2013 IEEE/WIC/ACM International Conferences on Web Intelligence (WI) and Intelligent Agent Technology (IAT)

Association between Selective Exposure and Attitude on Twitter

Yuki Ogawa
Rissho University
1700 Magechi, Kumagaya-shi, Saitama-ken, Japan
ogawa@ris.ac.jp

Hitoshi Yamamoto
Rissho University

Kakuko Miyata
Meijigakuin University

Ken'ichi Ikeda
Doshisha University

***Abstract*— In this study the analysis was carried out in order to find the condition under which the selective exposure occurs in Twitter. Specifically, the association between the selective exposure and the attitude was analyzed linking together the private activity log available on Twitter and the data obtainable through social research. As the result of analysis, it was found that the selective exposure tends to be strong to the retweetee, whereas the selective exposure tends to be weak to the followee and @tweetee. Furthermore, it became evident that the levels of the political participation, recognition on the importance of nuclear power plant, non-confidence in television news, anxiety, and the recognition of majority are associated with the strength of the individual selective exposure on Twitter.**

Keywords—twitter; selective exposure; attitude

I. INTRODUCTION

The social media such as Twitter and Facebook make it possible to transmit information of individuals and to acquire diverse information transmitted by other people. But, on the other hand, there is also concern about the problem, such as that the increase in individual selectivity against diversified informational environment will end up contributing to non-accommodating attitude toward the people with heterogeneous opinions as it encourages the selective exposure, whereby people select only the information which is homogeneous to their own opinions [6]. Particularly, the fragmentation of opinions due to avoidance of contact to excessively heterogeneous others in a scene where political issues are being discussed may possibly lead to deterioration in political/social tolerance, and the risk is also pointed out that it may interfere with democratic decision-making process [10].

Despite the increasing use of Twitter as the medium of political communication, the number of studies pertaining to the selective exposure is still limited [3][5]. In Japan, too, Twitter is also increasingly being penetrated in the wake of the Great East Japan Earthquake, and is expected to be used as a political communication medium for the future. In Twitter, is there the selective exposure wherein people selectively send, receive and impart the opinions which are in unison with their own, and what could be the conditions under which it occurs?

In this study, the individual private activity log by Twitter will be associated with the knowledge about the psychological attitudes obtained by social research so as to analyze the association between the selective exposure and the attitudes. Specifically, we will take up the pros and cons of the nuclear power plant, which has been at issue since the Fukushima nuclear disaster, and analyze the association between the individual psychological attitude toward this issue and the selective information contact on Twitter.

II. SELECTIVE EXPOSURE ON TWITTER

There are many studies on the association between the recognition of majority on online and the selective exposure can also be cited [1][3][4][5][6][7][12]. However there are studies which analyzed the activity logs on Twitter all right, but the number of studies in which analysis was made while associating behaviors with psychological attitudes is few.

In this study will be examined the connection between the selective exposure in Twitter and the psychological attitudes of individuals through the analysis wherein the private activity log available on Twitter was linked with the data obtainable through social research.

As regards the issue to serve as the object of analysis, we will take up the controversial political issue (pros and cons of nuclear power plant) which makes individuals at odds with each other. Also, the selective exposure behavior being defined variously by different studies with no index of definition, this study defines the "selective exposure" as an individual's own predisposition (pros or cons of a particular issue) and the tendency to contact selectively with homogeneous others. Specifically, of those others with whom an individual contacts on Twitter, it will be considered that higher the rate of those others who have the opinion (pros or cons of nuclear power plant) concurrent with that of the individual, higher the degree of the selective exposure being made.

A. Personal contactee on Twitter

Among the personal contactees on Twitter are the person who acquires information (the followee), the person with whom you communicate (@tweetee), and the person to whom you lend countenance (the retweetee). The personal

978-1-4799-2902-3/13 $31.00 © 2013 IEEE
DOI 10.1109/WI-IAT.2013.164

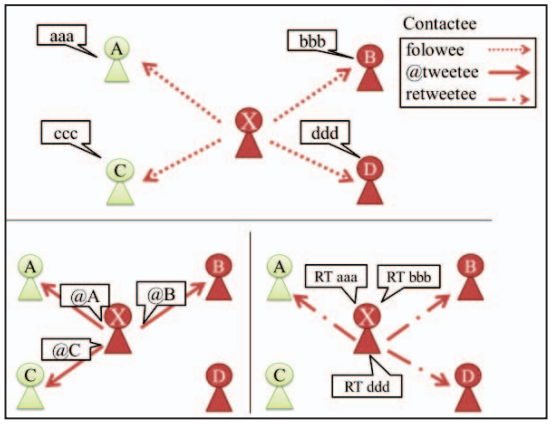

Fig.1. Personal contactee on Twitter (e.g. User X's contactees)

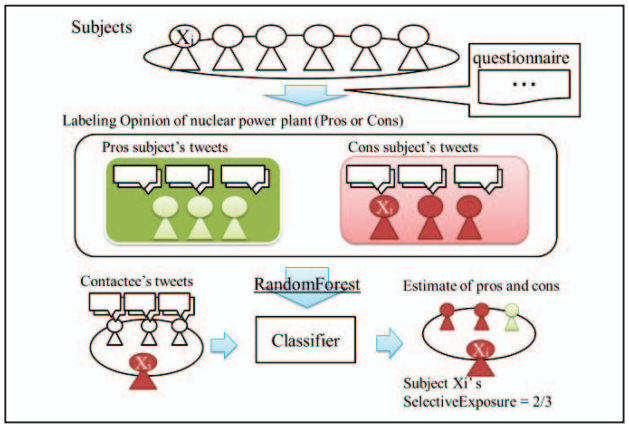

Fig. 2. Estimate of pros and cons of the contactee by machine learning

selective exposure with the following 3 types of contactees will be the object of this study (Fig.1). Although there is yet another contactee who transmits information (the follower), the follower being the one by whom to be joined and not supposed to be the one with whom to make selective exposure, it will be put outside the scope of this study.

III. METHODS

In the following is the summary of analysis made by this study.

A. Procedure

1) The data on the attitudes of the subjects will be obtained through questionnaire research. Then we ask the subjects for their Twitter accounts to obtain the information of Twitter activity log (statements and contactees).
2) In addition to the subject's opinion (the pros or cons of nuclear power plant) obtained from above (1), the opinions (the pros or cons of nuclear power plant) of the subject's contactees will be presumed based on the user's tweets by machine learning (Fig.2).
3) Based on the data on the pros and cons of nuclear power plant of the subject and of the contactees obtained by machine learning as stated in above (2), the index of selective exposure will be computed.
4) The association between selective exposure and subject's attitude made known in above (1) will be analyzed.

B. Subjects

A web survey was conducted through a research company on those. Of the total 1380 respondents, able to be collected, the 1,276 respondents (response rate 92.39%) who had the Twitter accounts and whose log data could be used were subjected to the analysis. The crawling of Twitter data was done using TwitterAPI to the 1,276 subjects of tweet data collection (Period: From March 1 to March 20, 2012). The users and tweets obtained are as Table 1.

TABLE I. TWITTER'S DATA COLLECTION

	Number of Users	Number of Tweets
Subjects	1,276	1,657,623
Subject's folowee	311,240	98,282,064
Subject's @tweetee	52,911	50,791,937
Subject's retweetee	83,419	32,105,769

C. Variables

The scale of attitude was made to consist of the following items.

Interest in politics: Regarding politics, they were made to evaluate on a 4-point scale i.e."1. I pay almost no attention","2. I pay scant attention","3. I pay a bit of attention", and"4. I pay considerable attention".

Anxiety: Regarding the evaluation objects i.e. "I feel uneasy whenever I think that another large-scale accident might happen at a nuclear power plant", and "I feel uneasy whenever I think that radioactive material might injure my health and the health of my family" , they were made to evaluate on a 4-point scale i.e. "1. I don't think so.", "2. I don't think so very much.", "3. I think so." and "4. I think so very much."

Confidence in media: Regarding the evaluation objects pertaining to the information on the nuclear accident after the East Japan Earthquake i.e. "The television coverage is credible", and "The information on Internet is credible", they were made to evaluate on a 4-point scale i.e. "1. I don't think so.", "2. I rather don't think so.", "3. I rather think so." and "4. I think so."

Opinion of nuclear power plant (Pros or Cons): Two opinions, A and B, were presented to them about nuclear

power plant and they were asked to evaluate on the extent to which their own views match the two opinions. The Opinion A was "Considering the stable supply of electricity, low CO2 emissions, and the eco-friendliness, we should maintain the nuclear power plants for the time being while studying on the alternative energy potential" (For nuclear power plant). On the other hand, the Opinion B was "Considering the potentially serious incident like the nuclear accident occurred after the Great East Japan Earthquake and the issue of radioactive waste disposal, the nuclear power plants should be shut down as quickly as possible" (Against nuclear power plant). Those who answered "1. It almost matches Opinion A" accounted for 23.7%, those who replied "2. It rather matches Opinion A" accounted for 26.8%, those who replied "3. It rather matches Opinion B" accounted for 24.5%, and those who replied "4. It almost matches Opinion B" accounted for 25.0%, dividing the answers roughly into quarters.

Recognition on the importance: Regarding the evaluation objects i.e. "How important to you is the nuclear power generation issue?" they were made to evaluate on a 4-point scale i.e. "1. It is almost unimportant.", "2. It is not so important.", "3. It is a bit important." and "4. It is very important."

Recognition of Majority on Twitter: As in the case of the evaluation of their own opinion about nuclear power plant, they were asked to evaluate on a 4-point scale to the question of "Which do you think is supported by more people in Twitter, Opinion A or Opinion B? ". For this, dummy variables were created i.e. "1" when the response coincided with their attitudes toward nuclear power plant and "0 when the response did not coincide with their attitudes.

Political Participation: As regards the political participation, they were made to evaluate on 14 items which pertain to their involvement in political participation activity (such as voting in election and involvement with community activities) in the last 12 months on a 2-point scale i.e. "1. Yes", and "2. No". (For the scale, additional value was used).

Use of Media: As regards the everyday use of media, they were asked about their frequency of watching the "political and social conditions in television" and the "political and social conditions on the Internet" and were made to evaluate on a 4-point scale i.e. "1. I don't watch.", "2. I don't watch so often.", "3. I watch.", and "4. I watch often.". In addition, they were also made to evaluate on a 4-point scale regarding "the information contact about nuclear power issue in television" and "the information contact about nuclear power issue on the Internet".

D. Selective Exposure

In this study, the opinions of the contactees (the pros and cons of nuclear power plant) will be estimated by means of machine learning based on the subjects' tweets obtained through the questionnaire research.

Estimate of pros and cons of the users by machine learning: There are several studies that presumed the genders and opinions of statement creators using the texts of tweet and blog entries [8][9][11]. In this study, we will also attempt to presume the pros and cons of the contactees with the users' pros and cons labels obtained by the questionnaire research and the tweets of those users as learning data. The learning data to be used will be the subjects' opinion labels on the pros or cons of nuclear power plant and the tweets obtained through the questionnaire research. Incidentally, for the purpose of presuming the pros and cost of nuclear power plant, the terms used were limited to those that include the words "nuclear power plant". Likewise, the tweets to be used for the learning as learning data were limited to the official Retweet that clearly lend users' countenance ("@tweets" and "non-official retweet" contain negative opinions as well). As regards the classification technique of the machine learning, we used the RandomForest [2] whose accuracy rate is said to be the highest. As for the terms, the top 2,000 terms of chi-square, whose accuracy rate by cross validation method was the highest, were used. The correction rate of the pros and cons classification by cross validation method in the machine learning was 74.8%.

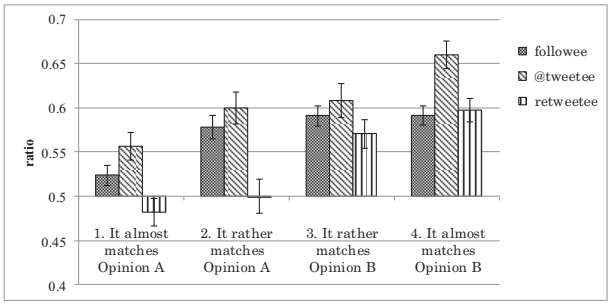

Fig. 3. Average of contactee's ratio of the opinion B (against nuclear power plant)

Variable of Selective Exposure: Based on the pros and cons data of contactees obtained, the indices of the percentage of those against the nuclear power plant and of the selective exposure are defined as follows.

- If user X*i*'s opinion is A (pros)

$$\text{Selective } Exposure = \frac{Number\ of\ Xi's\ contactees\ (Opinion\ A)}{Number\ of\ Xi's\ contactees\ (Opinion\ A\ or\ B)}$$

- If user X*i*'s opinion is B (cons)

$$\text{Selective } Exposure = \frac{Number\ of\ Xi's\ contactees\ (Opinion\ B)}{Number\ of\ Xi's\ contactees\ (Opinion\ A\ or\ B)}$$

IV. RESULTS

Fig.3 represents the ratio of the other side (the average of all subjects) to the Opinion B of the contactees. As for the retweetee, it can be noted from the figure that, while the average was less than 0.5 of the subjects with Opinion A, the average was more than 0.5 of the subjects with Opinion B. It shows that the subjects with Opinion A have more frequent contact with those others having Opinion A and the subjects with Opinion B have more frequent contact with those others having Opinion B; and that an individual tends to make selective exposure with the retweetee. Considering

TABLE II. PARTIAL CORRELATION BETWEEN SELECTIVE EXPOSURE AND THE ATTITUDE

	M (*SD*)	Selective Exposure (followee)	Selective Exposure (@tweetee)	Selective Exposure (retweetee)
Interest in politics	2.94 (0.76)	.05	.02	.03
Political participation	2.16 (1.83)	.12 **	.08 +	.13 **
Frequency of watching (TV)	1.95 (0.89)	-.01	-.02	-.06
Frequency of watching (Internet)	2.22 (0.72)	.04	.02	.03
Recognition on the importance	3.12 (0.77)	.00	.05	.12 *
Information contact about nuclear power issue (TV)	3.29 (1.03)	-.08 *	-.05	-.12 *
Information contact about nuclear power issue (Internet)	3.14 (0.88)	.07 +	.08 +	.04
Confidence in media (TV)	1.90 (0.76)	-.12 **	-.08 +	-.06
Confidence in media (Internet)	2.23 (0.75)	-.03	-.02	.04
Anxiety (Accident)	3.33 (0.78)	-.02	.07	.05
Anxiety (Health)	3.08 (0.91)	-.02	.09 **	.03
Recognition of majority (Society)	0.60 (0.49)	.07 +	.04	.05
Recognition of majority (Twitter)	0.64 (0.48)	.19 ***	.20 ***	.10 **
Num. of the tweet (log)	3.20 (0.75)	.01	-.06	.01
Num. of the tweet including "nuclear power" (log)	0.65 (0.65)	.17 ***	.12 ***	.22 ***

Notes. The subject's pros and cons opinion of nuclear power plant was set as the control variable.
+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$.

that the retreating is meant to lend countenance to the other person, the result can be said reasonable. On the other hand, as for the followee and @tweetee, the figures of both were more than 0.5. Thus, we can see that, as regards followee and @tweetee, the contact is not necessarily made with homogeneous others.

Table 2 shows the partial correlation between selective exposure and the attitude. The subject's pros and cons opinion of nuclear power plant was set as the control variable. From this result, it can be seen that the levels of the political participation, recognition on the importance of nuclear power plant, non-confidence in television news, anxiety, and the recognition of majority on Twitter are associated with the individual selective exposure on Twitter. It was also found that the strength of selective exposure and the number of tweets pertaining to nuclear power plant are highly correlated each other.

V. CONCLUSION

In this study the analysis was carried out in order to find the condition under which the selective exposure occurs in Twitter. Specifically, the association between the selective exposure and the attitude was analyzed linking together the private activity log available on Twitter and the data obtainable through social research. As the result of analysis, it was found that the selective exposure tends to be strong to the retweets, whereas the selective exposure tends to be weak to the followee and @tweetee. Furthermore, it became evident that the levels of the political participation, recognition on the importance of nuclear power plant, non-confidence in television news, anxiety, and the recognition of majority are associated with the strength of the individual selective exposure on Twitter.

As an issue in the future, it is necessary to make pros and cons classification and to analyze the contents of the tweets so as to find out if constructive discussion is made possible on Twitter. We also would like to conduct the analysis of covariance structure based on the result of the analysis discussed herein.

REFERENCES

[1] Adamic, L. A., and Glance, N. 2005. The political blogosphere and the 2004 U.S. election. In *Proceedings of the 3rd international workshop on Link discovery – LinkKDD '05*, 36-43.

[2] Breiman, L. 2001. Random Forests, *Machine Learning*, 45, 1, 5-32.

[3] Conover, M. D., Ratkiewicz, J., Francisco, M., Goncalves, B., Flammini, A., and Menczer, F. 2011. Political polarization on twitter. In *Proceedings of 5th International Conference on Weblogs and Social Media (ICWSM)*, 89-96.

[4] Johnson, T. J., Zhang, W., and Bichard,S.L. 2011. Voices of Convergence or Conflict? A Path Analysis Investigation of Selective Exposure to Political Websites. *Social Science Computer Review*, 29, 4, 449-469.

[5] Jungherr, A., and Jürgens,P. 2011. The Effect of Political Issues on the Structure of Communication Networks on Twitter. *Paper presented at the conference 'A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society'* organized by the Oxford Internet Institute, Oxford, UK, 21-24.

[6] Katz, J.E., and Rice, R.R. 2002. Social Consequences of Internet Use: Access, *Involvement, and Interaction. MA*: MIT Press.

[7] Price, V., Nir, L., and Cappella, J. N. 2006. Normative and informational influences in online political discussions. *Communication Theory*, 16, 47-74.

[8] Rao, D., D. Yarowsky, A. Shreevats, and M. Gupta. 2010. Classifying latent user attributes in twitter. In *Proceedings of the 2nd international workshop on Search and mining user-generated contents*, 37–44.

[9] Sriram, B., Fuhry, D., and Demirbas, M. 2010. Short text classification in twitter to improve information filtering. In *Proceedings of 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 841-842.

[10] Sunstein, C. R. 2001. Republic.com. *Princeton: Princeton University Press*.

[11] Turney, P. D. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 417–424.

[12] Yun, G.W., and Park, S. 2011. Selective Posting: Willingness to post a message online. Journal of Computer-Mediated Communication, 16, 2, 201-22