

Towards Learning Query Expansion

Authors : Ahlem Bouziri, Chiraz Latiri, Eric Gaussier

Abstract : The steady growth in the size of textual document collections is a key progress-driver for modern information retrieval techniques whose effectiveness and efficiency are constantly challenged. Given a user query, the number of retrieved documents can be overwhelmingly large, hampering their efficient exploitation by the user. In addition, retaining only relevant documents in a query answer is of paramount importance for an effective meeting of the user needs. In this situation, the query expansion technique offers an interesting solution for obtaining a complete answer while preserving the quality of retained documents. This mainly relies on an accurate choice of the added terms to an initial query. Interestingly enough, query expansion takes advantage of large text volumes by extracting statistical information about index terms co-occurrences and using it to make user queries better fit the real information needs. In this respect, a promising track consists in the application of data mining methods to extract dependencies between terms, namely a generic basis of association rules between terms. The key feature of our approach is a better trade off between the size of the mining result and the conveyed knowledge. Thus, face to the huge number of derived association rules and in order to select the optimal combination of query terms from the generic basis, we propose to model the problem as a classification problem and solve it using a supervised learning algorithm such as SVM or k-means. For this purpose, we first generate a training set using a genetic algorithm based approach that explores the association rules space in order to find an optimal set of expansion terms, improving the MAP of the search results. The experiments were performed on SDA 95 collection, a data collection for information retrieval. It was found that the results were better in both terms of MAP and NDCG. The main observation is that the hybridization of text mining techniques and query expansion in an intelligent way allows us to incorporate the good features of all of them. As this is a preliminary attempt in this direction, there is a large scope for enhancing the proposed method.

Keywords : supervised learning, classification, query expansion, association rules

Conference Title : ICMLA 2015 : International Conference on Machine Learning and Applications

Conference Location : Copenhagen, Denmark

Conference Dates : June 11-12, 2015