monte-carlo planning". In: *European conference on machine learning*. Springer. 282–293.

Kohli, P., M. Salek, and G. Stoddard. 2013. "A Fast Bandit Algorithm for Recommendation to Users With Heterogenous Tastes". In: *AAAI*.

Kohonen, T. 1998. "The self-organizing map". *Neurocomputing*. 21(1-3): 1–6.

Komiyama, J., J. Honda, H. Kashima, and H. Nakagawa. 2015. "Regret lower bound and optimal algorithm in dueling bandit problem". In: *Proceedings of The 28th Conference on Learning Theory*. 1141–1154.

Komiyama, J., J. Honda, and H. Nakagawa. 2016. "Copeland dueling bandit problem: regret lower bound, optimal algorithm, and computationally efficient algorithm". In: *Proceedings of the 33rd International Conference on International Conference on Machine Learning*. 1235–1244.

Komiyama, J., J. Honda, and A. Takeda. 2017. "Position-based Multiple-play Bandit Problem with Unknown Position Bias". In: *Advances in Neural Information Processing Systems*. 5005–5015.

Komiyama, J. and T. Qin. 2014. "Time-decaying bandits for non-stationary systems". In: *International Conference on Web and Internet Economics*. Springer. 460–466.

Konyushkova, K. and D. Glowacka. 2013. "Content-based image retrieval with hierarchical Gaussian Process bandits with self-organizing maps". In: *ESANN*.

Koren, Y., R. Bell, and C. Volinsky. 2009. "Matrix factorization techniques for recommender systems". *Computer*. 42(8).

Kuleshov, V. and D. Precup. 2014. "Algorithms for multi-armed bandit problems". *arXiv preprint arXiv:1402.6028*.

Kveton, B., C. Szepesvari, Z. Wen, and A. Ashkan. 2015a. "Cascading Bandits: Learning to Rank in the Cascade Model". In: *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*. 767–776.

Kveton, B., Z. Wen, A. Ashkan, and C. Szepesvari. 2015b. "Combinatorial cascading bandits". In: *Advances in Neural Information Processing Systems*. 1450–1458.

Kveton, B., Z. Wen, A. Ashkan, and C. Szepesvari. 2015c. "Tight regret bounds for stochastic combinatorial semi-bandits". In: *Artificial Intelligence and Statistics*. 535–543.

Lacerda, A. 2017. "Multi-Objective Ranked Bandits for Recommender Systems". *Neurocomputing*. 246: 12–24.

Lagrée, P., C. Vernade, and O. Cappe. 2016. "Multiple-play bandits in the position-based model". In: *Advances in Neural Information Processing Systems*. 1597–1605.

Lai, T. L. and H. Robbins. 1985. "Asymptotically efficient adaptive allocation rules". *Advances in Applied Mathematics*. 6(1): 4–22.

Langford, J., A. Strehl, and J. Wortman. 2008. "Exploration scavenging". In: *Proceedings of the 25th international conference on Machine learning*. ACM. 528–535.

Langford, J. and T. Zhang. 2008. "The epoch-greedy algorithm for multi-armed bandits with side information". In: *Advances in Neural Information Processing Systems*. 817–824.

Levine, N., K. Crammer, and S. Mannor. 2017. "Rotting bandits". In: *Advances in Neural Information Processing Systems*. 3077–3086.

Li, C., P. Resnick, and Q. Mei. 2016a. "Multiple Queries As Bandit Arms". In: *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. CIKM '16*. Indianapolis, Indiana, USA: ACM. 1089–1098. DOI: 10.1145/2983323.2983816.

Li, L., W. Chu, J. Langford, and R. E. Schapire. 2010a. "A contextual-bandit approach to personalized news article recommendation". In: *Proceedings of the 19th international conference on World wide web*. 661–670.

Li, L., S. Chen, J. Kleban, and A. Gupta. 2015. "Counterfactual estimation and optimization of click metrics in search engines: A case study". In: *Proceedings of the 24th International Conference on World Wide Web*. ACM. 929–934.

Li, L., W. Chu, J. Langford, T. Moon, and X. Wang. 2011a. "An Unbiased Offline Evaluation of Contextual Bandit Algorithms with Generalized Linear Models". In: *JMLR Workshop and Conference Proceedings vol. 26: On-line Trading of Exploration and Exploitation 2*. 19–36.

Li, L., W. Chu, J. Langford, and X. Wang. 2011b. "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. WSDM '11*. Hong Kong, China: ACM. 297–306. DOI: 10.1145/1935826.1935878.

Li, S., A. Karatzoglou, and C. Gentile. 2016b. "Collaborative Filtering Bandits". In: *The 39th International ACM SIGIR Conference on Information Retrieval (SIGIR)*.

Li, S., B. Wang, S. Zhang, and W. Chen. 2016c. "Contextual combinatorial cascading bandits". In: *International Conference on Machine Learning*. 1245–1253.

Li, W., X. Wang, R. Zhang, Y. Cui, J. Mao, and R. Jin. 2010b. "Exploitation and exploration in a performance based contextual advertising system". In: *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 27–36.

Lim, Y. J. and Y. W. Teh. 2007. "Variational Bayesian approach to movie rating prediction". In: *Proceedings of KDD cup and workshop*. Vol. 7. 15–21.

Liu, T.-Y. 2009. "Learning to rank for information retrieval". *Foundations and Trends® in Information Retrieval*. 3(3): 225–331.

Liu, T.-Y., J. Xu, T. Qin, W. Xiong, and H. Li. 2007. "Letor: Benchmark dataset for research on learning to rank for information retrieval". In: *Proceedings of SIGIR 2007 workshop on learning to rank for information retrieval*. Vol. 310. ACM Amsterdam, The Netherlands.

Losada, D. E., J. Parapar, and Á. Barreiro. 2016. "Feeling lucky?: multi-armed bandits for ordering judgements in pooling-based evaluation". In: *Proceedings of the 31st Annual ACM Symposium on Applied Computing*. ACM. 1027–1034.

Lu, T., D. Pál, and M. Pál. 2009. "Showing relevant ads via context multi-armed bandits". In: *Proceedings of AISTATS*.

Mahajan, D. K., R. Rastogi, C. Tiwari, and A. Mitra. 2012. "LogUCB: An Explore-exploit Algorithm for Comments Recommendation". In: *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*. *CIKM '12*. Maui, Hawaii, USA: ACM. 6–15. DOI: 10.1145/2396761.2396767.

Medlar, A., K. Ilves, P. Wang, W. Buntine, and D. 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*. *SIGIR '16*. Pisa, Italy: ACM. 1133–1136. DOI: 10.1145/2911451.2911455.

Medlar, A., J. Pyykko, and D. Glowacka. 2017. "Towards Fine-Grained Adaptation of Exploration/Exploitation in Information Retrieval". In: *Proceedings of the 22Nd International Conference on Intelligent User Interfaces*. *IUI '17*. Limassol, Cyprus: ACM. 623–627. DOI: 10.1145/3025171.3025205.

Mnih, A. and R. R. Salakhutdinov. 2008. "Probabilistic matrix factorization". In: *Advances in Neural Information Processing Systems*. 1257–1264.

Nakamura, A. 2014. "A UCB-Like Strategy of Collaborative Filtering". In: *ACML*. Vol. 39. 315–329.

Nazerzadeh, H., R. Paes Leme, A. Rostamizadeh, and U. Syed. 2016. "Where to sell: Simulating auctions from learning algorithms". In: *Proceedings of the 2016 ACM Conference on Economics and Computation*. ACM. 597–598.

Nguyen, T. T. and H. W. Lauw. 2014. "Dynamic Clustering of Contextual Multi-Armed Bandits". In: *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. CIKM '14*. Shanghai, China: ACM. 1959–1962. DOI: 10.1145/2661829.2662063.

Pandey, S., D. Agarwal, D. Chakrabarti, and V. Josifovski. 2007. "Bandits for Taxonomies: A Model-based Approach". In: *SDM*. SIAM. 216–227.

Qin, L., S. Chen, and X. Zhu. 2014. "Contextual combinatorial bandit and its application on diversified online recommendation". In: *Proceedings of the 2014 SIAM International Conference on Data Mining*. SIAM. 461–469.

Radlinski, F., R. Kleinberg, and T. Joachims. 2008a. "Learning diverse rankings with multi-armed bandits". In: *Proceedings of the 25th international conference on Machine learning*. ACM. 784–791.

Radlinski, F., M. Kurup, and T. Joachims. 2008b. "How does clickthrough data reflect retrieval quality?" In: *Proceedings of the 17th ACM conference on Information and knowledge management*. ACM. 43–52.

RicciLior, F., L. Rokach, B. Shapira, and P. B. Kantor. 2001. *Recommender Systems Handbook*. Springer.

Richardson, M., E. Dominowska, and R. Ragno. 2007. "Predicting clicks: estimating the click-through rate for new ads". In: *Proceedings of the 16th international conference on World Wide Web*. ACM. 521–530.

Robbins, H. 1985. "Some aspects of the sequential design of experiments". In: *Herbert Robbins Selected Papers*. Springer. 169–177.

Ruotsalo, T., J. Peltonen, M. J. Eugster, D. Glowacka, A. Reijonen, G. Jacucci, P. Myllymäki, and S. Kaski. 2015. "SciNet: Interactive Intent Modeling for Information Discovery". In: *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '15*. Santiago, Chile: ACM. 1043–1044. DOI: 10.1145/2766462.2767863.

Sanderson, M. 2010. "Test collection based evaluation of information retrieval systems". *Foundations and Trends® in Information Retrieval*. 4(4): 247–375.

Schuth, A., R.-J. Bruintjes, F. Buüttner, J. van Doorn, C. Groenland, H. Oosterhuis, C.-N. Tran, B. Veeling, J. van der Velde, R. Wechsler, *et al.* 2015. "Probabilistic multileave for online retrieval evaluation". In: *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM. 955–958.

Shen, W., J. Wang, Y.-G. Jiang, and H. Zha. 2015. "Portfolio Choices with Orthogonal Bandit Learning". In: *IJCAI*. Vol. 15. 974–980.

Slivkins, A., F. Radlinski, and S. Gollapudi. 2010. "Learning optimally diverse rankings over large document collections". In: *Proceedings of ICML*.

Slivkins, A., F. Radlinski, and S. Gollapudi. 2013. "Ranked bandits in metric spaces: learning diverse rankings over large document collections". *Journal of Machine Learning Research*. 14(Feb): 399–436.

Srinivas, N., A. Krause, S. M. Kakade, and M. Seeger. 2010. "Gaussian process optimization in the bandit setting: No regret and experimental design". In: *Proceedings of the 27th international conference on Machine learning*. 1015–1022.

Srinivasan, S., E. Talvitie, and M. H. Bowling. 2015. "Improving Exploration in UCT Using Local Manifolds". In: *AAAI*. 3386–3392.

Streeter, M. and D. Golovin. 2008. "An online algorithm for maximizing submodular functions". In: *Advances in Neural Information Processing Systems*. 1577–1584.

Streeter, M., D. Golovin, and A. Krause. 2009. "Online learning of assignments". In: *Advances in Neural Information Processing Systems*. 1794–1802.

Strehl, A., J. Langford, L. Li, and S. M. Kakade. 2010. "Learning from logged implicit exploration data". In: *Advances in Neural Information Processing Systems*. 2217–2225.

Sui, Y., V. Zhuang, J. W. Burdick, and Y. Yue. 2017. "Multi-dueling Bandits with Dependent Arms". *arXiv preprint arXiv:1705.00253*.

Sui, Y., M. Zoghi, K. Hofmann, and Y. Yue. 2018. "Advancements in Dueling Bandits". In: *IJCAI*. 5502–5510.

Sutton, R. S. and A. G. Barto. 1998. *Reinforcement learning: An introduction*. MIT press Cambridge.

Tang, L., Y. Jiang, L. Li, and T. Li. 2014. "Ensemble Contextual Bandits for Personalized Recommendation". In: *Proceedings of the 8th ACM Conference on Recommender Systems. RecSys '14*. Foster City, Silicon Valley, California, USA: ACM. 73–80. DOI: 10.1145/2645710.2645732.

Tang, L., Y. Jiang, L. Li, C. Zeng, and T. Li. 2015. "Personalized Recommendation via Parameter-Free Contextual Bandits". In: *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '15*. Santiago, Chile: ACM. 323–332. DOI: 10.1145/2766462.2767707.

Tang, L., R. Rosales, A. Singh, and D. Agarwal. 2013. "Automatic ad format selection via contextual bandits". In: *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*. ACM. 1587–1594.

Teevan, J., S. T. Dumais, and E. Horvitz. 2007. "Characterizing the value of personalizing search". In: *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM. 757–758.

Thompson, W. R. 1933. "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples". *Biometrika*. 25(3/4): 285–294.

Ueda, S., Y. Yamaguchi, and H. Kitagawa. 2017. "Collecting Non-Geotagged Local Tweets via Bandit Algorithms". In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM. 2331–2334.

Urvoy, T., F. Clerot, R. Féraud, and S. Naamane. 2013. "Generic Exploration and K-armed Voting Bandits". In: *ICML*. 91–99.

Valko, M., R. Munos, B. Kveton, and T. Kocak. 2014. "Spectral Bandits for Smooth Graph Functions". In: *ICML*. 46–54.

Vanchinathan, H. P., I. Nikolic, F. De Bona, and A. Krause. 2014. "Explore-exploit in top-n recommender systems via gaussian processes". In: *Proceedings of the 8th ACM Conference on Recommender systems*. ACM. 225–232.

Villar, S. S., J. Bowden, and J. Wason. 2015. "Multi-armed bandit models for the optimal design of clinical trials: benefits and challenges". *Statistical Science: a review journal of the Institute of Mathematical Statistics*. 30(2): 199–215.

Vorobev, A., 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. WWW '15*. Florence, Italy: ACM. 1177–1187. DOI: 10.1145/2736277.2741104.

Wang, H., Q. Wu, and H. Wang. 2016. "Learning Hidden Features for Contextual Bandits". In: *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. CIKM '16*. Indianapolis, Indiana, USA: ACM. 1633–1642. DOI: 10.1145/2983323.2983847.

Wang, H., Q. Wu, and H. Wang. 2017a. "Factorization Bandits for Interactive Recommendation". In: *AAAI*. 2695–2702.

Wang, X., Y. Wang, D. Hsu, and Y. Wang. 2014. "Exploration in Interactive Personalized Music Recommendation: A Reinforcement Learning Approach". *ACM Trans. Multimedia Comput. Commun. Appl.* 11(1): 7:1–7:22. DOI: 10.1145/2623372.

Wang, Y., H. Ouyang, H. Deng, and Y. Chang. 2017b. "Learning Online Trends for Interactive Query Auto-Completion". *IEEE Transactions on Knowledge and Data Engineering*. 29(11): 2442–2454.

Wang, Y., H. Ouyang, C. Wang, J. Chen, T. Asamov, and Y. Chang. 2017c. "Efficient Ordered Combinatorial Semi-Bandits for Whole-Page Recommendation". In: *AAAI*. 2746–2753.

Wen, Z., A. Ashkan, H. Eydgahi, and B. Kveton. 2015. "Efficient learning in large-scale combinatorial semi-bandits". In: *Proceedings of the 32nd International Conference on Machine Learning*. Vol. 37. 1113–1122.

Williamson, S. F., P. Jacko, S. S. Villar, and T. Jaki. 2017. "A Bayesian adaptive design for clinical trials in rare diseases". *Computational Statistics & Data Analysis*. 113: 136–153.

Wu, H. and X. Liu. 2016. "Double thompson sampling for dueling bandits". In: *Advances in Neural Information Processing Systems*. 649–657.

Wu, Q., H. Wang, Q. Gu, and H. Wang. 2016. "Contextual Bandits in a Collaborative Environment". In: *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '16*. Pisa, Italy: ACM. 529–538. DOI: 10.1145/2911451.2911528.

Xing, Z., X. Wang, and Y. Wang. 2014. "Enhancing Collaborative Filtering Music Recommendation by Balancing Exploration and Exploitation". In: *Ismir*. 445–450.

Yue, Y., J. Broder, R. Kleinberg, and T. Joachims. 2012a. "The k-armed dueling bandits problem". *Journal of Computer and System Sciences*. 78(5): 1538–1556.

Yue, Y. and C. Guestrin. 2011. "Linear submodular bandits and their application to diversified retrieval". In: *Advances in Neural Information Processing Systems*. 2483–2491.

Yue, Y., S. A. Hong, and C. Guestrin. 2012b. "Hierarchical exploration for accelerating contextual bandits". In: *Proceedings of ICML*.

Yue, Y. and T. Joachims. 2009. "Interactively optimizing information retrieval systems as a dueling bandits problem". In: *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM. 1201–1208.

Yue, Y. and T. Joachims. 2011. "Beat the mean bandit". In: *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. 241–248.

Zhao, X., W. Zhang, and J. Wang. 2013. "Interactive collaborative filtering". In: *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*. ACM. 1411–1420.

Zhou, L. 2015. "A survey on contextual multi-armed bandits". *arXiv preprint arXiv:1508.03326*.

Zhou, L. and E. Brunskill. 2016. "Latent contextual bandits and their application to personalized recommendations for new users". *arXiv preprint arXiv:1604.06743*.

Zoghi, M., Z. S. Karnin, S. Whiteson, and M. De Rijke. 2015a. "Copeland dueling bandits". In: *Advances in Neural Information Processing Systems*. 307–315.

Zoghi, M., T. Tunys, M. Ghavamzadeh, B. Kveton, C. Szepesvari, and Z. Wen. 2017. "Online learning to rank in stochastic click models". In: *International Conference on Machine Learning*. 4199–4208.

Zoghi, M., S. A. Whiteson, M. De Rijke, and R. Munos. 2014a. "Relative confidence sampling for efficient on-line ranker evaluation". In: *Proceedings of the 7th ACM international conference on Web search and data mining*. ACM. 73–82.

Zoghi, M., S. Whiteson, R. Munos, and M. De Rijke. 2014b. "Relative upper confidence bound for the k-armed dueling bandit problem". In: *Proceedings of ICML*. 10–18.

Zoghi, M., S. Whiteson, and M. de Rijke. 2015b. "MergeRUCB: A Method for Large-Scale Online Ranker Evaluation". In: *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. WSDM '15*. Shanghai, China: ACM. 17–26. DOI: 10.1145/2684822.2685290.

Zong, S., H. Ni, K. Sung, N. R. Ke, Z. Wen, and B. Kveton. 2016. "Cascading Bandits for Large-Scale Recommendation Problems". *arXiv preprint arXiv:1603.05359 - Proc. UAI*.