ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 6, June 2017

Analyzing tweets for Real-Time Event Detection and Development of Earthquake Reporting System

[1] Sumanth V, [2] Aman A. Saxena, [3] Jeevan B, [4] P. Vinay, [5] Vageesha N. J.
[1] Assistant Professor, Dept. of CSE, RRIT, Bangalore
[2][3][4][5] Dept. of CSE, RRIT, Bangalore

Abstract— **Twitter, a popular micro blogging service, has received much attention recently. An important characteristic of Twitter is its real-time nature. For example, when an earthquake occurs, people make many Twitter posts (tweets) related to the earthquake, which enables detection of earthquake occurrence promptly, simply by observing the tweets. As described in this paper, we investigate the real-time inter- action of events such as earthquakes, in Twitter, and pro- pose an algorithm to monitor tweets and to detect a target event. To detect a target event, we devise a classifier of tweets based on features such as the keywords in a tweet, the number of words, and their context. Subsequently, we produce a probabilistic spatiotemporal model for the tar- get event that can find the center and the trajectory of the event location. We consider each Twitter user as a sensor and apply Kalman filtering and particle filtering, which are widely used for location estimation in ubiquitous/pervasive computing. The particle filter works better than other com- pared methods in estimating the centers of earthquakes and the trajectories of typhoons. As an application, we construct an earthquake reporting system in Japan. Because of the numerous earthquakes and the large number of Twitter users throughout the country, we can detect an earth- quake by monitoring tweets with high probability (96% of earthquakes of Japan Meteorological Agency (JMA) seismic intensity scale 3 or more are detected). Our system detects earthquakes promptly and sends e-mails to registered users. Notification is delivered much faster than the announcements that are broadcast by the JMA.**

1. INTRODUCTION

Twitter, a popular microblogging service, has received much attention recently. It is an online social network used by millions of people around the world to stay connected to their friends, family members and co-workers through their computers and mobile phones [18]. Twitter asks one ques- tion, "What are you doing?" Answers must be fewer than
140 characters. A status update message, called a tweet, is often used as a message to friends and colleagues. A user can follow other users; and her followers can read her tweets. currently estimated as 44.5 million worldwide1 . Monthly growth of users has been 1382% year-on-year, which makes Twitter one of the fastest-growing sites in the world2 .
Some studies have investigated Twitter: Java et al. an- alyzed Twitter as early as 2007. They described the social network of Twitter users and investigated the motivation of Twitter users [13]. B. Huberman et al. analyzed more than 300 thousand users. They discovered that the relation between friends (defined as a person to whom a user has directed posts using an "@" symbol) is the key to under- standing interaction in Twitter [11]. Recently, boyd et al. investigated retweet activity, which is the Twitter-equivalent of e-mail forwarding, where users post messages originally posted by others [5].
brief text updates or micromedia such as photographs or au- dio clips. Microblogging services other than Twitter include Tumblr, Plurk, Emote.in, Squeelr, Jaiku, identi.ca, and so on3 . They have their own characteristics. Some examples are the following: Squeelr adds geolocation and pictures to microblogging, and Plurk has a timeline view integrating video and picture sharing. Although our study is applicable to other microblogging services, in this study, we specifically examine Twitter because of its popularity and data volume. An important common characteristic among microblog- ging services is its real-time nature. Although blog users typically update their blogs once every several days, Twit- ter users write tweets several times in a single day. Users can know how other users are doing and often what they are thinking about now, users repeatedly return to the site and check to see what other people are doing. The large num- ber of updates results in numerous reports related to events. They include social events such as parties, baseball games, and presidential campaigns. They also include disastrous events such as storm, fire, traffic jam, riots, heavy rainfall, and earthquakes. Actually, Twitter is used for various real- time notification such as that necessary for help during a large-scale fire

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 6, June 2017

emergency and live traffic updates. Adam Ostrow, an Editor in Chief at Mashable, a social media news blog, wrote in his blog about the interesting phenomenon of the real-time media as follows 4

1. http://www.techcrunch.com/2009/08/03/twitter-reaches-44.5-million-people-worldwide-in-june-comscore/
2. According to a report from Nielsen.com.
3. www.tumblr.com,www.plurk.com, www.emote.in,www.squeelr.com, www.jaiku.com, identi.ca
4. http://mashable.com/2009/08/12/japan-earthquake/

Japan Earthquake Shakes Twitter Users ... And Beyonce: Earthquakes are one thing you can bet on being covered on Twitter (Twitter) first, because, quite frankly, if the ground is shaking, you're going to tweet about it before it even reg- isters with the USGS and long before it gets re- ported by the media. That seems to be the case again today, as the third earthquake in a week has hit Japan and its surrounding islands, about an hour ago. The first user we can find that tweeted about it was Ricardo Duran of Scottsdale, AZ, who, judging from his Twitter feed, has been trav- eling the world, arriving in Japan yesterday.

This post well represents the motivation of our study. The research question of our study is, "can we detect such event occurrence in real-time by monitoring tweets?"

This paper presents an investigation of the real-time na- ture of Twitter and proposes an event notification system that monitors tweets and delivers notification promptly. To obtain tweets on the target event precisely, we apply se- mantic analysis of a tweet: For example, users might make tweets such as "Earthquake!" or "Now it is shaking" thus earthquake or shaking could be keywords, but users might also make tweets such as "I am attending an Earthquake Conference", or "Someone is shaking hands with my boss". We prepare the training data and devise a classifier using a support vector machine based on features such as keywords in a tweet, the number of words, and the context of target-event words.

Subsequently, we make a probabilistic spatiotemporal model of an event. We make a crucial assumption: each Twitter user is regarded as a sensor and each tweet as sensory infor- mation. These virtual sensors, which we call social sensors, are of a huge variety and have various characteristics: some sensors are very active; others are not. A sensor could be inoperable or malfunctioning sometimes (e.g., a user is sleep- ing, or busy doing something). Consequently, social sensors are very noisy compared to ordinal physical sensors. Regard- ing a Twitter user as a sensor, the event detection problem can be reduced into the object detection and location es- timation problem in a ubiquitous/pervasive computing en- vironment in which we have numerous location sensors: a user has a mobile device or an active badge in an environ- ment where sensors are placed. Through infrared commu- nication or a WiFi signal, the user location is estimated

as providing location-based services such as navigation and museum guides [9, 25]. We apply Kalman filters and parti- cle filters, which are widely used for location estimation in ubiquitous/pervasive computing.

As an application, we develop an earthquake reporting system using Japanese tweets. Because of the numerous earthquakes in Japan and the numerous and geographically dispersed Twitter users throughout the country, it is some- times possible to detect an earthquake by monitoring tweets. In other words, many earthquake events occur in Japan. Many sensors are allocated throughout the country. Fig- ure 1 portrays a map of Twitter users worldwide (obtained from UMBC eBiquity Research Group); Fig. 2 depicts a map of earthquake occurrences worldwide (using data from Japan Meteorological Agency (JMA)). It is apparent that the only intersection of the two maps, which means regions with many earthquakes and large Twitter users, is Japan.

Our system detects an earthquake occurrence and sends an e-mail, possibly before an earthquake actually arrives at a certain location: An earthquake propagates at about 3–7 km/s. For that reason, a person who is 100 km distant from an earthquake has about 20 s before the arrival of an earthquake wave.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 6, June 2017

We present a brief overview of Twitter in Japan: The Japanese version of Twitter was launched on April 2008. In February 2008, Japan was the No. 2 country with respect to Twitter traffic5 . At the time of this writing, Japan has the
11th largest number of users (more than half a million users) in the world. Although event detection (particularly the earthquake detection) is currently possible because of the high density of Twitter users and earthquakes in Japan, our study is useful to detect events of various types throughout the world.
The contributions of the paper are summarized as follows:

• The paper provides an example of integration of se- mantic analysis and real-time nature of Twitter, and presents potential uses for Twitter data.

• For earthquake prediction and early warning, many studies have been made in the seismology field. This paper presents an innovative social approach, which has not been reported before in the literature.

This paper is organized as follows: In the next section, we explain semantic analysis and sensory information, followed by the spatiotemporal model in Section 3. In Section 4, we describe the experiments and evaluation of event detection. The earthquake reporting system is introduced into Section
5. Section 6 is devoted to related works and discussion. Finally, we conclude the paper.

2. EVENT DETECTION

In this paper, we target event detection. An event is an ar- bitrary classification of a space/time region. An event might have actively participating agents, passive factors, products, and a location in space/time [21]. We target events such as earthquakes, typhoons, and traffic jams, which are visible through tweets. These events have several properties: i) they are of large scale (many users experience the event), ii) they particularly influence people's daily life (for that reason, they are induced to tweet about it), and iii) they have both spatial and temporal regions (so that real-time location estimation would be possible). Such events include social events such as large parties, sports events, exhibi- tions, accidents, and political campaigns. They also include natural events such as storms, heavy rainfall, tornadoes, typhoons/hurricanes/cyclones, and earthquakes. We des- ignate an event we would like to detect using Twitter as a target event.

2.1 Semantic Analysis on Tweet

To detect a target event from Twitter, we search from Twitter and find useful tweets. Tweets might include men- tions of the target event. For example, users might make tweets such as "Earthquake!" or "Now it is shaking". Con- sequently, earthquake or shaking could be keywords (which we call query words). but users might also make tweets such as "I am attending an Earthquake Conference", or "Some- one is shaking hands with my boss".

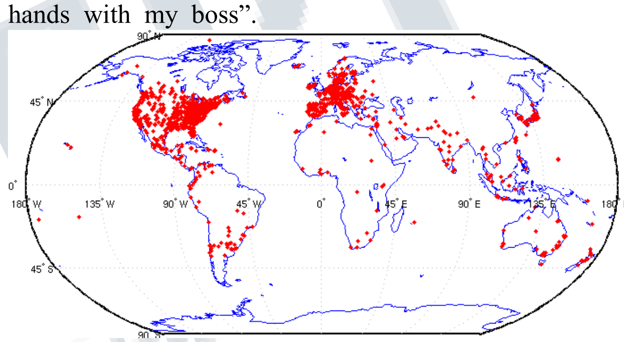

Figure 1: Twitter user map.

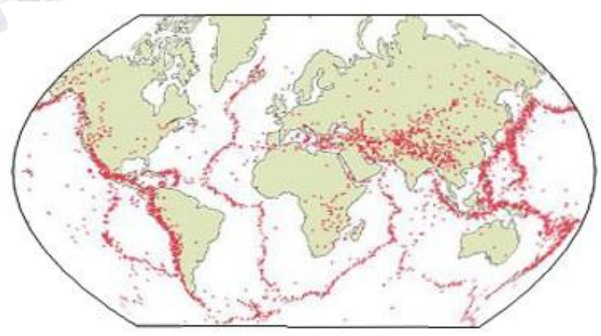

Figure 2: Earthquake map.

tweet is referring to the target event, it might not be appro- priate as an event report; for example a user makes tweets such as "The earthquake yesterday was scaring", or "Three earthquakes in four days. Japan

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 6, June 2017

scares me.” These tweets are truly the mentions of the target event, but they are not real-time reports of the events. Therefore, it is necessary to clarify that a tweet is actually referring to an actual earth- quake occurrence, which is denoted as a positive class.
To classify a tweet into a positive class or a negative class, we use a support vector machine (SVM) [14], which is a widely used machine-learning algorithm. By preparing pos- itive and negative examples as a training set, we can pro- duce a model to classify tweets automatically into positive and negative categories.
We prepare three groups of features for each tweet as fol- lows:

Features A (statistical features) the number of words in a tweet message, and the position of the query word within a tweet.

Features B (keyword features) the words in a tweet6 .

Features C (word context features) the words before and after the query word.

To handle Japanese texts, morphological analysis is con- ducted using Mecab7 , which separates sentences into a set of words. In the case of English, we apply a standard stop- word elimination and stemming. We compare the usefulness of the features in Section 4. Using the obtained model, we can classify whether a new tweet corresponds to a positive class or a negative class.

2.2 Tweet as a Sensory Value

We can search the tweet and classify it into a positive class if a user makes a tweet on a target event. In other words, the user functions as a sensor of the event. If she makes a tweet about an earthquake occurrence, then it can be considered that she, as an ”earthquake sensor”, returns a positive value. A tweet can therefore be considered as a sensor reading. This is a crucial assumption, but it enables application of various methods related to sensory information.

Assumption 2.1 Each Twitter user is regarded as a sen- sor. A sensor detects a target event and makes a report probabilistically.

The virtual sensors (or social sensors) have various char- acteristics: some sensors are activated (i.e. make tweets) only about specific events, although others are activated to a wider range of events. The number of sensors is large; there are more than 40 million sensors worldwide. A sen- sor might be inoperable or operating incorrectly sometimes (which means a user is not online, sleeping, or is busy do- ing something). Therefore, this social sensor is noisier than ordinal physical sensors such as location sensors, thermal sensors, and motion sensors.
A tweet can be associated with a time and location: each tweet has its post time, which is obtainable using a search API. In fact, GPS data are attached to a tweet sometimes, e.g. when a user is using an iPhone. Alternatively, each Twitter user makes a registration on their location in the user profile. The registered location might not be the current location of a tweet; however, we think it is probable that a person is near the registered location. In this study, we use GPS data and the registered location of a user. We do not use the tweet for spatial analysis if the location is not available (We use the tweet information for temporal analyses.).

Assumption 2.2 Each tweet is associated with a time and location, which is a set of latitude and longitude.

By regarding a tweet as a sensory value associated with a location information, the event detection problem is re- duced to detecting an object and its location from sensor readings. Estimating an object’s location is arguably the most fundamental sensing task in many ubiquitous and per- vasive computing scenarios [7].
Figure 3 presents an illustration of the correspondence between sensory data detection and tweet processing. The motivations are the same for both cases: to detect a target event. Observation by sensors corresponds to an observa- tion by Twitter users. They are converted into values by a classifier. A probabilistic model is used to detect an event, as described in the next section.

3. MODEL

In order for event detection and location estimation, we use probabilistic models. In this section, we first describe event detection from time-series data. Then, we describe the location estimation of a target event.

3.1 Temporal Model

Each tweet has its post time. When a target event oc- curs, how can the sensors detect the event? We describe the temporal model of event detection.

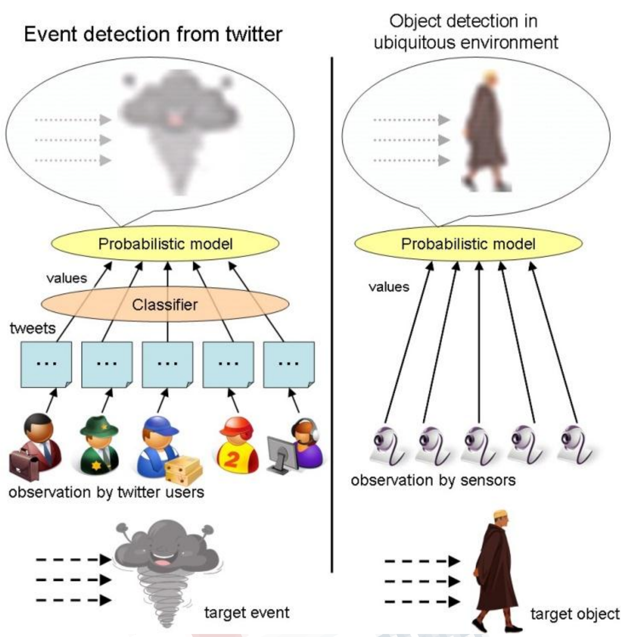

Figure 3: Correspondence between event detection from Twitter and ob ject detection in a ubiquitous environment.

To assess an alarm, we must calculate the reliability of multiple sensor values. For example, a user might make a false alarm by writing a tweet. It is also possible that the classifier misclassifies a tweet into a positive class.

3.2 Spatial Model

Each tweet is associated with a location. We describe how to estimate the location of an event from sensor readings.

To define the problem of location estimation, we consider the evolution of the state sequence {xt ,t ∈ N} of a target, an earthquake and a typhoon. It is apparent that spikes occur on the number of tweets. Each corresponds to an event occurrence. In the case of an earthquake, more than

10 earthquakes occur during the period. In the case of ty- phoon, Japan's main population centers were hit by a large typhoon (designated as Melor) in October 2009.

The distribution is apparently an exponential distribu- tion. The probability density function of the exponential distribution is f (t; λ) = λe−λt where t > 0 and λ > 0. The exponential distribution occurs naturally when describ- ing the lengths of the inter-arrival times in a homogeneous Poisson process.

In the Twitter case, we can infer that if a user detects an event at time 0, assume that the probability of his posting a tweet from t to Δt is fixed as λ. Then, the time to make a tweet can be considered as an exponential distribution. Even if a user detects an event, therefore, she might not make a tweet right away if she is not online or doing some- thing. She might make a post only after such problems are resolved. Therefore, it is reasonable that the distribution of the number of tweets follows an exponential distribution. Actually the data fits very well to an exponential distribu- tion; we get λ = 0.34 with R2 = 0.87 on average.

Presuming that p(xt−1 |zt−1) is available, the prediction stage uses the following equation: p(xt |zt−1) = R p(xt |xt−1)

p(xt−1 |zt−1) dxt−1 . Here we use a Markov process of order one. Therefore, we can assume p(xt |xt−1 , zt−1) = p(xt |xt−1). In update stage, the Bayes' rule is applied as p(xt |zt) = p(zt |xt)p(xt |zt−1)/p(zt |zt−1), where the normalizing constant

is p(zt |zt−1) = R p(zt |xt)p(xt |zt−1)dxt .

To solve the problem, several methods of Bayesian filters are proposed such as Kalman filters, multi-hypothesis track- ing, grid-based and topological approaches, and particle fil- ters [7]. For this study, we use Kalman filters and particle filters, both of which are widely used in location estimation.

3.2.1 Kalman Filters

The Kalman filter assumes that the posterior density at every time step is Gaussian and that it is therefore param- eterized by a mean and covariance. We can write it as $x_t = F_t x_{t-1} + v_{t-1}$ and $z_t = H_t x_t + n_t$. Therein, F_k and H_k are known matrices defining the linear functions. The covariants of v_{k-1} and n_k are, respectively, Q_{t-1} and R_k.

Algorithm 1 Particle filter algorithm

1. Initialization: Calculate the weight distribution $D_w(x, y)$ from twitter users geographic distribution in Japan.

2. Generation: Generate and weight a particle set, which means N discrete hypothesis.

(1) Generate a particle set $S_0 = (s_{0,0}, s_{0,1}, s_{0,2}, ..., s_{0,N-1})$ and allocate them on the map evenly: particle $s_{0,k} = (x_{0,k}, y_{0,k}, weight_{0,k})$, where x corresponds to the longitude and y corre- sponds to the latitude.

(2) Weight them based on weight distribution $D_w(x, y)$.

3. Re-sampling

(1) Re-sample N particles from a particle set S_t using weights of each particles and allocate them on the map. (We allow to re-sample same particles more than one.)

(2) Generate a new particle set S_{t+1} and weight them based on weight distribution $D_w(x, y)$.

Prediction: Predict the next state of a particle set S_t from the Newton's motion

4. EXPERIMENTS AND EVALUATION

we describe the experimental results and evaluation of tweet classification and location estimation We prepare a set of queries Q for an target event. We first search for tweets T including the query set Q from Twitter every s seconds. We use a search API11 to search tweets. In the earthquake case, we set Q = {"earthquake" and "shaking"} and in the typhoon case, we set Q = {"typhoon"}. We set s as 3 s. After determining a classification and obtaining a positive example, the system makes a calculation of a temporal and spatial probabilistic model. We consider that an event is detected if the probability is higher than a certain threshold ($p_{occur}(t) > 0.95$ in our case). The location information of each tweet is obtained and used for location estimation of the event. In the earthquake reporting system explained in the next section, the system quickly sends an e-mail (usually mobile e-mail) to registered users.

4.1 Evaluation by Semantic Analysis

. We used a linear kernel for SVM. We obtain the highest F -value when we use feature A and all features. Surprisingly, feature B and feature C do not contribute much to the classification performance. When an earthquake occurs, a user becomes surprised and might produce a very short tweet. It is apparent that the recall is not so high as precision. It is attributable to the usage of query words in a different context than we intend. Sometimes it is difficult even for humans to judge whether a tweet is reporting an actual earthquake or not. Some ex- amples are that a user might write "Is this an earthquake or a truck passing?" Overall, the classification performance is good considering that we can use multiple sensor readings as evidence for event detection.

4.2 Evaluation of Spatial Estimation

The location estimation of an earthquake on August 11. We can find that many tweets originate from a wide region in Japan. The estimated location of the earth- quake (shown as estimation by particle filter) is close to the actual center of the earthquake, which shows the efficiency of the location estimation algorithm.

location estimation for 25 earthquakes in August, September, and October 2009. We compare Kalman filter- ing and particle filtering, with the weighted average and the median as a baseline. The weighted average simply takes the average of latitudes and longitude on all the positive tweets, and median simply takes the median of them. Particle filters perform well compared to other methods. The poor perfor- mance of Kalman filtering implies that the linear Gaussian assumption does not hold for this problem. We can find that if the center of the earthquake is in the sea area, it is more difficult to locate it precisely from tweets. Similarly, it becomes more difficult to make good estimations in less- populated areas. That is

reasonable: all other things being equal, the greater the number of sensors, the more precise the estimation will be.
the trajectory estimation of typhoon Melor based on tweets. In the case of an earthquake, the center is one location. However, in the case of a typhoon, the center moves and makes a trajectory. The comparison of the performance is shown in Table 3. The particle filter works well and outputs a similar trajectory to the actual trajectory.

5. EARTHQUAKE REPORTING SYSTEM

We developed an earthquake reporting system using the event detection algorithm. Earthquake information is much an earthquake is usually made within a minute or so. The delay can result from the time for posting a tweet by a user, the time to index the post in Twitter servers, and the time to make queries by our system. We apply classification for
49,314 tweets retrieved by query words in one month; We can turn off a stove or heater in our house and hide ourselves under a desk or table if we have several seconds before an earthquake actu- ally hits. Several Twitter accounts report earthquake occur- rence. Some examples are that the United States Geological Survey (USGS) feeds tweets on world earthquake informa- tion, but it is not useful for prediction or early warning.
Vast amounts of work have been done on intermediate- term earthquake prediction in the seismology field (e.g. [23]). Various attempts have also been made to produce short- term forecasts to realize an earthquake warning system by observing electromagnetic emissions from ground-based sen- sors and satellites [3]. Other precursor signals such as iono- spheric changes, infrared luminescence, and air-conductivity change, along with traditional monitoring of movements of the earth's crust, are investigated.

The government has allocated a considerable amount of its budget to mitigating earthquake damage. An earthquake early warning service has been operated by JMA since 2007. It provides advance announcements of the es- timated seismic intensities and expected arrival times. It detects P-waves (primary waves) and makes an alert imme- diately so that earthquake damage can be mitigated through countermeasures such as slowing trains and controlling el- evators. In fact, P-waves are a type of elastic wave that can travel faster than the S-waves (secondary waves), which cause shear effects and engender much more damage.
The proposed system, called Toretter13 , has been operated since August 8 of this year. A system screenshot is depicted in Fig. 11. Users can see the detection of past earthquakes. They can register their e-mails to receive notices of future
earthquake detection reports. A sample e-mail is presented in Fig. 12. It alerts users and urges them to prepare for the earthquake. It is hoped that the e-mail is received by a user shortly before the earthquake actually arrives. An earthquake is transmitted through the earth's crust at about
3–7 km/s. Therefore, a person has about 20 s before its arrival at a point that is 100 km distant.

6. RELATED WORK

Twitter is an interesting example of the most recent social media: numerous studies have investigated Twitter. from the studies introduced in Section 1, several others have been done. Grosseck et al. investigated indicators such as the influence and trust related to Twitter [8]. Krish- namurthy et al. crawled nearly Our system sent e-mails mostly within a minute, sometimes within 20 s. The delivery time is far faster than the rapid broadcast of announcement of JMA, which are widely broadcast on TV; on average, a JMA announcement is broadcast 6 min after an earthquake classification of tweets might be done similarly to our algorithm. Web2express Digest17 is a website that auto- discovers infor- mation from Twitter streaming data to find real-time inter- esting conversations. It also uses natural language process- ing and sentiment analysis to discover interesting topics, as we do in our study.
Various studies have been made of the analysis of web data (except for Twitter) particularly addressing the spatial aspect: The most relevant study to ours is one by Back- strom et al. [2]. They use queries with location (obtained by IP addresses).
Dear Alice,
We have just detected an earthquake around Bangalore. Please take care.

AJVV Alert System
manually classified messages into nine categories [19]. The numerous categories are ”Me now” and ”State- ments and Random Thoughts”; statements about current events corresponding to this category.
Some studies attempt to show applications of Twitter: Borau et al. tried to use Twitter to teach English to English- language learners [4]. Ebner et al. investigated the ap- plicability of Twitter for educational purposes, i.e. mobile learning [6]. The integration of the Semantic Web and mi- croblogging was described in a previous study [20] in which a distributed architecture is proposed and the contents are aggregated. Jensen et al. analyzed more than 150 thousand tweets, particularly those mentioning brands in corporate accounts [12].
In contrast to the small number of academic studies of Twitter, many Twitter applications exist. Some are used for analyses of Twitter data. For example, Tweettronics16 provides an analysis of tweets related to brands and prod- ucts for marketing purposes. It can classify positive and negative tweets, and can identify influential users. a random search from this cell will be equal to the query under consideration. The framework finds a query’s geographic center and spatial dispersion. Exam- ples include baseball teams, newspapers, universities, and typhoons. Although the motivation is very similar, events
to be detected differ. Some examples are that people might not make a search query earthquake when they experience. Some works have targeted collaborative bookmarking data, as Flickr does, from a spatiotemporal perspective: Serdyukov et al. investigated generic methods for placing photographs on Flickr on the world map [24]. They used a language model to place photos, and showed that they can effectively estimate the language model through analyses of annota- tions by users. Rattenbury et al. [22] specifically examined the problem of extracting place and event semantics for tags that are assigned to photographs on Flickr. They proposed
scale-structure identification, which is a burst-detection method based on scaled spatial and temporal segments.
Location estimation studies are often done in the field of ubiquitous computing. Estimating an object’s location is arguably the most fundamental sensing task in many ubiq- uitous and pervasive computing scenarios. Representing lo- cations statistically enables a unified interface for location information, which enables us to make applications indepen- dent of the sensors used — even when using very different sensor types, such as GPS and infrared badges [7], or even Twitter. Well known algorithms for location estimation are Kalman filters, multihypothesis tracking, grid-based, and topological approaches, and particle filters. Hightower and Borriello made a study of applying particle filters to location sensors deployed throughout a lab building [10].

7. DISCUSSION

We plan to expand our system to detect events of various kinds using Twitter. We developed another prototype that detects rainbow information. A rainbow might be visible somewhere in the world; someone might be twittering about a rainbow. Our system can identify rainbow tweets using a similar approach to that used for detecting earthquakes. The differences are that in the rainbow case, the information is not so time-sensitive as that in the earthquake case.
Our model includes the assumption that a single instance of the target event exists. For example, we assume that we do not have two or more earthquakes or typhoons simulta- neously. Although the assumption is reasonable for these cases, it might not hold for other events such as traffic jams, accidents, and rainbows. To realize multiple event detec- tion, we must produce advanced probabilistic models that allow hypotheses of multiple event occurrences.
A search query is important to search possibly-relevant tweets. For example, we set a query term as earthquake and shaking because most tweets mentioning an earthquake occurrence use either word. However, to improve the recall, it is necessary to obtain a good set of queries. We can use advanced algorithms for query expansion, which is a subject of our future work.

8. CONCLUSION

As described in this paper, we investigated the real-time nature of Twitter, in particular for event detection. Seman- tic analyses were applied to tweets to classify

them into a positive and a negative class. We consider each Twitter user as a sensor, and set a problem to detect an event based on sensory observations. Location estimation methods such as Kalman filtering and particle filtering are used to estimate the locations of events. As an application, we developed an earthquake reporting system, which is a novel approach to notify people promptly of an earthquake event.
Microblogging has real-time characteristics that distin-guish it from other social media such as blogs and collabo- rative bookmarks. In this paper, we presented an example using the real-time nature of Twitter. It is hoped that this paper provides some insight into the future integration of semantic analysis with microblogging data.

9. REFERENCES

[1] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking. IEEE Transactions on Signal Processing, 2001.

[2] L. Backstrom, J. Kleinberg, R. Kumar, and J. Novak. Spatial variation in search engine queries. In Proc. WWW2008, 2008.

[3] T. Bleier and F. Freund. Earthquake warning system. Spectrum, IEEE, 2005.

[4] K. Borau, C. Ullrich, J. Feng, and R. Shen. Microblogging for language learning: Using twitter to train communicative and cultural competence. In Proc. ICWL 2009, pages 78–87, 2009.

[5] d. boyd, S. Golder, and G. Lotan. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In Proc. HICSS-43, 2010.

[6] M. Ebner and M. Schiefner. In microblogging.more than fun? In Proc. IADIS Mobile Learning Conference, 2008.

[7] D. Fox, J. Hightower, L. Liao, D. Schulz, and G. Borriello. Bayesian filters for location estimation. IEEE Pervasive Computing, 2003.

[8] G. Grosseck and C. Holotescu. Analysis indicators for communities on microblogging platforms. In Proc. eLSE Conference, 2009.

[9] J. Hightower and G. Borriello. Location systems for ubiquitous computing. IEEE Computer, 34(8):57–66, August 2001.

[10] J. Hightower and G. Borriello. Particle filters for location estimation in ubiquitous computing: A case study. In Proc. UbiComp04, 2004.

[11] B. Huberman and D. R. F. Wu. Social networks that matter: Twitter under the microscope. First Monday, 14, 2009.

[12] B. Jansen, M. Zhang, K. Sobel, and A. Chowdury. Twitter power:tweets as electronic word of mouth. Journal of the American Society for Information Science and Technology, 2009.

[13] A. Java, X. Song, T. Finin, and B. Tseng. Why we twitter: Understanding microblogging usage and communities. In Proc. Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007, 2007.

[14] T. Joachims. Text categorization with support vector machines. In Proc. ECML'98, pages 137–142, 1998.

[15] J. Leskovec, L. Adamic, and B. Huberman. The dynamics of viral marketing. In Proc. ACM Conference on Electronic Commerce, 2006.

[16] Y. Matsuo and H. Yamamoto. Community gravity: measuring bidirectional effects by trust and rating on online social networks. In Proc. WWW2009, 2009.

[17] Q. Mei, C. Liu, H. Su, and C. Zhai. A probabilistic approach to spatiotemporal theme pattern mining on weblogs. In Proc. WWW'06, 2006.

[18] S. Milstein, A. Chowdhury, G. Hochmuth, B. Lorica, and R. Magoulas. Twitter and the micro-messaging revolution: Communication,connections, and immediacy.140 characters at a time. O'Reilly Media,

2008.

[19] M. Naaman, J. Boase, and C. Lai. Is it really about me? Message content in social awareness streams. In Proc. CSCW'09, 2009.

[20] A. Passant, T. Hastrup, U. Bojars, and J. Breslin. Microblogging: A semantic and distributed approach. In Proc. SFSW2008, 2008.

[21] Y. Raimond and S. Abdallah. The event ontology, 2007. http://motools.sf.net/event/event.html.

[22] T. Rattenbury, N. Good, and M. Naaman. Towards automatic extraction of event and place semantics from flickr tags. In Proc. SIGIR 2007, 2007.

[23] E. Scordilis, C. Papazachos, G. Karakaisis, and V. Karakostas. Accelerating sesmic crustal deformation before strong mainshocks in adriatic and its importance for earthquake prediction. Journal of Seismology, 8, 2004.

[24] P. Serdyukov, V. Murdock, and R. van Zwol. Placing flickr photos on a map. In Proc. SIGIR 2009, 2009.

[25] M. Weiser. The computer for the twenty-first century. Scientific American, 268(3):94–104, 1991.