Information Processing and Management 53 (2017) 1026–1042

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

On dynamicity of expert finding in community question answering

Mahmood Neshati[a,*], Zohreh Fallahnejad[b], Hamid Beigy[b]

[a] Faculty of Computer Science and Engineering, Shahid Beheshti University, G.C, Tehran, Iran
[b] Computer Engineering Department, Sharif University of Technology, Iran

ARTICLE INFO

Article history:
Received 24 October 2016
Revised 17 February 2017
Accepted 11 April 2017

ABSTRACT

Community Question Answering is one of the valuable information resources which provide users with a platform to share their knowledge. Finding potential experts in CQA is beneficial to several problems like low participation rate of the users, long waiting time to receive answers and to the low quality of answers. Many research papers focused on retrieving the expert users of CQAs. Most of them are taking expertise into consideration at the query time and ignore the temporal aspects of the expert finding problem. However, considering the evolution of personal expertise over time can improve the quality of expert finding. In many applications, it is beneficial to find the potential experts in future. The proper identification of potential experts in CQA can improve their skills and the overall user participation and engagement. Considering dynamic aspects of the expert finding problem, we introduce the new problem of *Future Expert Finding* in this paper.Here, given the expertise evidence in current time, we aim to predict the best ranking of experts in future. We proposed a learning framework to predict such ranking on StackOverflow which is currently one of the most successful CQAs. We examine the impact of various features to predict the probability of becoming an expert user in future time. Specifically, we consider four feature groups; namely, *topic similarity, emerging topics, user behavior* and *topic transition*. The experimental results indicate the efficiency of the proposed models in comparison with several baseline models. Our experiments show that the performance of our proposed models can improve the MAP measure up to 39.7% in comparison with our best baseline method. Moreover, we found that among all of these feature groups, user behaviors have the most influence in the estimation of future expertise probability.

1. Introduction

Expert Finding is one of the challenging problems which attracted a lot of attention in Information Retrieval community in the past few years. The problem of expert finding concerns itself with identifying persons with relevant knowledge on a given topic and ranking them according to their expertise score. Several studies have been conducted to solve this problem in different areas. Most of the existing approaches for expert finding have been proposed to identify experts in the

* Corresponding author.
E-mail addresses: m_neshati@sbu.ac.ir, mahmood.neshati@gmail.com (M. Neshati), zfallahnejad@ce.sharif.edu (Z. Fallahnejad), beigy@sharif.edu (H. Beigy).

http://dx.doi.org/10.1016/j.ipm.2017.04.002
0306-4573/

environments such as academic (Deng, Han, Lyu, & King, 2012; Deng, King, & Lyu, 2008; Hashemi, Neshati, & Beigy, 2013; Neshati, Hashemi, & Beigy, 2014), organizations (Balog, Azzopardi, & de Rijke, 2009; Karimzadehgan, White, & Richardson, 2009), forums (Xu & Ramanathan, 2016), microblogs and social network (Neshati, Asgari, Hiemstra, & Beigy, 2013; Neshati, Hiemstra, Asgari, & Beigy, 2014; Zhang, Tang, & Li, 2007) and more recently question answering communities (Pal, 2015; , Farzan, Konstan, & Kraut, 2011; Riahi, Zolaktaf, Shafiei, & Milios, 2012; van Dijk, Tsagkias, & de Rijke, 2015). In these approaches, associated documents, social interactions and the personal activities of each candidate are considered as expertise evidence. Identification of knowledgeable persons in a specific topic has a great importance in many applications such as assigning a paper to reviewers (Liang & de Rijke, 2016; Neshati, Beigy, & Hiemstra, 2012; Neshati, Beigy, & Hiemstra, 2014), finding the right supervisor in university domain (Alarfaj, Kruschwitz, Hunter, & Fox, 2012), finding expert users in question answering community (Riahi et al., 2012) and expert team formation (Kargar & An, 2011; Lappas, Liu, & Terzi, 2009).

Most of these approaches are taking expertise into consideration in a single snapshot of the environment (at the query time) and ignore the temporal aspects of the expert finding problem. However, people usually change their interests and expertise topics over time. Modeling the evolution of personal expertise over time not only is an important field of Information Retrieval (IR), but it also can improve the quality of expert finding (Balog, Fang, de Rijke, Serdyukov, & Si, 2012; Daud, Li, Zhou, & Muhammad, 2010; Fang & Godavarthy, 2014; Rybak, Balog, & Nørvåg, 2014). Various factors can affect the dynamics of personal expertise. The uptrend or downtrend popularity of a topic, the background of a person and his/her behavior on exploring new areas, and finally, the similarity of topics and the probability of transition between them are some few important factors in modeling expertise in a dynamic environment.

Community Question Answering (CQA) websites such as *StackOverflow*[1] provide users with a useful platform for information sharing. Users can post questions and answers, leave comments, and provide feedback on the quality of others' posts by voting, commenting and selecting the accepted answer to their questions. Successful CQA websites include those general ones such as *Yahoo! Answers*[2] and *Quora*[3], and those domain-specific ones like *StackOverflow* and *Mathematics Stack Exchange*[4]. Finding relevant experts on CQA for a given question/topic can enhance the quality of answers and accordingly improves the user experience and happiness (Riahi et al., 2012). On the other hand, one of the main problems of CQA services is the low participation rate of the users. Developing an expert finding system for intelligently routing newly posted questions can dramatically reduce the ratio of unanswered questions (Asaduzzaman, Mashiyat, Roy, & Schneider, 2013). More importantly, the job listings (SOFJobList, 2016) and CV Search (SOFCandidateSearch, 2016) (i.e. expert search) are two main revenue streams in StackOverflows business model which indicate the importance of an efficient expert finding system for such CQAs.

Finding experts in a CQA is a challenging task because of the following reasons:

- Only a small portion of users are responsible for answering a notable number of questions. It makes finding rising stars and potential experts quite difficult in CQAs (Daud, Ahmad, Malik, & Che, 2015).
- Emerging technologies like mobile programming and even small changes in the specification of some programming technologies can affect the behavior of CQAs' users (Linares-Vásquez, Bavota, Di Penta, Oliveto, & Poshyvanyk, 2014). It means that the capturing and modeling the dynamical aspects of expert finding in CQAs are more crucial in comparison with other expert finding environments like a bibliographic network in which people change their interests more slowly.
- The quality of user generated content in CQAs is not uniform for all users. Thus the quality of expert finding algorithms which depend on the quality of documents (i.e. questions and answers) may be indirectly affected. As a result, there are many research studies on detection and prediction of high-quality content on CQAs like StackOverflow (Ravi, Pang, Rastogi, & Kumar, 2014; Toba, Ming, Adriani, & Chua, 2014; Yao et al., 2015).

CQAs are dynamic environments because of the massive daily posts, joining new users, changing in their activities and interests, emerging new topics and the uptrend or downtrend of topics. For example, emerging technologies may make the existing ones obsolete; consequently, people change their skills and expertise. On the other hand, the success of CQA platforms is highly dependent on the users that can provide high-quality answers to the most difficult questions posted, however, this type of user is very rare (Riahi et al., 2012). As a result, the *prediction* and the nurture of users with topical expertise is becoming an important research topic in recent years (Pal, Farzan, Konstan, & Kraut, 2011; Pal, Harper, & Konstan, 2012a; van Dijk et al., 2015).

While these recent studies (Pal et al., 2011; Pal et al., 2012a; van Dijk et al., 2015) focus on the *early expertise detection* problem on CQA, we introduce the new problem of *Future Expert Finding* which focuses on the ranking of potential topical experts in the future time given expertise evidence in the query time. Most of the expert finding approaches are taking expertise evidence (e.g. experts' documents, experts' actions, etc.) into consideration in a single snapshot of the environment (i.e. at the query time) and neglect the temporal aspects of the expert finding problem. Therefore, the most immediate problem is how to model the dynamic and temporal aspects of the expert finding problem. Modeling the temporal characteristics of CQAs, not only helps us to analyze the changes in an expert's interests and expertise over the life cycle of users

[1] stackoverflow.com
[2] answers.yahoo.com
[3] www.quora.com
[4] math.stackexchange.com