

Self-Understanding Support Tool using Twitter Sentiment Analysis

Harumi Murakami, Naoya Ejima, and Naoto Kumagai

Osaka City University, Sugimoto, Sumiyoshi, Osaka 558-8585, Japan
harumi@osaka-cu.ac.jp
<http://murakami.media.osaka-cu.ac.jp/>

Abstract. Self-understanding, which is important for such daily aspects as decision making, requires grasping one’s strengths and weaknesses, likes and dislikes, joys and disappointments. However, gathering such useful data for self-understanding is difficult. The aim of this research is to develop a system to support self-understanding using sentiment analysis of Twitter posts. We developed a tool that displays bar graphs, tweet log lists, pie charts, and word lists that indicate the positive/negative emotions of a user’s tweets. Preliminary experiments revealed that our approach and tool are promising for self-understanding, understanding emotions, and recalling memories.

Keywords: sentiment analysis, Twitter, graph, self-understanding

1 Introduction

Self-understanding is important for various daily elements such as decision making. It requires grasping one’s strengths and weaknesses, likes and dislikes, joys and disappointments. However, it is difficult to gather such useful data for self-understanding. On the other hand, the spread of smartphones, microblogging, and social networking services simplifies the storage of personal information. Among such services, Twitter posts (called tweets) often include a user’s personal experiences and emotions to them. We believe that extracting and visualizing the emotions included in tweets will increase self-awareness. The aim of this research is to develop a system that supports self-understanding using the sentiment analysis of a user’s tweets.

We conducted our research using Japanese tweets and translated the examples in this paper from Japanese into English for publication.

2 Approach

We visualize the emotions included in user tweets for self-understanding. The Google Natural Language API sentiment analysis (Google SA) is used to get a score and a magnitude for each tweet. We display the following: several types of bar graphs that show the transition of emotions, a list of tweets with emotional

scores, a pie chart showing the proportion of emotions, and word lists included in positive and negative tweets.

Google SA calculates a score and a magnitude for a given text. The score of the sentiment ranges between -1.0 (negative) and 1.0 (positive) and corresponds to the overall emotional leaning of the text. The magnitude indicates the overall strength of the emotion within the given text that ranges between 0.0 and infinity.

2.1 Bar Graphs

We developed two kinds of bar graphs with different units: tweets and periods. The bar graph by tweet displays the scores of tweets. We implemented two periods: day and month and developed three bar graphs with different values: (a) maximal values of the scores (both positive and negative); (b) an average value of the scores; and (c) a modified average value using the scores and magnitudes. In addition, there are two modes for bar graphs, log lists, and pie charts: (1) positive and negative and (2) positive, negative, and neutral. In mode (1), a score less than 0 is negative, 0 is neutral, and greater than 0 is positive. In mode (2), a score less than -0.25 is negative, -0.25 to 0.25 is neutral, greater than 0.25 is positive, identical as Google SA. We combined three types (a)-(c) and two modes (1)-(2) in the bar graphs. For example, bar graph (a)-2 displays the maximal positive, negative, positive-side-neutral, and negative-side-neutral scores using three colors.

Graphs (a) and (b) have individual advantages and disadvantages. Graph (a) emphasizes the highest valued tweets but hide other tweets in the designated period. Graph (b) summarizes the values in one value, but offsets the plusses and minuses and the values tend to be small, complicating the detection of tweets with high emotion content. To overcome the disadvantage of graph (b), we designed graph (c), which emphasizes stronger emotions using a modified average value (an *E-score*):

$$E\text{-score} = \begin{cases} \frac{\sum_{i=1}^n |Score_i \times Magnitude_i|}{\sum_{i=1}^n Magnitude_i} & \text{if } \sum_{i=1}^n Score_i \geq 0, \\ -\frac{\sum_{i=1}^n |Score_i \times Magnitude_i|}{\sum_{i=1}^n Magnitude_i} & \text{otherwise,} \end{cases}$$

where $Score_i$ is a Google's score for tweet t_i , $Magnitude_i$ is the magnitude for tweet t_i , and n is the number of tweets except for those whose scores are 0.

2.2 Tweet Log Lists, Pie Charts, and Word Lists

Tweets can be sorted by date or by score. A pie chart shows the ratio of the positive/negative/neutral scores of the tweets.

The tool displays the top five most frequent word lists extracted from the positive and negative tweets. First, we performed morphological analysis (MeCab was used for Japanese) and assigned the score of one tweet to the words contained in it. For each positive and negative category, the summation of the word's score is divided by its frequency.

2.3 Example

Figure 1 shows an example of a user’s monthly usage. After he chose the period (1/Jun/2018 - 30/Jun/2018) and selected “Per day: A” with “displaying neutral,” a bar graph (a)-2, a tweet log list, a pie chart, and positive/negative word lists are displayed. In this period, he enjoyed gaming and watching soccer and baseball. The pie chart shows that 42.9% (9/21) of the tweets were positive, 19% were negative (4/21), and 38.1% (8/21) were neutral. On June 18, 2018 (JST), he posted a very positive tweet about Mexico’s FIFA World Cup victory and a very negative tweet about a big earthquake (2018 Osaka earthquake). The positive word list includes *Japan*, *Mexico*, *Fujinami* (baseball player), and *this season*, all of which are related to soccer and baseball. The negative word list includes *aftershocks*, related to the earthquake, and *Columbia*, related to soccer.



Note: The colors (red: positive, blue: negative, and green: neutral) are based on preliminary investigation for Japanese people. Although we chose these colors based on Japanese preferences, they can be tailored to satisfy specific cultural requirements.

Fig. 1. Example of screen with bar graph (a)-2 per day

3 Preliminary Experiments

We conducted preliminary experiments to evaluate our approach and the tool and recruited subjects who satisfied the following conditions: (a) have used Twitter more than one year, and (b) post more than 15 tweets per month. The subjects were four Japanese males (average age, 22.0 years) in Experiment 1 and five Japanese males (average age, 22.8 years) in Experiments 2 and 3.

3.1 Experiment 1: Google’s Scores

We verified the validity of Google’s scores for the tweets. The subjects evaluated their emotions of 50 recent tweets by five values (1: very negative; 5: very positive). Two types of correlation analysis (Kendall and Spearman) were performed between Google’s scores and manual scores. The average values were over 0.2 in both types of analysis (0.242 in Kendall and 0.313 in Spearman). We believe Google’s scores for tweets are promising for our approach.

3.2 Experiment 2: Bar Graphs

We evaluated the usefulness of daily bar graphs. In Experiment 2, not the whole tool but only bar graphs only were displayed. After the subjects described their positive and negative experiences for a month, they operated three bar graphs, (a)-2, (b)-2, and (c)-2, to describe those monthly experiences again and answered the following questions: “Was the bar graph A (or B or C) useful for understanding your emotions?” by five values (1: very unuseful; 5: very useful). The results of the average values were 4.2 (A), 3.6 (B), and 3.8 (C). We also asked subjects to rank the graphs: “Which graph was useful for understanding your emotions?” All subjects ranked graph A the best. Graph B was ranked the worst by four of the five subjects. Both for the average values and by ranked scores, graph A was best, then graph C, and graph B. The bar graphs per day were useful, and bar graph A was the most useful among the three.

Next we evaluated the usefulness of the bar graphs per month. Subjects again operated three bar graphs, (a)-2, (b)-2, and (c)-2, for six months and answered the same questions as for the daily graphs. The results for “Was the bar graph A (B or C) useful for understanding your emotions?” were 3.4 (A), 2.0 (B), and 3.0 (C). Three subjects ranked graph C first and two ranked graph A first. Graph B was ranked worst by all the subjects. Among the three graphs, graph (a) was best for the average values and graph (c) was best for ranking. Bar graph (b) per month is not very useful.

From the results of Experiment 2, we identified the usefulness of bar graph A per day and per month. Since graph C is superior to graph B, it compensates for the latter’s disadvantage.

3.3 Experiment 3: Tool

Finally, we evaluated the utility of our entire tool. After using it, our subjects described the most positive and negative events for a month and answered the questions shown in Table 1.

The average values exceeded 4.0 except for Q3 and Q4. The bar graphs and the log lists were useful, but the words were less useful. The following are some comments from the subjects: “I had fun during the experiment,” “It was fun because I recalled some good memories,” “I recollected past memories.” “It might be useful for looking back at the end of the year.” We believe that our approach and tool are promising for self-awareness and emotions and remembering.

Table 1. Results in Experiment 3: Tool

Questions	Mean	SD
Q1 Was the bar graph useful for understanding your emotions?	4.0	0.00
Q2 Was the tweet log useful for understanding your emotions?	4.4	0.55
Q3 Were the words useful for understanding your emotions?	2.8	0.84
Q4 Was the pie chart useful for understanding your emotions?	3.0	0.70
Q5 Was the system useful for understanding your emotions?	4.4	0.55
Q6 Was the system useful for recalling your memories?	4.4	0.55
Q7 Was the system useful for understanding yourself?	4.0	0.71
Q8 Was using the system fun?	4.8	0.45
Q9 Do you want to use the system in the future?	4.6	0.55

4 Related Work

This research is a part of our work on storing personal information. Using Twitter data, Murakami et al. [1] developed a knowledge-space browser that displays a network and another work [2, 3] developed a tag browser that displays a tag cloud to support human memory recollection.

Much research assigns sentiment values to texts, some of which are publicly available. Google SA is the most popular such service. Our preliminary investigation tried some services and research and decided to use it for getting initial scores. We presented three types of scoring for the bar graphs: maximal, average, and modified average.

Other research also did sentiment analysis and visualization from tweets. Kumamoto et al. [4] proposed a web application system for visualizing Twitter users based on temporal changes in the impressions from tweets. Mitsui et al. [5] developed an emotion lifelog system called Emote that classifies the emotions in Twitter and Facebook texts into six categories and visualizes them on a calendar. Wang et al. [6] presented a system for the real-time analysis of public sentiment toward presidential candidates in the 2012 U.S. election from Twitter using graph representations. Applications are available for sentiment analysis and visualization from tweets, such as sentiment viz [7]. Our research’s purpose, method of calculation, and interface are different.

The following are the three main contributions of this paper: (a) presented an approach for using sentiment analysis from tweets for self-understanding; (b) presented a prototype tool; and (c) described how our approach and tool are promising for understanding self and emotions and recalling memories.

Giachanou and Crestani [8] surveyed sentiment analysis on Twitter and discussed fields related to sentiment analysis on it. They didn’t address our task: self-understanding.

5 Summary

We developed a tool that displays bar graphs, tweet log lists, pie charts, and word lists using Twitter sentiment analysis for self-understanding support. Preliminary experiments revealed that our approach and tool are promising for self-understanding, understanding emotions, and recalling memories. We need in the future to improve the algorithms that calculate the modified average values and extract positive/negative words and the tool, especially such interfaces as selecting bar graphs. We also need to evaluate the tool with more subjects and both genders.

References

1. Murakami, H., Mitsunashi K., Senba K.: Creating User's Knowledge Space from Various Information Usages to Support Human Recollection. In: Jiang H., Ding W., Ali M., Wu X. (eds.) IEA/AIE 2012. LNCS, vol. 7345, pp. 596-605. Springer, Heidelberg (2012) doi:10.1007/978-3-319-60045-1_61
2. Matsumoto, M., Matsuura, S., Mitsunashi, K., Murakami H.: Supporting Human Recollection of the Impressive Events using the Number of Photos. In: Proceedings of the 6th International Conference on Agents and Artificial Intelligence - Volume 1: ICAART, pp. 538-543. (2014)
3. Murakami, H., Murakami, R.: A System using Tag Cloud for Recalling Personal Memories. In: Wotawa, F., Friedrich, G., Pill, I., Koitz-Hristov, R., Ali, M. (eds.) IEA/AIE 2019. LNCS, vol. 11606. pp. 398-405. Springer, Cham (2019). doi:10.1007/978-3-030-22999-3_35
4. Kumamoto, T., Wada, H., Suzuki, T.: Visualizing Temporal Changes in Impressions from Tweets. In: Proc. iiWAS 2014. pp. 116-125. ACM Press, New York (2014). doi:10.1145/2684200.2684279
5. Mitsui, T., Ito, T., Nakanishi, H, Hamakawa, R.: Emote: Emotion Lifelog system using SNS, IPSJ SIG Technical Report, 2014-EC-32, 1, pp. 1-6. (2014) (in Japanese)
6. Wang, H., Can, D., Kazemzadeh, A., Bar, F., Narayanan, S.: A system for real-time Twitter sentiment analysis of 2012 U.S. presidential election cycle. In: Proc. ACL 2012, pp. 115-120. (2012).
7. Healey, C.: Tweet Sentiment Visualization App. https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/.
8. Giachanou, A., Crestani, F.: Like It or not: A Survey of Twitter Sentiment Analysis Methods, ACM Computing Surveys, 49, 2, No. 28. ACM Press, New York (2016). doi:10.1145/2938640