Exploring Recommenders for Improved Information Retrieval

Wei Li
CGLN, School of Computing,
Dublin City University, Ireland
wli@computing.dcu.ie

The continuing increase in the volume of information available in our daily lives is creating ever greater challenges for people to find personally useful information. One approach used to addressing this challenge is Personalized Information Retrieval (PIR) which develops personal user profiles with the objective of giving retrieval results which better meet individual user information needs. However, in many situations there may be no opportunity to build a personal profile, e.g. when this user is querying on a new topic. Under this condition, the experiences and behaviour of other previous users, who have made similar queries, could be used to build a model of general user behaviour in this domain. Our research proposes to focus on the development a new IR approach for topic domains. Our work seeks to combine recommender algorithms trained using previous search behaviour from different searchers with a standard ranked IR method to form a topic-specific IR model to improve the search effectiveness for a user entering a query without personal prior search history on this topic. The challenges of our work are: how to provide users better results; what data collection can be used to conduct experiment; and how to do the combination of recommender and IR components.

Since there are no suitable data collections available to enable us to explore this proposal, simulation plays a key role in our current research. We simulate users’ search interests and their search behaviour, then cluster them into topical groups to train multiple recommenders models. For each user query, the best matching recommender is selected and used to produced a ranked list of recommended item. This prediction is then used to re-rank the output of a standard ranked IR system by linearly combining these lists.

In an initial study we combined an extended SMART system with recommender models built using the Slope-1 algorithm for a subset of the INEX 2010 dataset using a simple simulation model, producing a MAP increase from 0.0744 to 0.1178 (+58%). This initial experiment demonstrates the potential of our proposal. In order to prove our approach works effectively, we need more realistic data to conduct our experiment. My current work is focusing on exploring the use of crowdsourcing to collect more real data from users, techniques to automatically identify and cluster topical domains, issues of sparse training of recommender models and more advanced models for component fusion methods, and examination of alternative recommender algorithms.

Topic Specific Information Retrieval, Recommender Algorithms, Fusion Methods