A Phased Ranking model for Information Systems

Rui Liu

Abstract

To effectively sort and present relevant information pieces (e.g., answers, passages, documents) to human users, information systems rely on ranking models. Existing ranking models are typically designed for a specific task and therefore are not effective for complex information systems that require component changes or domain adaptations. For example, in the final stage of question answering, information systems such as IBM Watson DeepQA rank all results according to their evidence scores and judge the likelihood that each is correct or relevant. However, as information systems become more complex, determining effective ranking approaches becomes much more challenging.

Prior work includes heuristic ranking models that focus on a particular type of information object (e.g. a retrieved document, a factoid answer) using manually designed features specific to that information type. These models, however, do not use other, non-local features (e.g. features of the upstream/downstream information source) to locate relevant information. To address this gap, my research seeks to define a ranking approach that should easily and rapidly adapt to any version of system pipelines with an arbitrary number of phases.

We describe a general ranking approach for multi-phase and multi-strategy information systems, which produce and rank significantly more candidate results than the single phase and single strategy information systems to achieve acceptable robustness and overall performance. Our approach allows each phase in a system to leverage information propagated from preceding phases to inform the ranking decision. By collecting ranking features from the derivation paths that generate candidate results, the particular derivation path chosen can be used to predict result correctness or relevance. Those ranking features can be detected from an abstracted system object graph which represents all of the objects created during system execution (e.g. provenance) and object dependencies. This ranking approach has been applied to different domains including question answering and biomedical information retrieval. Experimental results showed that our proposed approach significantly outperforms comparable answer ranking models on the two domains.