Chapter 7

Conclusion and Future Work

With the information overload over the web, it has become imperative that the user should be provided with the most optimal results from the vast amount of content availaible. Both recommender and search systems tasks have grown in popularity and are active area of research in the recent times. For recommender systems and ranking in web search it is essential to provide the most relevant items foremost in the result list. In our thesis, we utilized Reinforcement learning based approaches to model the ranking and recommendation tasks. We addressed the different issues in recommendations and rankings such as scalability in large item space, cold start, high variance and noise in existing RL Based approaches in large state-action space. Further, we also presented a Multi Agent RL framework to utilize the shared information between the documents for the LTR task.

Firstly, we proposed a Deep Reinforcement learning based method, DRLRank for the ranking task in web search systems. In recent years, reinforcement learning methods have been applied effectively to information retrieval tasks, however traditional RL algo- rithms suffer from lack of complexity required when the state space becomes extremely large with millions of items. In our solution, we modeled the ranking tasks as a Markov Decision process process. We combined the powerful function approximation capabilities of Deep NN with the policy gradient algorithm to address

the large state action space in ranking problem. We utilize clipped objective function approach to control the updates during learning, making it more robust to sudden large policy changes and thus helping to reduce variance. With experience replay buffer incorporated in actor critic methods, we can reuse past experiences. This can help in decorrelating the data and reducing the variance and noisy gradients in the updates. Also, with double q learning utilizing separate value function networks i.e. 2 critics instead of a single one, it helps in reducing overestimation bias in value function estimates, making the learning process more stable. This, together with the delayed update, leads in a more stable approximation with reduced bias. We performed experiments on three different Letor datasets. The experimental results demonstrates the effectiveness of our method, DRLRank in terms of different metrices like NDCG and MAP over various baselines methods.

Next, we present a Multi-Agent Deep Reinforcement Learning based model, MARL-Rank for large-scale learning to Rank tasks. While reinforcement learning methods have been effectively applied for LTR rank, the work on Multi agent RL for the same has been very limited. The existing approaches for Learning to Rank overlooks the correlation between the documents as they tend to ignore/overpass the shared information between different documents. We employ the Multi-Agent framework to capture the correlation between the documents by sharing information across multiple agents, where the centralized critic architecture has access to this global knowledge(shared information). Further, in large scale datasets the issues such as variance and noise are further aggravated. To address this problem, we employ the Deep RL actor-critic framework that learns the value estimates directly from the samples.

Lastly, we proposed a PPO based hybrid recommender system for large scale recommendations. This approach utilizes the actor-critic framework and thus mitigates the high variance in Policy Gradient methods. Further, we also addressed the cold start issue in Collaborative filtering with autoencoder-based content filtering in a hybrid setting. With a clipped objective surrogate function, PPO ensures minimal

variance during learning by ensuring that the updated policy isn't too different from the old policy. The reduction of variation contributes to the increased stability of the learning process. Furthermore, the PPO agent uses an advantage function to restrict the policy gradient step and lower the variance of the estimation so that it does not deviate too far from the initial policy, resulting in unduly large updates that frequently make the policy unstable. Extensive experiments on the popular Movielens dataset demonstrate that our method outperforms the baselines in terms of different evaluation metrics such as Precision and Recall.

7.0.1 Future Work

In future, our work can be extended to incorporate more diversified search results in the ranking task. Next, we can extend our work to learn from implict feedback mechanism such as from the clicks rather than the strict requirement of having the specified ground truth labels. The click based data is comparatively less expensive and more extensively availaible compared to explicit feedback. However, implicit data does require pre processing at the initial stage. Different methods to further utilize the correlation between the document scan be explored and integerated with Reinforcement Learning based solution. Further techniques such as counterfactual multi agents for credit assignment can also be effective for comparing the individual agent contribution in multi agent settings and rewarding them correspondingly. Another research work could be to add more user personalization for the recommended items, through creating a user profile by collecting and analyzing data related to user preferences, behavior, demographics, and interactions etc. Further, more recent advances in Reinforcement learning research such as inverse reinforcement learning framework, adverserail rl etc. can be used to model the ranking problem. Soft actor critic methods are other alternative for actor critic methods and be used to model the explore exploit dielemma in the RL problem effectively. We are interested in enhancing our work in recommender systems to other domains such as e-commerce products and deploying multi-agent reinforcement learning based recommender sys-

tem frameworks to model the recommender sys- tem process in the future. Techniques that can lead to more diversity in the recommendation process could be the focus of future research. Using the correlation between the items, we can modify the action in our MDP architecture from the proposed approach to recommend more than one item at a time. Furthermore, we can analyse various possibilities of the recommendation process, each modelled by a different agent, using a multi-agent method.