

Can Twitter Be an Alternative of Real-World Sensors?

Tetsuro Takahashi, Shuya Abe, and Nobuyuki Igata

Fujitsu Laboratories, Ltd.

1-1, Kamikodanaka 4-chome, Nakahara-ku, Kawasaki 211-8588, Japan
{takahashi.tet,abe.shuya,igata}@jp.fujitsu.com

Abstract. Twitter is the most famous on-line microblogging service now. People can post (tweet) what they are doing in 140 characters. Since Twitter posts (tweets) reflect what people are looking, hearing, feeling and so on, we can obtain information about Real-world phenomena through the large amount of tweets. In other words, Twitter can be regarded as a sensor of Real-world phenomena including natural phenomena such as hay fever. This motivated us to investigate whether can Twitter be an alternative of Real-world Sensor. In this paper, we first describe about our system which collects and analyzes tweets in order to generates a hay fever map just like as a weather report map. There are some difficulties such as location estimation and normalization of number of tweets. Using the output of the system, we discuss the comparison with actual pollen data gathered by real sensors. The result shows that Twitter can reflect natural phenomena in some particular areas.

Keywords: Twitter, crowd knowledge, social sensor.

1 Introduction

Twitter is the most famous online microblogging service. People can post messages but are restricted to 140 characters.¹ Since tweets reflect what people are doing, seeing, hearing and feeling, we can obtain information about real-world phenomena through the tweets. In other words, Twitter can be regarded as a sensor of real-world phenomena, including natural phenomena, such as hay fever. Once we prove the possibility of Twitter as a sensor for real-world phenomena, we will be able to measure various phenomena without actual sensors. Furthermore, we can also acquire immeasurable data, such as how happy or tired people might be. This motivated us to investigate whether Twitter could be an alternative to real-world sensor. Based on this background, we chose hay fever (pollinosis) as the target of our experiment. This target is convenient for evaluation, because we could obtain actual data from real-world sensors for pollen.

Twitter is categorized as Consumer Generated Media (CGM). The most typical CGM is blogs for which there are many services and a great deal of research has been done [1]. Compared with blogs, Twitter has at least two major characteristics:

¹ We call the posted text a “tweet” instead of another name like status, messages or posts.

- *Real-time nature*
While blogs are posted one to several times a day, tweets are posted many times a day. People post “what is happening now” in real-time rather than as messages posted each day.
- *Geolocation information*
Twitter can attach geolocation information (latitude and longitude) to each tweet. People often use mobile devices to tweet compared to blogs, because of the restriction on text length and the nature of the service where people input their status.

These characteristics tell us that Twitter is more appropriate than the former CGM as the information source for sensors.

There are many researches on Twitter that have examined Twitter from various aspects, such as usage [2], social network [3] and communication [4]. Regarding the idea of using Twitter as a sensor, the most related work of this paper is Sakaki et al. [5]. They have tried to detect earthquakes using a social sensor that is based on tweets. While their objective was event detection, we aimed to measure the degree of hay fever. Compared with their work, our work has a complexity in both the calculation and the evaluation of degree. This is the contribution of this paper.

We developed a system that extracts information about hay fever from tweets and that visualizes the condition of hay fever for the entire country (Japan). We provided the output of the system as the Web service, “Hay fever now!” from March 2010 to May 2010. We will introduce the system in the next section.

2 Hay Fever Observation System

The goal of the system is to generate a hay fever map similar to a weather report map. Fig. 1 is an example of output. The system first collects tweets that contain the key phrase “hay fever”² by using the Twitter search API³. Then, the system sorts tweets by location (prefecture) and visualizes them in a map.

There are some difficulties in the implementation, such as location estimation, normalization of tweet numbers and classification of tweets. We will describe them in the following sections.

2.1 Location Estimation

Every tweets are able to have geolocation information (latitude and longitude), which is very useful for the objective of our system, however, not all users give permission to provide the information. Our preliminary investigation showed that only 0.6% of tweets included geolocation information.⁴ Because this is not sufficient our purposes, we used published user profile information as an alternative to geolocation information. Users can optionally register their location when registering for a Twitter account.

² Written in Japanese in the system.

³ <http://apiwiki.twitter.com/>

⁴ We investigated tweets that include the phrase “hay fever”

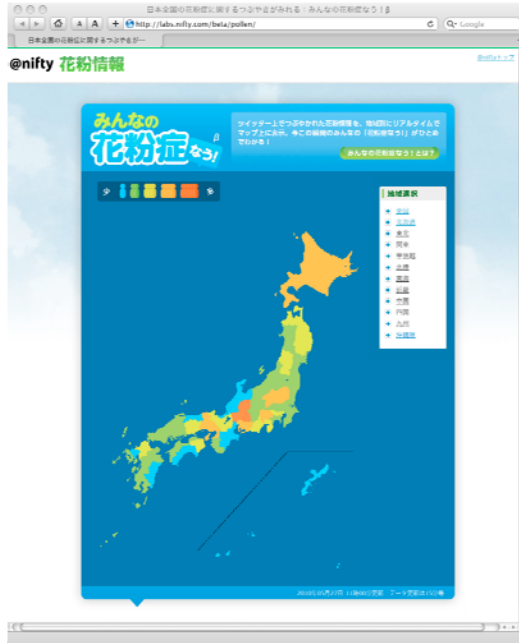


Fig. 1. Output of pollen observation system

Since users write the profile information in natural language, we needed to analyze the information and estimate the name of the prefecture. For the prefecture name estimation, we made a dictionary using a gazette edited by the Geospatial Information Authority of Japan⁵. The dictionary has the name of the city, ward, town, village, mountain, hill and lake for each corresponding prefecture. Most of the terms in the dictionary are written in kanji.⁶ We expanded them into hiragana and the English alphabet; Twitter users sometimes wrote their location in their profiles in hiragana “さっぽろ” (“Sapporo” in English) and the English alphabet “Kanagawa, Japan”. The number of terms in the dictionary was expanded from 3,196 to 11,584.

The 11,584 terms had 664 duplications. We used a Web search engine with the query “prefecture name & land name” to resolve the duplications. The prefecture name with the highest number of search results was selected as corresponding land name for each duplicate land name.

We attempted to use bi-gram similarity for the matching between the dictionary and profiles. Some users wrote their locations in the profiles as “i’m living tokyo now” which should be matched as “Tokyo”.

2.2 Normalization

The popularity of Twitter is different in each prefecture. Moreover, the average number of tweets posted by one person in a prefecture may vary from the others. Fig. 2

⁵ http://www.gsi.go.jp/KOKUJYOHO/gazetteer_j.html

⁶ Japanese has three character sets, hiragana katakana and kanji.

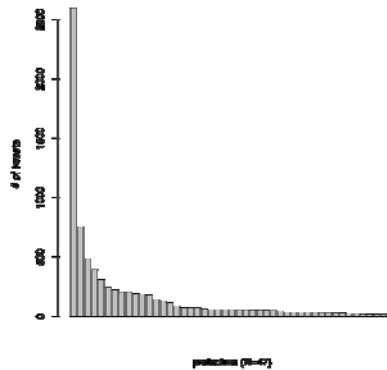


Fig. 2. Distribution of tweets over prefectures

shows the distribution of the number of tweets. Clearly we need to normalize the number. For the normalization, we collected 7,000 random tweets determined whether geolocation was added to the tweet or whether the location was estimated from the user's profile.

Using the distribution of tweets over all prefectures, we calculated a weight for each prefecture as the proportion to the max number, the number of tweets in Tokyo. The highest weight value was 259.4 for Shimane Prefecture. This meant that the number of tweets posted by users in Tokyo was 259.4 times as many as the number in Shimane Prefecture.

2.3 Classification of Tweet

If the system relied only on the key phrase “hay fever” for the data collection, it used many tweets in which the user does not suffer from hay fever. The tweets may be, “I do not want to get hay fever.” or “Am I getting hay fever? No I don't believe that”. We attempted to classify all target tweets into two classes: getting hay fever and not getting hay fever. In order to make the classifier, we first classified 1,000 tweets by hand into five levels as described in Table 1. It was a difficult task to classify tweets into the binary classes of “hay fever” or “no hay fever” even by person because many of the tweets were vague. Five classifications were more suitable for this reason.

We used a classifier developed by Iwakura et al. [6] for the classification. The classifier was trained with 1,000 examples of (1) and (2) as positive examples and (3), (4) and (5) as negative examples. This means that the classifier organized the tweets into two classes. The accuracy obtained after five fold cross-validation was 77.27%. The classification was not an easy task. It was difficult to discriminate between suffering from hay fever or not from only the surface information of text messages. While model selection or parameter tuning may give higher performance, we used the classifier for our experiment because the improvement was estimated as minimal because of the difficulty of the task.

Table 1. Category of manual classification

Level	Number	Description
(1)	430	Hay fever with symptoms
(2)	144	Might be hay fever with symptoms
(3)	119	Hay fever but no symptoms
(4)	67	Might be hay fever and no symptoms
(5)	240	Not hay fever and no symptoms

3 Evaluation

3.1 Evaluation in Location Estimation

Precision of Location Estimation. In the experiment, the system estimated location information for 70.4% of users with the algorithm described in Section 2.1. Table 2 shows the results of the estimation for 200 random sampled examples.

The precision of the estimation is 74.5% (149/200). The examples of “Fail” occurred for complex expressions as shown in Table 3. Example expressions of “Meaningless” are “here”, “town” and “in my home” from which the system should not estimate a location. In order to prevent an incorrect estimation, we can use (1) a threshold in the bi-gram matching score, and (2) the language model, which estimates the inadequacy of expression as location.

Recall of location estimation. In order to evaluate the recall of the location estimation, we checked location expressions for 29.6% users for which the system could not estimate their location. Table 4 shows the results of the evaluation for 200 random sampled examples.

The number of expressions for which the system should provide location information is 18, the sum of the number for “Detailed local” and “Slang” in Table 4. This rate is 9.8% out of the 200 samples.

Table 2. Evaluation of precision for location estimation

Category	Number
Correct	149
Fail	16
Meaningless	35

Table 3. Failure example of location estimation

Example expression	Corresponding expression in English
ポロサツシティ	“YorkNewCity” (for “New York City”)
(竜)宮城	brand new youk
大阪と京都といたりきたり	Back and forth between Osaka and Kyoto

Table 4. Evaluation of recall for location estimation

Category	Example expression	Number
Detailed local	Katsushika ward, Edogawa ward	15
Slang	しぞーか, ちーぼ ⁷	3
Wider area 1	Kanto region, Kyushu	13
Wider area 2	Japan, earth	40
Meaningless	strange adventure, “home”, “room”	129

The results of the above evaluation are as follows:

- Rate of appropriate location written in the user profile was $0.607 (0.704 * ((149+16)/200) + 0.296 * (18/200))$,
- Rate of correct estimation by the system was $0.524 (0.704 * ((149)/200))$,
- Recall of location estimation was $0.863 (0.524 / 0.607)$.

Consistency in Locations between geolocations and user profiles. The locations estimated from user profiles may be different from the locations where the user tweeted because Twitter is used often on mobile devices. In order to evaluate consistency, we compared actual locations obtained by geolocations with estimated locations from user profiles. Since 195 users posted actual geolocation, the comparison shown in Table 5 could be attempted.

In the comparison, actual locations might be multiple, and estimated location was one for one user. When (1) and (2) are regarded as correct in Table 5, the consistency is 72.8 % (142 / 195). Adding (3) as a correct, the consistency is 85.6% (167 / 195). Remembering that only 60.7% users indicated the appropriate location in their profiles, this ratio is acceptable.

Table 5. Location difference between geolocations and user profiles

Category	Number
(1) One location, and the same	132
(2) Multiple location, and includes the estimated one	10
(3) Different (neighbor)	25
(4) Different	28

3.2 Comparison with Real Sensor Data

Real Sensor Data. Using the system output generated from tweets, we discuss the comparison with actual pollen data gathered by real sensors. This analysis is based on the correlation between them. If a significant correlation exists, we can say that tweets can be used as an alternative to real sensors.

We used actual pollen data gathered by real sensors for the comparison. The pollen data were provided by a pollen observation system⁸ developed by the Ministry of the Environment Japan. This system provides pollen data gathered from observatories in

⁷ The corresponded expressions are; NYork for New York, boostoon for Boston

⁸ <http://kafun.taiki.go.jp/>

all 46 prefectures, except Okinawa. Because each prefecture has 2 to 10 observatories (average: 3.1), the average value of the multiple observations for each prefecture was used for the comparison.

Correlation between Pollen and Tweets. We first calculated the value of pollen data and tweet data by day. Then, we calculated the correlation coefficient between the two data sets over the experimental period (Feb. 14, 2010 – Mar. 5, 2010) for each prefecture. Table 6, Fig. 3 and Fig. 4 show the results. This analysis was conducted for the two data sets: without classification and with classification. Both results had a weak correlation for mean, 0.145 and 0.153, respectively. The classification seems to contribute to the correlation coefficient, however, the results of a statistical hypothesis test (T-test) showed that there is no difference between these two values. This result confirmed that text classification did not contribute to a correlation between real pollen data and tweet data.

Effect of the Number of Tweets and the Amount of Pollen. There is considerable divergence in the correlation coefficient among prefectures as shown in Fig. 3 and Fig. 4. In order to analyze the effect of the number of tweets and the amount of pollen on the divergence, we confirmed the correlation between them.

Table 6. Summary of correlation coefficient for prefectures

	w/o classification	w/ classification
Min.	-0.184	-0.202
1 st Q	0.00323	-0.0378
Median	0.0992	0.125
Mean	0.145	0.153
3 rd Q	0.248	0.249
Max	0.646	0.703

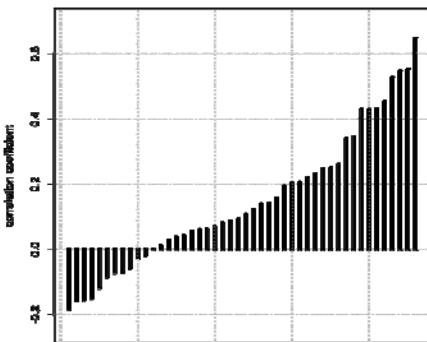


Fig. 3. Correlation coefficients for prefectures (w/o classification)

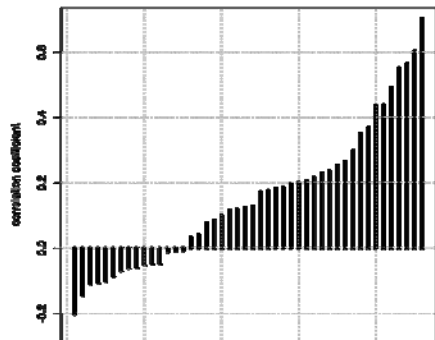


Fig. 4. Correlation coefficients for prefectures (w/ classification)

The result of Pearson's product-moment correlation test for the number of tweets was a 0.31 correlation coefficient with a 0.05 level of statistical significance for data without classification, and 0.33 for data with classification. This result confirmed that the more tweets that are posted, the higher the correlation coefficient between real pollen data and tweet data. In other words, the more tweets that are posted, the higher the possibility that Twitter can be used as an alternative for real sensors.

We also attempted the same analysis for the amount of pollen, however, there was no significant correlation between them.

Location Distribution of Correlation Coefficient. Fig. 5 shows the distribution of the correlation coefficient over a map of Japan for each prefecture. Dark color means high value of the correlation coefficient. While prefectures that have high correlation coefficients were centered around the Kanto and Chugoku regions, prefectures that have low correlation coefficients were centered around the Kansai region. The mean value of correlation coefficients in the Kanto region is 0.434, which is significantly higher than the mean value of all prefectures (0.153). On the other hand, the mean value of the correlation coefficients in the Kansai region is -0.0108 that is a significantly small value.

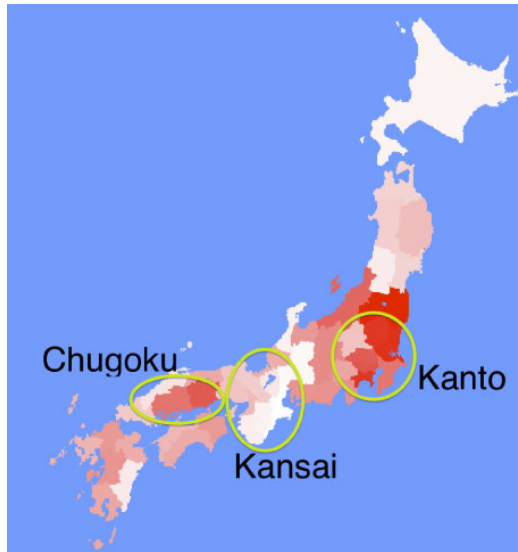


Fig. 5. Distribution of correlation coefficients on a map

4 Discussion

4.1 Method of Evaluation

Evaluating the output of our system was very difficult. We estimated the degree of hay fever in this experiment. This estimation was based on the hypothesis that the more people who showed symptoms of hay fever, the more tweets that would be

posted. In order to evaluate the estimation properly, it should be compared with data on the symptoms of hay fever instead of on pollen number. However, it was difficult to obtain data, such as the number of patients with hay fever. We used the pollen counts as an alternative indicator of symptoms of hay fever and investigated the correlation between the count and the degree, which was estimated from tweets. We found a correlation between them in several areas, however, the lack of a correlation in the rest of the areas might be due to the problem of the aforementioned hypothesis. Even if the pollen count were the same in two areas, the symptoms of hay fever might be different because of other conditions, such as temperature, humidity and wind. The evaluation should be refined by involving the different kinds of data.

4.2 Degree of Symptoms in One Tweet

In this experiment, all tweets were calculated equally, except document classification, which we mentioned in Section 2.3. This meant that the degree of symptoms in each tweet was not determined, however, actual tweets had different degrees as shown in the following tweet examples:

- My eyes are itchy from hay fever,
- Sneeze, mucous, itchy eyes, this is surely hay fever,
- My eyes are so irritated! Hay fever!

Since we tried to measure the degree of phenomena, such as hay fever instead of event detection as mentioned in Section 1, we should extract information about the degree from each tweet. Extending the classification described in Section 2.3 from the binary classification to a multi-class classification can carry out this objective, however, we need a more accurate classification method than the current one (77.27% accuracy) because the high number of classes causes a reduction in accuracy.

5 Conclusion

This paper verified that Twitter could be used as an alternative to real-world sensors. The results showed that there was a positive correlation between pollen data and tweet data, and Twitter has the possibility of being used as a real-world sensor at least in a particular area such as hay fever. Our experiments proved that the number of tweets affected the correlation. This result confirmed quantitatively an intuitive interpretation that the number of tweets leads to high performance as a sensor. Our experiments also showed the limit of our algorithm, such as the divergence of the correlation coefficient among regions. We must seek the reason of the divergence and must involve them into the system architecture in order to make the sensor accurate.

References

1. Pang, B., Lee, L.: Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* A1 2(1-2), 1–135 (2008)
2. Java, A., Song, X., Finin, T., Tseng, B.: Why we twitter: understanding microblogging usage and communities. In: *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web mining and Social Network Analysis, WebKDD/SNA-KDD 2007*, pp. 56–65. ACM, New York (2007)

3. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: Proceedings of the 19th international conference on World Wide Web, WWW 2010, pp. 591–600. ACM, New York (2010)
4. Naaman, M., Boase, J., Lai, C.H.: Is it really about me?: message content in social awareness streams. In: Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work, CSCW 2010, pp. 189–192. ACM, New York (2010)
5. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: WWW 2010: Proceedings of the 19th International Conference on World Wide Web, pp. 851–860. ACM, New York (2010)
6. Iwakura, T., Okamoto, S.: A fast boosting-based learner for feature-rich tagging and chunking. In: Proceedings of the Twelfth Conference on Computational Natural Language Learning, CoNLL 2008, pp. 17–24. Association for Computational Linguistics, Morristown (2008)