Research The Regional Differences in Person Sensibility Expression in Blog Document

[1] Abdunabi Ubul, [2] Hidekazu Kakei

[1] Faculty of Integrated Arts and Sciences, Tokushima University,
Minami josanjima 1-1. 770-8502 Tokushima, Japan

[2] Institute of Integrated Arts and Sciences, Tokushima University,
Minami josanjima 1-1. 770-8502 Tokushima, Japan

Abstract - Recently, the chance that the person sends information on Web has increased by the spread of the Internet. One of the backgrounds that the chance has increased includes the contribution of the blog. Because a lot of frank opinions of those of general people of the evaluation to the comment on the region and a certain commodity etc. are contributed though the diary is contributed as the main content in the blog. It is thought that it is used to express feelings plainly. Then, an individual sensibility expression difference that depended on the region was verified by executing the investigation for the blog user for the animated cartoon field, and analyzing the blog user's sensibility expression in the object region. There was a sensibility expression difference from the result of Tokyo 23 district and Tokushima City for the animated cartoon field among the object regions and it kept clear.

Keywords **- Blog Document**, **Sensibility Expression, Anime.**

1. Introduction

In a study on the use and understanding of emotional expression in the recent text communication, to extract a variety of emotional expression from electronic text information, such as the home page on the Internet, what is also often the analysis. However, emotional expression that appears in such text data is always not what many not have emerged as a result of with emotion.

Now, to extract the reputation information from the micro-blog has been numerous researches [1-5]. Studies exist with research [6-10] or bulletin board using a blog, but these techniques has not been studied regional differences in individual Sensibility expression in the region. In this paper, from the blog document, collecting post information about the Sensibility information keyword, the process for the blog-specific functions in addition to using a Sensibility expression extraction technique, by analyzing the Sensibility expressions of blog users in the study area, to investigate the Sensibility expression difference between the individuals in the region.

In Section 2, The survey overview will be explained.. In Section 3, Will be described Sensibility expression analysis technique. In Section 4, Will be described consideration and verification techniques. In Section 5, we summarize the paper and discuss future tasks.

2. Survey Overview

Based on using the Internet survey two cities result of the investigation into the animation field as the target of the blog user of (each 60 people), and the verification and considering the sensitivity expression difference between the area of the blog user.

2.1 Survey Content

For sensibility representation of individuals by region by posting information of the blog user, carry out the investigation. Survey content is shown in Table 1.

2.2 Surveyed Animation Field

For blog animation field specific using the field associated words [11-13] 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

Table 1: 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	120 members (Tokyo: 60 members, Tokushima City: 60 members,)
survey user age	16-year-old to 30-year-old
Analytical method	Emotional expression extraction method
survey field	Anime

writing the content and their impressions of the anime, to extract the field associative word about the animation 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 anime field: In the field specific, referring to the field associative language database at the time to extract the field associated words. To identify the animation field to calculate the score of the field associated words about animation that could be extracted

3. Analytical Method

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

3.1 Sensibility Representation

The sensibility representations, of the human cognitive ability, to the impression that feel directly through the sensory organs, the characteristics of the individual, have been regarded as subjective perceptions that reflect their experience.

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".

3.2 Sensibility Expression Extraction Method

Method [14] of extracting the Sensibility representation from blog sentence consists of five stages which will be described below. Showing a flow of process for extracting the Sensibility depicted in Figure 1.
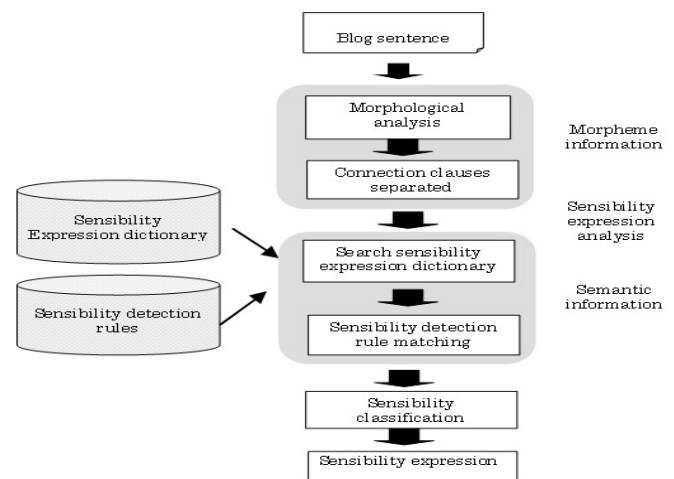

Fig. 1 Sensibility expression extraction processing

Step1: morphological analysis
Morphological analysis of the input blog sentence, to give the title and the part of speech information of each morpheme

Step2: connection clause delimiter
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: search of Sensibility expression dictionary
By using the morphological analysis result, to find the emotional expression dictionary. Here, if successful search obtain semantic information.

> The Sensibility dictionary, to its part of speech information Sensibility representation, it is a database to register by adding one of the semantic information.
> Semantic information and is, a representation of the meaning of individual expression, is added for the purpose of the concept of the Sensibility representation.

Step4: Sensibility detection rule matching
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.

> Sensibility classification: Sensibility the finest classification was classified by their meaning.
> Emotion classification:. Classification were classified Sensibility in the emotion of the principal.
> Rating classification: Sensibility "good", with three types of "bad", "Other" classified most simple classification.

Step 5: Sensibility expressed determination
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.

4. Verification and Discussion

In this research, Sensibility around the animation field from the official site of live door Blog [15] and Goo Blog [16], against local departure of the famous Tokushima in anime events and blog user 120 persons of the two areas, such as Tokyo's 23 wards posts data we were expression analysis survey. As a result by using the rank sum test and the chi-square test of Wilcoxon to carry out the discussion and verification.

4.1 The Wilcoxon Rank-Sum Test for Sensibility Expression Strength

1) Table 2 is a result of the Wilcoxon rank-sum test for Sensibility expression intensities in the two regions, P <0.01, to be highly Sensibility representation strength of Tokyo was confirmed.

Table 2: Rank sum test result of Wilcoxon

Statistic Target areas	Number of observation	Median	P value
Tokyo's 23 wards	60	0.1791	0.0012
Tokushima City	60	0.1685	

2) Figure 2 is a comparison of the emotion intensities in the two regions.
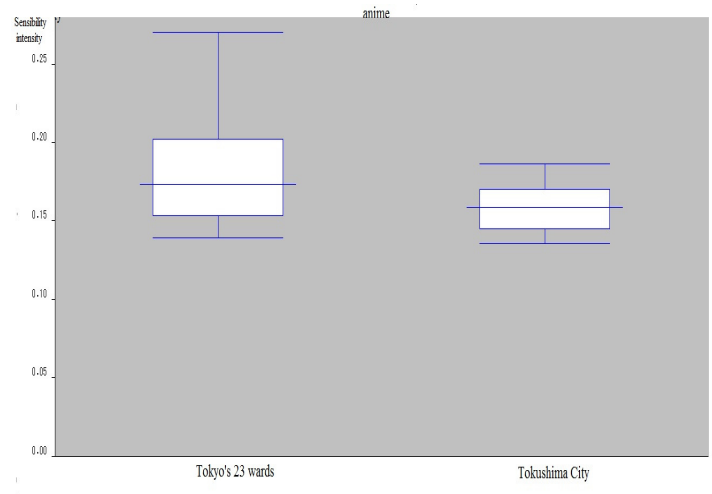

Fig. 2 Emotional intensity comparison of Tokyo's 23 wards and Tokushima user

4.2 Chi-Square Test for The Negative Expression And Positive Expression

1) Fig. 3, there is shown in the user-specific against the negative expression and positive expression in the two regions.
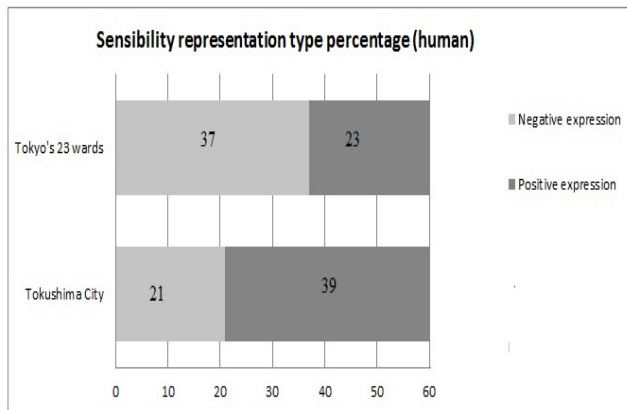
Fig. 3 Emotional intensity comparison of Tokyo's 23 wards and Tokushima user

2) Table 3 is a chi-square test analysis results that you made to the negative expression and positive expression of emotion in the two regions. The verification result, chi-square value = 3.55, p = 0.07> 0.05, Tokyo It was able to confirm that there is no difference to the negative expression and positive expression between the 23 wards and Tokushima.

Table 3: Chi-square test analysis results

χ^2 Test value	
$\chi^2 =$	8.5428
Degree of freedom =	4
Upper probability of χ^2 distribution (p) =	0.073

4.3 Consideration

Consider the survey two cities rank sum of Wilcoxon you have made to the Sensibility expression strength test and the chi-square test results with respect to animation field of (each 60 members) factors.

1) Discussion of the Wilcoxon rank sum test result
From rank-sum test results of the Wilcoxon you have made to the emotional intensity of Table 2, it was confirmed that there is a difference in the emotional intensity between Tokyo and Tokushima. The factors, number of occurrences of Sensibility representation words, importance is believed that central factors such as weight.

2) Consideration from the chi-square test verification result
From Table 3 of positive expression and the chi-square test verification result of to the balance of the negative representation, it was found that there is no difference between the 23 wards of Tokyo and Tokushima.

However, discomfort to post the contents of Tokyo's 23 wards user, anxiety, because the negative emotional expressions etc. were many, as the factor, the user's preference, the concept, is considered to psychological factors, such as intuition.

5. Conclusion

Conducted a study as a target the blog user surveyed two cities (each 60 members), to analyze the emotional expressions of blog users in the region, to assess the emotional difference between the individual depending on the region, the results of Wilcoxon It was verified by using a rank sum test method. From the results of the 23 wards of Tokyo and Tokushima in the target area, for the animation field, emotional intensity of a user of Tokyo's 23 wards have been high that to clear.

In future research goals, it is expected to carry out the verification in Tokushima City and Tokyo's 23 wards category other than the animation field for the user of the.

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] Lars Backstorm, Jure Leskovec, Supervised random walks:predicting and recommending links in social networks,WSDM, (2011)

[4] 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)

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

[6] Sergey Brin, Lawrence Page, The PageRank Citation Ranking:Bringing Order to the Web, 1998.

[7] Lise Getoor, Christopher P. Diehl, Link mining: a survey,SIGKDD, 2005.

[8]   Lars Backstorm, Jure Leskovec, Supervised random walks:predicting and recommending links in social networks,WSDM, 2011,

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

[10]   David Liben-Nowell, Jon Kleinberg, The Link-Prediction Problem for Social Networks, The American Society for Information Science and Technology, 2007.

[11]   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.

[12]   Abdunabi Ubul, Hidekazu Kakei, Jun-ichi Aoe: Research on Document Summary Generation Using Attribute Information、IJCAT Journal、1(1) 、557 - 569 （2014）

[13]   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.

[14]   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.

[15]   Live door Blog, http://blog.livedoor.com/.

[16]   Goo Blog, http://blog.goo.ne.jp/.

Author Profile:

Abdunabi Ubul   received his B. Sc. degree in economics and Management information from Xinjiang University, China in 2004. He has received his M. Sc. degree from Department of Economics, Faculty of Integrated Arts and Sciences, University Of Tokushima, Japan in 2008. Received his Ph. D. degree from Department of Information Science and Intelligent Systems. University Of Tokushima, Japan in 2012. His research interests include information retrieval, natural language processing and document processing.

Hidekazu Kakei   received his B.Eng. and M.Eng. Degrees in architecture from Nagoya University, Japan, in 1988 and 1990 respectively, and his Ph.D. in architecture in Kobe University in 2007. Since 2003 he has been an Assoc. Prof. in the Institute of Socio-Arts and Sciences, Tokushima University, Japan. His research interests include applying ICT to spatial and environmental design. He is a member of Architectural Institute of Japan and the Institute of Electronics, Information and Communication Engineers.