REFERENCES

[1] 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 *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM, 661–670.

[2] Shi-Yong Chen, Yang Yu, Qing Da, Jun Tan, Hai-Kuan Huang, and Hai-Hong Tang. 2018. Stabilizing reinforcement learning in dynamic environment with application to online recommendation. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 1187–1196.

[3] Jun Feng, Heng Li, Minlie Huang, Shichen Liu, Wenwu Ou, Zhirong Wang, and Xiaoyan Zhu. 2018. Learning to Collaborate: Multi-Scenario Ranking via Multi-Agent Reinforcement Learning. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 1939–1948.

[4] Yujing Hu, Qing Da, Anxiang Zeng, Yang Yu, and Yinghui Xu. 2018. Reinforcement Learning to Rank in E-Commerce Search Engine: Formalization, Analysis, and Application. *arXiv preprint arXiv:1803.00710* (2018).

[5] Wendi Ji and Xiaoling Wang. 2017. Additional Multi-Touch Attribution for Online Advertising.. In *AAAI*. 1360–1366.

[6] 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. *arXiv preprint arXiv:1802.09756* (2018).

[7] M Mahmud. 2010. Constructing states for reinforcement learning. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. 727–734.

[8] Andrew Kachites McCallum and Dana Ballard. 1996. *Reinforcement learning with selective perception and hidden state*. Ph.D. Dissertation. University of Rochester. Dept. of Computer Science.

[9] et al. Mnih, Volodymyr. 2015. Human-level control through deep reinforcement learning. *Nature 518, no. 7540 (2015): 529* (2015).

[10] Kevin P Murphy. 2000. A survey of POMDP solution techniques. *environment* 2 (2000), X3.

[11] Ronald Parr and Stuart Russell. 1995. Approximating optimal policies for partially observable stochastic domains. In *IJCAI*, Vol. 95. 1088–1094.

[12] Andres C Rodriguez, Ronald Parr, and Daphne Koller. 2000. Reinforcement learning using approximate belief states. In *Advances in Neural Information Processing Systems*. 1036–1042.

[13] Xuhui Shao and Lexin Li. 2011. Data-driven multi-touch attribution models. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 258–264.

[14] Pengfei Zhu, Xin Li, Pascal Poupart, and Guanghui Miao. 2018. On improving deep reinforcement learning for pomdps. *arXiv preprint arXiv:1804.06309*.