Is Twitter a Good Place for Asking Questions? A Characterization Study

Sharoda A. Paul[1], Lichan Hong[1], and Ed H. Chi[2]

[1]Palo Alto Research Center, Palo Alto, CA, USA. [2]Google Research, Mountain View, CA, USA
{spaul, hong}@parc.com, chi@acm.org

Abstract

People often turn to their social networks to fulfill their information needs. We conducted a study of question asking and answering (Q&A) behavior on Twitter. We found that the most popular question types were rhetorical and factual. Surprisingly, along with entertainment and technology questions, people asked personal and health-related questions. The majority of questions received no response, while a handful of questions received a high number of responses. The larger the askers' network, the more responses she received; however, posting more tweets or posting more frequently did not increase chances of receiving a response. Most often the 'follow' relationship between asker and answerer was one-way. We provide a rich characterization of Q&A in social information streams and discuss implications for design.

Introduction

With the rising popularity of online social networking sites like Facebook and Twitter, people are turning to their social networks to find answers to questions on a variety of topics (Morris et al. 2010; Efron & Winget 2010). Morris et al. (2010) conducted a survey of Microsoft employees' social Q&A behavior. They reported on motivations for asking and answering questions, the types and topics of questions asked, and the timing and usefulness of received responses. Inspired by their findings, we wondered: would a large-scale study of Q&A behavior of a more general population of social network users, using different methods, reveal similar or different results? We decided to study the Q&A behavior occurring on Twitter during users' normal use of the service. We were interested in examining: (1) What kinds of questions are Twitter users asking their friends? (2) Are users receiving responses to the questions they are asking? (3) How does the nature of the underlying social network affect Q&A behavior?

We conducted a sampled study of the questions asked on Twitter. First, we obtained 1.2 million tweets from the public Twitter stream and processed them to obtain candidate questions. Then, we used Amazon Mechanical Turk to identify 1152 questions from these candidate tweets. Finally, we analyzed these questions and their responses to provide a rich characterization of naturally-occurring Q&A behavior in an online social network.

Related Work

When searching for information online, people combine online sources with social sources of information (Evans & Chi 2008; Evans et al. 2010). Social Q&A refers to people asking questions to their social networks, either on social networking sites or using social search engines (Horowitz & Kamvar 2010). An emergent social Q&A behavior is people asking questions to their friends via their status updates on social networking sites. Morris et al. (2010) explored this phenomenon through a survey of 624 Microsoft employees. They found that people asked subjective questions to their friends whom they trusted to provide tailored responses. The researchers analyzed 249 self-reported questions from Facebook and Twitter by type and topic. They found that the majority of the questions were asking for recommendations and opinions, and pertained to technology and entertainment. Interestingly, participants were uncomfortable asking questions about health, religion, and dating as they were too personal.

Twitter is a popular micro-blogging social network. Even though Twitter is not designed for Q&A, people ask and answer questions during their natural use of Twitter. Efron and Winget (2010) analyzed 100 question tweets from the information retrieval community on Twitter to develop a taxonomy of questions in micro-blogging environments. They found similar types of questions as Morris et al. (2010) but did not report the relative proportions of different types of questions.

Q&A systems based on social networks are becoming popular. Recently, Facebook introduced a social Q&A feature called Facebook Questions (Facebook 2010). Questions asked on Facebook Questions appear in the Newsfeed of friends of the asker, but are also visible to all Facebook users. This helps the asker to not only leverage the expertise of friends , but also to tap into the collective knowledge of the Facebook community. Quora, an emerging social Q&A site, provides a knowledge repository of questions and answers created and organized by the community (Wadhwa 2011).

Methods

There were several challenges in identifying questions on Twitter. First, tweets are short (140-characters at most) and have a unique convention for the use of language. Second,

tweets often contain little context, since they are targeted at friends. Hence, from a third person perspective, it is often hard to determine which tweets are questions asked by a user. Third, in selecting question tweets, we were concerned about introducing our own bias regarding what is a question and what is not. Also, manually classifying thousands of tweets is time-consuming and not scalable.

Mechanical Turk has been used by researchers to collect large-scale human-labeled data (Kittur et al. 2008) and could help us deal with these challenges. We crowd-sourced the classification of tweets as questions, leveraging the scalability and low-cost of Mechanical Turk. To reduce the number of tweets to be classified by Turkers to find a reasonable number of questions, we first identified a set of candidate tweets that were likely to be questions, as described in the following section.

Identifying Candidate Questions

We collected a random sample of 1.2 million tweets from the public Twitter stream using Twitter's API. Then, we used several heuristics to filter these tweets to obtain candidate tweets. First, we removed retweets and tweets directed to a specific user (i.e., starting with @user) as we were interested only in questions that were asked by users to their social network in general. Second, we removed tweets containing URLs as it was difficult to judge whether the tweets were questions or not without inspecting the URLs. Third, we removed tweets containing obscene words since we did not want to expose Turkers to such content. Fourth, we removed non-English tweets.

Finally, we removed tweets that did not contain a '?' as previous studies (Morris et al. 2010) have found that most (81.5%) questions asked on social networking sites contained '?'. Also, rule-based methods of using only '?' to identify questions in online content achieves over 97% precision in detecting questions (Cong et al., 2008).

Tracking Responses to Questions. We tracked all responses to each candidate tweet. From the metadata of the candidate tweet, we extracted the name of the sender. We used Twitter's tracking API to capture all tweets containing the keyword "@sender" for the next 7 days. If the candidate tweet was later determined by Turkers to be a question, we identified all responses to the candidate tweet.

Classifying Questions Using Mechanical Turk

We designed a human-intelligence task (HIT) on Mechanical Turk to select question tweets from the candidate tweets. Here we briefly describe the Mechanical Turk study; details are provided in Paul et al. (2011).

Our HIT instructed Turkers to tell whether each of the tweets presented to them was a question posted by a Twitter user to his followers with the expectation of receiving a response. Each candidate tweets could be classified as *question*, *not question*, or *not sure,* and was rated by two Turkers. If both Turkers rated a tweet as a *question*, we then classified the tweet as a true question.

Quality Control. Turkers were required to be Twitter users; this helped ensure that they were familiar with the language of Twitter. Further, to deal with the problem of spam responses (Kittur et al. 2008), we inserted some control tweets along with the candidate tweets. Control tweets were tweets that were obviously *questions* or *not questions*. Each HIT had 25 tweets; 20 candidate tweets and 5 control tweets. For our analysis, we only included data from Turkers who rated all control tweets correctly.

Results

29% of Turkers rated all control tweets correctly and they together rated 4140 tweets. 32% (1351/4140) tweets were rated as questions by both Turkers. For our analysis, we examined these 1351 tweets.

Question Types and Topics

We coded the question tweets selected by Turkers by type and topic using the coding scheme from Morris et al. (2010). Two researchers independently coded the tweets and then discussed their differences to reach consensus. We modified the coding scheme to accommodate new types and topics of questions that emerged. We eliminated some (199) questions from our analysis because they were hard to understand. Thus, we analyzed *1152 questions*.

Question Types. Our results differ significantly from those of Morris et al. (2010) who found that the majority of the questions were recommendations (29%) and opinions (22%). In contrast, we found that the most popular question type on Twitter was *rhetorical* (42%), followed by *factual knowledge* (16%), and *polls* (15%). The high prevalence of rhetorical questions is not surprising considering that most people use Twitter for sharing daily chatter (Java et al. 2010). Many of the rhetorical questions were personal or philosophical in nature, such as, "*What ever happened to the little things in life?*"

A significant percentage (16%) of questions pertained to factual knowledge. Some of these factual questions were likely hard to answer using search engines since they were highly contextual, such as "*Can I still book grad photos?*" or pertained to real-time events, such as "*Where's these protests? Are they like everywhere in England? Or just one place?*" Another popular question type, which was not captured by Morris et al., was *poll* (15%). Polls were mostly related to entertainment and technology. For instance, one user asked, "*Has anyone been contacted to say that their Jedward tickets have been cancelled then?*"

Question Topics. While the most popular question topic in Morris et al.'s (2010) study was technology (29%), the most popular topic in our study was *entertainment* (32%). Surprisingly, the second most popular topic was *personal and health* (11%), in contrast to the findings of Morris et al. who reported that users were hesitant to ask friends questions about health, dating, politics and religion.

The personal questions were related to a variety of issues from body-piercings, hairstyles, and self-image to people's opinions about oneself. For instance, one user asked, "*Am I that hard to talk to? Do I not understand things? Is anything worth lyin to me about?*"

Health-related questions ranged from recommendations for healthy habits to asking for factual knowledge about health issues, such as "*Are you supposed to sweat out a head cold to dry out your sinuses? Cause I have one*." Dating questions were common and ranged from seeking opinions to asking for advice. For instance, "*My b/f is such a typical arab Iraqi, does something once like gets me a gift then refuses to do it again. How to fix this?*" We found that most personal questions were rhetorical. We also found questions related to topics not found by Morris et al., such as *greetings*, *time*, *weather*, and *general knowledge*.

Overall, these results suggest that natural Q&A behavior is quite different from self-reported Q&A behavior in surveys (Morris et al. 2010). It is also different from Q&A on social search engine Aardvark where advice and recommendation questions on travel, restaurants, and products were most popular (Horowitz & Kamwar 2010).

Responses to Questions

Response Rates. We found that only 18.7% (215/1152) questions received at least one response, which is much lower than the response rate (93.5%) found by Morris et al. (2010). Our response rate was also much lower than those found on community Q&A sites (84%, Harper et al. 2008) and Aardvark (88%, Horowitz & Kamvar 2010).

The low response rate to questions on Twitter is perhaps not surprising considering that, in general, only 23% of tweets receive a response (Sysomos 2010). Also, since community Q&A sites and Aardvark are specifically designed for asking and answering questions while Twitter is not, it is perhaps not surprising that the response rate for questions on Twitter is much lower than these sites.

Number of Responses. The 215 questions received a total of 624 responses, averaging 2.9 responses per question. In comparison, a question on community Q&A sites receives between 0.6 to 4.00 answers on average depending on the site (Harper et al. 2008). Though, the average could be a misleading measure for Twitter since the number of responses per question had a long-tail distribution; a handful of questions received a high number of responses and most questions received few responses.

The highest number of responses (147) was received by a technology-related question asking about factual knowledge: "*Can someone tell me how Skype makes their money please?*" The second highest number of responses (39) was received by a question about restaurant recommendations: "*I haven't spent much time in chicago recently... Anyone have good restaurant recommendations for downtown chicago?*" Some rhetorical questions received a relatively large number of responses such as the question "*What's everyone doing on this amazing warm day?*" which received 29 responses.

Number of responses was not significantly different by type or topic. Interestingly, personal and health-related questions had a low response rate (9%).

Speed of Responses. Due to the real-time nature of Twitter, we expected questions to receive responses quickly and found that this was indeed the case. 67% of all 624 responses came within 30 minutes and most (95%) responses appeared within 10 hours. The median response time for questions was 10.3 minutes. The fast response speed is consistent with findings about Twitter that 97% of replies to tweets happen within an hour (Sysomos 2010).

We also examined the *time to first response* and found that more than half (52%) of the first responses were posted within 5 minutes. The time to first response in our study was much lower than that found in previous studies. Only 24% participants in Morris et al. (2010) received a response within 30 minutes. Users on social networking sites expect low response times and Twitter seems to surpass these expected response times. The quick response time reflects the real-time nature of Twitter.

Relevancy of Responses. We examined the responses to see whether they were relevant to the question. For every question that received a response, we coded responses as *relevant*, *unrelated*, or a *clarification question* to the original question. We found that 84% of the responses were *relevant* to the questions and 9% were *unrelated*. Thus, while most questions did not receive a large number of responses, most responses were relevant.

The conversational nature of Twitter makes it possible for the answerer to request clarification information and for the asker to follow up with further reactions (Horowitz & Kamvar 2010). We found that 4% responses were *clarification questions*. Finally, 1% of people replied saying they would be interested in knowing the answer and we could not code 2% of responses due to missing context.

Social Network Usage and Characteristics

We examined what characteristics of the asker might improve her chances of receiving a response. We performed a logistic regression to see whether the following characteristics of the asker would predict whether she received a response: *number of followers*, *days on Twitter*, *number of tweets* she had ever posted, and the *frequency of use* of Twitter (which we defined as *number of tweets* divided by *days on Twitter)*. We found that the askers' number of followers (χ^2 (1, N=1152) = 31.8, p<0.0001) and days on Twitter (χ^2 (1, N=1152) = 18.4, p<0.001) were good predictors of whether their questions would get answered.

Furthermore, number of responses was positively correlated (r=0.28, p<0.0001) with the asker's number of followers; this was similar to previous findings (Morris et al. 2010a). However, the number of tweets the asker had posted or her frequency of use of Twitter did not predict whether her question would get answered and was not significantly correlated with the number of responses.

Relationship Reciprocity. We were interested in understanding whether the reciprocity of the follow relationship between the asker and the answerer determined whether she received a response. In general, the follow relationship on Twitter has low reciprocity; only 22% of relationships are mutual while 78% are one-way (Kwak et al. 2010). We found higher mutual reciprocity among question askers and answerers; 36% of relationships were reciprocal and 55% were one-way. Surprisingly, 9% of answerers were not following the

askers. In future work, we plan to investigate why users answered questions from those they were not following.

We examined whether relationship reciprocity affected the relevancy of responses received. The percentage of unrelated responses was the highest (13%) when the answerer was not following the asker and lowest (6%) when there was a mutual follow relationship between asker and answerer. This suggests that more relevant responses are received when there is a mutual relationship between askers and answerers. Intuitively, we would expect this, as mutual relationship would indicate stronger tie strength and hence, more number of relevant answers.

Discussion

We conducted a sampled study of the naturally occurring social Q&A on Twitter. We found that rhetorical questions were the most popular, followed by questions seeking factual knowledge, and polls. The most popular topic was entertainment, but a significant number of personal and health-related questions were being asked. However, most of these personal and health-related questions did not receive a lot of responses.

In contrast to previous studies, only a low percentage of questions received a response; even among those that did, most questions received 1-2 responses. The low number of responses might be due to the fact that Twitter users are exposed to a great number of tweets constantly, leading to a lack of visibility for question tweets. Questions get buried in Twitter streams, especially for users who follow a large number of people. The low response rate might also be due to the fact that a large percentage of questions are rhetorical. For questions that received responses, the responses came in quickly and were largely relevant.

A rich understanding of social Q&A behavior helps us explore how we can develop tools to support this behavior. A key challenge in social Q&A is improving user engagement for answering questions. The real-time nature of Twitter makes it critical that potential answerers see questions in time to answer them, as many questions pertain to current events. Making questions salient in the Twitter stream would bring them to the attention of potential answerers. This suggests that having separate Q&A features, such as Facebook Questions, might be useful additions to social networks like Twitter.

Also, since a considerable number of factual questions are being asked and answered, making such Q&A pairs available to third parties could be beneficial. Building a repository of Q&A pairs from Twitter would benefit users who might have similar information needs. For instance, while formulating a question, the user can be provided with similar questions asked previously and the relevant responses received to those questions. Finally, since answers came from users with one-way and no follow relationships with the asker, this suggest that making questions visible to the entire user community on a social networking site might be useful.

Limitations. By using the '?' to select question tweets, we may have missed some questions. We also could not judge the usefulness of the responses, as this is subjective and dependent on the asker. Finally, we may have missed some responses as we focused on only those responses that were made using the 'reply to tweet' feature of Twitter.

Conclusion

We contribute by providing a rich characterization of the types and topics of questions asked, the responses received, and the effects of the underlying social network on social Q&A behavior. We also provide insight into using micro-task markets to classify text from social streams. In future work, we hope to build upon our findings to design tools to support social Q&A behavior.

References

Cong, G., Wang, L., Lin, C., Song, Y., and Sun, Y. 2008. Finding Question-Answer Pairs from Online Forums. *Proc. SIGIR*, 467-474.

Efron. M., and Winget, M. 2010. Questions are Content: A Taxonomy of Questions in a Microblogging Environment. *Proc. ASIST*, 1-10.

Evans, B. and Chi, E.H. 2008. Towards a Model of Understanding Social Search. *Proc. CSCW,* 485-494.

Evans, B, Kairam, S., and Pirolli, P. 2010. Do your Friends Make you Smarter? An Analysis of Social Strategies in Online Information Seeking. *Information Processing and Management*, 679-692.

Facebook. 2010. Searching for Answers? Ask Facebook Questions. Retrieved 01/30/11 from http://www.facebook.com/blog.php?post= 411795942130

Harper, F., Raban, D., Rafaeli, S., and Konstan, J. 2008. Predictors of Answer Quality in Online Q&A Sites. *Proc. CHI*, 865-84.

Horowitz, D. and Kamwar, S. 2010. The Anatomy of a Large-Scale Social Search Engine. *Proc. WWW*, 431-440.

Java, A., Song, X., Finin, T., and Tseng, B. 2007. Why We Twitter: Understanding Microblogging Usage and Communities. *Proc. WebKDD/SNA-KDD*, 56-65.

Kwak, H., Lee C., Park, H., and Moon, S. 2010. What is Twitter, a Social Network or a News Media? *Proc. WWW*, 591-600.

Kittur, A., Chi, E.H., and Suh, B. 2008. Crowdsourcing User Studies with Mechanical Turk. *Proc. CHI*, 453-456.

Morris, M., Teevan, J., and Panovich, K. 2010. What Do People Ask Their Social Networks, and Why? A Survey Study of Status Message Q&A Behavior. *Proc. CHI*, 1739-1748.

Morris, M., Teevan, J., and Panovich, K. 2010a. A Comparison of Information Seeking Using Search Engines and Social Networks. *Proc. ICWSM*, 291-294.

Paul, S.A., Hong, L., and Chi, E.H. 2011. What is a Question? Crowdsourcing Tweet Categorization. *CHI 2011 Crowdsourcing and Human Computation Workshop*.

Sysomos. 2010. Replies and Retweets on Twitter. Retrieved 02/06/11 from http://sysomos.com/insidetwitter/engagement/

Wadhwa, V. 2011. Why I Don't Buy the Quora Hype. Retrieved 01/30/11 from http://techcrunch.com/2011/01/23/why-i-don't-buy-the-quora-hype/