

Query expansion based on user friends and interesting_libraries for social book search

Huang Hanjuan^{1, a}, Zhao Qiling^{2, b}

1. Wuyi University, Mechanical and electrical engineering college, Wuyishan354300, China;

2.Wuyi University, Mathematics and computer College, Wuyishan354300, China

^a215995504@qq.com, ^b946122846@qq.com

Keywords: social book search; query expansion;recommended degree

Abstract: In traditional book retrieval system, users may not provide accurate information about what they want to query because of the restrictions of expertise,so that they can not get a good result. We provide a new method.First,got set of user's friends and interesting_libraries from his profile.And then,used the probabilistic model to get the similar users whose interested books similar with the books that the user want to search and calculated the similarity.Got the co-occurrence words of original query from the personal digital library of similar users and calculated the co-occurrence rate using the Jaccard index.Finally,calculated recommend of each co-occurrence word.The one with the highest recommend is selected to combine with the original query and generated a new query.We conducted several experiments on a real-dataset collected from LibraryThing. It shows that our method can effectively expand the user original query and improve the accuracy of query.

Introduction

With the mass-digitization efforts of recent years, book knowledge become increasingly available online,as collections of digitised books in digital libraries and on the Web.A popular form of access to collections of digitised books is by online user services,such as Baidu Book and Amazon.User provide queries to enter these engines to search book information. In this case,as it is traditional in IR, the user's query is matched against a representation of a collection of books directly. However, user queries are usually too short to describe the book information need accurately and sometimes can't provide accurate information because of the restrictions of expertise.Query term is, however, the user provides the only way to express their intentions,So only through the analysis of the query words and extend to better characterize the user's query intention.

We propose a method of query expansion based on user friends and interesting_libraries for social book search.

Literature References

There is growing interest in social book search. [1]research on the impact of different document representations of book metadata for book search , they show that standard retrieval models perform better on user-generated metadata than on professional metadata. [2] Continue to investigate the complex nature of relevance in book search and the role of traditional and user-generated book metadata in retrieval.[3] experiment with traditional query expansion techniques, exploiting Wikipedia articles as rich sources of information that can augment the user's query.

Query expansion is used to modify a user's original query by adding extra terms drawn from documents in the searched collection. Pseudo-relevance feedback exploiting the retrieval result has been the most effective [4]. However, the traditional methods also highly depend on the original retrieval results. Currently, there are some popular analysis methods which are latent semantic indexing,such as the method based on WordNet[5].

It is the first study centered on using social bookmarking and tagging techniques for personalization of web search and its evaluation in a real-world scenario [6].[7] proposes a novel query expansion framework based on individual user profiles mined from the annotations and resources the user has marked.

In view of the above analysis, we propose a novel query expansion based on the user's friends and interesting_libraries for social book search, it may reflect the user intention more effectively and improve the precision.

Query expansion based on user friends and interesting_libraries for social book search

Similar users model

Similarity calculation

In this paper, we obtained set of user's friends and interesting_libraries from the user profile and then got the Similar users from the set. We think the greater probability the user query words appear in these personal digital library of the Similar users, the greater the similarity between the book type that the similar user interested and the book type that the user want to search.

We use probabilistic model to get the set of similar users from set of user's friend and interesting_libraries. More formally, given a query q . As the formular shown below:

$$M(q) = \frac{\text{frequency}(q)}{\sum_{s \in V} \text{frequency}(s)} (1)$$

In formulae (1), q is user original query, $\text{frequency}(q)$ is frequency of q appears in the personal digital library of user's friend or interesting_libraries, $\sum_{s \in V} \text{frequency}(s)$ is frequency of all book tags words appear in the personal digital library of user's friend or interesting_libraries, $M(q)$ is frequency of original query q appears in the personal digital library of user's friend or interesting_libraries. The higher $M(q)$, the greater the similarity between the book type what the user's friend or interesting_libraries interested and the book type what the user want to search.

Similar users set

In this section, we illustrate our approach to get the similar users set. Figure 1 shows these steps in a flow chart. Given a user U_i and his query q .

As the flow chart of getting the similar users set shown, we have the similar users set for the user U_i .

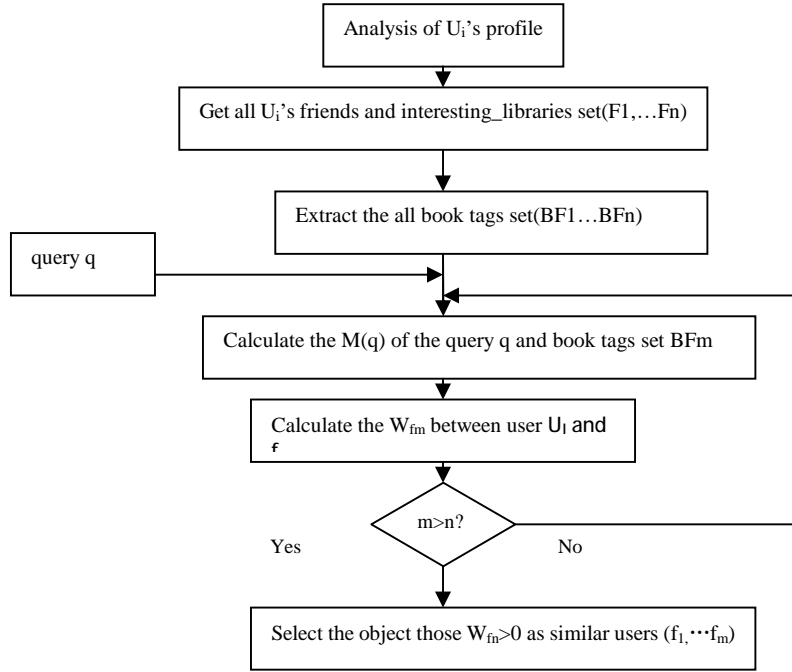


Figure 1 The flow chart of getting the similar users set

Calculating of the co-occurrence words and co-occurrence rate

Here, we used the co-word analysis of the co-occurrence analysis to co-occurrence analyze the original query word and the book tags of similar users. The co-word theory holds that two words (original query word and book tags) have the total relationship if they appear in the same book tags at the same time.

The co-occurrence rate

In this paper, Jaccard index is used for co-occurrence rate calculating .Jaccard index is calculated as:

$$J_{qt} = \frac{C_{qt}}{C_q + C_t - C_{qt}} \quad (2)$$

In formulae (2),Where J_{qt} represents the co-occurrence rate of the original query q and book tag word t and $0 \leq J_{qt} \leq 1$; C_{qt} represents the co-occurrence frequency of the original query q and book tag word t ; C_q represents the original query word q frequency. C_t represents the frequency of the books marked word t .

Recommended degree

We will combine Similarity with co-occurrence rate to choose recommended words for two reasons: one is that the Similarity between every user in the Similar users set and the user are different; the other one is the co-occurrence rate between every co-occurrence word of every similar user and the user original query are different. We define the recommended degree as follows:

$$S_{ft} = M_f \times J_{ft} \quad (3)$$

In the equation above, S_{ft} refers to the recommended degree that the book tag word t of object f recommend be to user's original query; M_{ft} refers to similarity degree between the object f 's interested book type and the book type users want to search. J_{ft} refers to co-occurrence rate of the initial query q and book tag word t .

Query recommended algorithm

The description of Query recommended algorithm is shown below

Input: LibraryThing account number of a user(U_i) and his original query (q)

Output: new query Q after expansion

The algorithm

- Analyze the profile of the U_i , then get the set (C_U) of friends and the interesting_libraries of U_i .

- Calculate the similarity W_f between the U_i and the users in C_U by using the formula(1)
- Select the user whose value of W_f is larger than 0 as the similar users.
- Obtain the personal digital library of the similar users, and extract the history book tags.
- Obtain the co-occurrence word between the original query and the book tags words, and calculate the rate of co-occurrence by using the formula (2).
- Calculate the recommended degree of every co-occurrence word by using formula (3).
- Select the two co-occurrence words with highest recommended degree as extend words and add them to the original query to rebuilt new query Q .

The experiment and the results

This corpus consists of a collection of 2.8 million records from Amazon Books and LibraryThing.com. The test collection contains 300 user queries taken from the LibraryThing discussion forums. We select one key word from the user query as the original query word. Two evaluation index were used as evaluation measures.

nDCG(The normalized of cumulative gain vector

$$nDCG[k] = \frac{DCG[k]}{DCG'[k]} \quad (4)$$

In the formula (4), $DCG[K]$ is discounted cumulative gain, $DCG'[K]$ is ideal discounted cumulative gain. The relevance assessments for the 300 queries of the test collection were collected at the book level from paid human judges. Assessment were made along a four point scale: Excellent, Good, Fair, Non-releva

Precision@10

We have transformed these graded judgements into binary judgements to be able to use standard measures like Precision at rank 10 ($P@10$).

We use lucene for indexing and retrieval, with stop words removal, content words stemming using the porter stemmer. All structure within the books was ignored. And then we do the following contrast experiments:

- Baseline: use the original query to do book search
- WordNet : query expansion based on WordNet
- New: query expansion based on user's friends and interesting_libraries

TABLE 1 COMPARISON OF $P@10$, $nDCG@10$ BETWEEN THREE METHODS (%)

Method	$P@n$	$nDCG@n$
Baseline	0.4714	0.3442
Query expansion based on WordNet	0.5429	0.4661
Query expansion based on user's friends and interesting_libraries	0.7571	0.6279

From table 1, we can see that compared to the Baseline, the performance with our method has been increased $P@10$ by an average of 28.57% and $nDCG@10$ by an average of 28.37%. Compared to the Query expansion based on WordNet, the performance with our method has been increased $P@10$ by an average of 21.42% and $nDCG@10$ by an average of 16.18%.

Found in the experiment, our method's effect is more apparent when users have more interesting_libraries in his personal digital library. Because it shows that the user may also like the type of books that the users who have been marked by the user as his interesting_libraries like, and then the user is likely to search for this kind of books in recent stage.

Conclusions

In this paper, we propose the method of query expansion based on the user's friends and interesting_libraries for social book search. Using the co-occurrence of the user query word and the

book tags word in the personal digital library of the similar users, it extended the original query effectively. The experimental results show that our method has solve two problems well: one is that users can not describe requirements well, for the limitation of professional knowledge; the other is the traditional query expansion technology can't reflect different user's query intention well. Experimental data show that our method greatly improve the satisfaction of user queries and precision.

Acknowledgements

This work was financially supported by the WUYI University Foundation (XL201405).

References

- [1] Koolen M, Kazai G, Kamps J, et al. Overview of the INEX 2012 Social Book Search Track[J]. Clef Evaluation Labs & Workshop Online Working Notes, 2012, 1180.
- [2] Koolen M, Kazai G, Kamps J, et al. Overview of the INEX 2013 Social Book Search Track[J]. Clef Evaluation Labs & Workshop Online Working Notes, 2013, 1180.
- [3] Hall M, Huurdemann H, Koolen M, et al. Overview of the INEX 2014 Interactive Social Book Search Track[J]. Clef Evaluation Labs & Workshop Online Working Notes, 2014, 1180.
- [4] Deveaud, Romain, Eric SanJuan, and Patrice Bellot. "Social Recommendation and External Resources for Book Search." *Focused Retrieval of Content and Structure*. Springer Berlin Heidelberg, 2012. 68-79.
- [5] Gong, Zhiguo, and Chan Wa Cheang. "Multi-term web query expansion using WordNet." *Database and Expert Systems Applications*. Springer Berlin Heidelberg, 2006.
- [6] Song, Min, et al. "Integration of association rules and ontologies for semantic query expansion." *Data & Knowledge Engineering* 63.1 (2007): 63-75.
- [7] Zhou, Dong, Séamus Lawless, and Vincent Wade. "Improving search via personalized query expansion using social media." *Information retrieval* 15.3-4 (2012): 218-242.