Asking Questions of Targeted Strangers on Social Networks

Jeffrey Nichols
IBM Research – Almaden
650 Harry Rd, San Jose, CA 95110
jwnichols@us.ibm.com

Jeon-Hyung Kang
USC Information Sciences Institute
4676 Admiralty Way, Marina del Rey, CA 90292
jeonhyuk@usc.edu

ABSTRACT
When people have questions, they often turn to their social network for answers. If the answer is obscure or time sensitive however, no members of their social networks may know the answer. For example, it may be difficult to find a friend who has experience with a particular feature or model of digital camera or who knows the current wait time for security at the local airport. In this paper, we explore the feasibility of answering questions by asking strangers. In this approach, strangers with potentially useful information are identified by mining the public status updates posted on Twitter, questions are sent to these strangers, and responses are collected. We explore feasibility in two ways: will users respond to questions sent by strangers and, if they do respond, how long must we wait for a response? Our results from asking 1159 questions across two domains suggest that 42% of users will respond to questions from strangers. 44% of these responses arrived within 30 minutes.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Experimentation, Human Factors

Author Keywords
Social Q&A, Twitter, qCrowd

INTRODUCTION
Unlike many other social networking sites, much of the content posted by users to Twitter is publicly accessible. This creates a unique opportunity for a Q&A system; users who may be able to answer questions can be identified from the content of their previous tweets and questions can be sent to those users directly using features such as the @reply. This approach is different from typical Q&A on social networks because it targets answerers who are strangers rather than friends. This approach is also different from other existing Q&A systems, which rely on a potential question answerer to find the question on a web site (e.g., Answers.com and StackOverflow) or on users to submit expertise profiles to help automatically identify potential answerers (e.g., Aardvark [1] and IM-an-Expert [7]).

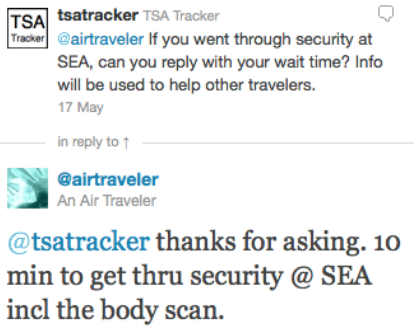

Figure 1. An example question/answer regarding the airport security wait time at SeaTac Airport.

We believe there are advantages of this "targeting strangers with questions" approach, at least for certain classes of questions. One particularly interesting class is *questions about an event that are best answered soon after the event*. The real-time nature of Twitter allows a question asker to identify strangers experiencing relevant events based on keywords in status updates and send relevant questions immediately. E.g., asking about the current wait time at an airport security checkpoint (see Figure 1). Another class is *questions for which there may be a diversity of opinion*. A Twitter-based Q&A system could target knowledgeable users across a range of different biases as detected from their social media posts (e.g., camera users that prefer Nikon or Canon, or political party affiliations [6]) and extract answers across that range.

We foresee at least two different implementations of this approach. A semi-automated dashboard implementation would allow users to enter a question, search for people that might be knowledgeable in that area, and send questions to a subset of those people. An automatic approach could also be useful in some situations. For example, a retailer's CRM system might identify complaints about service and send questions to learn more about the problem before engaging a human representative to provide a solution.

There are at least two challenges to building a Q&A system based on this approach. First, we must be able to identify strangers who can answer our questions based only on the content of their tweets. We have found that this varies

CSCW'12, February 11–15, 2012, Seattle, Washington, USA.

based on domain and is easier for heavily trafficked locations, such as airports, and popular technology products, such as high-end digital cameras. We expect growth in social networks and improvements in text analytics to make this more reliable for less popular topics in the future. Of course, we would not expect to be able to find strangers for topics that are not routinely discussed on social networks, such as personal health issues [4]. Second, strangers must be willing to respond when receiving a question. Here we focus on the second challenge and examine the following questions:

- Will users respond to unsolicited questions sent by strangers?
- How will providing an incentive for answering affect their response rate?
- If users respond, what is the duration between question and response?

To answer these questions, we conducted a series of experiments in which we watched a live stream of status updates from Twitter, manually sent questions to relevant users, and collected response rates and times. We asked questions in two different domains, airport security tracking and digital camera reviews, and in the airport domain we also tried questions with and without the incentive statement seen in Figure 1. Our overall finding from asking 1159 questions is that users responded to 42% of initial questions, and that 44% of all responses arrived within 30 minutes of sending the question. We also found that removing the incentive statement did not affect response rate, but seemed to result in a greater number of questions being marked as spam by recipients.

RELATED WORK

Numerous web services provide forums for posting questions and providing answers, including Yahoo! Answers, Answers.com, StackOverflow and Quora. All of these sites rely on experts to visit the site and provide answers to questions for which they are knowledgeable. There are also a myriad of question answering services that use Twitter, including TweetQA[1] and AskOnTwitter[2]. The Twitter-based services primarily offer two capabilities: aggregating questions asked by users on Twitter and tracking responses to those questions. We have not found a service that helps users direct questions to appropriate strangers, especially to strangers who have not previously opted-in to the service.

Several studies have been conducted of Q&A in a variety of settings. Zhang et al. [8] conducted a study of the Java Developers Forum and found the average response time to a question to be nearly 9 hours. Hsieh and Counts [2] found the average response time on Microsoft's Live QnA site to be 2 hours and 52 minutes, and that 20% of questions posted were never answered. Mamykina et al. [3] found that StackOverflow provides answers to software development questions in a median time of 11 minutes. Morris et al. [4] conducted a survey of Q&A within social networks and found that questions were answered in a comparable amount of time to other sites, such as those previously mentioned, but that users tended to trust the opinions of people they knew compared to those of strangers.

Paul et al. [5] conducted a study of question asking on Twitter. They found the response rate for questions in their corpus was only 18.7%, but the median response time was 10 minutes for questions that were answered. They also found that 67% of responses came within 30 minutes and 95% came within 10 hours. Later in the paper, we will discuss our results in the context of these numbers.

EXPERIMENTAL SETUP

Our goal for the design of the feasibility study was to maximize the chance that users would respond to our questions, because we expected the response rate might be very low even in favorable conditions. We also hoped to establish an upper bound on how well such a system might perform, which could be used in the future to judge the performance of automated systems. We built a simple dashboard application to help us choose which status updates to respond to and which questions to ask. We were conservative in choosing which users were sent questions.

We chose two question domains to which we believed strangers might be willing to respond: airport security wait time tracking and digital camera product reviews. We chose these two domains because:

- The desired information is not personal in nature (In the airport case we asked questions of people who had already reported their location publicly);
- There is enough traffic on these topics that it is easy to find relevant status updates;
- The initial question is easy to answer. Reporting a wait time can be two words (e.g., 10 minutes), and our initial camera question had a yes/no answer (the question was, "Do you have <camera model>?").

Airport Security Wait Time Tracking (TSA Tracker)

In this domain, we watch for users who report that they are at a US airport via a real-time stream from the Twitter streaming API. Once we have identified a relevant user, we send them one of two questions:

> **Q1:** If you went through security at <airport code>, can you reply with your wait time? Info will be used to help other travelers
>
> **Q2:** If you went through security at <airport code>, can you reply with your wait time?

The only difference between the two questions is the final sentence of Q1, which attempts to motivate users to respond because they will be helping others. We anticipated that users might view the profile of the Twitter account from which the message originated, so we used separate accounts

[1] http://www.tweetqa.com/
[2] http://www.askontwitter.com/

for each question: Q1 was asked by @tsatracker, whereas Q2 was asked by @tsatracking. Note that we use accounts with institutional sounding names. We plan to experiment with using accounts with personal names in the future.

For each response, we posted the wait time on a web site that we created to store the data: http://tsatracker.org/. We also sent back a thank you message with a link to the site.

Digital Camera Q&A

In this domain, we watch for users who mention specific model numbers for any of 8 different digital cameras (e.g., Nikon D7000, Canon 7D, etc.), again via a real-time stream. Once we have identified a user, we send a question asking whether the user owns whichever camera they mentioned. All questions were sent by @productqa.

A key difference between this domain and the previous is that we ask multiple questions. If the user responded that they did own the camera, we followed-up with a camera-related question such as, "How is the image quality?" We asked up to 3 follow-up questions based on their replies.

RESULTS

For our experiment, we have 3 conditions: 2 in the airport domain (Q1 vs. Q2) and 1 in the digital camera domain. Table 1 shows a summary of the results for the first question asked in each condition. Figure 2 shows a histogram of response times for each of the 3 conditions, coupled with line graphs showing the percentage of total responses received over time. Note that the response times follow a power law distribution.

To examine whether the response rate in any of the conditions differed, we performed 3 pairwise two-tailed Fisher's exact tests and found no significant difference between the groups. To test differences in response time, we performed a Kruskal-Wallis test that showed a marginal difference between the conditions ($H=5.7$, $df=2$, $p < 0.06$). This likely reflects the shorter tail in the distribution of response times in the Airport Q2 condition, which occurred because Twitter suspended the @tsatracking account within an hour or two after its last question. This prevented late responses and altered its response time distribution. No significant difference was found when comparing just the set of response times until the Q2 account was suspended across all 3 conditions. We were unable to collect any quantitative data that would allow a comparison to our other conditions where many more questions were asked. A Twitter representative told us the account exceeded a threshold for tweets marked as spam or user blocks.

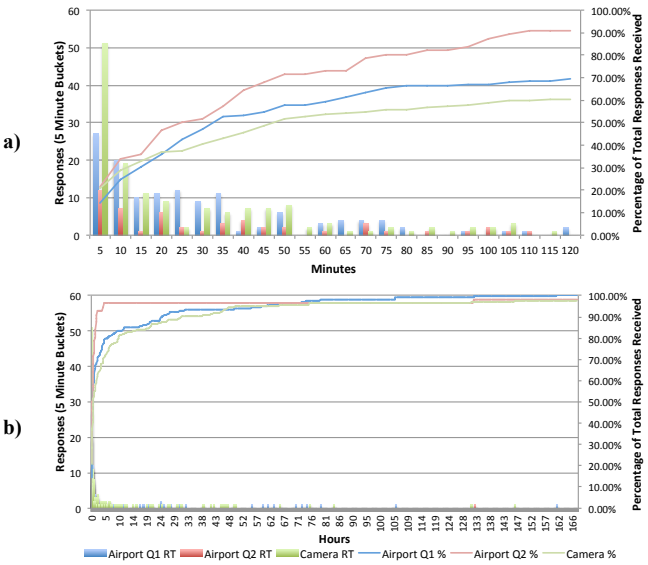

Figure 2. Total responses in 5 minute buckets and percentage of total responses received for each condition. a) The first two hours of data, b) All responses received within a week. Note that the percentage of total responses for Airport Q2 grows more quickly because that account was suspended, which artificially capped responses.

Table 2 shows a summary of the results for the follow-up questions asked in the camera condition. A pairwise two-tailed Fisher's exact test found significant differences in response rate between the first question and the first follow-up ($p < 0.0001$) and the first and second follow-up questions ($p < 0.0001$), but no significant difference between the second and third follow-ups. A significant difference in response time behavior was found between all four questions ($H=50.12$, $df=3$, $p < 0.0001$, Kruskal-Wallis) and just the three follow-ups ($H=25.46$, $df=2$, $p < 0.0001$, Kruskal-Wallis). This suggests that conversations were more likely to continue with users that responded quickly.

DISCUSSION

The first thing to note is that our results are different than those found by Paul et al. [5]. This is expected, because our questions have different properties from the questions they studied. In particular, each question in Paul et al.'s set were

Table 1. Summary of experimental results for the first question in each condition. *The Airport Q2 account was suspended after asking 150 questions, which distorts some of these numbers.

	Asked	Responses	Response Rate	Median Response Time (minutes)	% Responses Within 30 minutes	95% Responses Received In (hours)
Airport Q1	424	189	45%	25.5	47%	55.4
Airport Q2*	150	56	38%	33	52%	4.1
Cameras	585	245	42%	47	42%	47.0

Table 2. Summary of results for follow-up questions in the Camera condition

	Asked	Responses	Response Rate	Median Response Time (minutes)	% Responses Within 30 minutes	95% Responses Received In (hours)
1st follow-	245	151	62%	18.5	58%	17.5
2nd follow-	61	55	90%	2	84%	11.2
3rd follow-	24	23	96%	6	78%	2.0

sent by one user to all of that user's followers, as compared to our questions, each of which was seen by only one user. This difference in the number of potential answerers may explain the shorter response rate found by Paul et al. Our questions were also targeted at specific users who were likely to be able to answer the question and required a relatively basic response, whereas Paul et al.'s questions were more varied and may have required a more thoughtful response or expertise not possessed by anyone who saw the question. This difference may account for our much larger response rate (42% compared to 18.7% for Paul et al.).

Including the motivational sentence clearly had an effect on behavior, as not including the sentence resulted in suspension of the question asking account. Unfortunately, Twitter would not provide specific numbers on blocks or spam flags so it is not possible to compare these values across each of the 3 accounts used in our experiments. It does seem that the @tsatracking account exceeded some threshold for suspension whereas both of our other accounts, which together sent many times more questions, did not. We found it surprising that including the motivational sentence did not have a significant effect on response rate or time. Overall, these results suggest that the incentive sentence may not affect people who would be willing to respond to the question but does have an affect on those who are not willing to respond. Without the motivational sentence, users who would not be willing to respond may be more likely to mark the question as spam or block the user.

Despite being suspended, our question asking approach complies with Twitter's Terms of Service. With respect to unsolicited @replies, the Terms explicitly disallow sending large numbers of duplicate @replies or using @replies to spam a link or service. Our question tweets are all customized to the particular user they are sent to, so they are not duplicates. We also explicitly do not include a link in our initial question, which both complies with the terms and avoids the misperception that the question could be a phishing attack.

The responses that we received in all conditions were almost uniformly positive. The most negative comment that we received came from one user (in the non-incentive airport condition), who remarked that our question was "creepy," but also provided a wait time. We also received a lot of positive feedback, primarily in the airport conditions where we built a web site to display the data for everyone. Comments especially came from people who said they were frequent travelers and looked forward to making use of the tool in the future. The @tsatracker account, from which we asked many of our questions, also picked up 16 followers over the course of our experiment.

Our goal in this work was to determine the feasibility of building a Q&A system using our approach of asking targeted strangers, and we believe our results demonstrate that this approach can be feasible, at least in our two domains. Whether the approach can translate to other domains depends on the ease of finding strangers with potential to answer the question and the nature of the question being asked. Obviously, questions about personal information or other sensitive topics are likely to be ignored or blocked, and it will be difficult to find strangers to answer questions on topics that are not frequently discussed on Twitter. Although certainly not conclusive, our results suggest that asking questions of targeted strangers can be a viable approach if these other challenges are overcome.

CONCLUSION

In this paper, we have explored the feasibility of a system that obtains answers by sending unsolicited questions to people based on their public status updates on social networks. We found that 42% of people will respond to these questions, at least in the "best case" scenarios that we explored. Going forward, we plan to build technology that provides a dashboard to assist users in applying this approach in custom domains and an automated system that can be used to assist in customer support and crowd data collection scenarios.

REFERENCES

1. Horowitz, D. and Kamvar, S.D. (2010). "The anatomy of a large-scale social search engine," in *Proceedings of WWW*, 431–440.
2. Hsieh, G. and Counts, S. (2009). "A market-based real-time question and answer service." In *Proceedings of CHI'09*. 769-778
3. Mamykina, L., Manoim, B., Mittal, M., Hripcsak, G., and Hartmann, B. (2011). "Design lessons from the fastest q&a site in the west," In *Proceedings of CHI '11*. 2857-2866
4. Morris, M.R., Teevan, J., and Panovich, K. (2010). "What do people ask their social networks, and why?: a survey study of status message q&a behavior," In *Proceedings of CHI '10*. 1739-1748.
5. Paul, S.A., Hong, L., and Chi, E.H. (2011). "Is Twitter a Good Place for Asking Questions? A Characterization Study," In *Proceedings of ICWSM'11 Posters*.
6. Pennacchiotti, M., and Popescu, A. (2011) "A Machine Learning Approach to Twitter User Classification," in *Proceedings of ICWSM'11*. 281-288
7. White, R.W., Richardson, M., and Liu, Y. (2011). "Effects of community size and contact rate in synchronous social q&a," In *Proceedings of CHI '11*. 2837-2846
8. Zhang, J., Ackerman, M., Adamic, L. and Nam, K. (2007). "QuME: A Mechanism to Support Expertise Finding in Online Help-Seeking Communities," In *Proceedings of UIST'07*. 111-11