Information Processing and Management 000 (2016) 1–15

Contents lists available at ScienceDirect

Information Processing and Management

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

Identifying and predicting the desire to help in social question and answering

Zhe Liu[a,*], Bernard J. Jansen[b]

[a] IBM Almaden Research Center, San Jose, CA, USA
[b] Social Computing Group, Qatar Computing Research Institute, Doha, Qatar

ARTICLE INFO

Article history:
Received 3 September 2015
Revised 12 April 2016
Accepted 2 May 2016
Available online xxx

Keywords:
Social Q&A
Question routing
Online information seeking
Social networks
Twitter
Predictive model

ABSTRACT

The increasing volume of questions posted on social question and answering sites has triggered the development of question routing services. Most of these routing algorithms are able to recognize effectively individuals with the required knowledge to answer a specific question. However, just because people have the capability to answer a question, does not mean that they have the desire to help. In this research, we evaluate the practical performance of the question routing services in social context by analyzing the knowledge sharing behavior of users in social Q&A process in terms of their participation, interests, and connectedness. We collect questions and answers over a ten-month period from Wenwo, a major Chinese question routing service. Using 340,658 questions and 1,754,280 replies, findings reveal separate roles for knowledge sharers and consumers. Based on this finding, we identify knowledge sharers from non-sharers a priori in order to increase the response probabilities. We evaluate our model based on an analysis of 3006 Wenwo knowledge sharers and non-sharers. Our experimental results demonstrate knowledge sharer prediction based solely on non-Q&A features achieves a 70% success rate in accurately identifying willing respondents.

1. Introduction

Social networking sites have been widely adopted for online communication. Besides using these sites for relationship formation and maintenance (Zhang, Jansen, & Chowdhury, 2011), many people also rely on social networking sites (SNS) for seeking information (Jansen, Sobel, & Cook, 2011; Morris, Teevan, & Panovich, 2010). Although not intentionally designed for questioning and answering, people enjoy expressing their information needs in natural language questions on SNS (Liu & Jansen, 2012; Paul, Hong, & Chi, 2011; Shah, Oh, & Oh, 2008), a behavior referred to as social questioning and answering (social Q&A). By making use of social interactions online (Evans & Chi, 2008), social Q&A techniques provide individuals with simpler and more personalized search experiences over conventional information retrieval methods. Due to such advantages, social Q&A sites have attracted researchers' attention and has motivated the creation of models and tools to facilitate the social information seeking process (Jansen, Zhang, Sobel, & Chowdury, 2009).

Among the proposed methods are several question routing algorithms that mostly involve expert finding techniques to solve the problem of nonguaranteed responses in a social context. Many studies claimed that by routing questions to a larger audience base can effectively reduce the number of unanswered inquiries. However, we argue that even after finding people with required knowledge, we still do not know if they have the desire to help. In other words, we believe that it is

* Corresponding author.
E-mail addresses: liuzh@us.ibm.com, lztoby@gmail.com (Z. Liu), jjansen@acm.org (B.J. Jansen).

http://dx.doi.org/10.1016/j.ipm.2016.05.001
0306-4573/

very important to validate the feasibility of the question routing mechanism by measuring its practical effectiveness in real world circumstances. In addition, it would be also beneficial to develop methods to determine if an individual is willing to share his/her knowledge via responding to other's questions.

To address these issues, we analyze questions and answers posted during a 10-month period on Wenwo, a Chinese question routing service based on microblogging sites. We evaluate the real world performance of Wenwo from three perspectives, which are user's (a) participation (b) interest, and (c) connectedness. These measurements are chosen intentionally as they have been adopted as means to explore the patterns of user engagement in question answering communities (Adamic, Zhang, Bakshy, & Ackerman, 2008; Gyongyi, Koutrika, Pedersen, & Garcia-Molina, 2008; Shah, Oh, & Oh, 2009).

Next, we develop a predictive user engagement model based on a number of non-Q&A features, including: user profile, posting behavior, language style, and social activities. We did not rely on any Q&A related feature in this study in order to avoid the cold-start problem. Through our model, we find that less popular but more interactive individuals are more willing to respond to others in social Q&A environments. In addition, we also note that the psycho-linguistic characteristics of an individual's microblogging posts, such as their usage of verbs, pronouns, and cognitive expressions, are also indicative of their roles in the social Q&A sites. This research is beneficial because it provides a more in-depth understanding of the social Q&A process, especially the characteristics of people who are willing to engage in knowledge sharing. The findings can also be viewed as design guidelines for future question routing systems based on both capability and desire of the potential respondents.

In the next section, prior studies on social Q&A related to this work are presented. Research questions are proposed in Section 3 and followed by our data collection process. We examine the knowledge sharing behaviors and patterns among strangers in social Q&A in Section 5. Based on our results, we build a classifier of potential knowledge sharer in Section 6. We conclude in Section 7 with some future work.

2. Literature review

2.1. Social Q&A

Defined by Morris et al. (2010), social Q&A is the process of discovering information online with the assistance of social resources. It lies between the boundaries of technical and human-powered information seeking models. The Social Q&A technique outperforms the traditional information-seeking methods (e.g. search engine and online databases, etc.) for both more personalized search experience and results. Studies investigating motivations for participation in the social Q&A process suggested that people primarily search socially due to their trust in friends over strangers (Liu & Jansen, 2012; Morris et al., 2010; Yang, Morris, Teevan, Adamic, & Ackerman, 2011), weak beliefs on search engine performances (Morris et al., 2010), as well as non-urgent information needs (Teevan, Collins-Thompson, White, Dumais, & Kim, 2013).

2.2. User engagement in knowledge sharing

Although the perception of an individual's information needs in the social Q&A process is critical, it is equally important to measure the level of user engagement in knowledge sharing, since social Q&A services thrive on users' active involvement (Shah et al., 2009). Studies measured individual's knowledge sharing on Q&A sites mainly from three different perspectives: user engagement, user interest, and user connectedness. In terms of user engagement, Nam, Ackerman, and Adamic (2009) analyzed the Knowledge-IN website and found a significant separation between asker and answerer roles, with very little within-category reciprocity. While analyzing user engagement in both Yahoo!Answers and Google Answers, Shah et al. (2008) found that the majority of the population in Yahoo!Answers participated in both posting questions and answers, whereas in Google Answers, there were many one–time consumers and a small number of contributors. Gazan (2007) divided questioners into two types according to their involvement in follow-up discussions on Answerbag: Seekers and Sloths. Seekers tended to interact with others about their questions, while Sloths post their question and interact no further. Paul et al. (2011) noted that the majority of questions posted on Twitter received no response. When sending questions to strangers for help, Nichols and Kang (2012) indicated that less than half of the questions received answer.

Concerning user's topical interests in the process of Q&A, Liu and Jansen (2013) studied the questions posted on Sina Weibo, the largest Chinese microblogging site. They noticed a relatively higher response rate for questions posted on Sina Weibo than questions posted on Twitter. They also found that the question's topic could effectively affect its response rate. For instance, they noticed that questions on the topics of Entertainment, Society, Computer, etc., received fewer responses as compared to questions from the other categories. Lampe, Gray, Fiore, and Ellison (2014) analyzed a set of public status updates posted to Facebook and stated that mobilization requests got more responses than other kinds of posts. Adamic et al. (2008) analyzed a set of question-answer pairs from Yahoo!Answers and found different distributions of questioner/answerer overlap across topical categories.

Lastly, from the perspective of measuring user connectedness in the process of Q&A, Gyongyi et al. (2008) analyzed the interactions between individuals on Yahoo!Answers by analyzing undirected bipartite graphs generated by asker and answers under each topical categories. Their results showed that the vast majority of users were connected within a single large community on Yahoo!Answers. Zhang, Ackerman, and Adamic (2007) did the same kind of study by using a bow-tie structure analysis. They found that more than half of the users on Yahoo!Answers usually only ask questions without

Please cite this article as: Z. Liu, B.J. Jansen, Identifying and predicting the desire to help in social question and answering, Information Processing and Management (2016), http://dx.doi.org/10.1016/j.ipm.2016.05.001