Seoul Nat'l Univ.

Team Weekly Tent

Sawtooth Bubble Chart

Moojin Chae, Younghoon Lee, Joonhwan Lee

Department of Electrical and Computer Engineering / Department of Communication, Seoul National University

Introduction

The development of the text mining algorithm increases the attempts to extract and visualize human emotions expressed on the web or Social Networking Service (SNS) [2]. However, the visualization of human emotions conducted by previous researchers was limited because many researchers only used the scales of 'positive' and 'negative.'

The aforementioned methodology limits researchers to effectively indicate various human emotions. For example, 'excited' and 'serene' might be recognized as the same emotional status, even though they are, in fact, different emotional status. To solve such problems, James Russell proposed multidimensional model of human emotions. In the model, X-axis was designed to measure emotions from pleasant to unpleasant; Y-axis was designed to measure emotions from activation to deactivation.

All emotional words were matched with words in the emotional dictionaries, such as POMS (Positive of Mood States) or ANEW (Affective Norms for English Words), to digitize the levels of emotional status [3]. Those two dictionaries contain more than thousand words that measure the levels of valence and arousal with the scale (1 to 9).For example, when we find the word 'house' in ANEW, the valence and arousal levels of the word are 7.26 and 4.56, which indicate that 'house' is the word that expresses the feeling of pleasant. Thus, Russell's developed methodology enables researchers to digitize various human emotional status.

This study first utilized the emotional data that was digitized based on Russell's methodology. After that, it visualized the data using sawtooth bubble chart.

Sawtooth Bubble Chart

Sawtooth Formation by Valence Level

There are two steps for sawtooth formation. First, the shape of sawtooth should be decided. After that, the number of sawtooth on bubble should be decided. The shape of sawtooth indicates the valence levels (from unpleasant to pleasant). For example, the round shape of sawtooth represents pleasant; the sharp shape of sawtooth represents unpleasant; no sawtooth (only bubble) represents neutral. The height of sawtooth is one fifth of the radius of each bubble. Such the ratio is fixed (not changed by the size of bubbles) and does not make the area of sawtooth exceed the 25 percents of the area of bubbles. The next step is to decide the number of sawtooth on a bubble. As indicated in Figure 1, the number of sawtooth increases as the pleasant level increases.

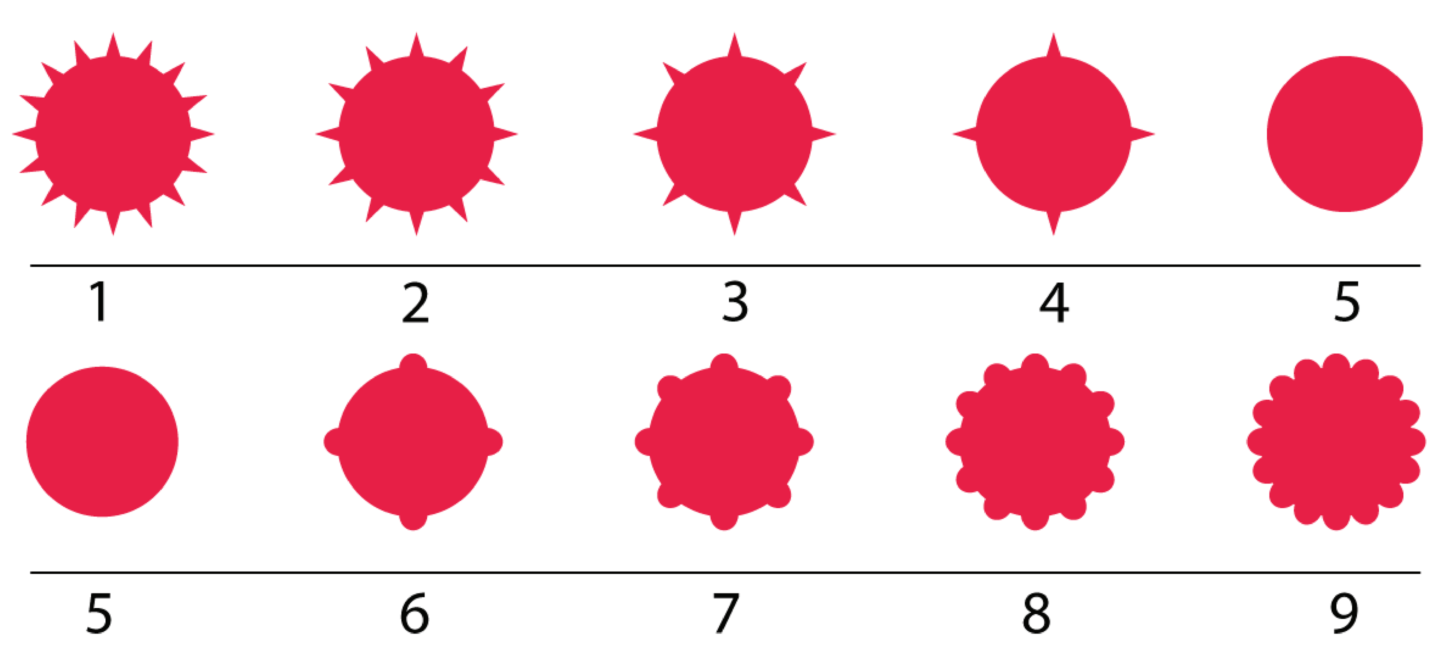

Figure 1. Sawtooth formation by valence level

Brightness Coding by Arousal Level

The colors of bubbles represent arousal levels. If the color of a bubble is dark and solid, it means they are deactivated while it means activated if the color of a bubble is light and less solid. Of course, the colors of bubbles can be controlled as indicated in Figure 2.

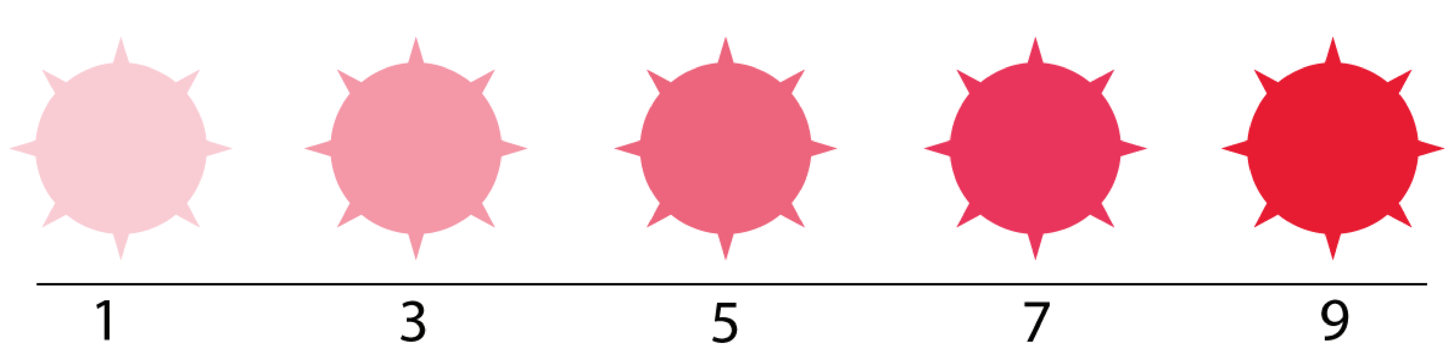

Figure 2. Brightness coding by arousal level

Applications

There are two possible applications of this methodology using twitter data. In both applications, the radius of each bubble indicates the number of tweets. All words in tweets were matched with words in ANEW dictionary in order to get each word's valence level and arousal level.

Time Series Twitter Data

In this application, data used for visualization were tweets which have '#worldcup' hash tag and were mentioned in June 23, 2014 from 1am to 11am in Korea. There was a football game between Algeria and Korea in 4am, and Algeria beat Korea. Extracted data show that arousal level of people rose between one hour before the game and one hour after the game. This result presented as lightened bubbles. Next, valence level scored high before the game, but the level dropped after the game because of Korean football team's defeat.

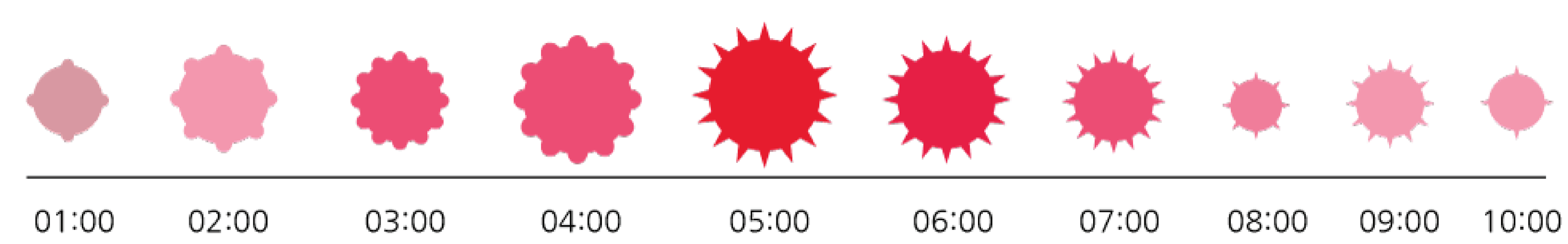

Figure 3. Visualization of tweets about '#worldcup' hash tag

Twitter Data Categorized by Age Group and Geolocation

This visualization application shows tweets that mentioned about "GeunHye Park"(the president of Republic of Korea) in May, 2014 and categorized by age group and geolocation. Arousal level scored high in the younger age group, and valence level scored low in the 'Gwangju. This result indicates that human emotions cannot be divided by unidimensional classification, but can be divided by multidimensional classification using valence level and arousal level.

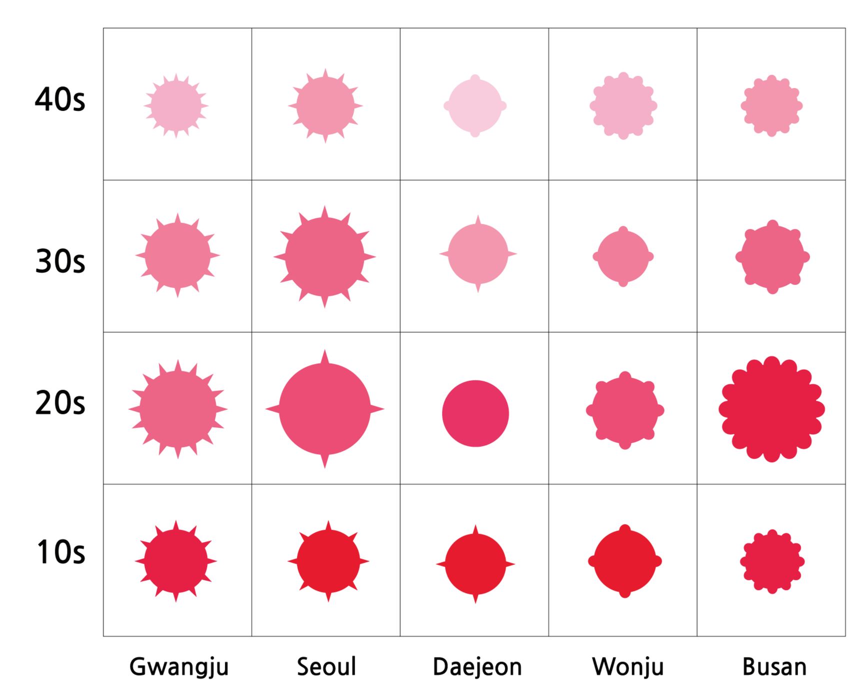

Figure 4. Visualization of tweets about "GeunHye Park"

Conclusion & Future Work

As the number of users of Social Networking Service rises, the demands for visualizing human emotions are increasing. Before this study, most of visualization methods were based on unidimensional classification of human emotions. However, extracting detailed human emotions are getting easier thanks to the development of the text mining algorithm. Therefore, this study proposes new visualization methodology which can express multidimensional human emotions more effectively. In future, studies for finding more applications to verify the effectiveness of this method are needed.

References

[1] James Russell. A circumplex model of affect. Journal of personality and social psychology, 39.6:1161, 1980.

[2] Ming Hao, et al. Visual sentiment analysis on twitter data streams. Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on. IEEE, 2011.

[3] M. Bradley and J. Lang. Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida, 1999.

Contact

chaesocool@gmail.com, younghoonlee.89@gmail.com, joonhwan@snu.ac.kr