Twitter Bursts: Analysis of their Occurrences and Classifications

Yuhiro Mizunuma, Shuhei Yamamoto, Yutaro Yamaguchi
Atsushi Ikeuchi, Tetsuji Satoh
University of Tsukuba
Tsukuba, Japan
Email:{yuhiro, atsushi}@slis.tsukuba.ac.jp
{yamahei, yamaguchi, satoh}@ce.slis.tsukuba.ac.jp

Satoshi Shimada
Housei University
Tokyo, Japan
Email: satoshi.shimada.57@hosei.ac.jp

***Abstract*—Twitter, a microblogging service launched in October 2006, has become one of the most popular social communication media. Because Twitter's characteristics are immediacy, ease of use, and bi-directionality, its timeline reflects the real world almost instantly. Once a major event happens, the number of tweets increases rapidly. In this article, this phenomenon is defined as a burst. The authors gathered Japanese tweets on a public timeline from Twitter API over a period of fifteen months starting from November, 2011, to February, 2013. We collected over 5 billion posts created by about 11 million users. Results of our analysises show that during the bursts, the total number of tweets showed a higher percentage of retweets, fewer replies, and fewer characters used per post than those during normal status. Cluster analysis revealed five types of bursts. Furthermore, we clarified that the scale of earthquakes in Japan and the distance from the quakes' epicenters to Tokyo significantly affected the occurrence of bursts on Twitter.**

Keywords-Twitter; Burst; Clustering

I. INTRODUCTION

In our increasingly technological world, people commonly and regularly use online social networking service to connect, communicate, and obtain information. As one of the most popular social networking and microblogging tools, Twitter enables users to send and read text messages, called tweets, of up to 140 characters via computers or mobile phones. Since its launch in 2006, Twitter has rapidly increased in terms of the number of users. In December 2012, Twitter Inc. reported having more than 200 million active users creating more than 400 million tweets daily [1]. In Japan, Twitter is more popular than any other social networking tool, for instance, mixi, Facebook, Linkedln, and Google+ [2]. The number of Twitter users in Japan is the third largest in the world, after the United States and Brazil. To put this status into perspective, Japan's population in 2011 was approximately 127.8 million; Brazil 196.7 million; United States 311.6 million [3]. Twitter has been widely regarded as an effective emergency communication tool, and after the Great East Japan Earthquake in March 2011, its users increased exponentially.

Twitter's technological characteristics—bi-directionality, immediacy, and ease of use—practically ensure increase in the number of tweets during or just after an event occurs. For example, in the popular animated television film *Castle in the Sky Laputa*, the word *balse* is spoken to cast a magic spell that devastates Laputa, the film's eponymous flying castle. On August 2, 2013, as "Balse!" was spoken on the film, a high number of Japanese users simultaneously tweeted "Balse!". In fact, Twitter Inc. reported that the world record of tweets per second (TPS) was broken with 143,199 TPS. We define this phenomenon as a "burst," and this study aims to examine and classify such bursts.

We examined the burst phenomenon through quantitative analysis, with the goal of answering the following three questions: (1) Why do bursts occur? (2) What are the characteristic features of tweets in burst status? (3) How is each burst classified? To accomplish our goal, we crawled 5,285,607,227 tweets for a span of 15 months, from November 16, 2011, to February 15, 2013.

Our report of the results is organized as follows: Next section relates our research to the perspectives of similar research. Section Ⅲ describes our data crawling methodology and our burst detection method. In Section Ⅳ, we apply the results of our analysis by examining tweets' characteristic features during burst status, comparing them with tweets during normal status (Ⅳ. A). After clarifying the characteristic features, we classify each burst (Ⅳ. B). Then, we verify factors affecting the earthquake burst (Ⅳ. C). Finally, in Section V, we summarize our findings and indicate future challenges.

II. RELATED WORK

With such widespread use of tweeting, studies have already been conducted to clarify exactly what Twitter is and how people use it. As pioneering researchers, Java et al. [4] collected tweets on a public timeline for 2 months, from April 1 to May 30, 2007, gathering 1,348,543 tweets posted by 76,177 unique users. Through this data, they examined the users' motivations for posting and the structures of Twitter. Java et al. clarified that the diameter of the network graph based on the follow relationship was 6. They also reported that 20% of all tweets were conversational with @, and 13% contained a URL sent to share information. Krishnamurthy et al. [5] crawled not only a public timeline but also user profiles and their tweets. They collected these data using two algorithms and performed an analysis similar to that of Java et al., describing the differences between data sets. In another

 ISBN: 978-1-61208-324-7

study, Poblete [6] collected 5,270,609,213 tweets by 4,736,629 users from 246 countries to reveal national differences in the Twitter network.

Numerous studies have examined Twitter from the viewpoint of information propagation and relationships between users. From June 6 to June 31, 2009, Kwak et al. [7] extracted 1.47 billion follow-follower relationships and 41.7 million user profiles. These researchers' results showed that 77.9% of follow-follower networks were one-way but that mutual follows accounted for only 22.1%. These features are unique to Twitter among social networking services. They suggest that Twitter's technological characteristics make it a stronger source for communicating and disseminating or obtaining information.

Many further studies of Twitter have related it to the real world. On one hand, some studies have attempted to relate tweets to later events, in other words, to predict the future through Twitter. Bollen [8] tried to predict stock prices; Asur [9] tried to predict movies' box office sales; and Tumasjan [10] tried to predict election results. On the other hand, some studies attempted to detect the actual condition of the world. From August to October 2009, Sakaki et al. [11] gathered tweet data that was used to detect earthquakes with a high probability: 96% of seismic intensity 3 earthquakes and 100% of more intense earthquakes were detected. Diao et al. [12] detected trends in event according to burst words in tweets. From September 1 to November 30, 2011, these researchers collected 3,967,927 tweets from users in Singapore. Using latent Dirichlet allocation (LDA) and two LDA improved algorithms (UserLDA, TimeLDA), they conducted automatic detection of topics from extracted words. Results showed that, using improved algorithms, their method can detect unique topics more precisely than conventional methods. Shirakihara et al. [13] obtained buzzwords from buzztter.com/. Then, using the algorithm proposed by Kleinberg, these researchers detected the time zone in which tweets including certain buzzwords increased rapidly.

In brief, most Twitter studies have focused on event detection in the real world rather than on users' information-gathering behavior and features of tweets. By focusing on the number of tweets as they increase through a certain time span, our study proposes to clarify why and how people tweet. Thus, we discuss the relationship between the real world and Twitter. In a research focused on the number of tweets, Inui [14] analyzed 179,286,297 tweets posted around the Great East Japan Earthquake that occurred on March 11, 2011, revealing that the tweets per minute (TPM) peaked in the week after the earthquake. The highest number of tweets was recorded on March 15, 2011, when the seismic intensity 6 earthquake occurred in Shizuoka Prefecture. The second highest number was recorded just after another earthquake that occurred on Sanriku coast.

III. Method

To analyze burst status, we must crawl tweet data and then set a threshold value for a burst. After detecting bursts, we analyze them. In Section III-A, we explain how we crawled tweet data and how we set the threshold value.

However, we first provide some explanation of how tweets work. To post a tweet to a particular user, one begins with "@username," and the tweet appears in the timeline of a recipient user or a user who follows both sender and recipient. This type constitutes about 40% of all tweets. The reply function, of course, makes a tweet go to a particular user-in this case, the sender of the tweet replied to. A "retweet" or a re-posting of someone else's tweet empowers a tweet receiver to spread information beyond the original tweet's followers. The retweet function is symbolized in a re-sent message by "RT@username." The rate of a retweet is a percentage of the text beginning with RT over the total number of texts; this type of retweet does not include a retweet with another user's comment, which is called the "classic retweet." Similarly, this type does not include tweets beginning with QT.

TABLE I. Data Collected

	All Data	Weekday	Weekend
The number of tweets	5,285,607,227	3,740,106,962	1,545,500,265
Average number of characters	45.76	46.15	44.81
Rate of Retweets（%）	8.82	8.94	8.60
Rate of Reply（%）	39.02	39.34	38.25

A. Data Collection

From November 16, 2011, to February 15, 2013, we collected public timeline tweets from Japan, written in Japanese, using Twitter Search API. We set the parameter language for '*ja*' (Japanese) and the geocode for a 2,000-km radius from Akashi-city, Hyogo, in order to cover only Japan. We collected 5,285,607,227 tweets posted by 10,918,410 unique users. Each tweet has its own identity (ID), the user's ID, the exact time of posting, the tweet's actual text, and so on. Table I displays fundamental statistics on the collected data. Of the 5,285,607,227 tweets harvested, 8.82% were retweets. The rate of reply was 39.02%. The mean for characters was 45.75, and the mode, the value that appears most often, for characters was 21.

B. Setting Threshold Value

For analysis of the phenomenon under consideration here, a sudden, large increase in tweets beyond the normal traffic is considered a "burst." For macroscopic analysis, of course, we must establish a quantitative burst threshold, a set value. We check the threshold every minute, and when the number of tweets rises above the threshold value, we judge that time to signal a burst. Figure 1 indicates average number of tweets according to day and time. In Figure 1, the average number of tweets at 4:00 is below 2,000 tweets par minute, and at 23:00, the average is more than 140,000 tweets par minute. This information suggested that we should not set the same threshold throughout the day, and thus, we decided to set the threshold by the minute-as Figure 1 illustrates. As indicated in Figure 1, the number of tweets also differs on weekdays and holidays. Holidays are weekends, and national holidays. On a usual weekday, Twitter traffic increases around 8 a.m. and around noon, indicating use after awakening, during the morning commute, and during lunch breaks. On holidays, however, tweeting steadily increases into the night hours. For these reasons, we set different thresholds for both the day and the time.

As Table II shows, the average number of tweets and unique users is increasing. If we applied the same threshold to all the

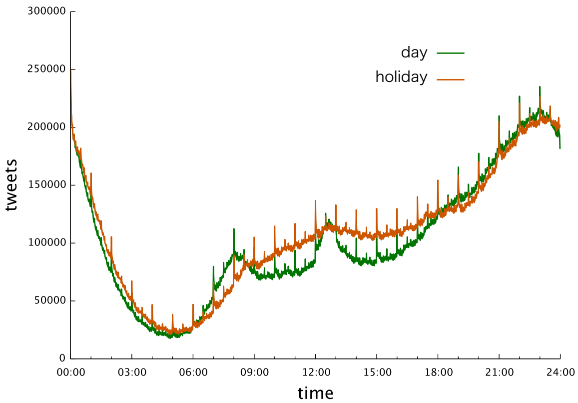

FIGURE 1. AVERAGE NUMBER OF TWEETS ACCORDING TO DAY AND TIME

TABLE II. CHANGING NUMBER OF TWEETS, USERS, AND TWEETS PER DAY

Span	Tweets	Unique Users	Tweets per User
Nov 16-Dec 15, 2011	10175500.7	113153.7	89.9
Dec 16-Jan 15, 2011	9959195.5	111298.1	89.4
Jan 16-Feb 15, 2012	10498451.5	113941.8	92.1
...	...	...	...
Nov 17-Dec 16, 2012	12302316.8	144700.2	85.0
Dec 17-Jan 16, 2013	12921195.3	145271.6	88.9
Jan 17-Feb 15, 2013	13555442.6	153905.0	88.0

data, it would be difficult to detect earlier bursts and easy to detect recent ones. Thus, we also set different thresholds for each month. To calculate the threshold value for a certain month, we used a dataset that included the month previous and the month after that under consideration. For instance, to calculate the threshold from March l6 to April 14, 2012, we used the dataset from February 14 to May 15, 2012. We detected bursts from December 16, 2011, to January 15, 2013, using data from November 16, 2011, to February 15, 2013. The threshold values for bursts were calculated using the following formula:

$$N_{nt}(t) = \overline{N}(t) + 3\sigma(t) \quad (1)$$

where $N_{nt}(t)$ represents the threshold value of a burst at a certain time(t), $\overline{N}(t)$ is average of the number of tweets per day at a certain time(t), and $\sigma(t)$ is the standard deviation at a certain time(t). For calculating the threshold value, two extreme values of tweet numbers for each time were removed from the dataset.

Based on this method, we detected 5,326 bursts from holiday dataset and 5,650 burst from weekday dataset. We detected burst events by checking tweet texts and times. Some bursts have relevance to television programs, for example, *Lupin III: The Castle of Cagliostro, Smile PreCure!,* and *Tetsuko's Room*, and bursts were caused by televised sports events as well. Justin Bieber's appearance on a Japanese television program caused a burst. In addition, a burst occurred 3 minutes after television news announced the arrest of Takahashi, the last Aum fugitive from the sarin gas attack on the Tokyo subway in 1995. All these examples suggest a strong association between bursts and television broadcasting. Moreover, bursts are relevant to other media; for example, Animation Song-Zanmai Z is a radio program that has caused bursts. In addition, Twitter has bursts unique to itself, such as "Twitter's server down!"

Furthermore, bursts have relevance to natural disasters, e.g., earthquakes, "bomb cyclones," tornadoes, heavy snow, and heavy rain. People experiencing a disaster post their situations on Twitter, and others use Twitter to disseminate information about the disaster.

IV. RESULTS AND DISCUSSION

A. *Features of Posting during Burst Status*

To clarify features of posting during bursts, we compared features of text during normal status and burst status. Table III represents the average number of characters, the rate of retweets compared to all tweets, and the rate of reply to all tweets during burst status, nonburst status, and all statuses, respectively.

TABLE III. COMPARISON OF TEXT FEATURES

	All Statuses	Burst Status	Nonburst Status
Average number of characters	45.8	42.2	45.8
Rate of retweets（%）	8.84	9.29	8.83
Rate of Reply（%）	39.02	33.83	39.15

In burst status overall, the average number of characters is fewer than that in normal status because users attempt to tweet as quickly as possible. Because time is of the essence, users make their posts short. In fact, during bursts caused by earthquakes, users posted very short texts in two or three Japanese characters, such as "Oh no!", "Earthquake," or "Shaking!" In burst status also, the retweet rates are higher than those in nonburst status, and the rate of reply in burst status is lower than that in nonburst status. Users try to spread information about burst events to many people, and thus, the retweet rates become higher. In burst status, people like to use Twitter's functions to diffuse rather than limit information. For example, during a burst on March 14, 2012, 20.9% of all tweets were retweets. On that date, an earthquake occurred, and users posted retweets of information about the disaster tweeted from a public office account: RT@zishin3255_2 Earthquake Early Warning (no.12) There was an earthquake in Sanriku offshore, 3 on the Japanese scale. [Detail] The 9.0 magnitude earthquake occurred at 18:08:29 on 14th March 2012, depth of 10 km. It will reach Tokyo at 18:11:26 [about 177 seconds later]. #EarthquakeEarlyWarning and RT@NHK_PR: There is a tsunami advisory for Iwate Prefecture and the Pacific Ocean coast in Aomori Prefecture. The Earthquake Early Warnings are issued mainly by the Japan Meteorological Agency (JMA), and NHK is Japan's national public broadcasting organization. Kwak [7] observed that Twitter is more a source of information than a social networking site, and our results confirm that, particularly during bursts, people tend to use Twitter as a source of information.

B. *Classifications of Bursts*

In the previous section, we explained that during bursts, tweets tended to be retweeted, less replies were received, and less characters than usual were contained. However, we

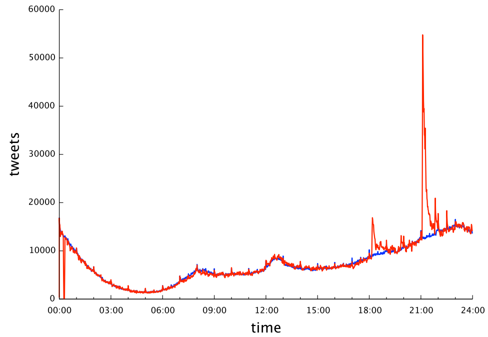

FIGURE 2. EARTHQUAKE BURST

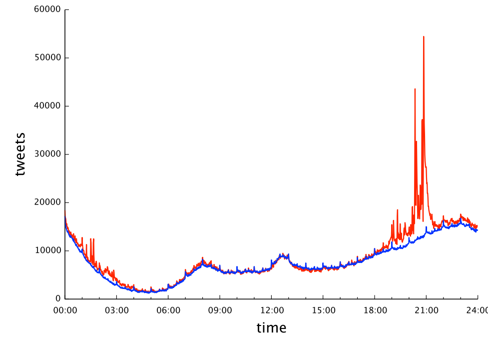

FIGURE 4. FOOTBALL (AUSTRALIA VS. JAPAN) BURST

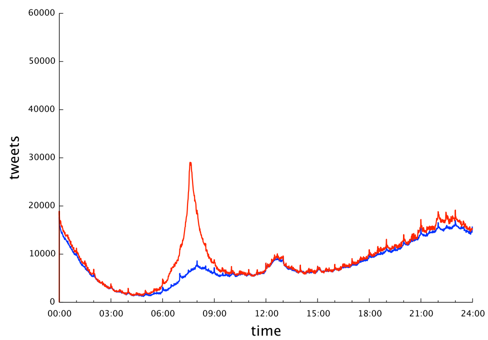

FIGURE 3. ANNULAR ECLIPSE BURST

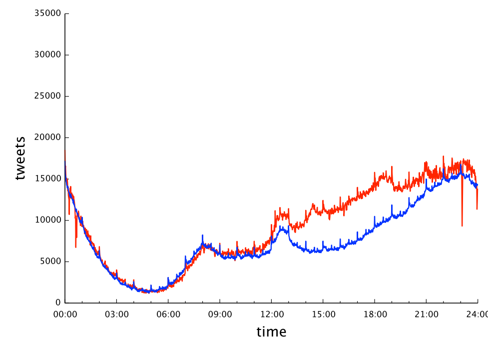

FIGURE 5. BOMB CYCLONE BURST

also framed the hypothesis that different events may cause different posting features. Hence, we tried to classify each burst according to its features, that is, the average number of characters, rates of retweets, and rates of replies. In addition, we found that the shape of a burst can indicate the type of event.

Figures 2-5 indicate the changing number of tweets according to event. For instance, Figure 2 shows the number of tweets increasing rapidly after an earthquake and then decreasing rapidly. In other words, an unpredictable event, such as an earthquake, causes an increase and then a decrease in the number of tweets within a short time span. Figure 3 illustrates the process of tweet numbers on May 21, 2012, the day an annular eclipse occurred. In this case, users knew when the eclipse would occur, and thus, the number of tweets increased and decreased moderately before and after the event. On June 12, 2012, the Fédération Internationale de Football Association (FIFA) World Cup qualifier with Japan versus Australia was played. As shown in Figure 4, a little before the game began at 19:00 and a little after the game ended at 20:50, the number of tweets was higher than usual, increasing particularly when goals were scored and when the game ended. Figure 5 illustrates tweet numbers on the day a bomb cyclone hit Japan. Different from other figures, span of increasing tweets was very long, although the distance to average was not so long. These examples, illustrated in the figures, show that Twitter users' reactions to various events can change the bursts' shapes.

We further framed the hypothesis that a burst's features reflect the event's nature. We classified the bursts and then detected the nature of events in each cluster. Each burst has data about shape, that is, (1) length of burst status and (2) distance to threshold. The *length* of burst status is the *total time between the number of tweets above and below the threshold value*. If the number of tweets goes above the threshold value at one measurement and then falls below the threshold value at the next measurement, the length of the burst is 1 min. In addition, each burst contains data regarding text features, that is, (3) average number of characters, (4) rates of retweets, and (5) rates of replies. Thus, using these five factors, we classified each burst through cluster analysis using Ward's method and the Euclidean distance of R2.15.1. Before clustering, we normalized all the data. Table IV shows the average of each feature in each cluster.

TABLE IV. RESULTS OF CLUSTER ANALYSIS

	Duration (minute)	Distance	Num Chars	Rate of RT (%)	Rate of @ (%)
1st cluster	21.32	567.44	44.87	9.42	38.28
2nd cluster	549.91	4110.08	43.41	10.95	36.10
3rd cluster	54.76	1245.37	40.35	6.99	30.86
4th cluster	51.57	5366.71	31.36	5.15	20.07
5th cluster	62.40	2333.45	48.86	22.21	27.94

Of the five clusters, the first has the shortest burst status and the shortest distance to threshold. Thus, the third cluster contains small bursts caused, in this case, by a seismic intensity 1 earthquake and unexpected strong rain. Both these events affected relatively few people in a small area. The third cluster was composed of such small bursts.

 ISBN: 978-1-61208-324-7

The second cluster reveals the peak of a big event, with both the longest distance to threshold and greatest length of burst status. Cluster five bursts are typified by participation of many people, such as celebrating the New Year or observing the annular eclipse.

The third cluster contains bursts previous to peaking. In this cluster, the distance to threshold is longer than that in the first cluster and shorter than that in the second cluster. Similarly, the length of burst status is longer than that in the first cluster and shorter than that in the second cluster. The beginning of the annular eclipse burst, the death of Kim Jong-il, and the televised Japanese animation My Neighbor Totoro were all in the first cluster. Thus, we concluded that this cluster is a type of burst in process.

The fourth cluster has a long threshold distance, although the burst length is comparatively short. Thus, this cluster contains a type of sudden, unpredictable event, for example, Olympic game victories, a goal at the FIFA World Cup, and earthquakes.

The fifth cluster is characterized by high rates of retweeting, and thus, we define this cluster as a type of information diffusion. Along with the many retweets, the number of characters is also the greatest, presumably to provide sufficient information. In this cluster, the bursts contained, for example, the news of a phantom killer in Shibuya and the arrest of the Aum suspect Takahashi.

To sum up, we classified bursts into five types: (1) small burst, (2) burst in process, (3) peak burst, (4) sudden burst, and (5) information diffusion burst.

C. Factors Affecting Earthquake Burst

Twitter's nature is one of immediacy and brevity. Users can post only 140 characters, and tweets appear on the timeline as soon as the user posts. When a disaster hits, then, Twitter can transfer information more expeditiously than other media. In this section, we discuss factors affecting earthquake bursts. Clarifying the relevance between disasters and Twitter can help provide the most rapid dissemination of information about the event. Some have studied using Twitter for the immediate spread of disaster information; for instance, Sakaki et al [11] detected disaster situations using locator information and tweet texts. However, no studies have clarified the relevance between disasters and bursts.

Throughout this investigation, bursts occurred many times in disaster situations, such as typhoons, heavy rains, earthquakes, and so on. In particular, earthquakes caused burst status 106 times. Therefore, we examined factors affecting earthquake bursts, and one factor is the earthquake's scale. Most earthquake tweets are posted when the user feels the shock of the quake. The higher the quake is on the scale, the more people notice it. Besides scale, the distance between urban centers and the earthquake may be relevant to bursts since the number of tweets increases along with the population density and numbers of Twitter users in the urban centers.

In Japan, earthquakes are assigned levels on a scale from 0 to 7, with 7 being the strongest, and we used the same scale in this study. Figure 6 shows changes in the number of tweets when earthquakes are registered in the upper 5 levels on the intensity scale. The figure contains the date of each earthquake, the name of the prefecture that recorded the maximum seismic intensity, and the distance from Tokyo.

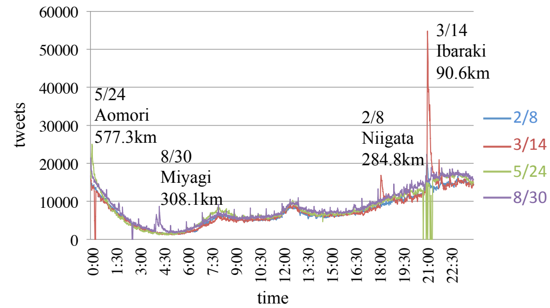

FIGURE 6. CHANGING TWEET NUMBERS IN ABOVE 5 SEISMIC INTENSITY EARTHQUAKES

This information reveals that earthquakes of the same seismic intensity do not cause the same number of tweets. The number of tweets on the March 14 is greater than that on other days. This is because of the quake's distance from the urban center. Ibaraki Prefecture is about 90 km from Tokyo, and the other prefectures are more than 100 km away. Similarly, the number of tweets after an intensity 4 quake in Aomori (May 24, 2012) was fewer than those after a less intense quake in Chiba (May 29, 2012). Now, Aomori Prefecture is about 577 km from Tokyo, but Chiba Prefecture is only about 40 km from Tokyo.

Therefore, we decided that the representative location of an urban center would be the Tokyo Metropolitan Government, with the closest seismograph station located at Kabukicho Shinjuku-ku, Tokyo. During the investigation, this seismograph station registered 50 earthquakes: 36 of intensity 1; 11 of intensity 2; and 3 of intensity 3. We detected bursts 46 times —a 92% rate of detection. These results indicate that bursts occur with high probability if the urban center experiences an earthquake.

On this basis, we adopted the hypothesis that an earthquake burst has relevance both to the scale of an earthquake and the distance from the urban center. Throughout this investigation, earthquakes of seismic intensity 3 or more occurred 341 times. We collected data for each earthquake: (1) time of occurrence, (2) epicenter, (3) maximum seismic intensity, (4) municipality recording maximum seismic intensity, and (5) distance between the urban center and the municipality. We used the earthquake database provided by the Japan Weather Association in order to collect time of occurrence, epicenter, maximum seismic intensity, and municipality that recorded maximum seismic intensity. For this study, we decided that the representative location of the urban center would be the Tokyo Metropolitan Government. The distance between the municipality and urban center was measured as the distance between the municipality's town hall and the Tokyo Metropolitan Government. We calculated the distance using Google Maps API. When more than one municipality recorded the same maximum seismic intensity, we chose the municipality closest to the Tokyo Metropolitan Government and calculated the distance. If the Twitter burst occurred within 3 min after the earthquake, we decided the earthquake caused the burst. However, if the burst occurred before the earthquake, we removed the earthquake data from our dataset. Earthquakes over seismic intensity 3 occurred 341 times. Within 3 min after earthquakes, 127 bursts occurred, but 11 of them had attained burst status before the earthquake occurred. Excluding those

11 earthquakes, we then calculated a rate of burst detection using 328 earthquakes and 106 bursts (Table V).

TABLE V. RATE OF BURST DETECTION

Distance from Urban Center	Intensity 3	Intensity 4	Above Intensity 5
Up to 100km	63.2%(24/38)	100.0%(16/16)	100.0%(5/5)
100-200km	14.0%(12/86)	57.1%(12/21)	60.0%(3/5)
200-300km	14.8%(4/27)	100.0%(3/3)	100.0%(3/3)
Over 300km	8.2%(8/98)	30.8%(8/26)	80.0%(4/5)

When the earthquakes registered seismic intensity 3, the more the proximity to the urban center, the higher was the rate of burst detection. This suggests relevance between the distance from the urban center and the burst. When earthquakes registered a seismic intensity of 5 or more, the rate of burst detection is very high, regardless of the distance from the urban center. This suggests relevance between an earthquake's scale and its resultant burst.

We performed logistic regression analysis to confirm these results. We used "burst or no burst" as the dependent variable, and "scale of the earthquake" and "inverse of distance from urban center" as independent variables. Tables VI reveal the results of logistic regression analysis using R2.15.1. McFadden's ρ is 0.26, Cox-Snell's R^2 is 0.366, and Negelkerle's R^2 is 0.488. Identification rate based on the regression equation is 80.5%.

TABLE VI. RESULTS OF LOGISTIC REGRESSION

	B	SE	Wald	p Value	Odds Ratio
Intensity	5.288161	0.703	56.657	5.2e-14**	197.98
Distance	1.354106	0.163	68.872	2e-16**	3.87

Intensity: maximum seismic intensity; Distance: Distance from urban center; B: partial Iregression coefficient; SE: Standard error

For each independent variable, a p value less than 0.05 was considered statistically significant. The correlation coefficient between each independent variable was below 0.1. Evidence for multicollinearity was absent because the variance inflation factor for independent variables in models was less than 2.0. The results suggest that the scale of the earthquake and the distance from the urban center are affecting earthquake bursts. In particular, the value of Wald suggests that the distance from the urban center more strongly influences a burst than the scale of the earthquake.

V. CONCLUSION

This study aimed to explore the media character of Twitter by focusing on the burst phenomenon. We clarified that burst tweets are more likely to be retweets, receive less replies, and contain fewer characters than usual. In burst status, in fact, Twitter becomes more a source of information than a social site. In addition, we classified each burst and clustered burst events into groups. According to certain features, we were able to classify five types of bursts (1) small burst, (2) burst in process, (3) peak burst, (4) sudden burst, and (5) information diffusion burst. Finally, we verified factors affecting earthquake bursts, namely, that the scale of earthquakes and the distance from an urban center affect earthquake bursts, with the latter having a stronger influence than the former.

In future research, we plan to focus more on individual users. To further clarify factors affecting earthquake bursts, we should separately consider two groups of users, those who tweet after perceiving the quake themselves and those who tweet after receiving news of an earthquake. To do so, we must more finely gather geocode and time-of-posting data. We will classify users according to network and profile data and then compare burst status between the two groups. For even finer research on this data, we should detect burst events through natural language analysis.

VI. ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Number 25280110.

REFERENCES

[1] k. Wickre, "Celebrating #Twitter7". The Official Twitter Blog. 2013-03-21. https://blog.twitter.com/2013/celebrating-twitter7. (accessed 2013-08-01).

[2] Internet Media Research Institute eds. Report of social media research 2011. impressR&D, 2011, 156p. [in Japanese].

[3] Semiocast. "Twitter reaches half a billion accounts More than 140 millions in the U.S.". Semiocast. http://semiocast.com/publications/2012_07_30_Twitter_reaches_half_a_billion_accounts_140m_in_the_US. (accessed 2013-08-01).

[4] A. Java, X. Song, T. Finin, B. Tseng, "Why We Twitter: Understanding Microblogging Usage and Communities" In Procedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007, 2007, pp. 56-65.

[5] B. Krishnamurthy, P. Gill, M. Arlitt, "A Few Chirps About Twitter," In Proceedings of the First Workshop on Online Social Networks, 2008, pp. 19-24.

[6] B. Poblete, R. Garcia, M. Mendoza, A. Jaimes, "Do All Birds Tweet the Same? Characterizing Twitter Around the World," In Proceedings of the 20th ACM International Conference on Information and Knowledge Management, 2011, pp. 1025-1030.

[7] H. Kwak, C, Lee, H. Park, S, Moon, "What is Twitter, A Social Network or a News Media?" In Proceedings of the 19th International Conference on World Wide Web, 2010, pp. 591-600.

[8] J. Bollen, A. Pepe, H. Mao, "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena," In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 2011, pp.450-453.

[9] A. Asur, B. A. Huberman, "Predicting the Future With Social Media," In Proceeding WI-IAT'10 Proceedings of the 2010 IEEEWICACM International Conference on Web Intelligence and Intelligent Agent Technology, 2010, pp. 492-499.

[10] A. Tumasjan, T. O. Sprenger, P. G. Sandner, I. M. Welpe, "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment," Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, 2010, pp. 178-185.

[11] S. Takeshi, O. Makoto, M. Yutaka. "Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors," In Proceedings of International Conference on World Wide Web, 2010, pp. 851-860.

[12] Q, Diao, J. Jiang, F. Zhu, E. P. Lim, "Finding Bursty Topics from Microblogs," In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, 2012, pp.536-544.

[13] S. Wataru, O. Tetsuya, H. Ryuzo, F. Hiroshi, K. Miyuki, "Report of Information Fundamentals and Access Technologies," 2010-IFAT-99(2), 2010, p. 1-8. [in Japanese].

[14] "FrontPage/Project311/trend analysis". Laboratory of Inui and Okazaki. 2012-10-13. http://www.cl.ecei.tohoku.ac.jp/index.php?Project%%E3%83%88%E3%83%AC%E3%83%B3%E3%83%89%E5%88%86%E6%9E%9. (accessed 2012-09-30).

 ISBN: 978-1-61208-324-7