

When Will You Answer This?

Estimating Response Time in Twitter

Jalal Mahmud, Jilin Chen, Jeffrey Nichols

IBM Research - Almaden
 650 Harry Rd, San Jose, CA 95120
 {jumahmud, jilinc, jwnichols}@us.ibm.com

Abstract

We present a study analyzing the response times of users to questions on Twitter. We investigate estimating these response times using an exponential distribution-based wait time model learned from users' previous responses. Our analysis considers several different model building approaches, including personalized models for each user, general models built for all users, and time-sensitive models specific to a day of the week or hour of the day. Our evaluation using a real world question-answer dataset shows the effectiveness of our approach.

Introduction

Recent years have seen a rapid growth in micro-blogging and the rise of popular micro-blogging services such as Twitter. One of the many uses of these services is to post questions and receive answers from friends or even strangers. Several researchers have investigated this phenomenon, both from the perspective of questions from friends (Morris et al. 2010, Paul et al. 2011, Teevan et al. 2011) and strangers (Nichols et al. 2012) and response rates have been reported for both scenarios (Paul et al. 2011, Nichols et al. 2012). However, no one has yet described a method to estimate the likely wait time at the moment a question is asked. This is particularly important for information collection in time-sensitive and emergency situations, such as during a terrorist attack or following a natural disaster. Estimates of wait time can also guide question askers in deciding when to ask their question and how many people to target with a specific question when a speedy response is required.

In order to estimate wait times for responses to questions on Twitter, we have developed predictive models from users' previous response times. Our models use an exponential distribution with the assumption that response events

follow a Poisson process. We have explored various alternatives for building our predictive models, including building a personalized predictive model for each user, building one general predictive model for all users, and building predictive models based on time of the day and day of the week. We evaluate our predictive models using a real world question answer dataset to demonstrate the effectiveness of our approach.

Related Work

There is previous work on response analysis on social media, social media activity modeling and general activity modeling. Paul et al. (2011) conducted a study of question asking on Twitter and reported an 18.7% response rate and 10 minute median response time for questions that were answered. In contrast, Nichols et al. (2012) describes a method of information collection that relies on asking questions to strangers on Twitter rather than friends as identified by the social network. They reported a 42% response rate with 44% of answers arriving within 30 minutes (Nichols et al. 2012). Zhang et al. (2007) studied Java developer forum and reported that when expert Java users posted questions, the average response time was 9 hours. Hsieh et al. (2010) reported that average response time on Microsoft's Live QnA site was 2 hours and 52 minutes.

There has also been work on classifying whether a user is likely to respond to a given question on social media (Mahmud et al. 2013). This work produced a model for estimating whether users would respond and a method for selecting the set of most likely users to respond subject to cost/benefit constraints, however it did not consider the temporal factor of when a response might be sent. Our work is complementary and might be used to improve the selection process of this previous system.

There are research efforts on social media activity modeling (Yang et al. 2012, Kumar et al. 2010) and general ac-

tivity modeling (Avrahami et al. 2006, Begole et al. 2002). Yang et al. (2012) presented an analysis of human behavior dynamics in online social media by analyzing users' inter-event time distribution between two consecutive actions. Kumar et al. (2010) described mathematical models to capture the patterns in social media conversations. Avrahami et al. (2006) build statistical models to predict one's responsiveness within a certain time interval to incoming instant messages. Begole et al. (2002) analyze users' desktop activity to predict availability. None of this work describes the estimation of response wait times for questions asked via social media such as Twitter.

Dataset

We obtained the dataset collected by Nichols et al. from the authors (Nichols et al. 2012). The dataset contains questions sent to strangers on Twitter in the context of two information collection scenarios: wait times at airport security checkpoints and product reviews for digital cameras. Human operators identified potential answerers by manually inspecting real-time Twitter streams and sending questions to users who either mentioned that they were at a US airport or mentioned owning one of a few digital camera models (e.g., Nikon D300). The dataset contains 1159 questions and 490 responses. For each user that received a question, the dataset also contains up to the 300 tweets that user had sent prior to receiving the question. From these tweets, we identified replies (i.e., tweets starting with @user), and used the Twitter API to access the original tweet for each reply. We use these replies in our analysis if the original tweets contained a question (?) mark¹. These pairs of tweets are assumed to be representative of questions and answers sent between friends on the social network. In total, we identified 13274 question-answer pairs (11.45 question-answer pairs per user). From these questions and answers, we computed response times. We found that 58% of such responses came within 30 minutes and average response time was 362 minutes.

Estimating Response Wait Time

We built predictive models for estimating response wait times using three alternative approaches.

Personalized wait time models

In this approach, our predictive strategy estimates the wait time of each question for a specific user based on the history of response wait times for only that user. For simplicity, we assume that each response event follows a Poisson process in the sense that each response event occurs continuously and independently at a constant average rate.

¹ Studies have found that most (81.5%) questions asked on social media contained a question (?) mark and rule based method to identify questions in online content achieved more than 97% accuracy (Cong et al. 2008).

With this assumption, we use an exponential distribution to model a user's response wait time with probability. The probability density function (pdf) of such an exponential distribution is

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

The distribution is supported on the interval from zero to infinite time. Here x is the future response wait time for the user for which the exponential distribution model returns probability $f(x; \lambda)$. λ is the rate parameter of the exponential distribution for each user, which is estimated as the inverse of the average response wait times of previous responses for that user. The cumulative distribution function is described by the following equation:

$$F(x; \lambda) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Figure 1a shows the probability distribution function and 1b shows the cumulative distribution function for the wait time exponential distribution for a random user chosen from the dataset. The rate parameter λ for this user was computed as 0.0833 per minute.

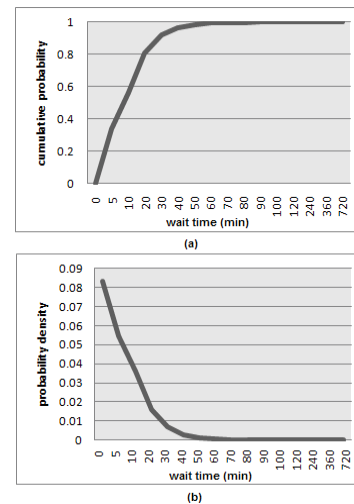


Figure 1. Wait time exponential distribution for a user (a) Cumulative probability (b) probability density

Generalized wait time models

In this approach, instead of building a separate model for each user, we build a single model from the previous responses to questions of all users in our dataset using the exponential distribution computation described above. Thus, the rate parameter λ was estimated from the responses of all users in our dataset.

Time-sensitive wait time models

In this approach, we incorporate into our models a sensitivity to the hour of a day or the day of the week when questions are sent to users. For a specific day or hour, we first identified the questions sent at that day or hour and their

responses. We then incorporate this information into both generalized wait time models and personalized wait time models. When building the generalized time-sensitive wait time models, we compute the rate parameter λ for the exponential distribution from all responses to the questions sent during the specific day or hour being modeled. When building the personalized time-sensitive models, we consider only users who have at least 5 responses to questions sent during the day or hour being modeled.

Experiments

We evaluated how accurately we can predict the wait time to respond in each of our model building approaches.

Experimental Setup

We tested each of our models for two settings: responding to friends and strangers. For the first setting, we trained from users' previous $N - 1$ responses and tested on the N th (most recent) response. For the second setting, we tested on the response to a human operator from the Nichols et al. study (a stranger) and trained on the previous N responses.

To evaluate our prediction algorithm, we reduce the problem to a binary classification problem of whether a user is sufficiently likely to respond (e.g., 80% or 0.8 probability in cumulative distribution function) within a given time period (e.g., 1 hour) and use the standard accuracy metrics recall (R) and precision (P). For a question and answer in our test data for a given user, we used the trained exponential distribution model to compute the cumulative probability of responding to the question within the specific time limit. A cut-off probability defines how much we are willing to tolerate the possibility that the user actually respond after the time limit. For example, if the rate parameter λ is 0.675 and the time limit is 1 hour, then the cumulative probability of responding is 0.77. If this probability is higher than the cut-off probability, we mark that user as "accept," otherwise we marked the user as "reject." For example, if the cut-off probability is 0.8, then this user will be marked "reject." We then compute precision and recall as follows:

Let N_1 denote the number of accepted users that actually replied within time limit (true positive), N_2 denote the number of users that actually replied within time limit (total positive) and N_3 denote the number of accepted users that actually did not reply within time limit (false positive). Then, precision, $P = N_1 / (N_1 + N_3)$ and recall, $R = N_1 / N_2$. The F measure is the harmonic mean of precision and recall.

Experimental Result

Here, we present the experimental result for each of our model building approaches. First, we present the experimental result for our personalized models using 1 hour as the time limit. We vary the cut-off probability in the cumulative distribution function from 0.1 to 0.9 at intervals of

0.1. The results are shown in Table 1. We observe that our predictive model achieves reasonable accuracy in estimating wait time when the 1 hour time limit is used. As the cut-off probability is increased, precision generally increases and recall decreases (with some exceptions). In addition, we observe that accuracies obtained for responding to friends and strangers were quite comparable.

cut-off probability in cumulative distribution function	Response to friends			Response to strangers		
	P	R	F	P	R	F
0.1	0.75	0.90	0.82	0.79	0.92	0.85
0.2	0.80	0.85	0.82	0.81	0.82	0.81
0.3	0.85	0.70	0.77	0.84	0.67	0.75
0.4	0.85	0.60	0.70	0.86	0.57	0.69
0.5	0.90	0.50	0.64	0.94	0.52	0.67
0.6	0.91	0.54	0.68	0.93	0.44	0.60
0.7	0.86	0.40	0.55	0.96	0.37	0.53
0.8	0.92	0.35	0.50	0.95	0.30	0.46
0.9	0.90	0.30	0.45	0.94	0.25	0.40

Table 1. Prediction accuracy for personalized wait time models, 1 hour time limit

We also varied the time limit while keeping a fixed cut-off probability at 0.5. The results are shown in Table 2. Our predictive approach can achieve over 80% F-measure when the time limit is 6 hours or more, and, as expected, predictive accuracy drops with very strict time limits. However, our model can achieve an F-measure of greater than 50% even when the time limit is 30 minute.

Time Limit	Response to friends			Response to strangers		
	P	R	F	P	R	F
15 min	0.65	0.30	0.41	0.65	0.25	0.36
30 min	0.90	0.40	0.55	0.91	0.38	0.54
1 hour	0.90	0.50	0.64	0.94	0.52	0.67
2 hour	0.92	0.68	0.78	0.90	0.65	0.75
6 hour	0.85	0.85	0.85	0.87	0.83	0.85
12 hour	0.88	0.90	0.89	0.88	0.87	0.88
18 hour	0.92	0.9	0.91	0.91	0.88	0.90
24 hour	0.98	0.93	0.95	0.97	0.91	0.94

Table 2. Prediction accuracy for personalized wait time models, cut-off in cumulative distribution function is 0.5

We also compare the performance of the personalized wait time models with our other alternatives, using the same variations in the cut-off probability and time limit as used above. Table 3 shows the average F-measure for each of the alternatives. We observe that a personalized model achieves higher prediction accuracy than the generalized model. This suggests that the variation between individual users is high and that sufficient data was available for each individual to make meaningful predictions. Adding time sensitivity to the models did not seem to substantially alter performance. For the general model, the time-sensitive variant achieved slightly higher accuracy, however the time-sensitive variants for the personalized models were not al-

ways better. This could be due to sparseness of the data in each time interval for creating the time-sensitive variants. We hope to investigate this further with a larger dataset.

	Average F	
	Response to friends	Response to strangers
Personalized model	0.74	0.72
General model	0.42	0.41
Personalized model-day	0.76	0.74
Personalized model-hour	0.73	0.73
General model-day	0.44	0.45
General model-hour	0.47	0.44

Table 3. Comparative accuracy for personalized, general and time-sensitive wait time models

	Average Error (min)	
	Response to friends	Response to strangers
Personalized model	214	219
General model	335	370
Personalized model-day	190	210
Personalized model-hour	220	215
General model-day	350	340
General model-hour	360	380

Table 4. Average error (min) for incorrect predictions

Error Analysis

We also investigated the incorrect predictions, by computing the expected wait times using the inverse cumulative distribution function, which is defined as:

$$F^{-1}(p; \lambda) = \frac{-\ln(1-p)}{\lambda}, \quad 0 \leq p < 1$$

We compared expected wait times with actual wait times and computed error minutes. We repeated this computation for different cut-off probabilities and time limits using the same intervals in our previous analyses. Table 4 shows the average errors for different models, and again the personalized wait time models resulted in the least average errors.

Conclusion and Future Work

We have presented a study on estimating wait time of responses in Twitter. Our wait time estimation is based on predictive models that follow an exponential distribution and are built from users' wait times from past responses to questions under different conditions. An evaluation using a real world question-answer dataset demonstrates the promise of our approach. Our predictive models for estimating wait times can be used for selecting people for question answering, whether or not the questions originate from friends or strangers. Our analysis is based on data collected from Twitter, however it may be applicable to other social media platforms where questions may be asked asynchronously. Our approach may also be reapplied on these platforms as long as temporal information from users' previous questions and answers are available. There are several ave-

nues for future research. First, we plan to develop more sophisticated wait time models (e.g., using an HMM) for representing different states of an individual in social network. Second, we hope to extend our findings for other types of social media activities beyond question asking, such as retweeting behavior on Twitter. Finally, we hope to validate our findings with a larger dataset and integrate our solution with a real world question-answering service.

Acknowledgement

Research was sponsored by the U.S. Defense Advanced Research Projects Agency (DARPA) under the Social Media in Strategic Communication (SMISC) program, Agreement Number W911NF-12-C-0028. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Defense Advanced Research Projects Agency or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

- Avrahami, D. and Hudson, Scott E. 2006. Responsiveness in Instant Messaging: Predictive Models Supporting Inter- Personal Communication, In *Proc. CHI 2006*.
- Begole, J., Tang, J.C., Smith, R.E., and Yankelovich, N. 2002. Work rhythms: Analyzing visualizations of awareness histories of distributed groups. In *Proc. CSCW'02*.
- Cong, G. Wang, L., Lin, C., Song, Y., Sun, Y. 2008. Finding Question-Answer Pairs from Online Forums. In *Proc. SIGIR 2008*.
- Hsieh, G. and Counts, S. 2009. mimir: A market-based real time question and answer service. In *Proc. CHI 2009*.
- Kumar, R., Mahdian, M. and McGlohon, M. 2010. Dynamics of conversations. In *Proc. SIGKDD, 2010*.
- Mahmud, J. Zhou, M. Megiddo, N., Nichols, J. and Drews, C. 2013. Recommending Targeted Strangers from Whom to Solicit Information in Twitter. In *Proc. IUI 2013*.
- 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. In *Proc. CHI 2010*.
- Nichols, J., Kang, J.H. 2012. Asking Questions of Targeted Strangers on Social Networks. In *Proc. CSCW 2012*.
- Paul, S.A., Hong, L. and Chi, E.H., 2011. Is Twitter a Good Place for Asking Questions? A Characterization Study. In *Proc. ICWSM'11 Posters*.
- Teevan, J. Morris, M., and Panovich, K. 2011. Factors Affecting Response Quantity, Quality, and Speed for Questions Asked via Social Network Status Messages. In *Proc. ICWSM 2011*.
- Yang, X., Zhang, Z., Wang, Ke. 2012. Human Behavior Dynamics in Online Social Media: A Time Sequential Perspective. In *Proc. SNAKDD 2012*.
- Zhang, J. Ackerman, M., Adamic, L. and Nam, K. 2007. QuME: A Mechanism to Support Expertise Finding in Online Help-Seeking Communities. In *Proc. UIST 07*.