Using Public Social Media to Find Answers to Questions

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

INTRODUCTION
On some social media platforms, such as Twitter, Youtube, Pinterest, and tumblr, much of the content generated by users is publicly accessible and communication can be easily initiated between strangers who have never previously communicated before. The communities that have risen up around these platforms, particularly on Twitter, can also be inclusive and supportive of interactions between strangers.

The public and open nature of these communities creates an opportunity to create a new kind of question answering system, where potential answerers are determined based on their published content and questions are sent through the mechanisms of the social media platform. For example, suppose that we are curious about the length of the line for a new movie at several local theaters. In the context of Twitter, we might search in real-time for people posting messages indicating that they are at one of the local theaters and then send @reply messages to those users inquiring about the length of the line. In another scenario, we might be trying to collect information about local restaurants. We could search in real-time for people tweeting about getting food from one of these restaurants, and have a system ask relevant questions about the restaurant to learn more about the food, service, or other aspects of the restaurants on which we’d like to gain a greater understanding.

This approach has some useful properties:

• An existing community of users from a social media platform can be bootstrapped into a community of question answerers. The nature of the new bootstrapped community can vary based on the needs of the question asking system. For example, in some cases it may be desirable for the answerers to be aware of other users and to see their answers, similar to existing community question and answer websites. In other cases it may be desirable for answerers to be unaware of the rest of the community, especially if the goal is to collect independent samples that are not biased by others’ responses.

• Questions are pushed to potential answerers, similar in form to the methods used by instant messaging-based services like Aardvark [1] and IM-an-Expert [7]. This helps to ensure that answerers are aware of the question and can speed responses depending on how the social media platform sends notifications. Depending on the scenario, the question can be asked as a response to a previous message that the potential answerer sent, which gives the answerer some context for why the question is being asked of them. For example, if we are collecting information about a restaurant, our question to a particular answerer could be in response to a message they sent naming that restaurant.

• Information that users have published about themselves on the social media platform can be utilized to make decisions about their suitability as a question answerer. Real-time status information can be useful for identifying people who are in the right place at the right time to answer a time and/or location-sensitive question, but historical information can also be used to rate a users’ expertise or to ensure that opinions are sampled across a range of demographics. For example, political party affiliation can sometimes be inferred from a user’s past tweets [6]. Based on these inferences, a policy question might be asked of people across a range of different political viewpoints to better understand the breadth of the issues involved.

There are disadvantages to this approach as well:

• Questions are asked out of blue by strangers, which may be disruptive or surprising to the potential answerers that receive the questions. Because questions are asked by strangers, there is likely to be little social incentive to respond.

• Although all questions are asked based on public information, the potential answerer may not have realized that information was public. They may feel that their privacy was violated.

• The social media platforms that are repurposed for question asking are, by definition, not designed for Q&A. This means that reputation and other incentive mechanisms may not exist or work in ways that are counter to producing quality and timely answers.
• If questions and answers are seen publicly, then it is unlikely that users will respond to questions that might reveal sensitive information. For example, we suspect it is unlikely that users would respond to health-related questions on Twitter (survey results seem to back this hypothesis [2] though we have not experimentally verified it).

RESEARCH QUESTIONS AND STUDIES
In our work, we are exploring the effectiveness of this question asking method and building systems to support its use. The research questions we were particularly interested in were:

- Will people respond to questions from strangers?
- What is the quality of the responses that we receive?
- How can we encourage more people to respond?
- How does this vary across different domains?

The first two of these questions we have addressed in two experiments in two different domains [4, 5]. The first experiment focused on the feasibility of the approach across two domains, while the second experiment focused specifically on quality in just one domain. The first domain, used only in the feasibility study, is wait time tracking at airport security checkpoints. The idea is to identify people who are at airports, send them a question about how long the security line was, and then collect the responses. The second domain is product reviews, which was used in both experiments. The idea in this case is to identify users who appear to own some particular product and send them questions about that product. For the products, we considered digital cameras, a popular tablet computer, and Los Angeles-area food trucks.

For the first study, most of our decisions were made to maximize the expected response rate from the people targeted with questions. We did this because we expected that response rates would be quite low, even in the most favorable conditions. Our choice of domains followed from this decision. Many people talk about airport security and electronics products, and we believed that most people wouldn't object to answering questions on these topics.

The results were interesting. We asked 574 total questions in the airport security domain and received X responses, giving a response rate of 42% overall. In the product review scenario, we asked 585 questions and received 245 answers, giving a response rate of 42%. We were surprised at the fairly high response rates in both domains, which we had expected to be closer to 20%.

Our second experiment explored the question of response quality. For this experiment we needed ground truth information to compare the responses to, and so we focused on product reviews exclusively (actual security checkpoint wait data was not available to us).

Over the course of this experiment, we asked nearly 1000 questions across two types of products (tablet computer and food trucks). The responses rates and durations were in general comparable to our previous studies (~40% response rate with around 50% of responses received within 30 minutes). We measured the quality of the responses by coding conducted by three researchers. These results can be seen in the table below.

<table>
<thead>
<tr>
<th>Table 1. Summary of Answer Coding Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response Count</td>
</tr>
<tr>
<td>Tablet</td>
</tr>
<tr>
<td>Food Truck</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2. Summary of Reasons for Low Quality Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td># Irrelevant Responses</td>
</tr>
<tr>
<td>Tablet</td>
</tr>
<tr>
<td>Food Truck</td>
</tr>
</tbody>
</table>

The most important result is that most answers were relevant, and the primary reason for an answer not being relevant was if the targeted answerer did not actually have an answer to the question. This was particularly problematic in the case of the tablet product reviews, where we had more difficulty targeting owners of the particular tablet that we were interested in and often accidentally sent questions to owners of phones or other tablets from the same manufacturer. This suggests that improved targeting algorithms will greatly improve the quality of the information received.

BUILDING A QUESTION ASKING SYSTEM
Ultimately, we are working towards building a system that assists with the process of asking questions of strangers. Currently this work is proceeding in three directions.

First, we are building interactive dashboards to help facilitate the question asking process. Part of this work is to explore the different ways that questions might be answered by interacting with strangers and better understanding how question askers might want to explore and interact with the available answerers. Figure X shows one example of a dashboard that we have constructed.

Second, we are developing modeling approaches to better understand social media users from their historical content. Based on these models, we can make decisions about whether or not a question might be appropriate (for exam-
ple, based on location [3]) or estimate the likelihood that an answerer will respond. Finally, we are continuing to conduct studies to better understand the properties of this question asking approach. For example, we are interested in conducting studies of question asking when time is not sensitive and thus questions might not be a response to a message that was just sent but based on a pattern behavior over a longer term. We are also interested in understanding differences that may arise based on who asks the question, for example an individual vs. an organization.

ACKNOWLEDGEMENTS
The work described in this paper has been a collaboration of many researchers, both at IBM Research – Almaden and other locations. People contributing to this work include Jilin Chen, Allen Cypher, Clemens Drews, Jeon Hyung Kang, Funda Kivran-Swaine, Kyumin Lee, Jalal Mahmud, Nimrod Meggido, Xiaohua Sun, Jenn Thom, Huahai Yang, and Michelle Zhou.

REFERENCES