Bulletin of Networking, Computing, Systems, and Software – www.bncss.org, ISSN 2186-5140
Volume 6, Number 1, pages 1–4, January 2017

A Research on sensibility expression analysis of individual in regional variation bused on Twitter

(A Research on sensibility expression analysis of individual in regional variation bused on SNS)

Abdunabi Ubul
Tokushima University
Faculty of Integrated Arts and Sciences
Tokushima, Japan
Abdunabi211@hotmail.co.jp

Hidekazu Kakei
Tokushima University
Institute of Integrated Arts and Sciences
Tokushima, Japan
kakei@tokushima-u.ac.jp

***Abstract*—**

Since Twitter is capable to quickly send information and a person can casually tweet anything not important enough to write in a diary or a personal feeling, it is ideal to analyze user's emotional feeling as a study subject. For analyzing an area difference in emotional expression between urban and rural area, we conducted a survey for young Twitter users in Tokyo 23 wards and Tokushima city regarding the field of politics as seemingly little interest mainly among young generations, then examined an individual emotional difference by area through analyzing emotional expression of Twitter users in the above-described study subject areas. There was a significant difference in the data median value between Tokyo 23 wards and Tokushima city from the survey results, and the study found a larger emotional intensity in the field of politics for Tokyo 23 wards compared to Tokushima city. The study also confirmed a stronger negative feeling by users in Tokyo 23 wards in terms of a balance between positive expression and negative expression.

Keywords*— *Twitter; sensibility expression; politics; Regional variation

I. INTRODUCTION

Recently, a SNS service called Twitter has been getting so much attention and also widely studied so far [1-4]. Twitter is ideal to analyze user's emotional feeling as a study subject since it highly holds a real time property compared to blog or message board and also allows a person to casually tweet anything not important enough to write in a diary or a personal feeling. Presently, extracting word-of-mouth information from microblog has been studied widely [5-6]. Although a study with use of blog[7] or message board [8] is available, an area difference in personal emotional feeling is not examined in these studies.

For analyzing a difference in emotional expression between urban and rural area through the present study, we examined and discussed an area difference in individual emotional feeling by implementing a survey for Twitter users in the study subject areas and analyzing their emotional expressions under the political environment with change of voting age while young generations seem mostly little interest in the field of politics.

II. SURVEY OVERVIEW

The study examines and considers whether there is an area difference in emotional expression between urban and rural area on the basis of the survey results for the field of politics with total 200 Twitter users through Internet (Survey area: Tokyo 23 wards and Tokushima city, Respondents: 100 Twitter users for each area)

A. Survey contents

The study conducts a survey for area difference in individual emotional expression by posted information of Twitter users. The study also selects Tokyo 23 wards or central city of politics/economy in Japan as urban survey area and Tokushima city or the author's place of residence as rural survey area. Defining a user's local area is decided by activity area according to user profile information and the detailed contents are shown in Table 1.

TABLE I. SURVEY CONTENT DETAIL

Investigation period	2014 years from September to 2015 years June
Survey area	Tokyo's 23 wards, Tokushima City
survey number of users	200 members (Tokyo: 100 members, Tokushima City: 100 members,)
survey user age	16-year-old to 30-year-old
Analytical method	Sensibility expression extraction method
survey field	political

B. *Surveyed political field*

For Twitter political field specific using the field associated words [9-11] described in the animation field extraction.

- Step 1. Field associated words of extraction: in order to perform the specific topic field of the document you are writing the content and their impressions of the political, to extract the field associative word about the political than in the document.
- Step 2. Aggregate score: Even field associated words that can be reminiscent of the same topic field, the different strengths that can be reminiscent of the topic field. Therefore, in the field associated words, it has set the score individually in accordance with the strength that can be associated, respectively.
- Step 3. Specific political field: In the field specific, referring to the field associative language database at the time to extract the field associated words. To identify the political field to calculate the score of the field associated words about political that could be extracted.

III. ANALYTICAL METHOD

Obtaining text information of the user from the Twitter sentence be analyzed, performed in the procedure sensibility representation extracted from the obtained character information for analysis.

A. *Sensibility representation*

As a category to classify the sensitivity expression, "positive expression" that the reputation information has a good reputation, "negative expression" that the reputation information has a bad evaluation, positive, other than negative, said the such requests and the fact opinion the case, including the defined as "other"

B. *Sensibility expression extraction method*

Method [12] of extracting the Sensibility representation from Twitter sentence consists of five stages which will be described below. Showing a flow of process for extracting the Sensibility depicted in Figure 1.

- Step1: Morphological analysis of the input blog sentence, to give the title and the part of speech information of each morpheme
- Step2: By using the morphological analysis results obtained in Step1, to divide the one sentence in the reverse connection of the connection clause. The subsequent processing is performed in the divided input sentence unit after the connection clause separated.
- Step3: By using the morphological analysis result, to find the emotional expression dictionary. Here, if successful search obtain semantic information.
- Step4: Sensibility to detect by matching the semantic information and sensitivity detection rules obtained in Step3, if it detects a Sensibility, get Sensibility classification, sentiment classification, the evaluation classification.
- Step5: Using a morphological analysis result and the unnecessary word dictionary, it extracts morphemes could become objective nouns as Sensibility candidates. The database you have registered the unnecessary word referred to as the unnecessary word dictionary. Then, for each Sensibility detected by Step4, to determine the Sensibility expressed evaluates classification and Sensibility representation classification.

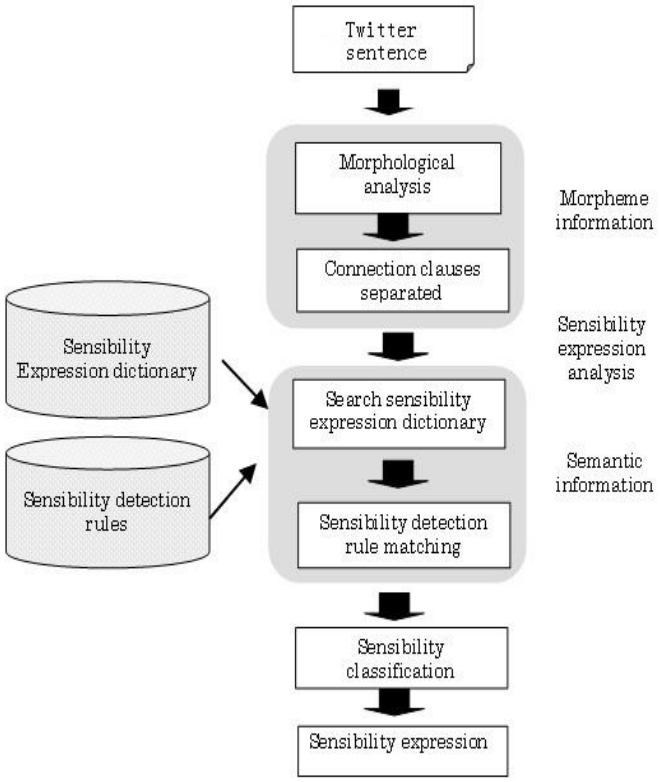

Fig. 1. Sensibility expression extraction processing

IV. VERIFIED METHODOLOGY

We conducted an analysis investigation for sensitivity expression on Twitter official site of [13], focused on the political field in Tokushima city and Tokyo 23 wards using 100 contributed article data. We set 10 % significance level for the verified methodology with Wilcoxon rank sum test and chi-square test.

A. Wilcoxon rank sum test analysis concerned with feeling intensity

We adopted Wilcoxon rank sum test method considering not able to apply specific distribution premise for the feeling intensity.

1) We conducted Wilcoxon rank sum test to determine the central value (median) difference of the feeling intensity data between Tokyo 23 wards and Tokushima-city in politics. Table.2 shows the results.

TABLE II. RANK SUM TEST RESULT OF WILCOXON

Statistic / Target areas	Number of observation	Median	P value
Tokyo's 23 wards	100	0.1936	0.00012
Tokushima City	100	0.0821	

2) The test result was p = 0.00012<0.01, then we could find a significant difference to the median data between Tokyo 23 wards and Tokushima-city, and found the feeling intensity of Tokyo 23 wards is greater than Tokushima-city on the political field. Figure.1 shows the comparison of these two areas.

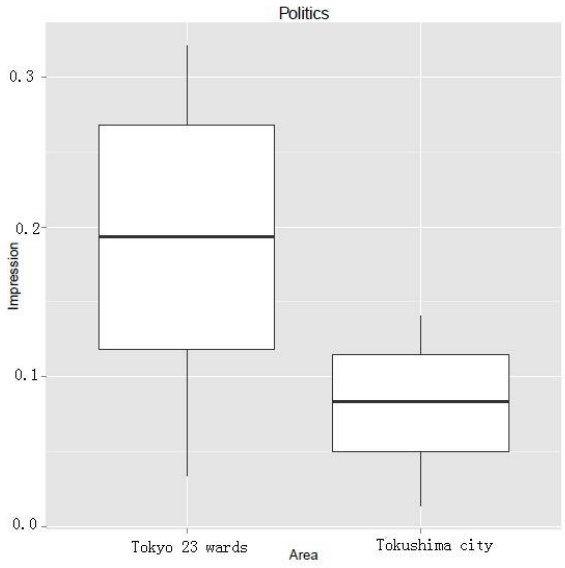

Fig. 2. Emotional intensity comparison of Tokyo's 23 wards and Tokushima user

B. Chi-square test analysis for positive and negative expression

We conducted chi-square test to verify the difference to positive and negative expression for sensitivity expression between two areas.

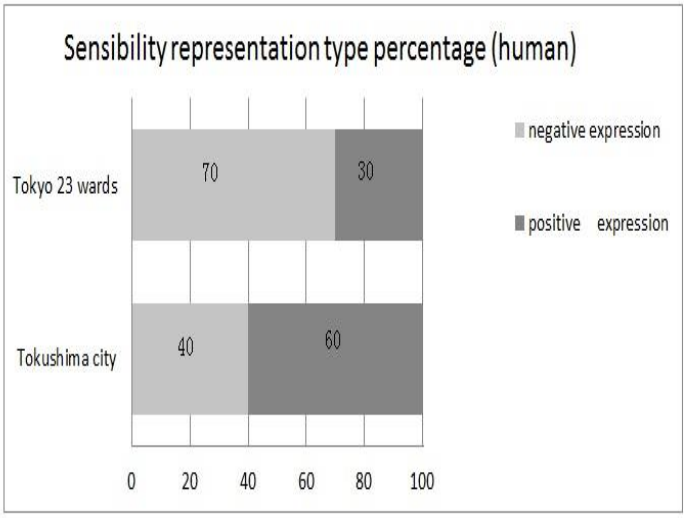

Fig. 3. Sensibility expression ratio number of people

TABLE III. CHI-SQUARE TEST ANALYSIS RESULTS

x^2 Test value	
x^2 =	16.28
Degree of freedom =	3
Upper probability of x^2 distribution (p) =	0.0134

Fig.3 shows the summary for the number of positive and negative expression in Tokyo 23 wards compare to that of Tokushima city, at the rate of positive and negative expression for sensitivity expression between two areas.

As the result of chi‐square test showed Table.3, the results were Chi square = 16.28, p value= 0.0134, and p = 0.0134<0.05, consequently, We could found the significant difference between Tokyo 23 wards and Tokushima city to positive and negative expression.

V. CONSIDERATION

We consider the factors of rank sum test and Wilcoxon chi-square test result with the feeling intensity of users in the political field in the two target cities (100 for each).

A. Consideration from the difference of feeling intensity

We confirmed the difference of feeling intensity between Tokyo 23 wards and Tokushima city with Wilcoxon rank sum test result. The results were considered mainly caused by the number of occurrence for the sensitivity expression word and importance.

B. Chi-square test analysis for positive and negative expression

We found the difference between Tokyo 23 wards and Tokushima city in Table3, from the chi-square test result for balance of positive and negative expression. We considered that Tokyo 23 wards users have strong negative sense due to so many key words such as distrust, disappointment, and problems with politics in the contributed articles.

We considered that caused by influence of psychological factors such as user intuition and effects of difference between average age of the users.

VI. CONCLUSION

We conducted investigation for the Twitter users in two target cities (100 for each), and investigated the difference of sensitivity expression for individuals by area with analyzing sensitivity expression of Twitter users in the areas, and verified the results using Wilcoxon rank sum test. From the user feeling intensity of Tokyo 23 wards and Tokushima city, high user feeling intensity of Tokyo 23 wards was clarified in political field. We confirmed the strong negative sense of Tokyo 23 wards users about a balance of positive and negative expression.

We are also going to verify categories other than political field of Tokushima city and Tokyo 23 wards users for the future studies.

REFERENCES

[1] D A Shamma, L Kennedy, E F Churchill, “Tweet the debates understanding community annotation of uncollected sources”, Proceedings of the first SIGMM workshop on Social media, 3-10 (2009)

[2] R. Plutchik, “The nature of emotions”, American Scientist, 89,344‐355 (2011)

[3] E. Aramaki, S. Maskawa and M. Morita,"Twitter Catches The Flu Detecting Influenza Epidemics using Twitter", Proceedings of the Conference on Empirical Methods in Natural Language Processing、1568-1576 (2011)

[4] Fouss F. "Random-Walk Computation of Similarities betweenNodes of a Graph with Application to CollaborativeRecommendation", IEEE Transactions on Knowledge andData Engineering, 2007.

[5] MJ. Paul and M. Dredze. "You Are What You Tweet: Analyzing Twitter for Public Health", Fifth International AAAI Conference on Weblogs and Social Media, 265‐272 (2011)

[6] L Backstorm, J Leskovec, "Supervised random walks:predicting and recommending links in social networks", WSDM, 2011

[7] Jonathan L. Herlocker, Joseph A. Konstan, Lorgen G. Terveen,John T. Riedl. " Evaluating collaborative filtering recommendersystems", ACM Transaction on Information Systems,2004,

[8] D Liben-Nowell, J Kleinberg. "The Link-Prediction Problem for Social Networks", The American Society for Information Science and Technology, 2007.

[9] A.Ubul, El.Atlam, H. Kitagawa, M. Fuketa, K. Morita, J. Aoe, "An Efficient Method of Summarizing Documents Using Impression Measurements", An Efficient Method of Summarizing Documents Using Impression Measurements, 2013, Vol.32, No.2, pp.371-391.

[10] A Ubul, H Kakei, J Aoe. " Research on Document Summary Generation Using Attribute Information",IJCAT Journal, vol1(1) ,pp.557 - 569 (2014)

[11] E.-S. Atlam, G. Elmarhomy, M. Fuketa, K. Morita and J. Aoe, "Automatic building of new field association word candidates using search engine", Information Processing & Management, 2006, Vol.42, No.4, pp.951–962.

[12] T. Yoshinari, E.-S. Atlam, M. Fuketa, K. Morita and J. Aoe, "Automatic acquisition for sensibility knowledge using co-occurrence relation", International Journal of Computing and Technology, 2003, Vol.33,No.3, pp.218–225.

[13] Twitter IP：https://twitter.com/