



User interest modeling and its application for question recommendation in user-interactive question answering systems

Xingliang Ni^{a,b,c,d}, Yao Lu^{b,c,d}, Xiaojun Quan^c, Liu Wenyin^{a,b,c,d,*}, Bei Hua^{b,d}

^a School of Computer and Information Engineering, Shanghai University of Electronic Power, Shanghai 200090, China

^b Department of Computer Science and Technology, University of Science and Technology of China, Hefei, China

^c Department of Computer Science, City University of Hong Kong, HKSAR, China

^d Joint Research Lab of Excellence, CityU-USTC Advanced Research Institute, Suzhou, China

ARTICLE INFO

Article history:

Received 13 August 2010

Received in revised form 8 May 2011

Accepted 16 September 2011

Available online 26 October 2011

Keywords:

Question answering

Topic model

User interest

ABSTRACT

In this paper, we propose a generative model, the Topic-based User Interest (TUI) model, to capture the user interest in the User-Interactive Question Answering (UIQA) systems. Specifically, our method aims to model the user interest in the UIQA systems with latent topic method, and extract interests for users by mining the questions they asked, the categories they participated in and relevant answer providers. We apply the TUI model to the application of question recommendation, which automatically recommends to certain user appropriate questions he might be interested in. Data collection from Yahoo! Answers is used to evaluate the performance of the proposed model in question recommendation, and the experimental results show the effectiveness of our proposed model.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

With the development of Web 2.0 technologies, various applications have been emerging to help users exchange views, opinions and knowledge. Among these Web 2.0 platforms, the User-Interactive Question Answering (UIQA) system, also known as the Community-based Question Answering (CQA) system such as Yahoo! Answers (Adamic, Zhang, Bakshy, & Ackerman, 2008), Live QnA and BuyAns (Liu, Hao, Chen, & Feng, 2009), has become a popular medium for askers to acquire information. In UIQA systems, users can obtain the information they need simply by asking a question and waiting for answers offered by other users. The data volume of the question-answer (QA) pairs in UIQA systems has been growing exponentially. For instance, there were approximately 23 million questions in Yahoo! Answers in 2008 (Adamic et al., 2008). The accumulated QA pairs can be viewed as a large knowledge base which stores the wisdom of crowds.

In this paper, we propose a new service in UIQA, question recommendation, which automatically recommends new questions to appropriate users in the UIQA System. The service is motivated by two facts. Firstly, the previous question askers may not be satisfied with the limited answers they obtained. By recommending new questions, they may track the new questions and find useful and satisfactory answers. Secondly, the current question askers will be the potential answerers/experts of similar questions in future. One of the reasons is that they may have known the correct answers when they encounter new questions similar to what they previously asked. Hence, we believe the recommendation is useful. Another reason is that askers should be interested in the knowledge related to their questions. They can be the potential answerers when they encounter questions related to the knowledge.

* Corresponding author at: Department of Computer Science, City University of Hong Kong, HKSAR, China. Tel.: +852 2784 4730; fax: +852 2788 8614.
E-mail addresses: xlni@mail.ustc.edu.cn (X. Ni), luyao59@mail.ustc.edu.cn (Y. Lu), xiaoquan@student.cityu.edu.hk (X. Quan), clsuiwy@cityu.edu.hk (L. Wenyin), bhua@ustc.edu.cn (B. Hua).

For the purpose of question recommendation, it is necessary to capture the different interests of users. In the following paragraphs, we refer to the interest of user as user interest without further explanation. Intuitively, user interest can be simply modeled by the bag-of-words (BOW) method, in which the terms in historical questions are used to represent user interest. However, the BOW method cannot recognize the semantic relation between terms. In reality, the questions asked by one user may not always have high term co-occurrence. For example, a computer game fan may ask several questions concerned with different games, while the co-occurrence between these questions might be very low. In this case, the semantic relation between terms should be introduced into user interest modeling. In this paper, we propose a new user interest model, Topic-based User Interest (TUI) model, by introducing a latent topic layer to narrow down the semantic gap between terms. The latent topic layer consists of a series of probabilistic topics to depict questions, such as topics on internet, health, education, traveling, and economics. In particular, the latent topic layer in TUI model is built upon three different types of data (historical questions, participated categories and relevant answer providers). The interest of each user is modeled as a probabilistic distribution over these topics, and the topics which the user is more interested in are with higher probabilities. Based on the proposed user interest model, we rank all the questions in the UIQA system for each user and recommend the top ranked ones to the user.

In the previous research of UIQA system, many efforts have been done on the issue of *answerer recommendation* (Guo, Xu, Bao, & Yu, 2008; Qu et al., 2009). It is worthy to clarify the difference between question recommendation and answerer recommendation. The task of answerer recommendation is to seek a user, who is likely to provide a high quality answer for an unsolved question. The key of answerer recommendation is to find the expert area of each user and recommend the unsolved questions to those expert users. Whereas a good question recommendation system should provide each user with questions he might be interested. Hence, answerer recommendation approaches cannot be applied to solve the problem of question recommendation directly.

The rest of the paper is organized as follows: Section 2 introduces some prior research related to our work. Section 3 states the framework of extracting user interest in the UIQA system. In Section 4, we briefly review the LDA model and illustrate the proposed TUI model. Section 5 gives the question rank method based on different models. The experimental results are shown in Section 6. Finally, we conclude our work and look forward to future work in Section 7.

2. Related work

In this section, we review three lines of research work which are closely related to our work: research on latent topic analysis, learning user interest for search engines and related work in the UIQA system.

2.1. Latent topic analysis

Vector Space Model (VSM) (Yates & Neto, 1999) is one of the most famous and widely used models for text retrieval. However, VSM leads to the problems of high dimension document representation and lack of semantic relationship between dimensions. In VSM synonyms are usually regarded as different words without any similarity. The Latent Semantic Indexing (LSI) model (Deerwester, Dumais, Landauer, Furnas, & Harshman, 1990) gives a way to reduce the dimension of VSM by Singular Value Decomposition (SVD) (Forsythe, Malcolm, & Moler, 1977). SVD clusters the terms in VSM and each cluster is thought to be a synonym set. However, the LSI model lacks of statistical foundation. Hofmann (1999) proposes the Probabilistic Latent Semantic Indexing (pLSI) model to leverage semantic between words in documents. Compared with LSI, the pLSI model has a solid statistical foundation and defines the generation process of documents as a Bayesian network. However, the pLSI model encounters the problem of overfitting because in pLSI the number of parameters grows linearly with the number of training documents. Blei, Ng, and Jordan (2003) propose an alternative generative model of documents to overcome the overfitting problem, namely the Latent Dirichlet Allocation (LDA) model.

2.2. User interest model for personalized search

In the context of search engine, some prior research has been done to extract user interest for personalize search (Chirita, Nejdil, Paiu, & Kohlschutter, 2005; Shen, Tan, & Zhai, 2005; Teevan, Dumais, & Horvitz, 2005). Chirita et al. (2005) characterize different user interest by a set of Open Dictionary Project (ODP)¹ categories selected by each user beforehand, and rerank search results according to their relevance to the selected categories. Shen et al. (2005) achieve the personalized search by exploring user history of search queries and results visited, and rerank the search results. Teevan et al. (2005) propose to build user interest models from various data sources containing not only search historical information but also other information about the user such as documents and emails the user has read and created.

2.3. Related work in the UIQA system

In the UIQA system, the search function can help the users find their interested questions. Many researchers try to improve the accuracy of question retrieval. Jeon, Croft, and Lee (2005a,b) evaluate four different retrieval methods of question

¹ Open directory project. <http://dmoz.org/>.

Download English Version:

<https://daneshyari.com/en/article/515651>

Download Persian Version:

<https://daneshyari.com/article/515651>

[Daneshyari.com](https://daneshyari.com)