Context-Aware Expert Finding in Tag Based Knowledge Sharing Communities

Hengshu Zhu, University of Science and Technology of China and Nokia Research Center, China
Enhong Chen, University of Science and Technology of China, China
Huanhuan Cao, Nokia Research Center, China
Jilei Tian, Nokia Research Center, China

ABSTRACT

With the rapid development of online Knowledge Sharing Communities (KSCs), the problem of finding experts becomes increasingly important for knowledge propagation and putting crowd wisdom to work. A recent development trend of KSCs is to allow users to add text tags for annotating their posts, which are more accurate than traditional category information. However, how to leverage these user-generated tags for finding experts is still underdeveloped. To this end, this paper develops a novel approach for finding experts in tag based KSCs by leveraging tag context and the semantic relationship between tags. Specifically, the extracted prior knowledge and user profiles are first used for enriching the query tags to infer tag context, which represents the user’s latent information needs. Specifically, two different approaches for addressing the problem of tag sparseness in authority ranking are proposed. The first is a memory-based collaborative filtering approach, which leverages non-negative matrix factorization (NMF) to find similar users for alleviating tag sparseness. The second approach is based on Latent Dirichlet Allocation (LDA) topic model, which can further capture the latent semantic relationship between tags. A large-scale real-world data set is collected from a tag based Chinese commercial Q&A web site. Experimental results show that the proposed method outperforms several baseline methods with a significant margin.

Keywords: Collaborative Filtering, Expert Finding, Knowledge Sharing Communities, Question Answering, Topic Models, User-Generated Tags

1. INTRODUCTION

In recent years, researchers have witnessed the rapid development of online Knowledge Sharing Communities (KSCs), such as blogs, discussion boards, and question answering (Q&A) communities. Users can share experiences and exchange ideas with others in such KSCs. Their sharing activities generate a large amount of knowledge and also attract many expert users of each domain to participate. As a result, more
and more people would like to use KSCs for problem solving. Some researchers have found that Q&A content is usually the largest part of content in KSCs (Cong et al., 2008; Feng et al., 2006). However, comparing with the large number of questions, the expert users are still scarce resources in KSCs. As a result, there are a lot of questions without satisfactory answers due to the lack of relevant experts. Thus, how to find the experts for an answer-lacking question becomes an important problem to be addressed.

The problem of expert finding for answer-lacking questions has been well studied. Some of the traditional works leverage content based approaches (e.g., Balog et al., 2006; Liu et al., 2005). In these works, researchers can utilize language models to rank user authority through the question textual distribution in each of the users’ historical records. However, these approaches are usually computationally intensive and are hardly applicable to large-scale data sets. Likewise, some novel KSCs are based on multimedia content and the textual information contained in questions are often not rich enough for building language models (Yeh & Darrell, 2008). Therefore, most of the state-of-the-art works leverage question categories as query inputs to find experts (e.g., Bouguessa et al., 2008; Jurczyk & Agichtein, 2007; Kao et al., 2010). A drawback of these works is that each question can only be classified into one category by them. Actually it is usually difficult to select the best category for a question because a question is usually related to multiple categories. For example, an inexperienced user cannot easily select a better category between “Mobile Device” and “Market” for the question “Where can I buy the Nokia new mobile phone N9?” As a result, the conventional category based expert finding approaches may have poor performances for a multiple-category question.

A recent trend of KSCs is to allow users to add text tags for their questions, such as Tianya Wenda (http://wenda.tianya.cn) and Douban (http://www.douban.com). In these web sites, users can use tags as descriptive labels to annotate the contents they post. To be specific, user can add tags like “N9”, “Mobile Market” and “Where” for the above question. Expert users can check the tags of a given question to decide whether to answer it. Compared with the textual information of question content, user-generated tags are simplified as query inputs and can be utilized on large-scale data sets for expert finding. Moreover, user-generated tags contain richer information of the user needs than category information and can be used for facilitating experts finding. However, because the tags are generated by users but not system, they are usually ambiguous and not regular. Therefore, how to leverage these user-generated tags for expert finding becomes a great challenge. The following motivating examples intuitively illustrate the challenges of using tags for expert finding.

- **Motivating Example 1.** Joy posts a question about the Sony video game console “Play Station” and adds a tag “PS” to annotate the question. However, in many contexts, the tag “PS” may be also referred to the Adobe software “Photo Shop”.
- **Motivating Example 2.** Kate wants to buy a new mobile phone and posts a question with tags “Mobile Phone”, “Market”. However, the latent information needs for Kate are about “Discount” and “Trustable store”.
- **Motivating Example 3.** Joy posts a question about computer devices with a tag “PC”, however, Kate may add a tag “Laptop” for the same question. Actually, the different tags may represent similar meanings.

Inspired by above observations, in this paper, we propose a novel tag based framework for expert finding by inferring the users’ latent information needs and uncovering the semantic relationship between tags. Specifically, we first introduce an effective tag extension method which enriches the original question tags with the question creator’s latent information needs. The latent information needs are modeled by tag concepts extracted from the users’ historical Q&A data and the question creator’s profile,
Related Content

Financial Trading Systems: Is Recurrent Reinforcement Learning the Way?
www.igi-global.com/chapter/financial-trading-systems/28382?camid=4v1a

FMAMS: Fuzzy Mapping Approach for Mediation Systems
www.igi-global.com/article/fmams/95957?camid=4v1a

Designing for Changing Work and Business Practices
www.igi-global.com/chapter/designing-changing-work-business-practices/4218?camid=4v1a

A Novel Prototyping and Evaluation Framework for NoC-Based MPSoc
www.igi-global.com/article/a-novel-prototyping-and-evaluation-framework-for-noc-based-mpsoc/95744?camid=4v1a