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).