Integrating expert profile, reputation and link analysis for expert finding in question-answering websites ☆

Duen-Ren Liu *, Yu-Hsuan Chen, Wei-Chen Kao, Hsiu-Wen Wang

Institute of Information Management, National Chiao Tung University, Hsinchu 300, Taiwan

ABSTRACT

Question answering websites are becoming an ever more popular knowledge sharing platform. On such websites, people may ask any type of question and then wait for someone else to answer the question. However, in this manner, askers may not obtain correct answers from appropriate experts. Recently, various approaches have been proposed to automatically find experts in question answering websites. In this paper, we propose a novel hybrid approach to effectively find experts for the category of the target question in question answering websites. Our approach considers user subject relevance, user reputation and authority of a category in finding experts. A user's subject relevance denotes the relevance of a user's domain knowledge to the target question. A user's reputation is derived from the user's historical question-answering records, while user authority is derived from link analysis. Moreover, our proposed approach has been extended to develop a question dependent approach that considers the relevance of historical questions to the target question in deriving user domain knowledge, reputation and authority. We used a dataset obtained from Yahoo! Answer Taiwan to evaluate our approach. Our experiment results show that our proposed methods outperform other conventional methods.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Question answering websites, also known as Q&A websites, are becoming a more and more popular knowledge sharing platform. The major reason for this is not only because people can post natural language questions, but also because they can share miscellaneous information or obtain the answers to their questions directly from the website. In addition, people sometimes just need opinions so that they will be more likely to seek for help through question-answering websites. The Yahoo! Answer Taiwan website (http://tw.knowledge.yahoo.com/), also named Yahoo! Knowledge plus, is a community-driven knowledge website; each user can share experience and exchange knowledge by asking and answering questions. On the website, users can browse the questions that other users have asked, search for answers to particular questions, or post questions and then wait for answers. Every solved question has a "best answer". To choose a best answer, the asker can either select an answer as the best answer or set the question–answer pair to a voted stage. The answer that receives the highest vote is chosen as the best answer. Moreover, users can give any solved question an evaluation (positive, neutral, or negative) regarding whether the question–answer pair is useful or not.

Unfortunately, such question answering mechanisms may run into some issues. When the quantity of questions waiting to be solved grows quickly, some questions may be skipped by users who can answer, and askers will waste a lot of time to

obtain answers. Even worse, askers may not obtain correct answers from appropriate experts. Consequently, knowledge sharing through question answering websites is interfered with such issues. Therefore, it is essential to automatically find appropriate experts for target questions, so as to shorten the waiting time, increase the quality of answers, and thus enhance the effectiveness of knowledge sharing.

Recently, various approaches have been proposed to automatically find experts in question answering websites. The link analysis approaches, including HITS (Kleinberg, 1999) and PageRank (Page, Brin, Motwani, & Winograd., 1998), have been adopted to find experts. Jurczyk and Agichtein (2007) adopt the HITS algorithm for author ranking. They represent the relationship of asker and answerer as a link network and calculate each user's hub and authority value, and then rank users according to their authority values. Zhang, Ackerman, and Adamic (2007) propose a PageRank-like algorithm called "ExpertiseRank" to rank experts in an expertise network considering how many people are involved and who has helped whom. Moreover, the user (expert) profiling approach (Liu, Croft, & Koll, 2005; Zhang, Ackerman, Adamic, & Nam, 2007), which built expert profiles from the contents of expert's questions and answers, is adopted to find experts without considering the reputations of experts and their authority values derived from link analysis.

In this paper, we propose novel methods to find appropriate experts to answer a given target question. A hybrid approach is proposed to effectively find experts for the category of the target question in question-answering websites. Different from the conventional approaches that only consider user profile or user authority, our approach considers user subject relevance, user reputation and authority of a category in finding experts. A user's subject relevance denotes the relevance of a user's domain knowledge to the target question. A user's domain knowledge for a specific category is represented as a user knowledge profile derived from the content and quality measures (e.g. voting/evaluation factor) of the user's historical question–answers in that category. A user's reputation in a category is derived from the user's historical question-answering records based on the ratio of the user's answers being adopted as best answers in that category, while user authority in a category is derived by applying link analysis to the category-based asker-answerer network.

In Yahoo! Answer, the range of a category domain is not small enough, so the domain experts may not be appropriate to answer the target question. We further extended our proposed category-based approach to develop a question dependent approach that considered the relevance of past questions to the target question in deriving user domain knowledge, reputation and authority. Our approach can enhance the quality of recommending experts through matching the content by relevance and taking the user reputation and user authority into account. The experiment result shows that our approach is better than other approaches that only use link analysis or user profile. Moreover, the question-dependent approach leads to a better result of finding experts for a target question.

The remainder of this paper is organized as follows. Section 2 presents the related literatures. Our proposed approach for expert finding is illustrated in Section 3. In Section 4, experiment evaluations are conducted to compare our approach with other methods. Conclusions and future research directions are finally presented in Section 5.

2. Related work

The widely used method to find the experts in community-driven question answering websites is link analysis or social network analysis. This method comprises of building a user social network first, and then using some kind of propagation algorithm to calculate each user's authority. Jurczyk and Agichtein (2007) build a link network based on the relation of asker and answerer between users in question–answer portals such as Yahoo! Answers. There exists an edge from user i to user j if user j has answered a question posted by user i. The link network is a multigraph in which multiple edges may exist between the two users if user j has answered several questions posted by user i. The authors do not consider the votes or best answers received by the answerer, and thus each edge is assigned the same weight. To discover the authorities in a particular category, the questions and their answers of that category are used to build the graph, and then the HITS algorithm is used to compute each user's authority value. Zhang, Ackerman, and Adamic (2007) address the issue of expert finding in an online help seeking community – the Java forums. An expertise (post-reply) network is constructed by viewing each user as a node and creating a directed edge from the user making the post to those users who replied to it. A PageRank-like algorithm called ExpertiseRank is adopted for expertise ranking. The edge of the network can be weighted according to the number of times one replies another. In addition, social network analysis has also been applied to derive users' reputations in a user-interactive question answering system – *CuiteAid* (Chen, Zeng, & Liu, 2006).

Besides the social network or link-based analysis, there are some other ways to find experts. Although link analysis methods have been adopted to find top-*K* users in a ranked list based on their expertise scores or authority values on subjects of interest, it is difficult to set appropriate values for *K*. Bouguessa, Dumoulin, and Wang (2008) address such issue and propose a probabilistic approach for automatic identification of authoritative actors in QA forums such as Yahoo! Answers. The proposed approach discriminate authoritative and non-authoritative users based on a mixture of gamma distributions. Zhang, Ackerman, and Adamic (2007) and Zhang, Ackerman, Adamic, and Nam (2007) address the issue of expert-finding in an online help seeking community – the Java forums. The vector space model in information retrieval (Manning, Raghavan, & Schütze, 2008) is used to represent the question and user profiles as term vectors. The proposed expert-finding method not only compares the similarity of questions and user profiles, but also considers the differences of expertise level, posting time of query, and the number of replies (status) to questions. Liu et al. (2005) adopt the language models in information retrieval (Manning et al., 2008) to build expert-profile models from the content of