AAAI Technical Report WS-12-04
When the City Meets the Citizen

Feasibility Study on Detection of Transportation Information Exploiting Twitter as a Sensor

Kenta Sasaki Shinichi Nagano Koji Ueno Kenta Cho

Corporate Research & Development Center, Toshiba Corporation
1, Komukai Toshiba Cho, Saiwai ku, Kawasaki 212 8582, Japan
{kenta.sasaki, shinichi3.nagano, koji3.ueno, kenta.cho}@toshiba.co.jp

Abstract

The concept of a smart community has recently been attracting great attention as a means of utilizing energy effectively. One of the modules constituting the smart community is an intelligent transportation system, in which various sensors track movements of people and vehicles in real time to optimize migration pathways or means. Social media have the potential to serve as sensors, since people often post transportation information on such media. This paper presents a feasibility study on detecting information, focusing on train status information, by exploiting Twitter as a sensor. We dealt with two issues: (1) for the ambiguity of textual information expressed in tweets, we utilized heuristic rules in text manipulation, and (2) for the differences in the numbers of tweets among train lines, we optimized parameter values in statistical analysis for each train line. The experimental results show that the F measure of detecting the information was more than 0.85 and the time taken to detect the information was less than 4 minutes. As a result we confirmed the high potential of detecting transportation information through Twitter.

Introduction

The concept of a smart community, which envisions the integration and management of diverse systems by exploiting information technology, has recently been attracting great attention as a means of utilizing energy effectively (Li et al. 2011). One of the modules constituting the smart community is an intelligent transportation system (ITS), which provides real-time transportation information to people by tracking movements of people and vehicles through various sensors such as GPS mounted on smartphones or cameras installed on roads. The goal of ITS is optimization of migration pathways or means, and reduction of energy usage (An, Lee, and Shin 2011). Social media have the potential to serve as sensors, since people often post transportation information on such media. Thus, in addition to physical sensors such as GPS or cameras, Twitter would be available as a sensor (Sheth 2009; Demirbas et al. 2010). Users of Twitter, the most popular social media service, can post text-based messages of up to 140 characters, known as tweets (Java et al. 2007), and the number of active users worldwide exceeded 100 million in September 2011 [1]. Therefore, it should be possible to obtain both information that physical sensors cannot monitor and information from the places where physical sensors are not installed.

Several studies on detection through Twitter of particular phenomena occurring in the real world have been reported. Sakaki, Okazaki, and Matsuo (2010) proposed a method of detecting earthquakes by utilizing Twitter. Their method detected earthquakes in Japan with high probability (96% of earthquakes of Japan Meteorological Agency (JMA) seismic intensity scale 3 or more were detected). Aramaki, Maskawa, and Morita (2011) proposed a method of detecting influenza epidemics by exploiting Twitter. Their method showed high correlation (0.89 correlation) between the number of tweets indicating that an influenza patient is present nearby and the number of patients suffering from influenza, outperforming Google Flu Trends [2] using aggregated Google search data. Tumasjan et al. (2010) investigated whether Twitter is utilized as a forum for political deliberation by conducting content analysis of over 100 thousand descriptions, and showed that Twitter is utilized extensively for that purpose.

This paper explores the possibility of detecting transportation information by exploiting Twitter as a sensor. If people instantaneously receive accurate information, they will be able to respond flexibly to incidents and optimize their behavior, e.g. select an alternate route, utilize another means of transportation or

[1] http://blog.twitter.com/2011/09/one hundred million voices.html
[2] http://www.google.org/flutrends/

change their schedules. In the event that transportation is disrupted owing to natural disasters such as earthquakes or typhoons, the information becomes especially important. In this paper, we focus on train status information, especially information about suspension or delay of services. The number of tweets posted when boarding trains is probably larger than that in the case of getting into cars since people cannot tweet when driving cars, and we consider the information about suspension or delay of services to be the most important detected. With a view to employing Twitter as a sensor, we confirm how accurately and quickly train status information can be detected by utilizing naïve methods.

The remainder of this paper is organized as follows: the next section describes the preliminary validation and the problem definition in detecting train status information through Twitter. This is followed by the explanation of the design of this feasibility study. Then, we report the results and discuss them. Lastly, we present conclusions.

Detection of Train Status Information Exploiting Twitter as a Sensor

Preliminary Experiment

To confirm the validity of exploiting Twitter as a sensor in the field of transportation information, we conducted a brief survey. We first collected tweets mentioning 68 train lines in Japan posted in a specific period (2011/07/28 - 2011/10/11), and then compared them with statistics on the numbers of train passengers [3]. Figure 1 is the scatter diagram of the numbers of train passengers per day and the numbers of tweets per day. We found a positive correlation of 0.83 between them, indicating that Twitter has the potential to represent the real-world context in the field of transportation information.

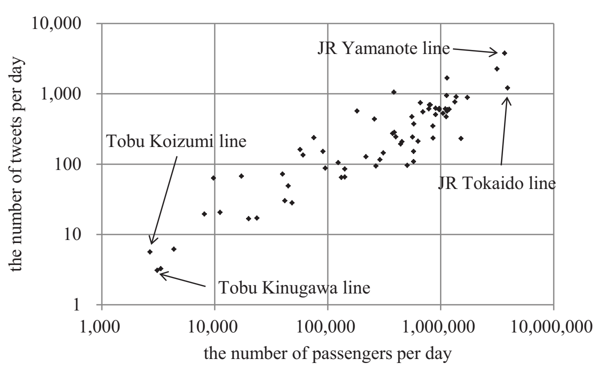

Figure 1: Scatter diagram of the numbers of train passenger per day and the numbers of tweets per day.

[3] http://www.train media.net/index.html

Problem Definition

We address two issues in detecting train status information through Twitter. One issue is the ambiguity of textual information expressed in tweets. The following sentences are examples of tweets retrieved by two keywords, “Yamanote Line” and “stopped”:

(1) The Yamanote Line has completely stopped.

(2) Since the Chuo Line has stopped, I’ll transfer to the Yamanote Line.

(3) Not again! RT @Tom the Yamanote Line has stopped.

(4) @Sara I killed time because the Yamanote Line had stopped.

Tweet (1) indicates that Yamanote Line services have probably been suspended. However, the other tweets do not necessarily describe those situations. Tweet (2) means that not Yamanote Line services but Chuo Line services have been suspended. Tweets (3) and (4) probably refer to the suspension of Yamanote Line services in the past. At the time that the users tweeted, the Yamanote Line might have been running normally. According to these observations, it is important to classify tweet (1) and tweets (2), (3), and (4), in order to detect train status information through Twitter. In this paper we define tweets related to train status information as follows:

Positive tweets: The tweet refers to real-time train status information of the line.

Negative tweets: The tweet does not necessarily refer to real-time train status information of the line.

While (1) is a positive tweet, (2), (3), and (4) are negative tweets.

The other issue is the differences in the numbers of tweets among train lines. As can be seen in Figure 1, whereas the numbers of tweets are large for train lines used by many passengers, they are small for train lines used by few passengers. Thus, when something occurs, large numbers of tweets are likely to be posted for train lines used by many passengers and small numbers of tweets are likely to be posted for those used by few passengers. Figure 2 shows both the distribution of the number of

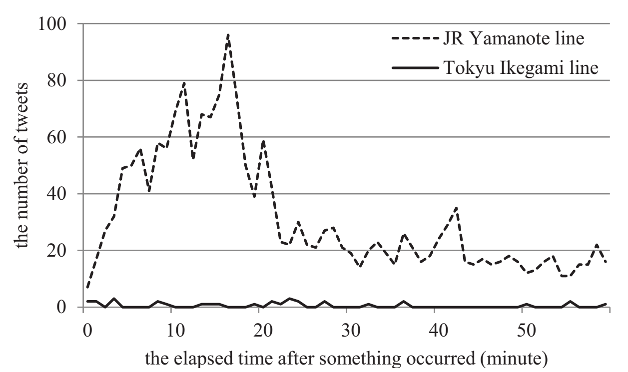

Figure 2: Distributions of the numbers of tweets after something occurred.

tweets including the keyword "Yamanote Line" when something occurs on the JR Yamanote Line, which is one of the most heavily used lines in Tokyo, and that of the number of tweets including the keyword "Ikegami Line" when something occurs on the Tokyu Ikegami Line, which is one of the least heavily used lines in Tokyo. As can be seen in Figure 2, whereas a large number of tweets were posted for the JR Yamanote Line, a small number of tweets were posted for the Tokyu Ikegami Line. It is important to consider these differences too, in order to detect train status information through Twitter.

Design of Feasibility Study

Our study consists of three steps. First, we collect tweets. Second, we extract positive tweets from those collected. Finally, we predict the existence of train status information by statistical analysis of the positive tweets. In this section, we describe these steps.

Step1: Collecting Tweets

We search and collect tweets referring to a specific train line via Twitter Search API[4]. We utilize the addresses of the train lines (which we call query keywords), including not only formal names but also abbreviated names, as search keywords. Table 1 shows some of the query keywords for the train lines. If two or more keywords exist, we collect tweets by OR search.

Train line	Query keywords
JR Yamanote Line	Yamanote Line
JR Chuo Line	Chuo Line
JR Sobu Line	Sobu Line
JR Yokosuka Line	Yokosuka Line, Suka Line
JR Takasaki Line	Takasaki Line
JR Kawagoe Line	Kawagoe Line
Keio Inokashira Line	Inokashira Line
Tokyo Metro Fukutoshin Line	Fukutoshin Line
Tobu Tojo Line	Tojo Line
Tokyu Ikegami Line	Ikegami Line

Table1: A part of the addresses of the train lines, including formal names as well as abbreviated names.

Step2: Extracting Positive Tweets

First, we choose the candidates from the tweets collected in step1, utilizing specific keywords closely related to train status information (which we call train status keywords). We choose only the tweets including not only query keywords but also train status keywords. We exploit fourteen kinds of train status keywords, such as "stopped", "delay" or "suspend". However, the candidates include negative tweets as explained in the previous section. Thus, we introduce an additional step that classifies the candidates into a positive class or a negative class by utilizing the three heuristic rules listed below. Only tweets satisfying all the rules are extracted as positive tweets.

(1) The query keyword is located in front of the train status keyword.
(2) The query keyword appears in the text area that users themselves compose (not quoted texts).
(3) The tweet is not a reply, i.e. the first character of the tweet is not "@".

Rule (1) eliminates the tweets referring to train status information of another line, such as "Since the Chuo Line has stopped, I'll change to the Yamanote Line." Whereas it is classified as a positive tweet for the JR Chuo Line, it is classified as a negative tweet for the JR Yamanote Line. Rules (2) and (3) exclude the tweets whose information is ambiguous. If the query keyword does not appear in the text area that users themselves compose, they are unlikely to board a train on the target line. Additionally, since reply tweets involve a time lag in communication, their information is likely to be old.

Step3: Predicting Train Status Information

In the case that a small number of positive tweets are obtained in step2, it is difficult to accurately predict the existence of train status information. This is because information in tweets is not necessarily correct. Indeed, it is not necessarily the case that people exactly express the facts of any train situation. Thus, we introduce a prediction step that estimates the existence or non-existence of the information by statistical analysis of positive tweets posted within a specific time window. If the number of positive tweets posted within the specific time window (N) exceeds a certain threshold (T), we consider train status information exists as shown in Figure 3.

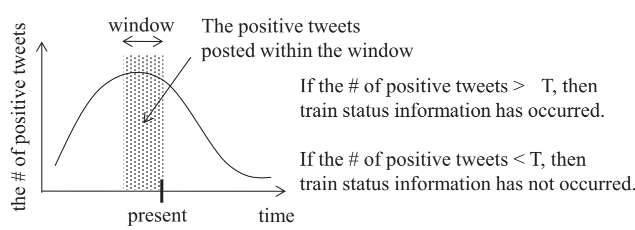

Figure3: Prediction of train status information.

Results of Feasibility Study

We evaluated how quickly and accurately we can detect train status information by exploiting Twitter as a sensor. In this section, we describe the evaluation dataset, the evaluation metrics, and the evaluation results in detail.

[4] https://dev.twitter.com/docs/api/1/get/search

Evaluation Dataset

We collected tweets of one day for 10 train lines in Tokyo and manually labeled each minute, indicating whether train status information existed or not. The target lines, the amount of time in which the information occurred, the number of information occurrences, and the number of tweets are shown in Table 2. We considered the period from the start of abnormality to the recovery of normality to be one information occurrence.

Train line	Time (min.)	The # of occurrences	The # of tweets
JR Yamanote Line	188	1	2,394
JR Chuo Line	518	3	3,198
JR Sobu Line	218	2	504
JR Yokosuka Line	82	1	306
JR Takasaki Line	157	1	266
JR Kawagoe Line	296	1	233
Keio Inokashira Line	77	1	338
Tokyo Metro Fukutoshin Line	15	1	16
Tobu Tojo Line	234	2	406
Tokyu Ikegami Line	118	1	57

Table2: Evaluation dataset.

Evaluation Metrics

We utilize two metrics: the accuracy that represents how accurately we can detect train status information, and the real time that represents how quickly we can detect the information. The accuracy depends on the overlapping of minutes between the predicted labels and the manual labels: the larger the overlapping of the periods, the higher the accuracy. The accuracy is expressed by recall, precision and F-measure. These values are calculated as follows:

Recall $TP / (TP + FN)$

Precision $TP / (TP + FP)$

F-measure $2 * Recall * Precision / (Recall * Precision)$

TP is the number of minutes that the system accurately detects the existence of train status information. FN is the number of minutes that the system does not detect the existence. FP is the number of the minutes that the system wrongly detects the existence. On the other hand, the real time depends on the time taken to detect the information in Figure 4: the shorter the time taken to detect the information, the higher the real time.

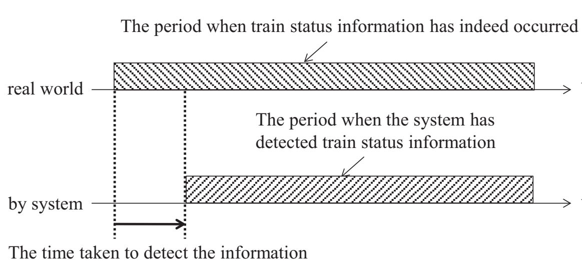

Figure 4: Real time of detecting train status information.

Evaluation Results

First, we show the evaluation results of the positive tweet extraction. Next, we show the evaluation results of the prediction of train status information.

Positive Tweet Extraction

The evaluation results of the positive tweet extraction are shown in Table 3. We used a comparative method, considering tweets including the train status keywords as positive tweets, without applying the three rules explained in the previous section. For the evaluations, we utilized a dataset of one thousand tweets randomly selected by the query keywords, which is different from the dataset explained in the previous subsection. We manually labeled each tweet as a positive tweet or a negative tweet. As seen in Table 3, though the recall decreased compared to the comparative method, the precision increased and the F-measure is higher than that of the comparative method. Therefore, in positive tweet extraction, the three rules are significant.

	Recall	Precision	F measure
Comparative method	0.80	0.74	0.77
Our proposed method	0.72	0.86	0.79

Table 3: Results of the positive tweet extraction.

Prediction of Train Status Information

In this section we illustrate two evaluation results in order to optimize N, the time window, and T, the threshold. We first show the evaluation results in which the same values are set for both parameters, and then show the evaluation results in which both parameter values are optimized for each train line.

Common Parameter Values

We prepared two values, 10 and 20 minutes, for N, and two values, 5 and 10, for T. Then, we conducted the experiments and calculated each value by four kinds of parameter value combinations. The overall evaluation results are shown in Table 4. The real time was coincident with the instinctive feeling that the time taken to detect train status information was shortened by lowering the threshold. The time window scarcely affected the real time. On the other hand, with regard to the accuracy, the recall decreased owing to the false negatives associated with shortening the window and raising the threshold. From result (3) in Table 4, both the accuracy and the real time were the highest in the case of setting 20 minutes for N and 5 for T, and each train line's evaluation results are shown in Table 5. For the JR Kawagoe Line and the Tokyu

Ikegami Line, the recall decreased. This is because the number of tweets, which was naturally small, became smaller and smaller toward the end of the period in which the information occurred, as shown in Figure 5. Additionally, for the Tokyo Metro Fukutoshin Line, both the recall and the precision decreased drastically. We consider this is because people could not tweet from their smartphones owing to the lack of electrical waves. Indeed, as seen in Table 2, very few people had probably posted tweets during the time when the information occurred for the Tokyo Metro Fukutoshin Line.

	N	T	Real time (minute)	Recall	Precision	F measure
(1)	10	5	5.6	0.68	0.95	0.80
(2)		10	7.8	0.44	1.00	0.62
(3)	20	5	5.2	0.88	0.80	0.84
(4)		10	8.2	0.69	0.96	0.80

Table4: Overall evaluation results. N represents the window time and T represents the threshold.

Train line	Real time (minute)	Recall	Precision	F measure
JR Yamanote Line	0.3	0.97	0.85	0.90
JR Chuo Line	1.3	0.98	0.81	0.89
JR Sobu Line	4.5	0.96	0.80	0.87
JR Yokosuka Line	2.0	0.98	0.63	0.77
JR Takasaki Line	5.0	0.90	0.65	0.76
JR Kawagoe Line	21.0	0.64	0.95	0.76
Keio Inokashira Line	2.0	0.97	0.73	0.86
Tokyo Metro Fukutoshin Line	13.0	0.13	0.15	0.14
Tobu Tojo Line	5.0	0.79	0.93	0.85
Tokyu Ikegami Line	9.0	0.56	1.00	0.72

Table5: Each train line's evaluation results (N 20, T 5).

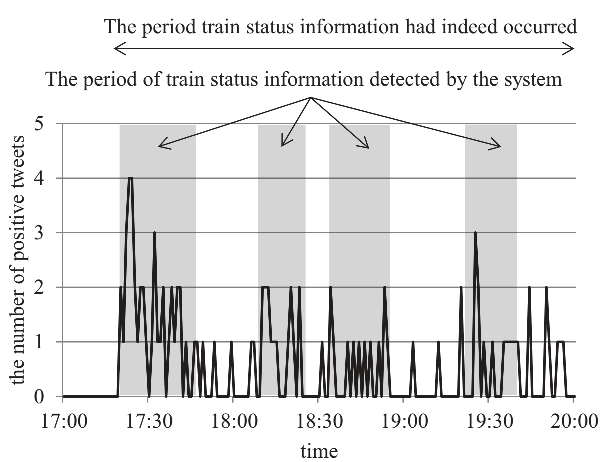

Figure5: An example of false negatives (the JR Kawagoe Line).

Optimized Parameter Values

From Table 4, we found that both the accuracy and the real time were the highest in the case of setting 20 minutes for N and 5 for T. Thus, we changed the parameter values for each train line based on the number of tweets. Since we considered the threshold affected much more than the time window, we changed only T for each train line (We fixed N at 20 minutes). We estimated that while false negatives would be decreased for the train lines where the numbers of tweets are small by lowering the threshold, positive negatives would be decreased for the train lines where the numbers of tweets are large by raising the threshold, and then we expected both the recall and the precision would be improved. The average numbers of tweets per day for each train line are shown in Table 6. We classified the train lines into two categories: those with over 1,000 tweets per day and those with under 1,000 tweets per day, and we explored the optimized value between 1 and 10 for every category. A part of each train line's evaluation results and the overall evaluation results are shown in Table 7. We obtained the optimized outcome in the case of setting 8 as the threshold for the train lines where the numbers of tweets per day were over 1,000, and 3 as the threshold for

Train line	The # of tweets per day (avg.)
JR Yamanote Line	3,761
JR Chuo Line	2,262
JR Sobu Line	1,212
JR Yokosuka Line	567
JR Takasaki Line	283
JR Kawagoe Line	85
Keio Inokashira Line	473
Tokyo Metro Fukutoshin Line	438
Tobu Tojo Line	618
Tokyu Ikegami Line	128

Table6: Average numbers of tweets per day for each train line.

Train line	T	Real time (minute)	Recall	Precision	F measure
JR Chuo Line	5	1.3	0.98	0.81	0.89
	8	3.3	0.93	0.94	0.94
JR Sobu Line	5	4.5	0.96	0.80	0.87
	8	5.0	0.94	0.90	0.92
JR Kawagoe Line	5	21.0	0.64	0.95	0.76
	3	6.0	0.89	0.88	0.88
Tokyu Ikegami Line	5	8.0	0.66	0.99	0.79
	3	3.0	0.78	0.94	0.85
Overall Average	5	5.2	0.88	0.80	0.84
	3,8	3.5	0.92	0.80	0.85

Table7: Evaluation results in the case of setting optimized parameter values for each train line.

the train lines where the numbers of tweets per day were under 1,000. On average, we improved the recall by 4 points while maintaining the precision and shortening the time taken to detect train status information to less than 4 minutes. The precision was not improved, since the number of false positives was originally small. In the case that we lowered the threshold for the JR Kawagoe Line and the Tokyu Ikegami Line, the recall significantly improved and the precision decreased only slightly. This is because we could reduce the false negatives as shown in Figure 5. Consequently the real time was also significantly improved. On the other hand, in the case that we raised the threshold for the JR Chuo Line and the JR Sobu Line, the precision improved and the recall decreased only slightly. However, the improvement was slight compared to that of the train lines for which we lowered the threshold to reduce false negatives, since the number of false positives was originally small. These results indicate that changing the parameter values for each train line, based on the number of tweets, is effective for improving both the recall and the real time.

Discussion

We confirmed that by exploiting Twitter as a sensor we could detect train status information with a high degree of both the accuracy, the F-measure was more than 0.85, and the real time, the time taken to detect the information was less than 4 minutes, provided we took account of the differences in the numbers of tweets among train lines. Indeed, most railway companies in Japan make announcements only in the case of delays of more than 30 minutes [5], and sometimes notifications are published several hours after they occur. Therefore, we consider the values obtained in this work are sufficient. For further improvement, we should consider the difference in the number of tweets depending on the scale of the occurrence. For instance, even for the same line, the number of tweets differs between a day-long suspension of services and services being 10 or 20 minutes behind schedule. In order to further improve both accuracy and real time, it will be necessary to take these differences into consideration.

It is also desirable to detect transportation information other than that focused on in this paper. For instance, we could probably detect train status information about the resumption of services by applying the same approach, utilizing keywords such as “resume” or “start”. On the other hand, it is extremely difficult to detect zones where trains are suspended or running behind schedule through Twitter. Indeed, in some cases, while trains are suspended in one zone, they may be running in another zone on the same line. In order to detect this information, we should also utilize the location information provided by GPS mounted on smartphones. In addition, it is also difficult to detect information about traffic congestion, since few tweets are posted when getting in cars. Thus, it is desirable to develop a service enabling drivers to tweet by voice and inducing passengers to tweet, or it is necessary to utilize the information captured by physical sensors such as cameras installed on roads. In other words, if we use a combination of textual information through Twitter and real-world information through GPS or cameras, we will be able to detect zones where train status information has occurred as well as the traffic information.

[5] http://traininfo.jreast.co.jp/train_info/e/service.aspx

Conclusion

We confirmed the high potential of exploiting Twitter for detecting transportation information, focusing on train status information about the suspension or delay of services. In future work, we intend to explore the possibility of detecting information other than that focused on in this paper, exploiting not only Twitter as a sensor but also GPS or cameras as physical sensors.

References

An, S. H.; Lee, B. H.; and Shin, D. R. 2011. A Survey of Intelligent Transportation Systems. In Proc. of the 3rd International Conference on Computational Intelligence, Communication Systems and Networks, 332 337.

Aramaki, E.; Maskawa, S.; and Morita, M. 2011. Twitter Catches The Flu: Detecting Influenza Epidemics using Twitter. In Proc. of the 2011 Conference on Empirical Methods in Natural Language Processing, 1568 1576.

Demirbas, M.; Bair, M. A.; Akcora, C. G.; Yilmaz, Y. S.; and Ferhatosmannoglu, H. 2010. Crowd Sourced Sensing and Collaboration Using Twitter. In Proc. of the 2010 IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks, 1 9.

Java, A.; Song, X.; Finin, T.; and Tseng, B. 2007. Why We Twitter: Understanding Microblogging Usage and Communities. In Proc. of the Joint 9th WEBKDD and 1st SNA KDD 2007 Workshop on Web Mining and Social Network Analysis, 56 65.

Li, X.; Lu, R.; Liang, X.; Shen, X.; Chen, J.; and Lin, X. 2011. Smart Community: An Internet of Things Application. *IEEE Communications Magazine* 49(12): 68 75.

Sakaki, T.; Okazaki, M.; and Matsuo, Y. 2010. Earthquake Shakes Twitter Users: Real time Event Detection by Social Sensors. In Proc. of the 19th International World Wide Web Conference, 851 860.

Sheth, A. 2009. Citizen Sensing, Social Signals, and Enriching Human Experience. *IEEE Internet Computing* 13(4): 87 92.

Tumasjan, T.; Sprenger, T. O.; Sandner, P. G.; and Welpe, I. M. 2010. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In Proc. of the 4th International AAAI Conference on Weblogs and Social Media, 178 185.