Modeling Exploration of Intrinsically Diverse Search Tasks as Markov Decision Processes

Razieh Rahimi
Department of Computer Science
Georgetown University
Washington D.C. 20057
razieh.rahimi@georgetown.edu

Grace Hui Yang
Department of Computer Science
Georgetown University
Washington D.C. 20057
huiyang@cs.georgetown.edu

Abstract

Online learning of ranking models based on user interactions has attracted considerable attention in recent years. However, these models are mainly designed to optimize search results for one-time queries, while intrinsically diverse search tasks where the goal is to explore all aspects of an information need generally require search page navigation. The goal in such search scenarios is to utilize user interactions with results in order to optimize the whole search session. To provide the user with the next page of search results, the search system, on the one hand, should exploit all feedback information about the query, obtained from the user. On the other hand, to fully satisfy the user's information need, the search system needs to cover all aspects of the information need in the search results, while the available feedback may be on some, but not all, aspects. The search system thus needs to explore and provide a diversified search results. with the aim of covering all aspects of the information need. Therefore, in each interaction, the search system has to resolve the exploration-exploitation tradeoff. We hypothesize that this tradeoff depends on the state of the search session, defined by properties such as the number of aspects identified thus far by the user. We thus propose a new approach that casts the defined search task as a Markov decision process (MDP). The parameters of MDP are initially unknown to the search engine. It therefore has to learn these parameters through user interactions, thus resulting in an *approximate* MDP, and then find the optimal policy for the learned MDP. Alternatively, the search engine might put its efforts to directly learn (an approximately) optimal policy from the interactions. The first approach is usually referred to as *model-based* learning in MDPs, whereas the second is called *model-free*. In this paper, we adopt the second approach. This research is supported by DARPA grant FA8750-14-2-0226 and NSF CAREER AWARD IIS-1453721.

Keywords: Multi-page exploratory search, Markov decision process