Predicting Potential Responders in Social Q&A Based on non-QA Features

Zhe Liu

College of Information Sciences and Technology

The Pennsylvania State University

zul112@ist.psu.edu

Bernard J. Jansen

College of Information Sciences and Technology

The Pennsylvania State University

jjansen@acm.org

Abstract

Given the recent advancement of online social networking technologies, social question and answering has become an important venue for individuals to seek and share information. While studies have suggested the possibilities of routing questions to potential answerers for their help and the information provided, there is little analysis proposed to identify the characteristics that differentiate the possible responders from the nonresponders. In order to address such gap, in this work we present a model to predict potential responders in social Q&A using only non-QA-based attributes. We build the classifier using features from two different aspects, including: features extracted from one's social profile and style of posting. To evaluate our model, we collect over 20,000 questions posted on Wenwo, a social Q&A application based on Weibo, along with all their responders. Our experimental results over the collected dataset demonstrate the effectiveness of responder prediction based on non-QA features and proposed potential implications for system design.

Author Keywords

Social question and answering, social Q&A, social search, social networks, information seeking, Weibo

ACM Classification Keywords

H.3.5 [Information Storage and Retrieval]: Online Information Services; J.4 [Social and Behavioral Sciences]: Sociology

Introduction

With the rapid development of Web 2.0 technologies, social networking sites (SNSs), such as Facebook, Twitter, and Google+, have been widely adopted for online communication. Besides using it for relationship formation and maintenance, many people also rely on SNSs for questioning and answering, to which we refer as social Q&A. As social Q&A is gaining popularity as an alternative source to find information, many researches and implications have been actively performed in this area. Studies have suggested the possibilities of routing

Examples of social Q&A:

- Anyone knows how to fix blinking monitor?
- Can anyone recommend any good places to go for afternoon tea in central London?
- #healthadvice Twitter I need help - how can I kick a cold/flu illness quickly?

Screenshots of Social Q&A on Twitter

questions to potential answerers based on their expertise [1]. However, another important question is that even after we identified people with required knowledge, will they respond to questions sent by the social Q&A system? Can we predict the potential responders of a routed question by using their non-QA characteristics? Here by "non-QA" we mean features related only to one's social activities, such as his/her SNS profile, as well as posting behaviors. QA-related attributes such as the difficulty of the question asked, and the knowledge and expertise level of the answerer, are not considered in this study.

To answer these questions, we analyzed a number of non-QA features extracted from a large social Q&A dataset. We collected the data from an automatic question routing application based on Weibo. Using those selected features we built a classifier to predict whether a person will answer the question directed to him/her or not. We found that answerer's profile and style of posting to some extent can reveal their other's questions. In addition, we also noticed that the characteristics of the questioner, such as gender, popularity and activeness on SNS, are not that important to the potential responders in social Q&A.

We think this research will benefit the HCI community by providing a more in-depth understanding of social Q&A, especially what are the characteristics of people voluntarily answering stranger's questions. It can also be viewed as a design guideline for future question routing systems based on SNSs.

RELATED WORK

Many of the prior studies in social Q&A investigate factors that motivate people to seek information via

social platforms [2], highlighting elements such as inter-personal trust and relatively easier access to timely information. Besides, other research has been conducted to understand the taxonomy of questions asked on SNS. [3] conducted their study using Broder's [4] taxonomy of traditional search. The authors presented a social search model of user activities before, during, and after search and proved the value of social interactions in information seeking tasks Through comparisons between social Q&A, traditional search and real time search, [5] found that social information seekers show more personalized requirements and more timely needs.

In addition to studies of questions asked in social Q&A, there are other works examining the answers received. Analyzing real world tweets, [6] noted that the majority of questions received no response on Twitter. [7] conducted analytical studies on Weibo posts and summarized eight factors that affect the response rate, including the number of followers, the number of @mentions, the usage of hashtags, etc. [1] proposed a question routing framework based on the evaluation of user's historical QA behavior, expertise and availability. There are two studies most related to our work. Through online experiments, [8] explored the feasibility of users responding to questions sent by strangers. They found that less than half of the people will answer to questions posted by strangers, however, they failed to indicate the characteristics of those responders. [9] offered a more in-depth analysis on potential answerers by leveraging users' non-QA social activities. Through their analysis of an inter-organizational CQA site, they found that some of the non-QA features can effectively predict the likelihood of one's answering of others' questions. The only limitation of their work is that given
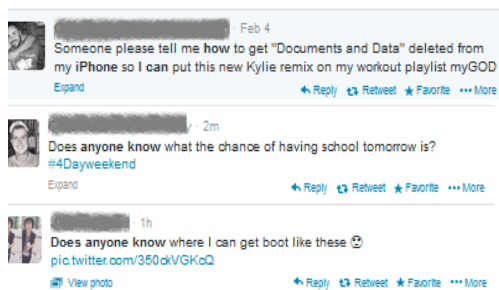
Screenshots of Wenwo

the differences between organizational CQA site and SNSs, their framework may meet with some difficulties when generalized to social Q&A context. Considering the lack of study and the importance of identifying potential responders in social Q&A, we conducted this work to understand what kinds of users are more willing to respond to stranger's questions on SNSs. Given the very low response rate reported in [8], we believe that without a more comprehensive evaluation of factors affecting whether or not a question will be answered, SNS's power in social information seeking may never be achieved.

Research Question

This study was motivated to predict the potential responders in social Q&A. In order to achieve this goal, we addressed the following research questions:

- Given a potential answerer to a stranger's question (predicted based on his/her relevant expertise and experience), can we tell whether he/she will respond using only non-QA features?

- If so, how much can those non-QA features infer in predicting potential responders?

- What are the most effective non-QA features in predicting potential responders in social Q&A?

Research Design

Data Collection
In this study, we collected data from a social Q&A application based on Sina Weibo[1], called Wenwo[2]. First, Sina Weibo is China's largest microblogging site,

---

[1] http://www.weibo.com

[2] http://www.wenwo.com
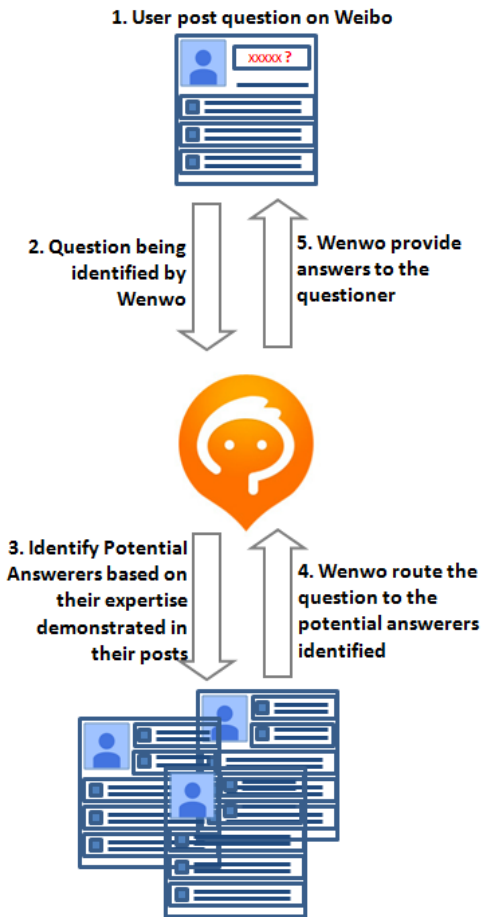
attracted nearly 30 million users within only three years, with a current average of 10 million Weibo statuses published every day[10]. Weibo essentially adopts the same operating concept and provides very similar functions to its users as Twitter. Second, Wenwo is social Q&A application based on Chinese microblogging sites, including Sina Weibo and Tencent Weibo. Wenwo aimed to route one's question to the most appropriate experts on SNSs identified using some internal machine learning mechanisms. In addition to presenting the answers received, under most cases, Wenwo also informed the users to whom the question has been routed. This allows us to know who has responded a question and who hasn't, and enables us to build the classifier with both positive and negative instances.

We first crawled 21,753 questions from Wenwo in a month period from September 18th to October 18th, 2013, together with all their replies. On average each question received 2.82 answers. We eliminated all answers received from either Tencent Weibo or Wenwo website, kept only answers posted through Sina Weibo. In addition, to ensure the quality of our classification, we also removed questions without indicating non-responders. In final, we kept 20,479 answers posted by 9,840 distinct responders through Sina Weibo, as well as 189,818 non-responders. To avoid classification bias and get better results, we balanced our data set by randomly selecting 20,479 non-responders for the same question set.

Then using Sina Weibo's API, we collected the profile information of all questioners, responders and non-responders. Due to the limitation of Sina Weibo's API, we also wrote a script to crawl the most recent 200 posts of each collected responder and non-responder.

1. User post question on Weibo

2. Question being identified by Wenwo

5. Wenwo provide answers to the questioner

3. Identify Potential Answerers based on their expertise demonstrated in their posts

4. Wenwo route the question to the potential answerers identified

Non-QA Features

In this section, we discuss how we can automatically determine whether a user will respond to a stranger's question routed to him/her based on their social activities. Since Wenwo has already helped us identify the most appropriate experts based on their knowledge levels, in this study we only proposed non-QA features from two perspectives: the profile-based ones, and the posting style-based ones. All features explored in this study are listed in Table 1, except the verification type. Unlike Twitter, Sina Weibo defined different verified types of users, including: celebrities, business, government, education, media, intro-level expert, and advanced-level expert. The profiled based features are relatively straight forward. The posting style-based features can be further divided into two categories: comments and likes a user gets in all his/her recent 200 posts. We assumed this features can indicate the popularity and sociability of a user on Weibo. In addition, we also incorporated features like the number and the percentage of retweeting posts, and posts containing at least one mention, one url, one location and one hashtag. We believed these posting patterns can somehow reveal one's personal characteristics that can lead them to answer other's questions. All features were extracted for both questioner and potential answerers (both responders and non-responders).

Relevant Feature Selection

Before conducting the classification, we performed feature selection using the Chi-square evaluator implemented in WEKA with default parameters and 10-fold cross validation. As one the most effective approaches of feature selection for classification, the Chi-square method helped us to identify the most informative and relevant features in the classification

process and to reduce noise from irrelevant or inaccurate features. The top 10 selected features of all three datasets are shown on the side bar of the next page.

Features	Description
Profile-based	
TD	Total days on Weibo
AVG_PPD	Average posts per day
FD	Number of friends
FL	Number of followers
BIFL	Number of bi-directional relationship
FV	Number of favorite
GD	gender
VF	Whether is a verified user or not
VFT	Verified type (e.g. celebrities, business, etc.)
DP	Whether has a description
Posting Style-based	
AVG_RT	Average number of times a post being retweeted
AVG_CM	Average number of times a post being
AVG_LK	Average number of times a post being liked
NUM_RT / PT_RT	Total number / percentage of posts retweeting
NUM_MT/ PT_MT	Total number / percentage of posts mentioning
NUM_URL/PT_URL	Total number / percentage of posts sharing urls
NUM_LC / PT_LC	Total number / percentage of posts sharing
NUM_HT / PT_HT	Total number / percentage of posts using

Table 1. Profile and posting style-based features

From that Table, we saw that none of the top 10 features selected were from the questioner's perspective, which means that the characteristics of the questioner, such as gender, popularity and activeness on SNS, will have very limited effect on the responders' decisions to provide their answers. In addition, we

#of questions	21,753
# of responds	20,479
# of non-responds	20,479
# of distinct questioners	13,482
# of distinct responders	9,840
# of distinct non-responders	20161

Descriptive Statistics of Data Collected
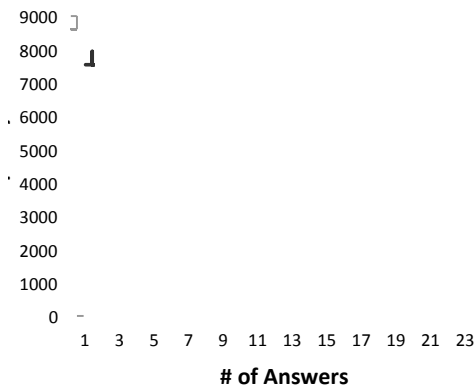
Distribution of Answers Received

found that both profile-based features, including: answer's total days on Weibo, average posts per day, number of followers, friends, and posting style-based features, including: average number and percentage of times a post being retweeted, commented, and the percentage of posts mentioning others, sharing urls.
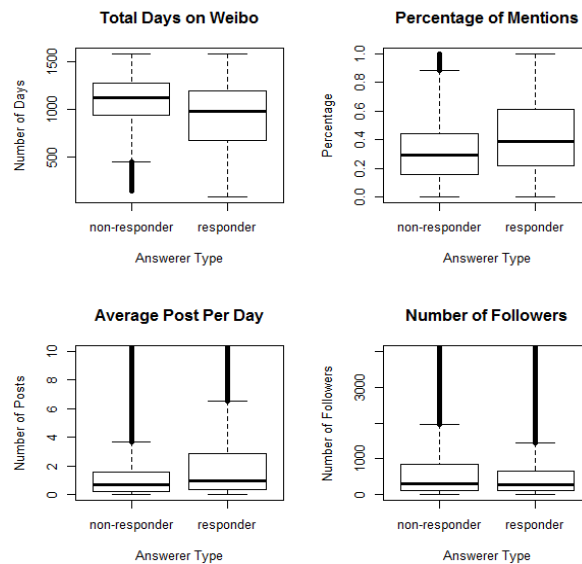

Figure 1. Boxplot of top 4 selected features of responder and non-responder classes

We next further demonstrated the discrimination power of the selected power, we used the boxplot representation to show the distribution of the top-ranked user-based features for both responders and non-responders (due to space limit, we only demonstrated the top 4 features selected). The discrimination capacity of those features can be estimated based on the lower quartile, median and

upper quartile values as presented in the boxplot. As observed in Figure 1, we found that new and more active and interactive users were more willing to respond to stranger's questions, while users with more followers actually are relatively less willing to help others in general.

Classification of Responders and Non-responders
Based on all features proposed, we next built a binary classifier to automatically label cross and within-ideological relationship in an interaction. We tested classification algorithms implemented in Weka[3], such as naïve Bayes, Bayes network, SVM and J48 decision tree etc, all with default values for all parameters using 10-fold cross-validation. Since the goal of our study is not comparing different learning algorithms, we only reported the best results obtained. We adopted the evaluation metrics provided by Weka, including true-positive rate (TP), false-positive rate (FP), precision, recall, F-measure, and ROC Area to assess the performance of our classifier. The majority induction algorithm, which simply predicts the majority class, was applied to determine the baseline performance of our classifier. Since we have balanced our data into 50% positive samples versus 50% negative ones as mentioned in the data collection section, with this approach we get a baseline accuracy of 50%.

For all three classification tasks using only features from the questioners, answerers, and questioners plus answerers, the J48 decision tree algorithm achieved the best performance. A summary of the results were shown in Table 2. Again, we saw that questioner-related features showed very limited effect on the

---

[3] http://www.cs.waikato.ac.nz/ml/weka/

Feature	Mean	StdDev
A_ TD	984.00	352.76
A_ PT_MT	0.38	0.26
A_ AVG_PPD	1.97	3.37
A_ FL	3367.01	41786.40
A_PT_URL	0.22	0.30
A_FD	397.06	454.94
A_ AVG_RT	2.42	95.46
A_ NUM_URL	34.79	52.74
A_NUM_MT	66.98	50.78
A_ AVG_CM	1.62	30.41

Top 10 selected features using Chi-sqaure

prediction task as compared to the baseline. However, prediction based potential answerer's profile and posting patterns demonstrated the best results. Given that only non-QA related features were included in our classifier, we deem our classifier achieved satisfactory performance in identifying responders from non-responders.

Features	TP	FP	Prec	Reca	F1	RO
Questioners	0.5	0.4	0.57	0.56	0.5	0.5
Answerers	**0.7**	**0.2**	**0.76**	**0.76**	**0.7**	**0.7**
Questioners+Answer	0.7	0.2	0.76	0.76	0.7	0.7

Table 2. Classification Results

Discussion and Implication

In this study we proposed an approach to automatically predict responders to stranger's questions routed in social Q&A context. We adopted only non-QA features while building our classification model. We trained and tested our classifier using profile and posting style-based features from both the questioner and answerer's perspectives. We showed that the characteristics of a questioner can hardly affect the answerer's decision on whether to respond or not. However, we further demonstrated that newer and more active users are more willing to respond to routed questions. We believe our study is a first step in designing and implementing automatic question routing tools in social context. For future work, we will extend from this study to include QA-related features, such as expertise, availability, and social relationship into our predicting model.

Reference

[1] Li, B. and I. King. Routing questions to appropriate answerers in community question answering services. in Proceedings of the 19th ACM international conference on Information and knowledge management. 2010. ACM.

[2] Morris, M.R., J. Teevan, and K. Panovich. What do people ask their social networks, and why?: a survey study of status message q&a behavior. in Proceedings of the SIGCHI conference on Human factors in computing systems. 2010. ACM.

[3] Evans, B.M. and E.H. Chi. Towards a model of understanding social search. in Proceedings of the 2008 ACM conference on Computer supported cooperative work. 2008. ACM.

[4] Broder, A. A taxonomy of web search. in ACM Sigir forum. 2002. ACM.

[5] Liu, Z. and B.J. Jansen, Almighty Twitter, what are people asking for? Proceedings of the American Society for Information Science and Technology, 2012. 49(1): p. 1-10.

[6] Paul, S.A., L. Hong, and E.H. Chi. Is Twitter a Good Place for Asking Questions? A Characterization Study. in ICWSM. 2011.

[7] Liu, Z. and B.J. Jansen. Factors influencing the response rate in social question and answering behavior. in Proceedings of the 2013 conference on Computer supported cooperative work. 2013. ACM.

[8] Nichols, J. and J.-H. Kang. Asking questions of targeted strangers on social networks. in Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. 2012. ACM.

[9] Pan, Y., et al. To answer or not: what non-qa social activities can tell. in Proceedings of the 2013 conference on Computer supported cooperative work. 2013. ACM.

[10] Sina Weibo Statistics Available from: http://pcedu.pconline.com.cn/softnews/yejie/1205/2782402.html.