Suggestion of Operation Method of the Interest Shift Model of the Twitter User

Yusuke Ueda and Yumi Asahi

Shizuoka University

Abstract. Being interested in Twitter is modeled. The model is called a concern shift model. The concern shift model is constructed based on the linear differential equation model and the Kermack-Mckendrick model. This model has the feature in which it is interested. It analyzes it for the case to be interested by using the concern shift model on Twitter and CGM (Consumer Genelated Media). The tendency to analyze by the analysis is quantitatively shown. In doing the comparison between a proposal of the marketing technique and existing analysis service from the feature and the analysis result of the model, The model's view in the future is shown.

1 Introduction

Twitter is SNS site to contribute the short sentence within 140 characters. The user reach 13,920,000 people, it is the second scale next to Facebook in Japan. It is said that 14% of the amount of tweet that is frequently used in Japan and exchanged all over the world is Japanese [8]. The business use was also active, and in the realities of the use of social media of the enterprise in 2011, Twitter was 58% and 1st place [5]. It keeps taking the user communications as a use case with Twitter and the construction coming of an excellent relation. The satisfaction is given to the user and the case where the brand image royalty is improved rises [8]. The enterprise wants to contribute tweet and to learn the size and the directionality of user's interest from these cases. The real experience and the true opinion become important sources because a commodity, a frank opinion of service, and the evaluation are written as a user, and it reflects it. In a word, the element that bears sales and people's interests becomes it , saying that "Very good word of mouth is generated, and it is diffused". First of all, it is necessary to contribute tweet to give birth by word of mouth good, and to diffuse it. It is necessary to know timing to which the size and tweet of people interested in the matter that it wants you to tweet it are contributed.

The index that measures the size of the interest is assumed to be a number of tweet including the word that shows the object case. It is possible that tweet is not contributed and the user who has a potential interest that has the interest exists. Therefore, it is called the user who has the interest including them. The strength is not considered and only the interest existence pays attention to the size of the interest. When announcing of the new item and service and putting it

S. Yamamoto (Ed.): HIMI 2014, Part I, LNCS 8521, pp. 664–677, 2014.

on the market, a definite moment in a big event etc. are the most comprehensible at time that tweet is contributed. These events can be paraphrased as the boom. In the present study, it pays attention to the transition of the tweet number including the word of the object case at the boom, and making the size of the interest and the timing of the contribution visible is tried.

The differential equation model has been widely treated from the previous work to the transition of the state and the spread of information. The model whom Nakagiri and others [7] are proposing analyzes two or more cases, is done the comparison with real data, and shows the consumer's state transition in the boom quantitatively. Ueda and others [1] and Shirai and others [3] are analyzing the spread of the boom and information by using Kermack-Mckendrick model. Therefore, the present study tries whether to function in the information medium formed by the exchange of one user a person such as Twitter by using these models as an existing model. The model to make them adjust to Twitter user's interest is constructed.

The purpose of the present study constructs Twitter user's interest shift model, The case with the boom is analyzed, the transition of user's interest is read, and the factor to rouse the contribution of tweet is derived, and it proposes the marketing technique that uses Twitter.

2 Existing Model

2.1 Model of Liner Differential Equation

Standard Model. The standard type of model of linear differential equation that Nakagiri and others produced is called standard model.

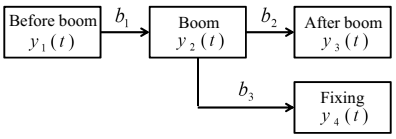

Fig. 1. Standard model

Fig. 1 shows the shift of the state of interest. Those states are shown below.
Before boom : The state that don't purchase products because of the boom
Boom : The state that purchase products because of the boom
After boom : The state that don't purchase because of tired of products
Fixing : The state that continue purchasing the product regardless of a boom
The authors postulates the total of the population of consumers always keeps constant value N.

$$y_1(t) + y_2(t) + y_3(t) + y_4(t) = N \tag{2.1}$$

Indicate the differential equation means shift of consumer.

$$y_1'(t) = -b_1 y_1(t) \tag{2.2}$$
$$y_2'(t) = b_1 y_1(t) - (b_2 + b_3) y_2(t) \tag{2.3}$$
$$y_3'(t) = b_2 y_2(t) \tag{2.4}$$
$$y_4'(t) = b_3 y_2(t) \tag{2.5}$$

(2.2)∼(2.5) mean changing number of consumer in unit time. b_1 ∼b_3 mean rate of consumer that shift next state.
About initial condition,

$$y_1(0) = (1-k)N \tag{2.6}$$
$$y_2(0) = 0 \tag{2.7}$$
$$y_3(0) = 0 \tag{2.8}$$
$$y_4(0) = kN \ (t \leq T) \tag{2.9}$$

It is said that I always take constant values until a boom begins. T is the time when a beginning boom, $k(\geq 0)$ points at the rate of consumers who have already become the Fixing. When present initial conditions (2.6) to (2.9), The number of consumers who are in each state obtaind by differential equation (2.2) to (2.5).

Sudden Model. Sudden model is the model what expended Standard model. It is suitable for an example to attract the interest of people rapidly. Show this in Fig. 2.

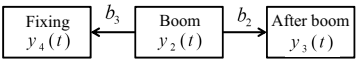

Fig. 2. Sudden model

Indicate differential equation to express the shift of consumer.

$$y_1'(t) = 0 \tag{2.10}$$
$$y_2'(t) = -(b_2 + b_3) y_2(t) \tag{2.11}$$
$$y_3'(t) = b_2 y_3(t) \tag{2.12}$$
$$y_4'(t) = b_3 y_2(t) \tag{2.13}$$

It takes the value that b_1 has a very big ($b_1 \to \infty$) to become the (2.10), therefore the consumers before the boom shift to the boom instantly In addition, the number of consumers who are in each state in T becomes (2.14) to (2.17) at the boom start time.

$$y_1(T) = 0 \tag{2.14}$$
$$y_2(T) = (1-k)N \tag{2.15}$$
$$y_3(T) = 0 \tag{2.16}$$
$$y_4(T) = kN \tag{2.17}$$

When present initial conditions (2.10)∼(2.13), The number of consumers who are in each state obtaind by differential equation (2.14) ∼ (2.17).

2.2 Kermack-Mckendrick Model

Sir Model. Kermack-Mckendrick model expresses the shift of a population infected with an epidemic. In previous reserch, simulation of the social boom and the spread of the information. Thus, the authors can expect an estimate, the prediction of the shift of the interest of the user. This section shows SIR model [1] from Kermack-Mckendrick model.

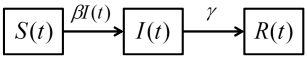

Fig. 3. SIR model

From Fig. 3, replace symptoms of individual with the state of user's inrterest in Twitter
S(t) : The number of users without the interest in object
I(t) : The number of users that the interest become obvious in object
R(t) : The number of users whom the interest in object was settled down
That user's interest become obvious mean the contribution of a tweet. After all, $S(t)$ points the user who contributes a tweet. In this reserch, the content of a tweet doesn't matter. The total of a user belonging to each state is N.

$$S(t) + I(t) + R(t) = N \tag{2.18}$$

The state shift of a user is (2.19) to (2.21).

$$\frac{dS(t)}{dt} = -\beta S(t)I(t) \tag{2.19}$$

$$\frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t) \tag{2.20}$$

$$\frac{dR(t)}{dt} = \gamma I(t) \tag{2.21}$$

$S(t)I(t)$ in (2.19) to (2.20) points at the number of times that a user who is interested comes into contact with an uninterested user. Express it in other words with "The user who did not yet show interest in an object is followed" and "Read tweet about object". The parameter β, γ is as follows.
β : The transmission of a interest
γ : The rate that interest drops

Seir Model. SEIR model is the model who added state E of the user who had a potential interest in SIR model. Potential interest is added to the process that

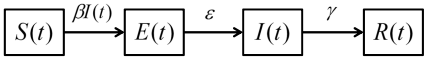

Fig. 4. SEIR model

a user is interested, Thus, this model can expect expression by a realistic shift of a interest.

Added state E points that there is the interest, but does not contribute the tweet.

$E(t)$: The number of users having the interest in object potentially

By this model, estimate the potential interest of an invisible user. The total of each state becomes (2.22) from Fig. 4 and (2.18).

$$S(t) + E(t) + I(t) + R(t) = N \tag{2.22}$$

The state shift of a user is (2.23) to (2.26).

$$\frac{dS(t)}{dt} = -\beta S(t)I(t) \tag{2.23}$$

$$\frac{dE(t)}{dt} = \beta S(t)I(t) - \epsilon E(t) \tag{2.24}$$

$$\frac{dI(t)}{dt} = \epsilon E(t) - \gamma I(t) \tag{2.25}$$

$$\frac{dR(t)}{dt} = \gamma I(t) \tag{2.26}$$

Parameter ϵ is the rate that the potential interest of the user become obvious. It shows that a contribution of tweet affects other users.

3 Interest Shift Model

3.1 Summary

The authors build the model who expressed the shift of the interest of the Twitter user based on an existing model.

The SEIR model considers potential interest. However, this model is made so that potential interest always shift a interest becoming obvious. The existing model is not suitable for a change of the interest of real people including Twitter user. The user whom read among users of Twitter and CGM exist 33% in Twitter as of September, 2011 [4]. As explained above, the authors suggest the interest shift model that expressed a change of the interest of the Twitter use.

Show below each state of the Figkanshin.

Indifference : The state without the interest in object

Interest(1) : The state that the interest become obvious in object

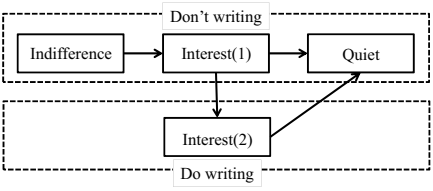

Fig. 5. Interesr shift model

Interest(2) : The state having the interest in object potentially
Quiet : The state whom the interest in object was settled down
It allowed a user of Interest(1) to diverge in Interest(2) and Quiet in the interest shift model. The interest shift model has two superiority and novelty.

1. Reflection of the action of the social media user whom worked as including Twitter
2. Visualize potential interest by considering a user receiving information generated in Twitter

In consideration of a user only for reading, the authors realize the change that is almost real social media and raise the precision of the analysis. In expressing the change of the user with the potential interest again, the authors catch the domain that was invisible with the tendency of the example until now. Make use for marketing and a trendy prediction.

The interest shift model builds two of "the linear model that assumed a linear differential equation model the basis" and "the non-linear model that assumed a Kermack-Mckendrick the basis".

3.2 Linear Interest Shift Model

The authors make three assumption on building the linear interest shift model.

1. The number of contributed tweet is the size of an appearing user's interest.
2. The contribution of tweet is suddener than the consumption activity in the existing model.
3. The change of the number of tweet before the boom does not consider it.

The linear interest shift model built by the above becomes Fig. 6.

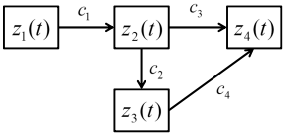

Fig. 6. Linear interest shift model

The authors assume the ratio of user belonging to each state $z_1(t) \sim z_4(t)$ in parameter of the time t

$$z_1(t) + z_2(t) + z_3(t) + z_4(t) = 1 \tag{3.1}$$

Expression of the state change of the user

$$z_1'(t) = -c_1 z_1(t) \tag{3.2}$$
$$z_2'(t) = c_1 z_1(t) - (c_2 + c_3) z_2(t) \tag{3.3}$$
$$z_3'(t) = c_2 z_2(t) - c_4 z_3(t) \tag{3.4}$$
$$z_4'(t) = c_3 z_2(t) + c_4 z_3(t) \tag{3.5}$$

Parameter to be given $c_1 \sim c_4$
c_1 : Percentage of the users who begin to get interested
c_2 : Percentage of the users who occur surfaced interest
c_3 : Percentage of the users who settle down before surfaced interest
c_4 : Percentage of the users who cool down gradually after surfaced interest
The value that is in each state at the time of the boom

$$z_1(T) = (1 - l - m) \tag{3.6}$$
$$z_2(T) = l \tag{3.7}$$
$$z_3(T) = m \tag{3.8}$$
$$z_4(T) = 0 \tag{3.9}$$

l is ratio of user of "Interest(1)", and m is ratio of user of "Interest(2)".

Give differential equation (3.2) $\sim$(3.5) condition (3.6)$\sim$ (3.9), and the solution of the differential equation is found.

3.3 Non-linear Interest Shift Model

The authors show the model what let SEIR model adapt to the interest shift model.

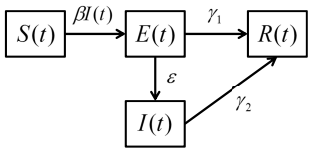

Fig. 7. Non-Linear interest shift model

By Fig. 7, the grand total that is in each state equals expression (2.22). The state change of the user becomes (3.10) to (3.13).

$$\frac{dS(t)}{dt} = -\beta S(t)I(t) \tag{3.10}$$

$$\frac{dE(t)}{dt} = \beta S(t)I(t) - (\epsilon + \gamma_1)E(t) \tag{3.11}$$

$$\frac{dI(t)}{dt} = \epsilon E(t) - \gamma_2 I(t) \tag{3.12}$$

$$\frac{dR(t)}{dt} = \gamma_1 E(t) + \gamma_2 I(t) \tag{3.13}$$

γ_1 and γ_2 point at recovery rate. the rate that γ_1 calms without interest being appeared, γ_2 is the same as γ in Fig. 4. Each parameter sets it as follows. The authors apply β for average of the number of the followers. γ, γ_1 and γ_2 are found from the reciprocal number of the mean infection period. It is the average of the period when interest lasts during the mean infection period, the authors estimates it from a change of the number of tweet. ϵ is found from the reciprocal number of the period awaiting mean infection. It is the average during the period to take before potential interest is appeared during the period awaiting infection. This period is elected optionally by a change of the number of tweet, adopt the most suitable value.

Next is given as an advantage of the non-linear interest shift model. It does not need to set the outbreak time of the boom. The authors can analyze it without dividing an object into two after the boom before a boom. Correspondence is possible for the phenomenon that is complicated because it is a non-linear model.

4 Analysis by Linear Interest Shift Model

4.1 Analysis Method

Whether Nakagiri and others linear differential equation model and linear interest shift model correspond to the boom on Twitter is analyzed. The domination of the linear interest shift model is shown from the comparison of the analysis results. Moreover, the case is analyzed and the tendency to the case is requested.

The data of the analysis used and acquired real-time, high index site TOPSY[6] around Twitter. This is called and real data is called. The object of the analysis greatly interests the user, and elects the case where the aspect of the boom is shown. It is provided the place word after the ranking of the Google retrieval rise retrieval word the first half of 2012 (The word ranking in 2012) and the nuclear power plant. The word ranking is July, 2011 to September, 2012 in 2012 in nuclear power plant September, 2010 to May, 2012 for the acquisition period of data. High-ranking word of word ranking is "sutema, siri, annular eclipse, annular eclipse of the sun, hikarie, sky tree, and comp-gatya" in 2012. The one that the user had been greatly interested in the tweet number of word rankings [2] calls the boom from showing expanding rapidly and getting depressed within a short term and is all right.

Comparison of the Model. The domination of the model is judged from the comparison of fitting in the application about real data and estimation. The analysis object is assumed to be "Nuclear power plant". It is clarified that the linear interest shift model is a model the goodness of the application fitting in by both models is compared that is appropriate for Twitter. The index of the application fitting in used for the comparison is assumed to be coefficients of determination $R^2$to which an eye measurement and real data of the graph are derived by the regression analysis of which the explanatory variable the objective variable, and is estimation.

Analysis of the Example. The analysis of the case clarifies interesting in each case by using the linear interest shift model. The tweet number is presumed according to the interest shift model. The validity of the state transition of user's interest is judged by the value of $R_2$and measuring the graph with eye. The tendency is considered from the content of the background and tweet that surrounds the value and the case with the set parameter about the transition of the interest and the classification and the factor are derived.

4.2 Analysis Result

Real data and standard, broken break out $Tw \times y3$ indicate estimation to *data* recorded in the graph. The backgrounds of the nuclear power plant that is the investigation object are a chain of accidents of the first nuclear power plant in Fukushima generated along with a East Japan great earthquake. The analysis results of the tweet number by an existing model become Fig. 8.

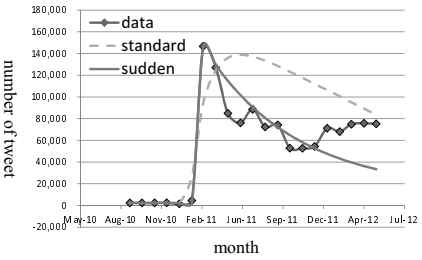

Fig. 8. Estimated result by existing model (Nuclear power plant)

Estimation cannot catch up with the transition of real data, and can be said that it is incompatible in a standard model. The explanation attaches from 0.007 the value of $R^2$to it. The result by a model broken because of one side shows an intimate transition to real data from taking of the eye measurement of the graph and the value of $R^2$0.605 compared with a standard model and it is understood that accuracy has improved. However, even the phenomenon of the decreasing tweet number's rising again was not able to be reproduced.

The estimated results of the tweet number by the interest shift model are shown in Fig. 9.

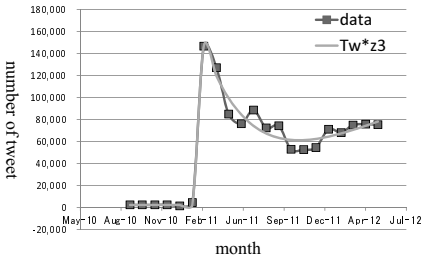

Fig. 9. Result by interest shift model (Nuclear power plant) $R^2 = 0.895$

Showing in Fig. 8 was able to show an impossible part. It can explain the accuracy according to the eye measurement of the graph and the size of the value of R^2 Among three models used for the comparison from Fig. 8 and Fig. 9, the accuracy of the presumption of the interest shift model is the most excellent. Therefore, it can be said that the interest shift model is a model that is appropriate for Twitter.

It was possible to classify it from the transition of the state of the interest of the user who had obtained it from the analysis of the tweet number as shown in table 1.

Table 1. Location of case

Case	Target Example (Coefficients of determination)
CONTINNUATION	genpatsu(0.896), sutema(0.877), hikarie(0.919)
THE SECOND BOOM	siri(0.718), sky-tree(0.884)
ONE-SHOT	kinkan-nissyoku(0.908), kinkan-syoku(0.780), comp-gacha(0.920)

The value of $R^2$of each case can be said that it is overall excellent, and there validity in the state transition of the interest with 0.7 to 0.9. The tendency is considered from the analysis result of the case where it belongs to the classified each case.

Continuance is a case to keep keeping the interest of a constant amount after the boom is generated. A constant amount indicates the interest of a high level is kept compared with before the boom is generated. The value of the parameter is $c_4 > c_1 > c_2 \geq c_3$. Big bias are not, and can be said the case with an active change of the interest to the size of the value. The user to whom the interest is actualized has the tendency to make quietly at the early stage while the period when do the latency of the interest is long.

The second boom is a case to show an exponential increase again when it has settled down the transition of the tweet number. The transition of the interest shows the aspect that looks like continuance against the tweet number in which a sudden change is repeated. The value of the parameter is $c_4 >> c_1 > c_2 >> c_3$. ">>" indicates the difference of the value is great. It is gotten tired of the interest that actualizes the interest easily and actualizes at once compared with in the case of continuance.

It is a case where the tweet number and the interest get depressed similar before it is generated if the boom ends. The appearance in which it is interested after the boom is generated is not shown. It is guessed the case forgotten as it is as a word that doesn't get into the news and a phenomenon that has been concluded. The value of the parameter is $c_4 >> c_3 > c_2 > c_1$. The rate that the interest to which the ratio with the interest begin is actualized small unlike in the case of current is made quiet is very large. Therefore, the loss boom ends the interest one after another as for the user who had the interest when the boom is generated. It gets depressed so that user's interest may respond to the tweet number, too and it makes it quietly.

5 Analysis by Non-linear Interest Shift Model

5.1 Analysis Method

The analysis process is done in order of the analysis by the extraction and the model who acquires of data and analyzes it.

The method of acquiring data makes the script using Twitter Streaming API by Python and acquires data. Neither the key word nor the user are squeezed from the public time line and data is acquired at random. It acquires it concurrently including the content of tweet at the account name and the contribution time. The time zone on July 4th to July 21st is made 8:00-24:00 for the period in 2013 that is the House of Councilors election campaign period of 2013.

GNU R and morphological analysis engine MeCab are used at the extraction to be analyzed. The transition of each word is derived by dividing the content of tweet acquired by using the morphological analysis into a significant, minimum word, and showing the occurrence rate of each word from the frequency analysis. As a result, the object case is extracted. It squeezes it to the noun that indicates the name of the event when the object case is extracted.

Odesolve Package is introduced into GNU R, the nonlinear interest shift model is mounted, and the object case requested from the frequency analysis is presumed. The number of users that has the interest actualized by using the account name acquired with tweet of the object case is requested. The tendency to the case is led from the transition of user's interest as well as the analysis of the case and the precision of analysis of the model is confirmed.

5.2 Analysis Result

As for the acquired data, the contributor's of tweet 200161.3 matter, each day average became the average a day tweet number with 188818 person and user's

people of number 54.4 of average followers. The acquired data extracts the event that the aspect of the boom is shown from data because it cannot use it for the analysis as it is by the diversity of the contribution number and the content.

The frequent occurrence word by the morphological analysis and the frequency analysis became the result of the word's without the meaning in the unit like the sign and the particle, etc. that composed the numerical value, URL, and the emoticon occupying the high rank. The word that was able to be caught here as an event was only "Election. "

It presumes according to the SIR model and the SEIR model who is an existing model intended for the election.

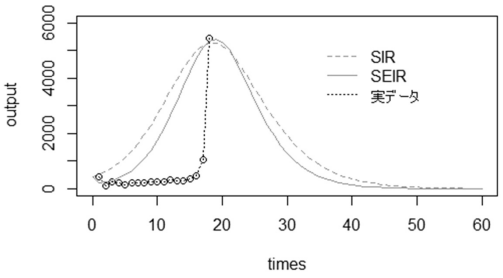

Fig. 10. Transition of the user interested in election by SIR model and SEIR model

Fig. 10 As for the transition of real data, an abrupt increase is shown as well as the object case that has been analyzed up to now, and not very different is understood even if taking the place from the number of months the unit of time of days the shown transition. A lot of users are having the interest actualized on the boundary of the 21st that it is an election year final day and is the ballot counting day. The estimation before the boom and the unbridgeable gulf of real data became large results. It will not be possible to correspond to a sudden change of the value. Estimation by the SEIR model is approached the transition of the event of the reality to consider the user who has a potential interest from approaching to real data and leads to the precision enhancement compared with the estimation of the SIR model.

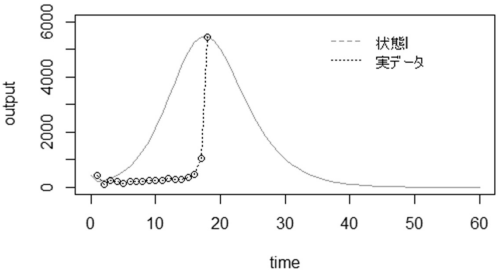

Fig. 11. Transition of the user who did tweet of election by nonlinear interest shift model

It presumes continuously according to the nonlinear interest shift model.

The estimated result became a result just like the SIR model and the SEIR model of figure from Fig. 11.

6 Conclusion

Two operation techniques of the interest shift model rise. First, proposal of marketing technique that uses factor to rouse contribution of tweet. Second, comparison with existing Twitter analysis service that bases novelty and domination of model.

Marketing Technique. Three tendencies to obtain from the analysis of the case are classified into two. The case where user's interest continues is Continuance and The second boom. The case where user's interest doesn't continue is One-shot. The factor divided into two is big and small of two parameter c_1 and c_3. The factor to decide the bigness and smallness of c_1 and c_3 can be guessed according to the background of the content and the case with tweet as follows. It derived to various topics like the case with the nuclear power plant around the case, and in the case where it belongs to continuance and the 2nd boom that became $c_1 > c_3$the tweet number and it was interested. The topic did not derive after generating the boom and the case with single-engined that became $c_1 < c_3$ was made quiet. It becomes a factor that this rouses the contribution of tweet. It proposes the marketing technique by using the factor.

Comparison with Existing Twitter Analysis Service. The part that should be paid attention is up to to limit the object of the analysis from existing analysis service only to the user who contributed tweet. 1/3 or more of the Twitter user is a user only of inspection like being in the meaning that the interest shift model constructs, and the size cannot be disregarded. The improvement of accuracy by which user's reaction and needs are read becomes possible by taking their interests into consideration by the quantification of a potential interest that is the novelty of the interest shift model.

References

1. Hiroshi, U., Youhei, O., Itsuki, K., Syu, S., Taku, Y., Katsuhide, W.: Mathematical principle of propagation phenomenon in society. In: MODELing Seminar: Spread Team of Remark Theory, Information, and Culture (2006)
2. GoogleJapanBlog. Retrieval key word ranking that became topic in the first half of 2012, `http://googlejapan.blogspot.jp/2012/06/google-2012.html`, final indorsement: 2013/9/18
3. Takashi, S., Takeshi, S., Fujio, T., Kosuke, S., Kazuhiro, K., Itsuki, N., Masasuki, N., Satoshi, K.: Estimation of False Rumor Diffusion Model and Estimation of Prevention Model of False Rumor Diffusion on Twitter. In: The 26th Annual Conference of the Japanese Society for Artificial Intelligence (2012)

4. Ministry of Internal Affairs and Communications. White paper on telecommunications in Heisei 23, http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h23/html/nc232330.html, final indorsement: 2013/9/18
5. Chen, Y.: One consideration about an effect of Collective Intelligence and the company use of social media, pp. 241–266. Osaka Sangyo University
6. TOPSY, http://topsy.com/, final indorsement: 2012/9/30
7. Yuko, N., Osamu, K.: On a Differential Equation Moel of Booms. Transactions of the Operations Research, 83–105 (2004)
8. Yoshio, H.: A Study on the business and take advantage of social media, pp. 155–179. Osaka Sangyo University. (2012)