Dissertations

Query-Dependent Selection of Retrieval Alternatives

Download

Niranjan Balasubramanian, *University of Massachusetts - Amherst*

Follow

Date of Award
9-2011

Document Type
Open Access Dissertation

Degree Name
Doctor of Philosophy (PhD)

Degree Program
Computer Science

First Advisor
James Allan

Second Advisor
W. Bruce Croft

Third Advisor
David Jensen

Keywords
Information Retrieval, Learning to Rank, Machine Learning

Subject Categories
Computer Sciences

Comments
The main goal of this thesis is to investigate query-dependent selection of *retrieval alternatives* for Information Retrieval (IR) systems. Retrieval alternatives include choices in representing queries (query representations), and choices in methods used for scoring documents. For example, an IR system can represent a user query without any modification, automatically expand it to include more terms, or reduce it by dropping some terms. The main motivation for this work is that no single query representation or retrieval model performs the best for all queries. This suggests that selecting the best representation or retrieval model for each query can yield improved performance. The key research question in selecting between alternatives is how to estimate the performance of the different alternatives. We treat query dependent selection as a general problem of selecting between the result sets of different alternatives. We develop a relative effectiveness estimation technique using retrieval-based features and a learning formulation that directly predict differences between the results sets. The main idea behind this technique is to

Included in

Computer Sciences Commons

SHARE

aggregate the scores and features used for retrieval (retrieval-based features) as evidence towards the effectiveness of the results set. We apply this general technique to select between alternatives reduced versions for long queries and to combine multiple ranking algorithms. Then, we investigate the extension of query-dependent selection under specific efficiency constraints. Specifically, we consider the black-box meta-search scenario, where querying all available search engines can be expensive and the features and scores used by the search engines are not available. We develop easy-to-compute features based on the results page alone to predict when querying an alternate search engine can be useful. Finally, we present an analysis of selection performance to better understand when query-dependent selection can be useful.

Recommended Citation

Balasubramanian, Niranjan, "Query-Dependent Selection of Retrieval Alternatives" (2011). *Dissertations*. Paper 429.

http://scholarworks.umass.edu/open_access_dissertations/429