Enhanced Web Log Based Recommendation by Personalized Retrieval

Xueping Peng

FACULTY OF ENGINEERING AND INFORMATION TECHNOLOGY

UNIVERSITY OF TECHNOLOGY, SYDNEY

A thesis submitted for the degree of

Doctor of Philosophy

February 2015

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Student

_____

Acknowledgements

First I would like to thank my supervisors. Prof Chengqi Zhang, Prof Zhendong Niu and Dr Ling Chen who introduced me to the wonderful world of research. They not only gave me invaluable academic advice, but also helped my transition into a different culture. Prof Chengqi Zhang has been a great mentor and collaborator, being both energetic and full of ideas. Prof Zhendong Niu, who works in the School of Computer Science, Beijing Institute of Technology, guided me in terms of information retrieval and web log mining. Dr Ling Chen helped me by asking insightful questions, and giving me thoughtful comments on the thesis. I enjoyed working with them, and benefited enormously from my interactions with them.

I spent nearly two and a half years in the University of Technology, Sydney. I thank the collaborators, faculty staff, fellow students and friends in the QCIS centre, who made my graduate life a very memorable experience. In particular, I thank you if you are reading this thesis.

I thank my family. My parents richly endowed me with curiosity about the natural world. Last but not least, my deep gratitude is extended to my dear wife Suling Niu who brings me so much love and happiness. It is no exaggeration to say that she helped to make my thesis writing an enjoyable endeavor.

Contents

List of Figures

List of Tables

Abstract

With the rapid development of the Internet and WWW, it is more and more important for people to access quality web information. Thus the problem of enabling users to quickly and accurately find information has become an urgent issue. As one of the basic ways to solve this problem, personalized information services have been focusing on fulfilling the personalized information requirements of different users based on their actual demands, preference characteristics, behaviour patterns, etc.

This thesis focuses on enhancing web log based recommendation by personalized retrieval, and its main works and innovations include:

- For personalized retrieval, the thesis proposes two models to improve user experience and optimize search performance. The first is a query suggestion model based on query semantics and click-through data. This model calculates the subject relevance between queries, and then combines the semantic information and the relevance of the query-click matrix model as this can effectively eliminate the ambiguity and input errors of reminder queries. The second is a collaborative filtering retrieval model based on local and global features. By the integration of the local and global characteristics of the accessed information, this model overcomes the limitations of a single feature, and increases the degree of application of the retrieval model.
- For recommendation by personalized retrieval, we propose two recommendation models based on the web log. The first is based on the user's atomic

retrieval transaction sequence and the browse characteristics. This model decomposes search transactions, and calculates the user's degree of interest on the search term, which allows users to query information more clearly. Further, it incorporates the user feedback on the search results evaluation value, which overcomes the shortcomings of the model based on content filtering. The second model is based on user interests association findings, which can be used to: find the relationship between resources accessed by users, extract the associations of user interests, and address the problem of user interests isolation.