

Prediction of the Concern of People Using CGM

Yusuke Ueda and Yumi Asahi

Shizuoka University

1 Introduction

It is that the spread of CGM (consumer generated media) is been noticed in recent years. There are 42,890,000 Japanese SNS users at the end of 2011, and it is say that 56,430,000 people will be user of it at the end of 2014. The month-long user of a buzz marketing site has reached to 83,560,000 people, and it being increasing. Recent market scale is 200,500 million yen in Japan. This will be expanded 396,300 million yen in 2013. It originates in the background of growth at (A), (B), and (C). (A) It can use for free. (B) Not only a general user but a company can do promotion and communication. (C) Receiving physical restrictions, such as a place and time, are decreased by development of a communications infrastructure and a portable terminal. On the other hand, pretending to be it a negative campaign and false rumor information by the anonymity of the sender occurs a lot. This becomes the inveterate problem that the net literacy of users is asked. Still ,CGM is becoming indispensable to our life up to everyday communication and interchange, purchasing activity [1]. We show some concrete example. It is used for Web promotion activity or the questionnaire survey. Because its give a user easiness of friendly feeling and get the true intention. Subsequently, we give an example of the further possibility of CGM. kabu.com Securities Co. Ltd develop a system that collect and analyze the huge information from SNS such as Twitter or mixi. It may be said that this is one of the big data business. And we expect further value of CGM. This time, we select Twitter from CGM and I build a model to observe the interest of people, and to estimate and inspect it. Twitter is SNS site to contribute the short sentence within 140 characters to. The user reach 13,920,000 people, it is the second scale next to Facebook in Japan. Also it has four characteristics. First, it holds the organization which can be connected without needing each other's approval. Second, it is easy to aim the opinion leader from the number of the followers and the number of retweet. Third, full retrieval service. Last, observation of the number of tweet is easy, and it is easy to treat it as quantitative data. Thus, We paid attention to Twitter.

The goal of this researches to prediction in transition of new objects (a product and a phenomenon). For that purpose, we build a model to express a change of the interest, and carry out numerical value simulation. From a simulation result, I examine to grouping the tendency of research object.

2 Concern Shift Model

The model is built based on "ON A DIFFERENTIAL EQUATION MOEL OF BOOMS" [3]. This model is built by paying attention a boom. A boom is said

“A product and a phenomenon spread among people instantly, and it is forgotten afterwards during a short term.” The interest of people greatly changes by a boom. In a word, it is easy to read a change if it greatly works, and is suitable for surveying a change of the interest. In the construction of the model. I use the number of tweet to measure interest for numerical value. The existing model collects the number of the words and target consumers population from various information media (ex Newspaper, data yearbook ,statistics data on web). And it estimates condition of consumption and the recognition. The change of the number of tweet before the boom does not consider. Because, there is more Twitter in the appearance number of the word after boom than these information media.

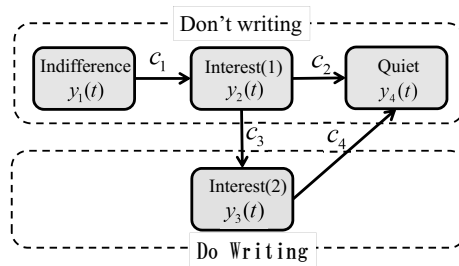


Fig. 1. concern shift model

From fig. 1, State of the user

Indifference : State of the user who does not recognize to an object

Interest(1) : State of the user who has not surfaced although it is interested to an object

Interest(2) : State of the user who it is interested and has surfaced to an object

Quiet : State of the user for whom the concern about an object settled down

I assume the ratio of user belonging to each state $z_1(t)$ ‘ $z_4(t)$ in parameter of the time t

$$z_1(t) + z_2(t) + z_3(t) + z_4(t) = 1 \tag{2.1}$$

Expression of the state change of the user

$$\dot{z}_1(t) = -c_1 y_1(t) \tag{2.2}$$

$$\dot{z}_2(t) = c_1 y_1(t) - (c_2 + c_3) y_2(t) \tag{2.3}$$

$$\dot{z}_3(t) = c_2 y_2(t) - c_4 y_3(t) \tag{2.4}$$

$$\dot{z}_4(t) = c_3 y_2(t) + c_4 y_3(t) \tag{2.5}$$

Parameter to be given c_1, c_4

c_1 : Percentage of the users who begin to get interested

c_2 : Percentage of the users who occur surfaced concern

c_3 : Percentage of the users who settle down before surfaced concern

c_4 : Percentage of the users who cool down gradually after surfaced concern

The value that is in each state at the time of the boom

$$z_1(T) = (1 - l - m) \tag{2.6}$$

$$z_2(T) = l \tag{2.7}$$

$$z_3(T) = m \tag{2.8}$$

$$z_4(T) = 0 \tag{2.9}$$

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

Give differential equation (2.2) (2.5) condition (2.6) (2.9), and the solution of the differential equation is found.

($T \leq t$)

$$z_1(t) = Pe^{-c_1(t-T)} \tag{2.10}$$

$$z_2(t) = \frac{c_1 P}{C_1} (-e^{-c_1(t-T)} + e^{-(c_2+c_3)(t-T)}) + le^{-(c_2+c_3)(t-T)} \tag{2.11}$$

$$z_3(t) = \frac{c_1 c_3 P}{C_1 C_2} (-e^{-c_1(t-T)} + e^{-(c_2+c_3)(t-T)}) + \frac{lc_3}{C_2} (e^{-(c_2+c_3)(t-T)} - e^{-c_4(t-T)}) + \frac{c_1 c_2}{C_2(c_1 - c_4)} P(e^{-c_1(t-T)} - e^{-c_4(t-T)}) + me^{-c_4(t-T)} \tag{2.12}$$

$$z_4(t) = \left(c_3 + \frac{c_2 c_4}{C_2} \right) \left(\frac{P}{C_1(c_2 + c_3)} \right) E_2 + \frac{lc_2}{C_2(c_2 + c_3)} \left(-c_4 e^{-(c_2+c_3)(t-T)} + (c_2 + c_3) e^{-c_4(t-T)} \right) + \frac{c_2 P}{C_2(c_1 - c_4)} (-c_4 e^{-c_1(t-T)} + c_1 e^{-c_4(t-T)}) \tag{2.13}$$

$$- \frac{lc_3}{c_2 + c_3} e^{-(c_2+c_3)(t-T)} - me^{-c_4(t-T)}$$

$$+ \frac{c_2 P}{C_2} \left(1 + \frac{c_4}{c_2 + c_3} \right) + \frac{1}{c_2 + c_3} \{c_3 P - c_3 l + c_2 l\} + m$$

$$P = (1 - l - m), C_1 = (c_1 - c_2 - c_3), C_2 = (c_4 - c_3 - c_2)$$

$$E_2 = (c_2 + c_3) e^{-c_1(t-T)} - c_1 e^{-(c_2+c_3)(t-T)}$$

3 Data Analysis

3.1 Analysis Method

The process of analysis method is as follows.

1. The numbers of the tweet containing the name of investigation contents are observed and collected.
2. Simulation result has been compared with collected data. We value it at the eye measurement of graph and the coefficient of determination decided by regression analysis.
3. The potential concern of the user who cannot measure from the number of the tweet is predicted. (The user's concern makes the number of the tweet the index.)
4. It is inspected whether a tendency is looked at a user's concern predicted transition of the number of the tweet.

There is not the numerical index to measure usefulness in procedure 4. Thus, we judge it from the user reaction (The number and contents of the tweet) and the environment of the target word. The target word elects from upswing in Google searching search word ranking in the first half in 2012. In the election of the word, it is thought that a thing collecting much tweet is desirable as subjects of survey. The thing which greatly collected the interest of the user showed sudden growth during a short term. Understood that it was similar to a boom.

Target words are "genpatsu (nuclear power generation), sutema (stealth-mareting), siri, kinkan-nissyoku (annular solar eclipse), kinkansyoku (Synonym of kinkan-nissyoku), hikarie (One of the shopping centers), nottv (net media), sky-tree, comp-gacha (complete-gatya)". The investigation period is Jul-2011 - Sep-2012. The use site is TOPSY [2].

3.2 Setting of the Parameter

We set the parameter to perform the simulation by the model. The setting of the parameter uses least squares method by the Solver of Excel. When we set it, set the number of tweet per unit time of $Tw \times z_3(t)$ Tw expresses the number of total tweet per unit time. We suppose the number of tweet per month to be 1,200 million and analyze.

4 Result of Analysis

We perform inspection and consideration of the analysis. $Tw \times y_3$ is indicated estimated value, *data* is indicated measured value. The number of tweet of the target word and change of the interest state of the user are provided than simulation. We can classify it in three cases from them.

Continuation

The case which continues keeping the interest of the fixed quantity after a boom. The fixed quantity is that the interest of the standard is higher than before boom. The revival that the number of tweet increases again is classified in this case too. In the change of the interest, Interest(1) continues lengthening after a boom.

The Second Boom

The case which shows sudden increase again when a change calmed down. The increase and decrease of the number of tweet is confirmed before a boom. The change of the interest is similar to continuation.

One - Shot

The case which suddenly falls in when a boom is over. After a boom, there are few signs that interest rises to. It is supposed that the phenomenon that has concluded without what is used on Twitter is just forgotten. The number of tweet decrease, and the interest of the user calms at the same time.

Table 1. Positioning of the target word

	Target word
CONTINUATION	genpatsu , sutema hikarie , nottv
THE SECOND BOOM	siri sky-tree
ONE - SHOT	kinkan-nissyoku kinkansyoku comp-gacha

Inspect the precision of the model as a specific example from three cases which were classified.

4.1 STUEMA

An official name is called stealth-mareting with sutema. Point to publicizing it with an article so that it is not noticed with an advertisement by consumers. The words existed than before. However, it became if in late years it would be used frequently led by a net user.

It is said that the simulation result is good from Fig. 2 and R^2 . But it did not come true to reproduce the increase and decrease of the number of tweet fully. It is necessary for precision of the simulation to give a thought to a standard to what extent.

The change of the user who is in each state becomes Fig. 3.

The user of the state of Indifference occupies most in a change . In this situation, the user of other states is uncommon. I will omit a change of Indifference

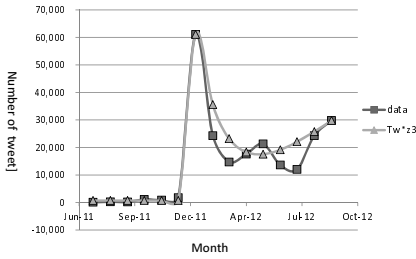


Fig. 2. Change of the number of tweet (sutema) ($R^2 = 0.877$)

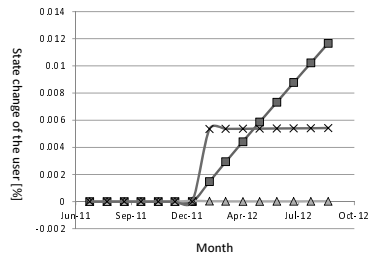


Fig. 3. State change of the user (sutema)

in a similar case from notation in future. When we pay attention to a change of Fig. 2, Interest(1) increases steadily, and Quite shows presence of mind. When interest gradually penetrates from there, we can read it. On the other hand, it is thought that this is supported only by a digital native and youth good at net slang.

4.2 SIRI

SIRI is secretary function applications software for iOS. It provides a reply and the Web Service for the question of the user by the sound recognition. It was put on iPhone 4S on October 4, 2011, came to support Japanese on March 8, 2012, and attracted attention by announcement of iOS6 of September, 2012 again.

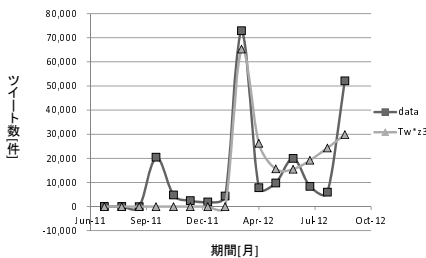


Fig. 4. Change of the number of tweet (siri) ($R^2 = 0.718$)

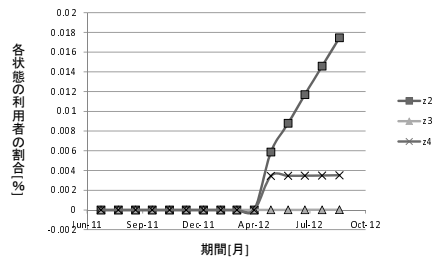


Fig. 5. State change of the user (siri)

We set March, 2012 when a Japanese edition was released with a boom start from Fig. 4. We have ignored the big upsurge at the time of the announcement in the previous year in October. Not having been able to follow a rapid increase of the number of tweet of September, 2012, it may be said that it is the factor that a value of R^2 has lowered. These will become the future problem.

The change of the user who is in each state becomes figure 5.

We were able to confirm that there was not the influence to the whole Twitter like stemma. The user of Interest(1) shows big growth for the littleness of the ratio of a tweeting user. We can consider this growth from contents of the tweet. For example "A new story came out when I said that I told siri that it was interesting. <http://t.co/4zC6BpBF>". He(she) does not merely declare the convenience of the verge using SIRI, suggests play to let the interest of the partner go down. It may be said that the organization where a user offers new value to is a characteristic of CGM.

4.3 Comp-Gacha

Comp-gacha calls an official name complete-gatya. (The meaning of gacha is capsule toy.) It is one of the charging methods of the social game for cell-phone, smartphones. It purchases an item by a lottery method like a capsule toy. And complete several kinds of decided items. The system what obtain a rare item by doing so it. Consumer Affairs Agency forbade comp-gacha by the charging large amount request of the social game on May 5, 2012. The contents that accounted for important weight in the profit of the industry were abolished. Therefore a social game-related brand of the Tokyo business market makes a sudden drop. DeNA and GLEE of industry major became the stop-low, and aggregate market value of 200 billion yen became extinct in a day.

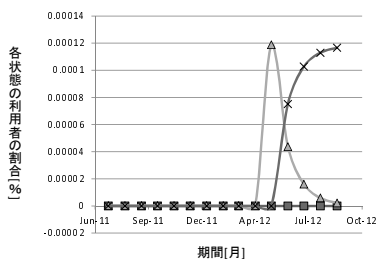
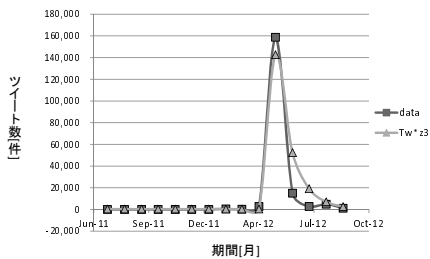


Fig. 6. Change of the number of tweet **Fig. 7.** State change of the user (comp-gatya) ($R^2 = 0.928$)

There was not the tweet about comp-gacha while service is carried out. Detailed structure was transmitted many users after Consumer Affairs Agency decided it illegal. In that way we understand that it followed that collect big interest. The stock plunge acts as a cause to attract interest. Change of the user who is in each state is Fig. 7

We collect interest at the time of the outbreak of the phenomenon, but calm if a boom is over because there is not an additional impact. The case concluded early, It is thought that it becomes the aspect like this time.

4.4 Observations

Cf. analysis, If the pattern of continuation and pattern oh one-shot, An estimate of the precision that R^2 is high in like about 0.9 is possible. The problems that the model had became clear from the second boom.

Specially, because the control of the boom is early as for the single-engine pattern, a stable estimate is possible. Conversely, Interest(1) continues lengthening, and do not get control in other cases. Thus, it is necessary to build the structure which can be equivalent about a future prediction.

About a Change of the Interest

It may be said that the judgment of the right or wrong of the purpose “visualization of size and the change of the interest” is difficult. Really, the proof of the estimated result is poor just method to consider it from contents of the tweet, and to plan adjustment. The measures thought about are qualitative investigations using the questionnaires at the very beginning. It is thought that sequential adjustment is difficult by Twitter observing a continuous change and the questionnaire to capture a change discretely. In other words We cannot assert with the interest of the user by an estimated result as now. However, it is a fact that one featuring it of the target words became clear by digitizing a process before leading to the estimate of the change of the number of tweet.

Corectively, We was able to visualize invisible interest. A problem is left in the proof of the result, and the right or wrong is divided by how to receive estimate results. In a future study, you must define it some other time how you treat it.

Text Mining

We examine qualitative analysis by the text mining. In this reserch, we express the influence that the qualitative elements such as contents of tweet used for only the proof in the change of the interest give.

The factor that comp-gatya is classified in a single-engine case is ”lacked versatility of word” and ”stamina of the topic is short”. In the case of nuclear power generation, number of tweet and volume of interest are bigger than other target words. Not only because this is derived from interest of the nuclear power generation in various elements, it is explained, but also it is pointed to be problem connected directly with our life. It is thought with the example that a sense of impending crisis supports than interest. We can ask about size of the influence that the qualitative factor and background in the target word. Then, we give the tendency of the word to classify from contents of the tweet.

1. Sutema, siri, kinkansyoku, kinkan-nissyoku, sky-tree and comp-gacha are liven up by user.
2. Hikarie and nottv are liven up by user and service-provider. They work hard not to run out of interest.
3. Genpatsu is liven up by user, the media, learned men and society as a whole.

We can classify these with volume of growth of Interest(1). The case to write in with the service-provider has a bigger surge of the interest than the case to write

in with user. Furthermore, we understood size of the influence by the media and learned men. From the above, we derive a tendency from contents of the tweet about the target word. And we promise improvement of the precision by it. K

5 Conclusion

Express the thing that became clear and the consideration about the inflection method of the model in this reserch.

- The simulation result by the model left a problem in the second boom, but appearance of the graph, decision coefficient R^2 were good together.
- In Twitter, there was a big difference in the number of tweet before and after a boom. Thus, it worked to have built a model based on an existing model.
- There is a tendency in the number of tweet and the interest of the user each and can sort it. In addition, We seemed to be able to grasp a tendency from an attribute derived than consideration by contents of the tweet.
- There is not a numerical index about the proof of the change of the user interest. Instead using contents of the tweet inspected it. We were able to get consistency by reading the background that surrounded a target word. However, it is not interest itself of the user under the present conditions. The side as the specification of the process before tweet showing interest is stronger.

References

1. Chen, Y.: One consideration about an effect of Collective Intelligence and the company use of social media, pp. 241–266. Osaka Sangyo University
2. TOPSY, <http://topsy.com/>
3. Yuko, N., Osamu, K.: On a Differential Equation Moel of Booms. Transactions of the Operations Research, 83–105 (2004)