Evaluation of the Diffusion Phenomenon using Information from Twitter

Kohei Otake[1]
School of Information and Telecommunication Engineering
Tokai University, Tokyo, Japan

Takashi Namatame[2]
Faculty of Science and Engineering
Chuo University, Tokyo, Japan

***Abstract*—Social media services, including social networking services (SNSs) and microblogging services, are gaining prominence. SNSs have a variety of information on products and services, such as product introductions, utilization methods, and reviews. It is important for companies to utilize SNSs to understand the various ways of engaging with them. Against this backdrop, numerous studies have focused on marketing activities (e.g., consumer behavior and sales promotion) using information on the internet from sources such as SNSs, blogs, and news sites. In particular, to understand the dissemination of information on the Internet, various researchers have undertaken studies pertaining to the diffusion phenomenon occurring in the real world. Here, topic diffusion is a phenomenon whereby a certain topic is shared with several other users. In this study, we aimed to evaluate the diffusion phenomenon on Twitter. In particular, we focused on the state of a targeted topic and analyzed the estimation of the topic using natural language processing (NLP) and time series analysis. First, we collected tweets containing four titles of animation broadcasts using hashtags. Approximately 250,000 tweets were posted on Twitter in a month. Second, we used NLP methods such as morphological analysis and N-gram analysis to characterize the contents of each title. Third, using the time series data for the tweets, we created a mixture model that replicated the diffusion phenomenon. We clustered the diffusion phenomenon using this model. Finally, we combined the features related to the content of the tweets and the results of the clustering of the diffusion phenomenon and evaluated them.**

Keywords*—Twitter; diffusion phenomenon; natural language processing; mixture model

I. Introduction

In recent years, social media has become highly prominent. Social media is the collection of online communication channels with community-based input, content sharing, interaction, and collaboration [1]. Websites and smartphone applications dedicated to forums, social networking, social bookmarking, and wikis are some of the different types of social media.

In particular, the number of users of social networking services (SNSs), such as Twitter and Facebook, is on the rise. These users can share a variety of information, such as their preferences and favorites related to services, products and so on, with their friends through SNSs. In such a scenario, marketing campaigns using SNSs have received increasing attention from electronic commerce (EC) suppliers. SNSs have a wide range of information about products, such as product introductions, utilization methods, and reviews. In addition, these pieces of information are posted by the consumers themselves, a phenomenon that did not happen in the past. In general, the information available on SNSs has several features. The representative features are "wide reach" and "high-speed communication". Using information on SNSs effectively is important for EC suppliers.

Consequently, there have been numerous studies that focused on marketing activities (e.g., consumer behavior and sales promotion) using information from sources such as SNSs, blogs, and news sites [2][3][4]. Several studies have focused on how tweets are shared with users.

To understand the dissemination of information on the Internet, various researchers have conducted studies pertaining to the diffusion phenomenon occurring in the real world. Here, topic diffusion is a phenomenon whereby a certain topic is shared with several other users.

Mane and Borner [5] proposed maps that supported the identification of major research topics and trends through the analysis of a complete set of papers published in Proceedings of the National Academy of Sciences of the United States of America between 1982 and 2001. The authors demonstrated the utilization of Kleinberg's burst detection algorithm [6], word co-occurrence analysis, and graph layout techniques. In addition to this study, Takahashi et al. [7] proposed a method for measuring bursts of topics estimated by a topic model. The authors analyzed two ways to model information flow in news streams, namely, Kleinberg's burst modeling and topic modeling such as the dynamic topic model.

Dipak and Bijith [8] proposed a model that predicted movie sales using word-of-mouth publicity by Twitter users. Specifically, the authors classified movie reviews into four types (strongly negative, negative, positive, and strongly positive), and created a model that predicted the intention to watch a movie. As a result, the authors suggested that utilization of online reviews on Twitter will help the movie industry's marketing strategies.

Matsuzawa et al. [9] proposed a method for analyzing the statistical characteristics of a time series of retweets extracted from the actual logs of tweets. Specifically, the authors suggested a burst detection model, assuming a lognormal distribution based on the results of Sartwell [10] using statistical data about infectious diseases. The time series clustering classification caused most of the retweets to extend over a short time scale of approximately one day.

Ueda and Asahi [11] proposed a model that replicated a sudden increase (boom) in interest among Twitter users. Specifically, the authors created a model that had two features: reflection of the behavior of social media users (including Twitter users) and estimation of the potential interests of the users. Based on these features, the authors analyzed the factors of a boom in interest. From the results of the analysis, the authors evaluated the differences between transient and secondary booms.

From these studies, it was clear that understanding the diffusion phenomenon is very important for understanding social trends. To understand social trends, we believe that it is necessary to use the information from SNSs properly.

The remainder of this paper is organized as follows. Section 2 describes the purpose of this study. Section 3 presents the dataset used in this study. Section 4 discusses the characterization analysis used to target the content of tweets using natural language processing (NLP). Section 5 describes the modeling of the diffusion phenomenon using time series data and presents an evaluation of the diffusion phenomenon based on the results mentioned in Sections 3 and 4. Finally, Section 6 summarizes the paper and discusses future work.

II. Purpose of the Study

In this study, we attempted to evaluate the diffusion phenomenon on Twitter. We examined the content of tweets using NLP. In addition, based on the results of previous studies, we tried to analyze the tweets using time series data. In particular, we analyzed the diffusion phenomenon by building a mixed normal distribution model using retweeted data, and then evaluated the results. For the analysis, we targeted animations in Japan. Animation information is shared widely across SNSs. Therefore, it was considered appropriate for the analysis.

First, we used posted tweets as data and attempted to characterize the content of the tweeted data by using NLP. From this analysis, we extracted characteristic expressions (characteristic words) included in the tweets for each target animation title. Second, using the time series of the tweeted data, we built a model that reflected the diffusion phenomenon. We also performed clustering of the diffusion phenomenon using the model. Finally, we combined the features related to the tweeted content with the results of the diffusion phenomenon clustering, and then evaluated it.

III. Dataset

First, we collected tweets pertaining to Japanese animations. In this study, we focused on four animation titles based on broadcasting time and evaluation by ranking site. We collected tweets posted over the course of one month for these four titles. We used hashtags and keywords (animated titles) and collected data using the Twitter application programming interface. Consequently, we collected approximately 250,000 tweets. In addition, at the time the tweets were collected, we also acquired information about the users, such as their number of follows, followers, favorites, and retweets. Table I summarizes the acquired tweeted data, and Table II presents a summary of the contents of each animation title and the number of tweets posted in a month. Data were selected from these datasets and analyzed.

TABLE I. Summary of the Acquired Tweeted Data

Number of tweets and retweets	245,146
Number of unique IDs	224,967
Number of unique tweets (except retweets)	115,104

TABLE II. Content Summary of each Animation Title and Number of Tweets

Title	Content summary	No. of tweets and retweets
A	*A* is an animation created by mixed media of game maker and production. *A* performs various content development activities such as animation, game applications, and card games. Category: idol group (male); Main target: adult women	48,728
B	*B* is an animation based on the fifth work of a series of games. Category: serious fantasy; Main target: youth	43,690
C	*C* is an animation based on a toy project and is also serialized over the same period in magazines for elementary school students. Category: comedy; Main target: children	31,190
D	*D* is a cartoon from a popular magazine for youth, presented in original animation. Category: serious fantasy; Main target: youth	121,541

IV. Characterization Analysis Targeting Tweet Content using NLP

In this section, we discuss the NLP analysis performed to characterize the content of the tweets. Specifically, the analysis was conducted using the following procedure. From Table II, it can be seen that the number of tweets is highly skewed for each title. In this case, the number of tweets for a title will affect the characteristic extraction. We deemed it necessary to adjust the number of tweets to target to acquire the characteristic amount of the number of tweets more accurately by using NLP. Therefore, we extracted 27,000 tweets per title by random sampling, and we adjusted the number of tweets. Here, 27,000 is the maximum number of tweets that could be acquired for each animation title while excluding duplicate tweets.

1) Parts of speech decomposition based on morphological analysis.
2) Word weighting by $tfidf$ value.
3) Calculation of co-occurrence frequency by N-gram (we use the bigram model).

We first decomposed the tweets by morphological analysis into units according to parts of speech. Morphological analysis was used to divide natural language (text data) into columns of morphemes (minimum elements that constitute sentences) and to discriminate parts of speech and the like for each morpheme. Information such as parts of speech for words defined grammatically and in the dictionary, was used for morphological analysis. In this study, a morphological analysis was performed using the R language. In addition, Mecab [12] was used as a Japanese morpheme dictionary. Table III summarizes the frequency of appearance for the parts of speech obtained by the morphological analysis of each title.

Nouns and proper nouns appear frequently in all titles, as seen in Table III. Therefore, in the analysis, we focused on this category.

Next, using the results of the morphological analysis, we calculated the importance of words using the $tfidf$ method. The $tfidf$ method is a type of index for word weighting, which can be calculated by the product of tf (term frequency) and idf (inverse document frequency) [13].

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (1)$$

$$idf_i = \log \frac{|D|}{|\{d : d \ni t_i\}|} \quad (2)$$

$$tfidf_i = tf_{i,j} \times idf_{i,j} \quad (3)$$

where $n_{i,j}$ is the occurrence frequency of word t_i in sentence d_j, $\sum_k n_{k,j}$ is the summation of the count of all the words in sentence d_j, $|D|$ is the total number of documents, and $|\{d: d \ni t_i\}|$ is the number of documents that contain word t_i. Here, the sentence is one tweet, and the document is the tweet group of each animation title.

Here, tf is considered to have a higher degree of importance as the number of occurrences in a sentence becomes larger. In addition, according to idf, words that are used across several documents are not important. The reason for calculating the importance of words is that the degree of expression of a characteristic differs based on the title, even though it is the same word. We believe that it is possible to obtain a summary of the title by calculating the importance of words. In addition, in this study, the tweet group of each title was created as a separate document. We calculated the importance of words across the four documents.

Based on the results, words with the top 100 $tfidf$ values were selected as characteristic words. We investigated the characteristic words of each title. In addition, to consider the relationship among the characteristic words, we calculated the co-occurrence frequency using the N-gram method. The N-gram is a type of language model that investigates how often N character strings or word combinations appear in a certain character string. In this study, we evaluated the co-occurrence relation with N equal to 2 (a bigram) [14].

Table IV summarizes the analysis results of $tfidf$ and the N-gram method for each title.

From the above results, it is clear that the words that have high values of importance differ depending on the title. In addition, some of the features of each title were identified. For example, for title A, the importance of Animation is the lowest compared with other titles. We believe that this is because of the variety of related product development in addition to animation for A. Conversely, for title C, Animation has a higher importance when compared with other titles. C is a toy project, but television broadcasting is its main field of activity. D is a cartoon from a popular magazine for youth in original animation. Therefore, it can be inferred from the appearance of the expected value for the production company that the importance of the production company is high.

TABLE III. FREQUENCY OF APPEARANCE OF PARTS OF SPEECH UNIT

Part of Speech \ Title	*A*	*B*	*C*	*D*
Noun, proper noun	651,973	779,838	567,162	772,288
Verb	100,167	61,122	108,464	57,962
Symbol	127,261	128,996	96,528	87,153
Adjective	9,468	6,896	17,306	8,113
Total	1,128,610	1,174,366	1,082,699	1,143,288

TABLE IV. SUMMARY OF THE ANALYSIS RESULTS OF TFIDF AND N-GRAM FOR EACH TITLE

Title	Features obtained from NLP results
A	The importance of a word in the content of the card game is high. The importance of Animation is the lowest in comparison with importance of other titles. Words that suggest a product (e.g., Badge, Ice Cream, Commission, Price, Lottery, Cards) have high importance. Similar animation titles have high importance.
B	Words that suggest products have high importance. In particular, Releases, Distribution, Restrictions, and Quantities have high importance in comparison with other titles. The importance of a word with the character name and broadcast contents of animation are high.
C	The importance of Animation is highest in comparison with importance of other titles. The importance of a word containing a character name and broadcast contents of animation is high. Similar animation titles and mixed media product have high importance.
D	The importance of words related to broadcast content of animation is high. The importance of the word that is a character name and Production Company of animation is high. Similar animation titles have high importance.

From the overall trends, it is conceivable that a tweet composed of the following elements represents well the characteristics of the animation:

- Participants in the content: character, producer, production company, and related animation.
- Media mix or related product: product name, product category, and other media (events, radio, and books).
- Review of the animation broadcast: impression, criticism, and next thoughts.

V. MODELING OF THE DIFFUSION PHENOMENON USING TIME SERIES DATA

In this section, we discuss the construction of a model of the diffusion phenomenon using time series data. Specifically, we focus on the phenomenon of how a tweet spreads across users by being retweeted. We also conducted a clustering of the diffusion phenomenon using the model.

A. Summary of Data

First, we performed a basic aggregation and estimated the trend. Of the 115,104 unique tweets that were collected, we used 9,378 tweets that had been shared with (retweeted to) others at least once. Fig. 1 shows the distribution of the number of shared tweets. The vertical axis of the figure represents the frequency, and the horizontal axis represents the number of retweets. The maximum number of retweets is 12,317. In

Fig. 1 shows the distribution of the number of tweets according to the number of retweets (up to 40). According to Fig. 1, approximately 40% of tweets were retweeted only once, and the distribution has a very long tail.

There are various definitions of the diffusion phenomenon, but in this study, we focus on 11 original tweets for which the number of retweets is more than 1,000. Tweets with IDs 2, 5 and IDs 10, 11 are transmitted by the same user. Table V shows the number of follows and followers, the targeted title of the animation, and the user attributes for 11 tweets. Based on an investigation of the attributes of the user who posted the 11 tweets, 6 tweets (from 4 accounts) were posted by the official account of a company and the other tweets were posted by personal accounts.

Initially, we aggregated and visualized the transition of the number of retweets for the 11 targeted tweets. We calculated the width of time point t using the following three approaches:

1) Changes in the number of retweets per day from the day of the tweet until the end of the collection period (2016/10/2).

2) Changes in the number of retweets per hour from the time of the tweet until the end of the collection period (2016/10/2).

3) Changes in the number of retweets per hour that were made within 48 hours of the tweet.

Based on the results of 1, it was clear that the changes in the number of retweets per day did not differ between general and official accounts. We tabulated the change in retweets per hour for each tweet. We investigated the rate of the number of retweets and found that about 95% of the retweets were posted within 72 h (92% of retweets were posted within 48 hours) of the original tweet. From the above results, it can be inferred that the remaining period was approximately 2–3 days when the tweets are spreading.

Based on the results of 2, that is, the changes in the number of retweets per hour, we found multiple peaks. Specifically, there were tweets that peaked quickly after the tweet was posted, and some that reached a peak after some time elapsed. As a whole trend toward zero within 48 h, but the occurrence of the time of the peak differed depending on the tweets themselves.

B. Analysis of the Diffusion Phenomenon Model

From the basic analysis, it can be inferred that the state of diffusion is not uniform. In this section, we describe the application of a mixed normal distribution to the time series change in retweets. In a previous study, burst detection used a log normal distribution was performed [9]. However, in this study, a phenomenon occurred whereby the retweeting began a considerable time after the original tweet. In addition, the purpose of this analysis was to evaluate the number of peaks. Consequently, we used a normal distribution in this study. For validation, we also performed burst detection using a log-line normal distribution. It was found that the BIC (Bayesian Information Criterion) [15] was almost the same. When a log normal distribution was used, it was necessary to estimate the number of peaks after estimating the bursting start point. Therefore, a two-step estimation was necessary. For these reasons, we attempted to estimate the number of peaks using a mixed normal distribution in this study.

The number of tweets was counted every hour.

$$f(t) = \sum_{i=1}^{n} \pi_i f_i(t) \quad (4)$$

$$f_i(t) \sim N(\mu_i, \sigma_i^2) \quad (5)$$

Here, π_i is the composition ratio of the number of i. In addition, although the data were collected over 400 hour or more, only the data after the 60 hour following the tweets were used for analysis. The EM (Expectation–Maximization) algorithm [16] was used for parameter estimation. In addition, the number of peaks (double or triple) were selected based on the BIC.

In the estimation results, single and double peaks were observed. The number of peaks for each tweet is described in the next section.

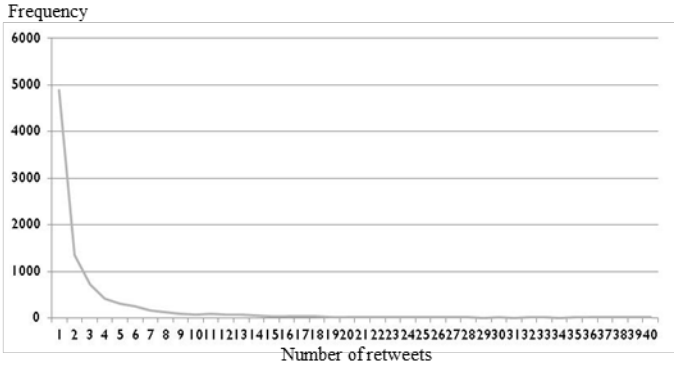

Fig. 1. Distribution of the Number of Shared Tweets (All Tweets).

TABLE V. DISTRIBUTION OF THE NUMBER OF SHARED TWEETS (ALL TWEETS)

User ID	Tweet ID	No. of Follows	No. of Followers	Target Title	User Type
a	1	396	1,637	B	General
b	2,5	0	117,275	D	Official
c	3	310	3,787	A	General
d	4	229	211	A	General
e	6	29,198	305,188	B	Official
f	7	287	2,808	C	General
g	8	182	624,503	D	General
h	9	36,588	566,091	B	Official
i	10,11	9	129,605	B	Official

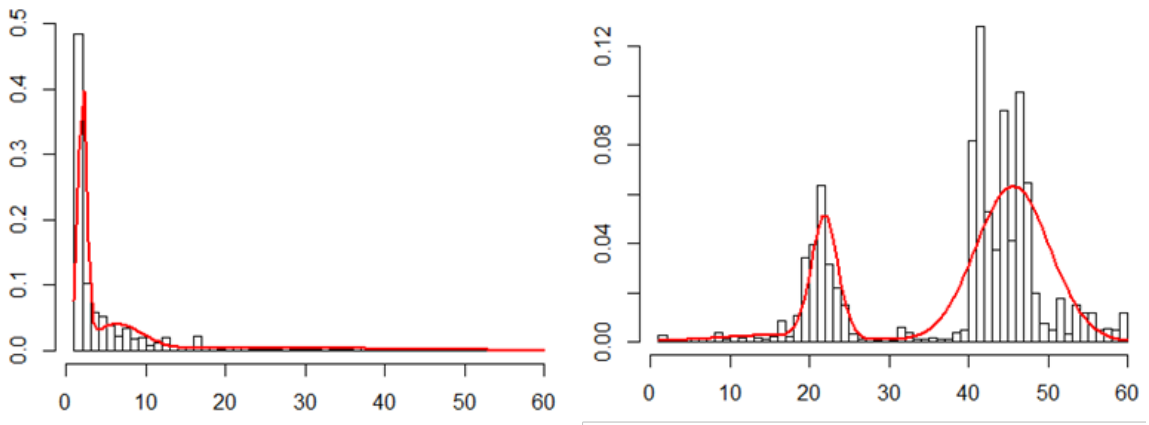

Fig. 2. Example of the Time Trends of Retweets and their Estimation Graphs.
Left: "Transient Topic" (ID 13), Right: "Secondary Spread Occurred Topic" (ID 4).

VI. Evaluation of Diffusion Phenomenon

Using the results from Sections IV (NLP analysis) and V (modeling using time series data), we attempted to evaluate the diffusion phenomenon. First, we classified each tweet as a "transient topic" or a "topic with secondary spread". Here, a "transient topic" refers to a topic that peaked soon after being posted, and then decreased. The peak lasted for approximately 1–3 hour (maximum 6 hour). In contrast, a "topic with secondary spread" refers to a topic where the diffusion phenomenon occurred again, only after the first peak had passed. For the classification, we used the mixture ratio of the estimated results that fit with the mixed normal distribution, and the shape of the established density when it fit with the mixed normal distribution. Fig. 2 shows an example of a "transient topic" and "topic with secondary spread". Comparing both graphs, it is clear that the states of the diffusion phenomenon are different.

Next, we calculated the characteristic value (c_j) for each tweet to evaluate how characteristically the title was expressed. First, we obtained the $tfidf$ value ($w_{i,j}$) for word i of tweet j. To evaluate that tweet j expresses the characteristic, the characteristic value (c_j) is given by equation (6).

$$c_j = \sum_{i \in d_j} w_{i,j} \quad (6)$$

For the calculation of the characteristic value, we used the $tfidf$ value obtained from the document word matrix for each animation title (Section IV). The classification of the diffusion phenomenon and the relative ranking of the characteristic value of the tweets are shown in Table VI.

In Table VI, there is no noticeable relationship between the number of peaks and the ranking of the characteristic values. Therefore, it can be inferred that a tweet has no distinctive characteristic (whether it expresses content) in terms of the presence or absence of reinfusion. Furthermore, it can be stated that most of the information transmitted through the official account was classified as "transient topic". Conversely, approximately 60% of the tweets of general accounts were "topic with secondary spread".

Next, the specific content of each tweet was described, as summarized in Table VII.

TABLE VI. Summary of the State of the Diffusion Phenomenon and Relative Ranking of the Characteristic Value

Tweet ID	User type	Title	No. of peaks	Ranking of characteristic value
1	General	*B*	2	6
2	Official	*D*	1	4
3	General	*A*	1	10
4	General	*A*	2	5
5	Official	*D*	1	3
6	Official	*B*	1	11
7	General	*C*	2	1
8	General	*D*	1	2
9	Official	*B*	2	7
10	Official	*B*	1	9
11	Official	*B*	1	8

TABLE VII. Summary of the Contents of Tweets

Tweet ID	Summary of the contents
1	Reviews and impressions
2	New announcement
3	Reviews and impressions
4	Reviews and impressions
5	Event public relations
6	New announcement
7	Reviews and impressions
8	New announcement
9	Mix Media news
10	New announcement
11	Event public relations

In terms of the content of tweets, "reviews and impressions" and "event public relations" had high characteristic values. Thus, the information posted by the official account was found to be related, in terms of content, to

new product announcements such as new animation, information about characters, goods, and mixed media. Generally, information on animation was posted from time to time in official accounts on Twitter and was used as advertising media. When an announcement is made, it can be inferred that it will spread among anime fans, and the diffusion will dissipate over time. Therefore, in the case of official accounts, it can be inferred that there are many "transient topics".

The tweets from general accounts were reviews conveying the impression made by animation broadcasting, characters, goods, and voice actors/actresses. As mentioned above, 60% of the general account tweets had two peaks. Consequently, further analysis is necessary, but if animation broadcasting and posting reviews and impressions are supported for Twitter users, it can be inferred that subsequently, trends such as continuous spreading until the next broadcast can be evaluated. In addition, according to the next broadcast, past posts are reevaluated and diffused. Based on the results of the analysis, we can evaluate the tendency of the diffusion phenomenon.

VII. Conclusion and Future Work

In this study, we attempted to evaluate the diffusion phenomenon on Twitter. We focused on the content of tweets and the diffusion phenomenon. Specifically, we analyzed the content of tweets using NLP. We also analyzed the diffusion phenomenon by generating a mixed normal distribution model using retweeted data, and we evaluated it using the results of the analysis. On the other hand, we modeled and evaluated the diffusion process of specific topics in this study. From the perspective of social marketing, there are also assessment of business relationships in companies [17][18][19] and research focusing on the rise of topics such as flaming phenomena [20]. It is also our important research theme to consider these phenomena based on the diffusion process of this research.

In the future, it will be necessary to increase the number of cases we study. In this study, we targeted animation. However, we are planning to investigate other content in future studies. We intend to study not only the normal distribution but also other distributions for modeling the diffusion phenomenon.

Acknowledgment

We thank Rooter Inc. for permission to use valuable datasets and for useful comments. This work was supported by JSPS KAKENHI Grant Number 19K01945and 17K13809.

References

[1] WHITE PAPER Information and Communications in Japan, (2018).

[2] E. Ioanăs and I. Stoica, "Social Media and its Impact on Consumers Behavior," International Journal of Economic Practices and Theories, Vol. 4, No. 2, pp. 295-303 (2013).

[3] D. Boyd, S. Golder and G. Lotan, "Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter," Proceedings of the 43rd Hawaii International Conference on System Sciences, pp. 1-10 (2010).

[4] B. J. 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, Vol. 60, pp. 2169-2188 (2009).

[5] K. Mane and K. Borner, "Mapping Topics and Topic Bursts in PNAS," Proceedings of the National Academy of Sciences of United States of America, Vol. 101, Suppl. 1, pp. 5287-5290 (2004).

[6] J. Kleinberg, "Bursty and Hierarchical Structure in Streams," Proceedings of 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 91-101 (2002).

[7] Y. Takahashi, D. Yokomoto, T. Utshuro and M. Yoshioka, "Analyzing Burst of Topics in News Stream," The Special Interest Group Technical Reports of Information Processing Society of Japan, Vol. 204, No. 6, pp. 1-6 (2011). (in Japanese).

[8] D. Gaikar and B. Marakarkandy, "Product Sales Prediction Based on Sentiment Analysis Using Twitter Data," International Journal of Computer Science and Information Technologies, Vol. 6, No. 3, pp. 2303-2313 (2015).

[9] Y. Matsuzawa, S. Saeyor Santi, F. Toriumi Y. Chen and H. Ohashi, "Analysis of Retweets Time Series by Mixture Model of Three-Parameter Lognormal Distribution," Proceedings of the 27th Annual Conference of the Japanese Society for Artificial Intelligence, pp. 1-4 (2013). (in Japanese).

[10] P. E. Sartwell, "The Distribution of Incubation Periods of Infectious Diseases," American Journal of Hygiene, Vol. 51, pp. 310-318 (1950).

[11] Y. Ueda and Y. Asahi, "Construction and Verification of Transition Model of Interest of Twitter Users," Communications of the Operations Research Society of Japan, Vol. 59, No. 4, pp. 219-228 (2014). (in Japanese).

[12] MeCab, http://taku910.github.io/mecab/ (2017/11/23, author checked).

[13] R. A. Baeza-Yates and B. A. Ribeiro-Neto, Modern Information Retrieval: the Concepts and Technology behind Search (2nd Edition), Addison-Wesley Professional (2011).

[14] C. E. Shannon, "A Mathematical Theory of Communication," The Bell System Technical Journal, Vol. 27, pp. 379-423 (1948).

[15] G. J. McLachlan and T. Krishnan, The EM Algorithm and Extensions, Wiley (2007).

[16] G. E. Schwarz, "Estimating the Dimension of a Model," Annals of Statistics, Vol. 6, No. 2, pp. 461-464 (1978).

[17] H. Tsurumi, J. Masuda and A. Nakayama, "Analysis of Relationship between Communication on Twitter about an Item and its Sales," Communications of Operations Research Society of Japan, Vol, 58, No. 8, pp. 436-441 (2013). (in Japanese).

[18] G. Mishne and N. Glance, "Predicting Movie Sales from Blogger Sentiment," in AAAI 2006 Spring Symposium on Computational Approaches to Analysing, (2006).

[19] R. Dijkman, P. Ipeirotis, F. Aertsen and R. van Helden, "Using Twitter to Predict Sales: A Case Study," arXiv:1503.04599 (2015).

[20] N. Takahashi and Y. Higaki, "Flaming Detection and Analysis using Emotion analysis of Twitter," IEICE Technical Report, Vol. LOIS2016-86, pp. 135-141 (2017). (in Japanese).