Calculation and Comparison of Sentimental Values for Tweeting Message during Multiple Years

Shunsuke DOI Masakazu Kyōkane Yoshiro Imai Kazuaki Ando
Koji Kagawa Yusuke Kawakami Tetsuo Hattori
Faculty of Engineering, Kagawa University, 2217-20 Hayashi-cho, Takamatsu, 761-0396, Japan
{s13t243@stmail. , imai@, ando@, kagawa@ }eng.kagawa-u.ac.jp

ABSTRACT

In this research, we have performed the following manipulations to realize the above targets; deciding to select and cheese suitable Twitter accounts who tweet interesting messages and acquire a possible mount of their relevant message, specifying and defining value for word of messages for sentimental dictionary, categorizing and listing such a calculated value for each word or expression as item of the dictionary, using these values as "Sentimental Values" for calculation of tweeting messages, calculating Sentimental Values for tweeting message from students of selected universities, checking statistical relation between the above "Sentimental Values" and published data. With results of the above calculated sentimental values of relevant messages, we have tried to visualize statistically relation between these values and some events/phenomena from the published documents.

KEYWORDS

Sentimental Analysis, Visualization of Relation, Data Mining for Twitter.

1 INTRODUCTION

In recent years, young people use social media such as Twitter. And information processing for social media can bring several kinds of effective results to us in the domains of industries, education, social science, economics, and so on. We can expect that our current environment will bring us useful facilities to analyze messages from Twitter based on Big Data Analysis and their attractive results to visualize normally hidden relations among several kinds of events and phenomena.

This study is to perform some kind of data mining as an example of visualization of sentimental words extracted from messages of Twitter and to analyze hidden or unknown relation between sentimental value and the relevant behavior/phenomena. We have defined translation scheme from words in Twitter's message into sentimental values, applied the scheme into data mining with calculation of sentimental values for tweeting messages acquired from Twitter, and demonstrated relation between sentimental values and categorized human behavior.

2 RELATED WORK

This section introduces some suitable related works to design and implement our data mining approach for message from Social Media.

Yang Yu and Xiao Wang [2] of Rochester Institute of Technology USA, collected real-time tweets from U.S. soccer fans during five 2014 FIFA World Cup games using Twitter search API and used sentiment analysis to examine U.S. soccer fans' emotional responses in their tweets, particularly, the emotional changes after goals. They found that during the matches that the U.S. team played, fear and anger were the most common negative emotions and in general, increased when the opponent team scored and decreased when the U.S. team scored. Anticipation and joy were also generally consistent with the goal results and the associated circumstances during the games. Their project revealed that sports fans use Twitter for emotional purposes and that the big data approach to analyze sports fans' sentiment showed results generally consistent with the predictions of the disposition theory when the fanship was clear and showed good predictive validity.

Apoorv Agarwal and his team [3] of Columbia

University USA, presented results for sentiment analysis on Twitter. They used previously proposed state-of-the-art unigram model as their baseline and reported an overall gain of over 4% for two classification tasks: a binary, positive versus negative and a 3-way positive versus negative versus neutral. They presented a comprehensive set of experiments for both these tasks on manually annotated data that is a random sample of stream of tweets. They tentatively concluded that sentiment analysis for Twitter data was not that different from sentiment analysis for other genres.

The article of Jeremy W. Crampton et.al. [4] presented an overview and initial results of a geoweb analysis designed to provide the foundation for a continued discussion of the potential impacts of big data for the practice of critical human geography. The principal case study focused on the widely reported riots following the University of Kentucky men's basketball team's victory in the 2012 NCAA championship and its manifestation within the geoweb. Drawing upon a database of archived Twitter activity including all geotagged tweets since December 2011they analyzed the geography of tweets that had used a specific hashtag (#LexingtonPoliceScanner) in order to demonstrate the potential application of our methodological and conceptual program. By tracking the social, spatial, and temporal diffusion of this hashtag, they showed how large databases of such spatially referenced internet content could be used in a more systematic way for critical social and spatial analysis.

Lewis, S. C., Zamith, R., & Hermida, A [5]explained Massive datasets of communication are challenging traditional, human-driven approaches to content analysis. Computational methods present enticing solutions to these problems but in many cases are insufficient on their own. They argued that an approach blending computational and manual methods throughout the content analysis process might yield more fruitful results, and draw on a case study of news sourcing on Twitter to illustrate this hybrid approach in action. Careful combinations of computational and manual techniques could preserve the strengths of traditional content analysis, with its systematic rigor and contextual sensitivity, while also maximizing the large-scale capacity of Big Data and the algorithmic accuracy of computational methods.

Kumamoto and Tanaka [6] in Japan, focused on the impressions people got from news articles, and propose a method for determining impressions of these news articles. Their proposed method consisted of two main parts; One part involved building an 'impression dictionary' that described the relationships among words and impressions. Another of the method involved determining impressions of input news articles using such an impression dictionary. The impressions of a news article were represented as scale values in user-specified impression scales, like 'sad - glad' and 'angry - pleased'. Each scale value was a real number between 0 and 1, and was calculated from the words (common nouns, action nouns, verbs, adjectives, and katakana characters) extracted from an input news article using the above impression dictionary.

3 OUTLINE OF OUR SYSTEM

This section describes our system configuration in the first half. And it shows system flow of data mining of Tweeting messages and sentimental analysis with dictionary of sentiment value in the second one.

3.1 System Configuration

Our system is configured with three major parts, namely acquisition of data from Web (Social Media), information processing scheme (data mining for acquired message from Web), and generation of documents by our system. Figure1 shows Schematic Diagram for System Configuration (another expression has been illustrated in the paper: "A Study of Visualization for Hidden Relation between Published Documents and Message from Twitter by means of Sentimental Analysis Approach [1]")

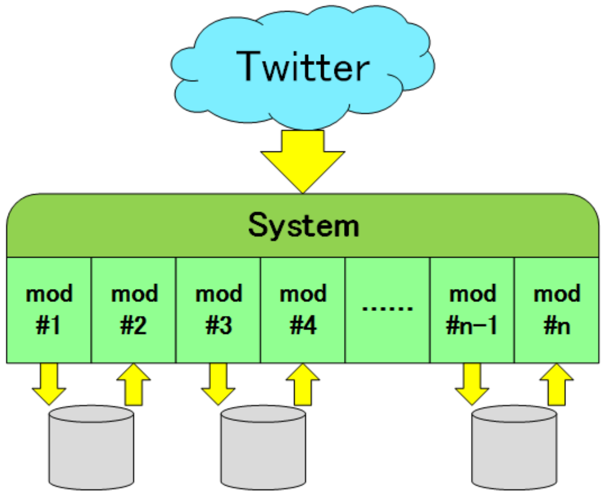

Figure 1. Schematic Diagram for System Configuration of Data Mining of Tweeting Message and Sentimental Analysis with Dictionary of Sentiment Value

3.2 Data Mining Flow of our system

With the system shown in Figure1 we can perform our sentimental analysis as one of data mining approach for tweeting messages we have previously focused on. Schematic diagram of our sentimental analysis flow can be illustrated and explained in Figure2. It is surveyed as follows;

- We and our system have focused on specific data based on speakers' attributes as target users.
- After connecting Twitter as Web-based Social Media, our system has acquired an amount of data to be analyzed with "twpro API" and "Twitter API".
- With these data, the middle part of our system performs a series of manipulations for information processing from data into generated results which include some temporary files.
- From top to bottom in Figure2, such a part called the 'Manipulation' one can continue to carry out a series of manipulations for information processing for sentimental analysis sequentially sometimes by automatically and otherwise by manually.
- This part is the main body of our system, showing "System Configuration and Processing Flow" and can accomplish some kind of sentiment analysis and generate our interesting results. We have obtained related relation between the above files based on messages from Twitter.

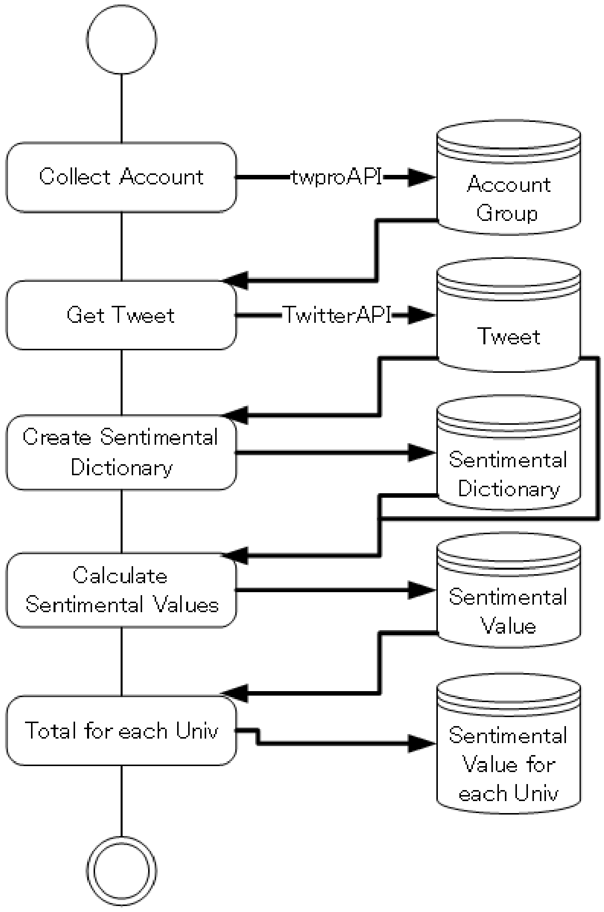

Figure 2. System Flow of Data Mining of Tweeting Message and Sentimental Analysis with Dictionary of Sentiment Value

4 SAMPLE DATA MINING RESULT

This section demonstrates a sample results of our sentimental analysis as data mining approach. It illustrates two-years dot-plotting graphs for relation between Sentimental Value and the number of a famous coffee shops called "Starbucks[1]" of the west side of Japan in 2014 and in 2015. And it compares with them.

4.1 Result for 2014

Figure3 shows dot-plotting graph for relation between calculated sentimental value of tweet-

[1] https://en.wikipedia.org/wiki/Starbucks

ISBN: 978-1-941968-40-6 ©2016 SDIWC

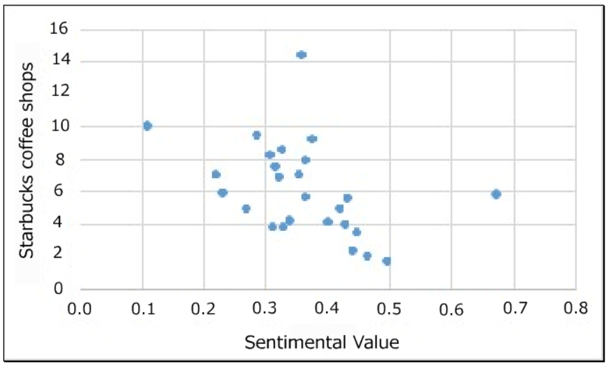

Figure 3. Plot Graph for Relation between Sentimental Value and the number of Starbucks coffee shops in 2014

ing message from the Twitter's users we have already focused and the numbers of Starbucks coffee shops in 2014. Although existing some potential noises, we can probably recognize that there had been some kind of *negative* correlation between the relevant sentimental values and coffee shop numbers in 2014.

4.2 Result for 2015

Figure4 shows dot-plotting graph for relation between calculated sentimental value of tweeting message from the *same* Twitter's users and the numbers of Starbucks coffee shops in 2015. Even if there may exist some *other types of* potential noises, we can also probably recognize that there had been some kind of *positive* correlation between the relevant sentimental values and coffee shop numbers in the case of 2015. After visualization of relation between sentimental value and the numbers of Starbucks coffee shops in 2014 as well as in 2015, we could not understand why Figure4 for 2015

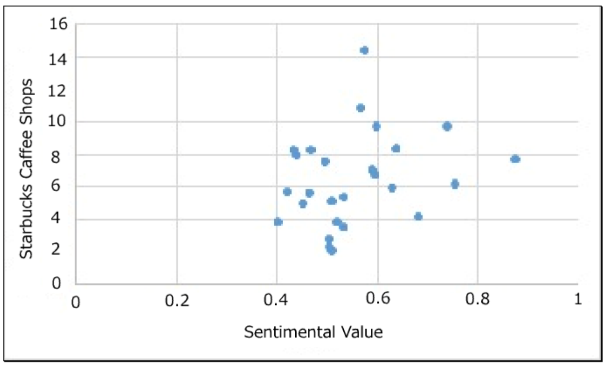

Figure 4. Plot Graph for Relation between Sentimental Value and the number of Starbucks coffee shops in 2015

was quite different from Figure3 for 2014. Of course, the number of Starbucks coffee shops increased year-by-year so it seems that such a number must be changed for each prefecture where the focused users were belonging during 2014 and 2015. But difference of numbers of coffee shops in 2014 and 2015 is not so large that it can not play a deterministic role to generate such a difference of Figure3 for 2014 and Figure4 for 2015. Therefore, we must investigate whether the trend of users tweeting messages in 2015 has differed from one in 2014 or not.

5 DISCUSSIONS ABOUT RESULTS

This section describes discussions about results of sentimental analysis. In the first half, we show comparison of sentimental values in 2014 and 2015. Secondarily, we look at result of comparison in the view of correlation of sentimental values in 2014 and 2015.

5.1 Comparison of Sentimental Values in 2014 and 2015

Table1 shows results of sentimental analysis of Tweeting messages in 2014 and 2015, where ID is Id Number of Comparison to protect the anonymity of targets. Figure5 illustrates bar chart of the results of sentimental analysis of Tweeting messages in 2014 and 2015.

Table 1. Sentimental Values of universities in 2014 and 2015 (ID is Id Number of Comparison Targets)

ID	2014	2015	ID	2014	2015
001	0.398	0.682	014	0.462	0.509
002	0.362	0.420	015	0.439	0.503
003	0.328	0.402	016	0.352	0.591
004	0.447	0.533	017	0.374	0.739
005	0.109	0.567	018	0.672	0.595
006	0.321	0.875	019	0.431	0.464
007	0.325	0.638	020	0.495	0.504
008	0.418	0.451	021	0.338	0.509
009	0.229	0.628	022	0.311	0.518
010	0.364	0.439	023	0.306	0.434
011	0.428	0.532	024	0.268	0.753
012	0.218	0.467	025	0.284	0.599
013	0.313	0.497	026	0.357	0.574

ISBN: 978-1-941968-40-6 ©2016 SDIWC

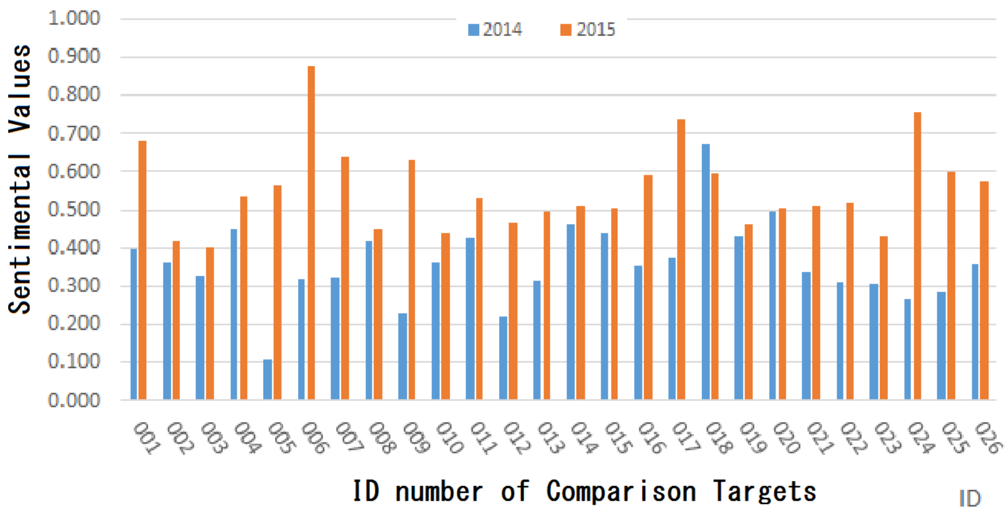

Figure 5. Bar Chart of Sentimental Value in 2014 and 2014 for each university

5.2 Discussion

We discuss consistency about our approach with sentimental analysis. About two year 2014 and 2015, we had applied our approach based on Twitter messages so that it had been assumed that sentimental analysis must generate similar sentimental values between Twitter message in 2014 and 2015. Figure6 shows relation between sentimental values from analysis for 2014 and 2015 Twitter messages.

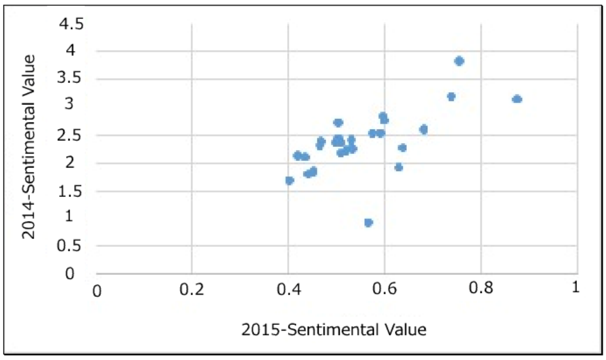

Figure 6. Plot Graph of Sentimental Values in 2014 and 2015

We can confirm that there is a certain relationship between sentimental values from 2014 and 2015 in the relevant dot-plotting graph.

6 CONCLUSION

We have tried to demonstrate and visualize normally hidden or unknown relations between messages from Twitter and information of published documents by means of sentimental analysis. For example, we have been analyzing relationship between sentimental analyzed results and several information from the above documents, such as population of district, number of governmental universities' students, number of Starbucks coffee shops, and so on. During multiple years, we recognize that it is important to confirm relationship between sentimental values year by year.

REFERENCES

[1] M. Kyokane, et al., "A Study of Visualization for Hidden Relation between Published Documents and Message from Twitter," International Journal of E-Learning and Educational Technologies in the Digital Media Vol.2, No.1, pp.217– 224, 2016.

[2] Y.Yu, et al., "World Cup 2014 in the Twitter World," Computers in Human Behavior, Vol.48, pp. 392–400, 2015.

[3] A. Agarwal, et al., "Sentiment Analysis of Twitter Data," Workshop on Language in Social Media (LSM 2011), pp. 30–38, June 2011.

[4] J. W. Crampton, et al., "Beyond the geotag: situating big data and leveraging the potential of the geoweb," Cartography and Geographic Information Science, Vol. 40, No. 2, pp. 130-139, 2013

[5] S. C. Lewis, et al., "Content Analysis in an Era of Big Data: A Hybrid Approach to Computational and Manual Methods," Journal of Broadcasting & Electronic Media, Vol. 57, No. 1, pp. 34 –52, 2013

[6] T. Kumamoto, et al., "Proposal of Impression Mining from News Articles", Lecture Notes in Computer Science Volume 3681, pp 901–910, 2005