Interactive Online Learning from Incomplete Knowledge

Qingyun Wu (qw2ky@virginia.edu)
Department of Computer Science, University of Virginia

Interactive online learning is vital in many information service systems. For example, in search engines and recommender systems, due to the heterogeneity and dynamic nature of a large population of users, a generic offline trained algorithm can hardly satisfy each individual user's need, which calls for interactive online learning solutions. Online learning solutions explore the unknowns by sequentially collect individual user's feedback to evaluate the quality of interactions while monitoring changes in their values. It helps address the notorious explore/exploit dilemma during sequential decision making.

My thesis focuses on developing online learning algorithms, and more specifically multi-armed bandit algorithms, to make the service systems more interactive and natural. Specifically, the research of my thesis can be understood from the following two perspectives, which are two sides of the same coin. On one hand, I study online learning solutions to sequentially estimate the information need of users in a collaborative manner, which enables information sharing across users, and thus expedites the learning process in a dynamic environment. On the other hand, instead of passively responding to users' request, I propose to proactively choose the most representative users and information to initiate or incentivize the interaction for the most beneficial feedback, which further improves the system's utility in the long run. By combining the two parts, an information service system can provide right information to the right user at the right time, and improve users satisfaction in the long run. This will become a new paradigm for human-machine interaction: both systems and users win in this interactive game. More importantly, the proposed solutions can be applied to a wide spectrum of applications including not only the aforementioned information service systems, but also crowdsourcing, human-machine interactions in cyber physical systems, clinical trials in healthcare, sequential treatment design in psychology and many more. Ultimately, the proposed research will lead to an intelligent and personalized learning system, which 'talks' to users for interactive knowledge acquisition and 'learns' from them for adaptive model update.

First of all, from the machine learning perspective of my proposed research, the challenges stem from the noisy and incomplete feedback from a group of heterogeneous and dependent users, and the highly dynamic nature of both users interest and information need. To conquer these challenges, we will study solutions to perform collaborative learning, which enables information sharing among users, and solutions to detect the potential changes during the sequential interaction process. **Capture user dependency in online learning:** In real-world scenarios, users targeted by learning systems are often not independent but connected by networks. The network structure provides an important source of information, revealing potential affinities between pairs of users, especially when the observations are sparse but users have strong social ties, e.g., being friends on online social networks. Base on this insight, I proposed a collaborative bandit learning framework in [4, 1]. The insight of my solution is to capitalize on the information propagation among users to expedite online learning. A remarkable improvement in learning efficiency comparing to the case where users are modeled independently was rigorously proved. Empirical evaluation on a large scale news recommendation dataset also showed the effectiveness the proposed collaborative learning schema. The current solutions assume the availability of user dependency information. My ongoing research will further advance them by removing the dependency availability assumption: we plan to automatically estimate user dependency from the interaction data accumulated on the fly and utilize the learnt user dependency to perform collaborative learning. This could make the proposed collaborative online learning solutions more realistic and appealing in real world scenarios. **Conquer a non-stationary environment:** Temporal dynamics is an intrinsic characteristic of many information systems, but it is largely ignored in most of the traditional stochastic online learning formulations, which usually assume a static environment. However, this static assumption rarely holds in reality as users' preferences can be influenced by various internal or external factors, which lead to slow or dramatic shift in users' preferences. This temporal dynamics make traditional learning approaches incompetent. In such a non-stationary environment, to provide up-to-date information, a learning algorithm need to be adjusted in time. In my thesis, I propose a set of dynamic bandit algorithms to capture the potential dynamics in the environment. Currently, we have proposed solutions [2, 5] to automatically detect and adapt to abruptive changes by maintaining a suite of bandit models during identified stationary periods based on its interactions with the environment. Our proposed solutions can handle not only context-free

changes but also context-dependent changes. Both theoretical analysis and empirical evaluation verified the effectiveness of the proposed algorithms in capturing the temporal dynamics. The current ongoing research is to further extend our solutions from addressing an abruptly changing environment to a more general continuously changing environment, which can make our solutions more general and capable to capture the gradual changes of users. In addition, in such a changing environment, information sharing between learners could be particularly useful because a learner could benefit from previous experience of another learner to adapt to their new environment. This information sharing inspired me to further explore the possibility of expediting the change detection process by leveraging social ties among users.

The second major research question I will explore in my thesis is how to efficiently perform proactive information acquisition to further expedite the learning process while not affecting user engagement. The motivations of performing proactive information acquisition are two-folds: First, in many applications, users might be myopic or conservative and thus are reluctant to explore new options, which will inevitably make the collected feedback not well-rounded and make the model uncertainty unbalanced. In this case, incentivizing users for feedback on options that can reduce the learning agent' uncertainty will be particularly beneficial. This proactive information acquisition will help improve the learning agent's estimation quality and in turn better serve the users in the long-run. Second, among a large group of users, some users could be more informative than the others. In this case, in a collaborative learning environment where users have dependency on each other, if the system could start with users whose feedback can mostly reduce the system's uncertainty about the rest users, less exploration would be needed to optimize the system's utility for all users. In the process of information acquisition, we also need to pay attention to its non-negligible effects on the users: too aggressive information acquisition may frustrate users and negatively affect users' engagement. In my research, long-term user engagement improvement [3] has been studied from the perspective of reinforcement learning. I will continue to explore how to perform efficient proactive information acquisition with engagement constraints in my thesis.

To summarize, my thesis focuses on developing online learning solutions to make information systems truly interactive and more useful. My research will generate impacts from the following perspectives: First, the proposed collaborative online learning solutions can reduce learning complexity and enable the algorithms to interact with users. More importantly, the proposed proactive information acquisition can help further improve the learning systems without affecting user experience. The research in my thesis can be applied to a wide spectrum of applications that involve sequential decision making.

References

[1] Huazheng Wang, Qingyun Wu, and Hongning Wang. Factorization bandits for interactive recommendation. In *The 31th AAAI Conference on Artificial Intelligence (AAAI 2017)*.

[2] Qingyun Wu, Naveen Iyer, and Hongning Wang. Learning contextual bandits in a collaborative environment. In *SIGIR 2018*.

[3] Qingyun Wu, Hongning Wang, Liangjie Hong, and Yue Shi. Returning is believing: Optimizing long-term user engagement in recommender systems. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, CIKM '17, pages 1927–1936, New York, NY, USA, 2017. ACM.

[4] Qingyun Wu, Huazheng Wang, Quanquan Gu, and Hongning Wang. Contextual bandits in a collaborative environment. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 529–538. ACM, 2016.

[5] Qingyun Wu, Huazheng Wang, Yanen Li, Naveen Iyer, and Hongning Wang. Dynamic ensemble of contextual bandits to conquer a non-stationary environment. In *NIPS 2018 (Under Review)*.