Deep Reinforcement Learning for Information Retrieval: Fundamentals and Advances

Weinan Zhang
Shanghai Jiao Tong University
wnzhang@sjtu.edu.cn

Xiangyu Zhao
Michigan State University
zhaoxi35@msu.edu

Li Zhao
Microsoft Research
lizo@microsoft.com

Dawei Yin
Baidu
yindawei@acm.org

Grace Hui Yang
Georgetown University
huiyang@cs.georgetown.edu

Alex Beutel
Google
alexbeutel@google.com

ABSTRACT

Information retrieval (IR) techniques, such as search, recommendation and online advertising, satisfying users' information needs by suggesting users personalized objects (information or services) at the appropriate time and place, play a crucial role in mitigating the information overload problem. Since the widely use of mobile applications, more and more information retrieval services have provided interactive functionality and products. Thus, learning from interaction becomes a crucial machine learning paradigm for interactive IR, which is based on reinforcement learning. With recent great advances in deep reinforcement learning (DRL), there have been increasing interests in developing DRL based information retrieval techniques, which could continuously update the information retrieval strategies according to users' real-time feedback, and optimize the expected cumulative long-term satisfaction from users. Our workshop aims to provide a venue, which can bring together academia researchers and industry practitioners (i) to discuss the principles, limitations and applications of DRL for information retrieval, and (ii) to foster research on innovative algorithms, novel techniques, and new applications of DRL to information retrieval.

KEYWORDS

Deep Reinforcement Learning, Information Retrieval

ACM Reference Format:
Weinan Zhang, Xiangyu Zhao, Li Zhao, Dawei Yin, Grace Hui Yang, and Alex Beutel. 2020. Deep Reinforcement Learning for Information Retrieval: Fundamentals and Advances. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '20), July 25–30, 2020, Virtual Event, China.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3397271.3401467

1 BACKGROUND AND MOTIVATIONS

Recent years have witnessed the increased popularity and explosive growth of the World Wide Web, which has generated huge amounts of data and leaded to progressively severe information overload problem [4]. In consequence, how to extract information (products or services) that satisfy users' information requirements at the proper time and place has become increasingly important. This motivates several information retrieval mechanisms, such as search, recommendation, and online advertising. These retrieval mechanisms could generate a set of objects that best match users' explicit or implicit preferences [10]. Efforts have been made on developing supervised or unsupervised methods for these information retrieval mechanisms [21]. However, since the widely use of mobile applications during the recent years, more and more information retrieval services have provided interactive functionality and products [41], these conventional techniques typically face several common challenges. First, most of traditional approaches consider information retrieval tasks in a static environment and extract information via a fixed greedy strategy, which may fail to catch the dynamic characteristics of users' preferences (or environment). Second, the majority of existing methods aims to maximize user's immediate satisfaction, while completely overlooking whether the generate information can benefit more to user's preference in the long run [24]. Thus, learning from interaction becomes a crucial machine learning paradigm for interactive IR, which is based on reinforcement learning [27].

Driven by recent advances in reinforcement learning theories and the prevalence of deep learning technologies, there has been tremendous interest in resolving complex problems by deep reinforcement leaning methods, such as the game of Go [25, 26], video games [16, 17], and robotics [14]. By integrating deep learning into reinforcement learning, DRL is not only capable of continuing sensing and learning to act, but also capturing complex patterns with the power of deep learning. Under the DRL schema, complex problems are addressed by acquiring experiences through interactions with a dynamic environment. The result is an optimal policy that can provide decision making solutions to complex tasks without any specific instructions [13]. Introducing DRL to information retrieval community can naturally tackle the above-mentioned challenges. First, DRL considers information retrieval tasks as sequential interactions between an RL agent (system) and users (environment), where the agent continuously update the information retrieval strategies based on users' real-time feedback, so as to generate information best match users' dynamic preferences. Second, the DRL-based techniques targets to optimize users' long-term satisfaction or engagement. Therefore, the agent could identify information to achieve the trade-off between users' short-term and long-term satisfaction.

SIGIR '20, July 25–30, 2020, Virtual Event, China
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-8016-4/20/07…$15.00
https://doi.org/10.1145/3397271.3401467

Given the advantages of reinforcement learning, there have been tremendous interests in developing RL based information retrieval techniques [3, 5, 11, 12, 15, 38–40]. While these successes show the promise of DRL, applying learning from game-based DRL to information retrieval is fraught with unique challenges, including, but not limited to, extreme data sparsity, power-law distributed samples, and large state and action spaces.

Therefore, in this workshop, we will provide a venue for both academia researchers and industry practitioners to discuss the fundamental principles, technical and practice limitations, and observations and lessons learned from applications of DRL for (interactive) information retrieval. We also aim to foster research on novel information retrieval algorithms, techniques and applications of DRL.

2 REVIEW OF EXISTING WORK

In this section, we briefly review related work of deep reinforcement learning in main IR scenarios i.e., search, recommendation and online advertising.

Search targets at retrieving and ranking a set of items (e.g. documents, records) according to a user query [7, 29]. For query understanding, Nogueira et al. [19] proposed a reinforcement learning based query reformulation task, which maximizes the number of recalled relevant documents via rewriting a query. In [18], a multi-agent reinforcement learning framework is proposed to increase the diverse query reformulation efficiency, where each agent learns a local policy that performs well on a subset of examples, and all agents are trained with parallelism to make the learning faster. For relevance ranking, conventional methods typically optimize the evaluation metric before a predefined position (e.g. NDCG@K), which ignores the information after rank K. MDPRank [31] is proposed to address this problem by using the metrics calculated upon all the positions as reward function, and the model parameters are be optimized via maximizing the accumulated rewards for all decisions. Beyond relevance ranking, another important goal is to increase the diversity of search results [8, 22], which needs to capture the utility of information users have perceived from the preceding documents in a sequential document selection. MDP-DIV [34] formalized diverse ranking as a continuous state Markov decision process, and policy gradient algorithm of REINFORCE is leveraged to maximize the accumulated long-term rewards in terms of the diversity metric.

Recommender systems aim to learn users' preferences based on their feedback and suggest items to match their preferences. User's preference is assumed to be static in traditional recommendation algorithms such as collaborative filtering, which is usually not true in real-world recommender systems where users' preferences are highly dynamic. Bandit methods [33, 37] usually utilizes a variable reward function to delineate the dynamic nature of the environment (reward distributions). Another solution is to introduce the MDP setting [6, 9, 42], where *state* represents user's preference and *state transition* depicts the dynamic nature of user's preference over time. In [39], a user's dynamic preference (state) is learned from her browsing history and feedback. Each time a user provides feedback (skip, click or purchase) to an item, the recommender system will update the state to capture user's new preferences. Conventional recommender algorithms also suffer from the exploitation-exploration dilemma, where exploitation is to suggest items that best match users' preferences, while exploration is to randomly suggest items to mine more users' possible preferences. The contextual bandit method is introduced to achieve the trade-off between exploitation and exploration with strategies such as ϵ-greedy [30], EXP3 [2], and UCB1 [1].

Online advertising is to suggest the right ads to the right users so as to maximize the click-through rate (CTR) or return on investment (ROI) of the advertising campaign, which consists of two main marketing strategy, i.e., guaranteed delivery (GD) and real-time bidding (RTB). In guaranteed delivery setting, ads that grouped into campaigns are charged on a pay-per-campaign basis for the pre-specified number of deliveries [23]. A multi-agent reinforcement learning approach [32] is proposed to derive cooperative policies for the publisher, where impression allocation problem is formulated as an auction problem, and publishers can submit virtual bids for impressions. In Real-Time Bidding setting, an advertiser submits a bid for each impression in a very short time frame. The ad selection task is typically modeled as multi-armed bandit (MAB) problem [20, 28, 35, 36], which neglects the fact that bidding actions would continuously occur before the budget running out. Thus, the MDP setting is introduced. For example, a model-based RL framework is proposed in RTB setting [3], where the state value is approximated by neural network to address the scalability problem of large auction amounts and the limited budget. In [12], a multi-agent bidding model is proposed to jointly consider all the advertisers' biddings in the system, and a clustering approach is introduced to deal with a large number of advertisers.

3 KEY CHALLENGES AND OPEN QUESTIONS

While there has been a plenty of research work on deep reinforcement learning for information retrieval published during the recent years, several problems are still unsolved or remain as key challenges. Here we list some of them for broad discussion.

Sample Efficiency. Sample efficiency has been of key problem for DRL since most DRL methods are model-free. Given sufficient training data from the agent interacting with the environment, the paradigm of model-free RL is suitable for training deep neural network based value functions or policies. However, most IR systems directly interact with the users and collect the training data, which is normally insufficient to train the model-free RL solutions.

Sparse & Biased Feedback Data. Feedback data is commonly biased. In RL view, the experience data is sampled from the occupancy measure distribution of the policy-environment interaction. Although off-policy training method can still help improve the policy based on biased data, the sample efficiency is seriously reduced. Moreover, in some IR scenarios such as online advertising, the positive reward is highly sparse (e.g., 0.3% click-through rate for display ads), which results in very low efficiency or failure of RL.

Online Deployment. From industry perspective, deploying DRL solution onto production IR platform is challenging. The common model pipelines in IR platforms center on relevance estimation models, e.g., relevance or CTR estimation, while the DRL model pipelines center on the policy module. Bridging gap between two generations of model pipelines should be positioned as high priority for applied research team in this field.

4 PROGRAM SKETCH

4.1 Workshop Format

The workshop is planned to be host a whole day, with 2 keynotes, 4 invited talks and 6 oral research talks. The keynote speakers should be well-recognized professors or scientists working on the area. There are two encouraged types of invited talks and peer reviewed oral research talks: (i) the academic talk on fundamental research on reinforcement learning with an attempt of application on IR; (ii) the industrial talk on practice of designing or applying deep reinforcement learning techniques for real-world IR tasks.

Each talk is expected to be presented as a lecture with slides. There will be a QA session at the end of each talk.

4.2 Online Materials

A website (http://drl4ir.github.io) for this workshop will be made available online right before the lecture is presented. All the relevant materials will be made available on this website, including the talk information, presentation slides, referred papers, speaker information and related open source projects etc.

5 RELATED WORKSHOPS

The Deep Reinforcement Learning Workshop at NeurIPS (2015-2019)[1] and IJCAI (2016)[2] focused on the techniques to combine neural networks with reinforcement learning, and domains like robotics, strategy games, and multi-agent interaction. The Deep Reinforcement Learning Meets Structured Prediction at ICLR (2019)[3] focused on leveraging reinforcement learning paradigm on tasks of structured predictions. The workshops at ICML (2019)[4] and KDD (2019)[5] focus on a wide range of real life reinforcement learning applications. The proposed workshop is *the first* to focus on deep reinforcement learning for information retrieval. This workshop will bring together experts in information retrieval and reinforcement learning. Our proposed workshop will have invited keynotes and talks, paper presentation, poster session, and panel discussion to help interested researchers gain a high-level view about the current state of the art and potential directions for future contributions. Real datasets and codes will also be released for attendees to practice in the future.

6 ORGANIZERS INFORMATION AND QUALIFICATION

Dr. Weinan Zhang, the workshop lead organizer, is currently a tenure-track associate professor in Shanghai Jiao Tong University. His research interests include machine learning and big data mining, particularly, deep learning and reinforcement learning techniques for real-world data mining scenarios, such as computational advertising, recommender systems, text mining, web search and knowledge graphs. He has published over 80 papers on first-tier international conferences and journals, including KDD, SIGIR, ICML, ICLR, JMLR, IJCAI, AAAI, WSDM, CIKM etc. He won the Best Paper Honorable Mention Award in SIGIR 2017, the Best Paper Award in DLP Workshop in KDD 2019, ACM Rising Star Award, Alibaba DAMO Young Scholar Award etc. Weinan has organized workshops and tutorials in SIGIR, KDD, CIKM and ECIR etc.

Xiangyu Zhao is a senior Ph.D. student of computer science and engineering at Michigan State University (MSU). His supervisor is Dr. Jiliang Tang. Before joining MSU, he completed his MS (2017) at USTC and BS (2014) at UESTC. He is the student member of IEEE, SIGIR, and SIAM. His current research interests include data mining and machine learning, especially (1) Reinforcement Learning and AutoML for E-commerce; (2) Urban Computing and Spatio-Temporal Data Analysis. After joining MSU, he has published his work in top journals (e.g. SIGKDD, SIGWeb) and conferences (e.g., KDD, SIGIR, CIKM, ICDM, RecSys). He was the recipients of the KDD'18/19, RecSys'18, SDM'18, and CIKM'17 Student Travel Award.

Dr. Li Zhao is currently a Senior Researcher in Machine Learning Group, Microsoft Research Asia (MSRA). Her research interests mainly lie in deep learning and reinforcement learning, and their applications for text mining, recommendation, finance and games. She has co-organized the 3rd Asian Workshop on Reinforcement Learning (AWRL'18), and is one of the invited speakers for AWRL'19. She obtained her Ph.D. degree majoring in Computer Science in July, 2016, from Tsinghua University, supervised by Professor Xiaoyan Zhu. During her Ph.D. studies, she has conducted research on sentiment extraction, text mining and weakly supervised learning. She published several research papers in top conferences, including NeurIPS, KDD, IJCAI, AAAI, EMNLP and CIKM.

Dr. Dawei Yin is Engineering Director at Baidu inc.. He is managing the search science team at Baidu, leading BaiduâĂŹs science efforts of web search, question answering, video search, image search, news search, app search, etc.. Previously, he was Senior Director, managing the recommendation engineering team at JD.com between 2016 and 2020. Prior to JD.com, he was Senior Research Manager at Yahoo Labs, leading relevance science team and in charge of Core Search Relevance of Yahoo Search. He obtained Ph.D. (2013), M.S. (2010) from Lehigh University and B.S. (2006) from Shandong University. From 2007 to 2008, he was an M.Phil. student in The University of Hong Kong. His research interests include data mining, applied machine learning, information retrieval and recommender system. He published more than 80 research papers in premium conferences and journals, and was the recipients of WSDM2016 Best Paper Award, KDD2016 Best Paper Award, WSDM2018 Best Student Paper Award, and ICHI 2019 Best Paper Honorable Mention.

Dr. Grace Hui Yang is an Associate Professor in the Department of Computer Science at Georgetown University. Dr. Yang is leading the InfoSense (Information Retrieval and Sense-Making) group at Georgetown University, Washington D.C., U.S.A. Dr. Yang obtained her Ph.D. from the Language Technologies Institute, Carnegie Mellon University in 2011. Dr. Yang's current research interests include deep reinforcement learning, dynamic information retrieval, search

[1] https://sites.google.com/view/deep-rl-workshop-neurips-2019/home

[2] https://sites.google.com/site/deeprlijcai16/

[3] https://sites.google.com/view/iclr2019-drlstructpred/

[4] https://sites.google.com/view/RL4RealLife

[5] http://www.cse.msu.edu/~zhaoxi35/DRL4KDD/

engine evaluation, privacy-preserving information retrieval, internet of things, and information organization. Prior to this, she has conducted research on question answering, ontology construction, near-duplicate detection, multimedia information retrieval, and opinion and sentiment detection. Dr. Yang has co-chaired SIGIR 2013 and 2014 Doctoral Consortiums, SIGIR 2017 Workshop, WSDM 2017 Workshop, ICTIR 2017 Workshop, CIKM 2015 Tutorial, ICTIR 2018 Short Paper and SIGIR 2018 Demonstration Paper Program Committees. Dr. Yang served on the editorial board of Information Retrieval Journal from 2014 to 2017.

Dr. Alex Beutel is a Staff Research Scientist in Google Brain SIR, leading a team working on responsible and fair ML, as well as researching neural recommendation and ML for Systems. He received his Ph.D. in 2016 from Carnegie Mellon University's Computer Science Department, and previously received his B.S. from Duke University in computer science and physics. His Ph.D. thesis on large-scale user behavior modeling, covering recommender systems, fraud detection, and scalable machine learning, was given the SIGKDD 2017 Doctoral Dissertation Award Runner-Up. He also received the Best Paper Award at KDD 2016 and ACM GIS 2010, was a finalist for best paper in KDD 2014 and ASONAM 2012, and was awarded the Facebook Fellowship in 2013 and the NSF Graduate Research Fellowship in 2011. More details can be found at alexbeutel.com.

Acknowledgement. Weinan Zhang is supported by "New Generation of AI 2030" Major Project (2018AAA0100900) and NSFC (61632017, 61702327, 61772333).

REFERENCES

[1] 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.

[2] Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. 2002. The nonstochastic multiarmed bandit problem. *SIAM journal on computing* 32, 1 (2002), 48–77.

[3] Han Cai, Kan Ren, Weinan Zhang, Kleanthis Malialis, Jun Wang, Yong Yu, and Defeng Guo. 2017. Real-time bidding by reinforcement learning in display advertising. In *WSDM*. 661–670.

[4] Chia-Hui Chang, Mohammed Kayed, Moheb R Girgis, and Khaled F Shaalan. 2006. A survey of web information extraction systems. *TKDE* 18, 10 (2006), 1411–1428.

[5] Haokun Chen, Xinyi Dai, Han Cai, Weinan Zhang, Xuejian Wang, Ruiming Tang, Yuzhou Zhang, and Yong Yu. 2019. Large-scale interactive recommendation with tree-structured policy gradient. In *AAAI*, Vol. 33. 3312–3320.

[6] Minmin Chen, Alex Beutel, Paul Covington, Sagar Jain, Francois Belletti, and Ed H Chi. 2019. Top-K Off-Policy Correction for a REINFORCE Recommender System. In *WSDM*. ACM, 456–464.

[7] W Bruce Croft, Michael Bendersky, Hang Li, and Gu Xu. 2011. Query representation and understanding workshop. In *ACM SIGIR Forum*, Vol. 44. ACM New York, NY, USA, 48–53.

[8] Marina Drosou and Evaggelia Pitoura. 2010. Search result diversification. *ACM SIGMOD Record* 39, 1 (2010), 41–47.

[9] Jun Feng, Minlie Huang, Li Zhao, Yang Yang, and Xiaoyan Zhu. 2018. Reinforcement learning for relation classification from noisy data. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

[10] Hector Garcia-Molina, Georgia Koutrika, and Aditya Parameswaran. 2011. Information seeking: convergence of search, recommendations, and advertising. *Commun. ACM* 54, 11 (2011), 121–130.

[11] Li He, Liang Wang, Kaipeng Liu, Bo Wu, and Weinan Zhang. 2018. Optimizing Sponsored Search Ranking Strategy by Deep Reinforcement Learning. *arXiv preprint arXiv:1803.07347* (2018).

[12] Junqi Jin, Chengru Song, Han Li, Kun Gai, Jun Wang, and Weinan Zhang. 2018. Real-time bidding with multi-agent reinforcement learning in display advertising. In *CIKM*. 2193–2201.

[13] Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. 1996. Reinforcement learning: A survey. *Journal of artificial intelligence research* 4 (1996), 237–285.

[14] Henrik Kretzschmar, Markus Spies, Christoph Sprunk, and Wolfram Burgard. 2016. Socially compliant mobile robot navigation via inverse reinforcement learning. *The International Journal of Robotics Research* 35, 11 (2016), 1289–1307.

[15] Feng Liu, Ruiming Tang, Xutao Li, Weinan Zhang, Yunming Ye, Haokun Chen, Huifeng Guo, and Yuzhou Zhang. 2018. Deep reinforcement learning based recommendation with explicit user-item interactions modeling. *arXiv preprint arXiv:1810.12027* (2018).

[16] Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. 2016. Asynchronous methods for deep reinforcement learning. In *ICML*. 1928–1937.

[17] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. 2013. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602* (2013).

[18] Rodrigo Nogueira, Jannis Bulian, and Massimiliano Ciaramita. 2018. Learning to coordinate multiple reinforcement learning agents for diverse query reformulation. *arXiv preprint arXiv:1809.10658* (2018).

[19] Rodrigo Nogueira and Kyunghyun Cho. 2017. Task-oriented query reformulation with reinforcement learning. *arXiv preprint arXiv:1704.04572* (2017).

[20] Alessandro Nuara, Francesco Trovo, Nicola Gatti, and Marcello Restelli. 2018. A combinatorial-bandit algorithm for the online joint bid/budget optimization of pay-per-click advertising campaigns. In *AAAI*.

[21] Yanru Qu, Bohui Fang, Weinan Zhang, Ruiming Tang, Minzhe Niu, Huifeng Guo, Yong Yu, and Xiuqiang He. 2018. Product-based neural networks for user response prediction over multi-field categorical data. *TOIS* 37, 1 (2018), 1–35.

[22] Razieh Rahimi and Grace Hui Yang. [n. d.]. Modeling Exploration of Intrinsically Diverse Search Tasks as Markov Decision Processes. ([n. d.]).

[23] Konstantin Salomatin, Tie-Yan Liu, and Yiming Yang. 2012. A unified optimization framework for auction and guaranteed delivery in online advertising. In *CIKM*. 2005–2009.

[24] Guy Shani, David Heckerman, and Ronen I Brafman. 2005. An MDP-based recommender system. *JMLR* 6, Sep (2005), 1265–1295.

[25] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. 2016. Mastering the game of Go with deep neural networks and tree search. *nature* 529, 7587 (2016), 484.

[26] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. 2017. Mastering the game of Go without human knowledge. *Nature* 550, 7676 (2017), 354.

[27] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.

[28] Liang Tang, Romer Rosales, Ajit Singh, and Deepak Agarwal. 2013. Automatic ad format selection via contextual bandits. In *CIKM*. 1587–1594.

[29] Zhiwen Tang and Grace Hui Yang. 2017. A Reinforcement Learning Approach for Dynamic Search.. In *TREC*.

[30] Christopher John Cornish Hellaby Watkins. 1989. *Learning from delayed rewards*. Ph.D. Dissertation. King's College, Cambridge.

[31] Zeng Wei, Jun Xu, Yanyan Lan, Jiafeng Guo, and Xueqi Cheng. 2017. Reinforcement learning to rank with Markov decision process. In *SIGIR*. 945–948.

[32] Di Wu, Cheng Chen, Xun Yang, Xiujun Chen, Qing Tan, Jian Xu, and Kun Gai. 2018. A multi-agent reinforcement learning method for impression allocation in online display advertising. *arXiv preprint arXiv:1809.03152* (2018).

[33] Qingyun Wu, Naveen Iyer, and Hongning Wang. 2018. Learning contextual bandits in a non-stationary environment. In *SIGIR*. 495–504.

[34] Long Xia, Jun Xu, Yanyan Lan, Jiafeng Guo, Wei Zeng, and Xueqi Cheng. 2017. Adapting Markov decision process for search result diversification. In *SIGIR*. 535–544.

[35] Min Xu, Tao Qin, and Tie-Yan Liu. 2013. Estimation bias in multi-armed bandit algorithms for search advertising. In *NIPS*. 2400–2408.

[36] Hongxia Yang and Quan Lu. 2016. Dynamic contextual multi arm bandits in display advertisement. In *ICDM*. IEEE, 1305–1310.

[37] Chunqiu Zeng, Qing Wang, Shekoofeh Mokhtari, and Tao Li. 2016. Online context-aware recommendation with time varying multi-armed bandit. In *KDD*. 2025–2034.

[38] Xiangyu Zhao, Long Xia, Liang Zhang, Zhuoye Ding, Dawei Yin, and Jiliang Tang. 2018. Deep Reinforcement Learning for Page-wise Recommendations. In *Proceedings of the 12th ACM Recommender Systems Conference*. ACM, 95–103.

[39] Xiangyu Zhao, Liang Zhang, Zhuoye Ding, Long Xia, Jiliang Tang, and Dawei Yin. 2018. Recommendations with Negative Feedback via Pairwise Deep Reinforcement Learning. In *KDD*. ACM, 1040–1048.

[40] Xiangyu Zhao, Liang Zhang, Zhuoye Ding, Dawei Yin, Yihong Zhao, and Jiliang Tang. 2017. Deep Reinforcement Learning for List-wise Recommendations. *arXiv preprint arXiv:1801.00209* (2017).

[41] Xiaoxue Zhao, Weinan Zhang, and Jun Wang. 2013. Interactive collaborative filtering. In *CIKM*. 1411–1420.

[42] Xiangyu Zhao, Xudong Zheng, Xiwang Yang, Xiaobing Liu, and Jiliang Tang. 2020. Jointly Learning to Recommend and Advertise. *arXiv preprint arXiv:2003.00097* (2020).