Relationship Between Emotional Words and Emoticons in Tweets

Kaito Soranaka
Kansai University
Osaka, Japan.
s_kaito_k@hotmail.com

Mitsunori Matsushita
Kansai University
Osaka, Japan.
mat@res.kutc.kansai-u.ac.jp

Abstract—Asynchronous communication using text messaging is a major mode of online communication. It is simple and easy to use; however, there is often an inconsistency between the sender's intended tone and how the recipient perceives it. Emoticons, additional textual expression using icons for facial expressions, are often used to supplement or adjust the verbal part of the text, though the problem persists. The goal of our research is to solve the problem by developing a system that supports fluent communication. The system would estimate the emotions in the sender's text and note whether different intentions may be conveyed. For that purpose, we analyzed concurrences between emotional words and emoticons in a text. We observed the following: (1) There are cases when the emotion represented in words and the emotion represented by the emoticon are inconsistent; (2) The expressed emotion can change between positive and negative emotion according to the co-occurring words.

Keywords: component; Twitter; text mining; emotional words; emoticon

I. INTRODUCTION

In recent years, IT tools such as e-mail, blogs, and social networking services (SNSs) are becoming primary media for personal communication. The major modality in such computer-based media is text. However, such the text-based communication often results in an emotional inconsistency between a sender and recipient, compared to face-to-face communication (Kato [3]).

A communication modality can be classified as one of two types: verbal or nonverbal. Verbal communication conveys content to a partner by text and utterance using words. Nonverbal communication, in contrast, relays feelings and intentions using such means as gestures and postures. The problem with textual communication arises because one can only use verbal communication channels.

To overcome this problem, people often uses an "emoticon," which is a sequence of characters such as :-) or (^_^) , that are intended to represent a facial expression to convey an emotion. Within the text medium, such emoticons make up for nonverbal information, but it is still difficult to read the feelings that the user wants to convey. The primary reason is that the meaning of the emoticon often changes according to personal interpretation (Murakami [5]). In brief, to estimate the feelings of the user and communicate smoothly requires a context-dependent understanding of the relationship between emotional words and emoticons.

The goal of our research is to estimate the emotion of the sender with the aim of creating a support system that would reduce inconsistency between the intention conveyed and the intention as understood by the addressee.

For that purpose, this paper tries to demonstrate an understanding of the emotional words and the emoticon of the user as the foundation for such a system. In this paper, we focus on Twitter tweets and analyze words that create or convey emotion and the emoticons that appear in the tweets.

II. RELATED WORK

Several studies have been conducted to understand the relationship between emotions and emoticons. For example, Kawakami created a database of emoticons [4]. He then evaluated emoticons that were preregistered on a cell phone and created a database of emoticons classified according to six emotions: joy, anger, sadness, pleasure, surprise, and anxiety. However, his experiment simply assessed the emotions of the emoticons, not how they change in relation to the context.

In addition, Emura et al. built a system to recommend emoticons based on their reflection of emotional communication as it relates to the context of the user's sentence [2]. The system assumed a keyword for each emoticon and analyzed the sentence containing the emoticon. Their system estimated the emotion based on the context in which it was applied in a user's sentence and then recommended an appropriate emoticon. On a questionnaire, participants using this system stated that the negative sentence read more precisely than the positive sentence because of the emoticon recommendation. However, the system needs to improve its precision by increasing entries in the emoticon clue word dictionary based on an expanded database. In addition, adaptation to a new emoticon might pose a problem since the dictionary assumes one particular emoticon for a keyword.

Cui et al. aimed to make a dictionary with a polar structure and positive and negative emoticons based on the idea that the emotion of an emoticon is collocated with the emotion of the tweet; emoticons are frequently used in tweeting [1].

These studies aim to estimate intended feelings from fixed, simple words, or fixed simple emoticons. However, they may fail to identify the correct feeling when an emoticon is used with words that express a different feeling. Moreover, the meaning of a particular emoticon is often contextual (Kawakami [4]). Therefore, it is necessary to

analyze the collocation of words and emoticons to estimate an intended feeling from a sentence.

III. ANALYSIS

Emotion is generally thought to be classifiable into four types: joy, anger, sadness, and pleasure. According to Ekman, however, the psychology of emotion can be classified into six types: joy, anger, sadness, pleasure, surprise, and anxiety [7]. Taking it further, the *Dictionary of Emotional Expression* includes 10 types: joy, anger, sadness, fear, shame, like, hate, excitement, calm, and surprise (Nakamura, [6]). In this way, the classification of emotion is vague, making it difficult to assign one-to-one classifications. Therefore, this study unified the kinds of emotional words and emotions of the emoticons using the *Dictionary of Emotional Expression* (Nakamura, [6]) and a study by Kawakami [4] to classify emotion.

A. Targets for Analysis

At first, we chose 20 emotional words at random, belonging to five categories (joy, anger, sadness, surprise, and excitement) using the *Dictionary of Emotional Expression* (Nakamura [6]). We did this with synonyms in each category of emotion as shown in TABLE I.

Then we chose 30 emoticons belonging to six emotional categories (joy, anger, sadness, pleasure, surprise, and anxiety) using a precedent study on the five most frequently used emoticons in each category. However, because few identical emoticons were used in the classifications joy and pleasure, we unified these under joy. As a result, we defined 25 known emoticons as related to known emotions. TABLE II displays the known emoticon meanings in these categories (where bracket A shows the repetition of emoticons for joy and pleasure). In addition, we treated the word "anxiety," classified in the study documents (Kawakami [4]), as the same emotion as excitement, as given in the *Dictionary of Emotional Expression.* This study defined it as a class of emotion to treat joy (pleasure), anger, sadness, surprise, and anxiety (excitement) as found in the *Dictionary of Emotional Expression* (Nakamura [6]) and the study by Kawakami [4].

We built a data set to analyze the collocation of emotional words and emoticons using emotional word relationships as defined above. We collected tweets that included emotional words and built the data set, which ultimately employed five emotional categories.

TABLE I. Target of Emotional Words (Nakamura [6])

Joy	嬉しい (glad)	幸せ (happy)	楽しい (pleasant)	満足 (satisfaction)
Anger	怒り (anger)	腹立つ (get angry)	不愉快 (unpleasant)	むかつく (piss off)
Sadness	悲しい (sad)	泣く (cry)	号泣 (wailing)	しくしく (sob sob)
Surprise	驚く (surprise)	ショック (shock)	意外 (unexpected)	動揺 (unrest)
Excitement	焦る (anxiety)	緊張 (strain)	興奮 (excitement)	うずうず (sorely tempted)

TABLE II. Target of Known Emoticons (Kawakami [4])

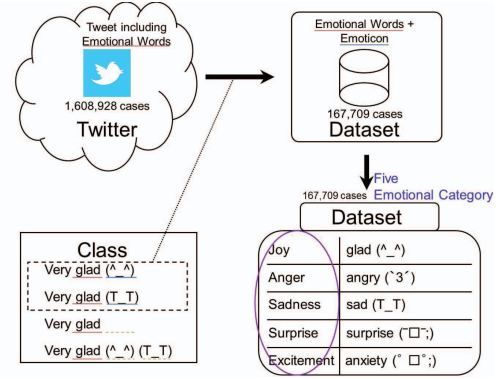|					
Joy	(≧▽≦)	(oˆoˆo)	(>∀<)	(ˆ▽ˆ)	(ˆoˆ)／
Anger	(｀Д´)	(｀ヘ´)	(- -#)	(＿＿#)	(＿□＿;)
Sadness	(;_;)	(T_T)	(ToT)	(つД｀)	(>_<)
Pleasure	(≧▽≦)	(oˆoˆo)	(ˆ▽ˆ)	(>∀<)	(ˆoˆ)／
Surprise	Σ(＿＿)	(￣Д￣;)	(￣□￣)	(￣□￣;)	(￣□￣)
Anxiety	(￣Д￣;)	(＿□＿;)	(￣□￣;)	(ˆ_ˆ;)	Σ(￣□￣)

Figure 1. Dataset collection

B. Construction of the Data Set

From tweets in three sections that included emotional word relationships, we collected 1,608,928 tweets from November 2, 2011, to January 6, 2012, using a search feature of Tween (http://www.tweenapp.org/en/), a Twitter client. Among the collected tweets, we thought that the emotional words were more likely to correspond to the emoticon if they were in close proximity. Therefore, for this analysis we chose tweets in which an emoticon came within three characters before or after emotional words. In addition, we prevented plural types of emotional word compositions by limiting the number of emoticons in a tweet to one. As a result, we targeted 167,709 tweets for analysis of five kinds of emotion —joy, anger, sadness, surprise, and excitement—based on emotional words, and we decided to treat them as a five-part data set. The structural procedure is shown in Figure 1.

C. Analysis of the Collocation

We analyzed the collocation of emotional words and the emoticon using our five-part data set. The words that belong to joy collocated with a known emoticon to express sadness. We accounted for the possibility that the emotion of the words and the emoticon in one sentence might not agree. Therefore, we analyzed the emotion of emotional words and the known emoticon based on a viewpoint of agreement.

At first, we extracted a tweet with a known emoticon from each of the five parts of the data set to analyze the collocation of emotional words and emoticons. We then checked the appearance ratio in every category of emotion words to collocate. Then we determined whether the emotion that the emoticon expressed agreed with the emotion that the words expressed. We display the appearance ratio of known emoticons co-occurring with emotional words in Figure 2. Some common emoticons such as (￣□￣;) are included in plural emotional categories in
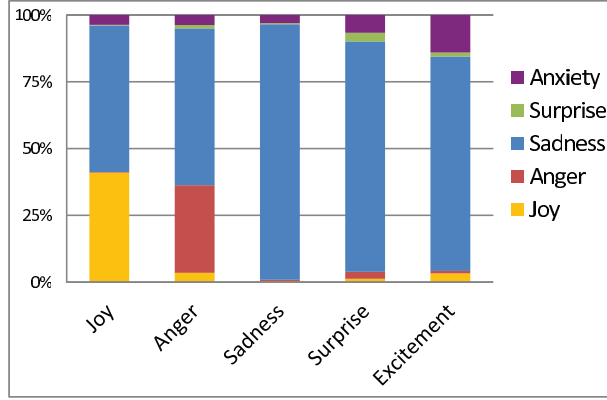
Figure 2. Emoticons known to co-occur with emotional words

TABLE I if, for instance, they appeared in relation to both anger and anxiety. One can observe that a known emoticon for sadness (a figure in blue) often collocates with the word in that emotional category, as shown in Figure 2.

We predicted that emotional words and the known emoticon to collocate with them might not belong to the same emotional category. As a result of analysis of the uses of emoticons known to express sadness, 54.9% of their uses collocated with emotional words belonging to the category of joy, 58.8% with anger, 95.7% with sadness, 86.3% with surprise, and 80.3% with excitement. In short, it was suggested that an emoticon collocates with emotional words that belong to categories other than sadness, as addressed in TABLE II.

Next, we analyzed how an emoticon collocated with emotional words belonging to a particular category. We extracted the 50 emoticons that appeared most frequently in each of the five parts of the data set, 117 units total, and settled upon a corresponding emotional category to develop each emoticon. As a result, it was confirmed that 13 emoticons (11.1%) collocated with five emotional words; 14 emoticons (12.0%) collocated with four; 9 emoticons (7.7%) collocated with three; 22 emoticon (18.8%) collocated with two; and 59 emoticons (50.4%) collocated with only one. In short, it was suggested that half of the 117 emoticons that collocated with emotional words varied according to context. In addition, of the emoticons to collocate with two emotional words in TABLE III, nine emoticons were found in the category of joy and excitement, and nine emoticons collocated with either sadness or surprise. However, none of the emoticons collocated with both joy and surprise or sadness and excitement. TABLE IV shows one emoticon collocated with both joy and excitement. Furthermore, we observed that the emoticons that express joy and anger collocated with the emotional words for surprise and excitement, as shown in Figure 2. As a result, there were fewer emoticons to express joy in the context of surprise than to express anger. There were also more emoticons to express joy in the context of excitement than to express anger. In short, we found that emoticons for surprise co-occur close to negative expressions, and emoticons for excitement co-occur close to positive expressions.

D. Construction of the Data Set

Based on the results in the foregoing paragraph, we analyzed the polarity of positive and negative emoticons collocated with emotional words for surprise and excitement. We observed that the range of emoticons close to words such as "unexpectedness" or "shock" was strange.

As content analysis, we checked the appearance of the 10 most frequently used emoticons out of 20 that collocate with words of surprise and excitement; this resulted in the classification of eight emotional words in total. Figure 3 graphs the appearance of this ratio. For example, the ratio of the word "unexpectedness" is low, and the ratio of the word "shock" is high in F of the graph, but the ratio of unexpectedness is high, and the ratio of shock is low in G of the graph. In addition, the ratio of emoticons to co-occur with words such as excitement or "strain" was a transition of a ratio of words, as seen in Figure 4. For example, the ratio of the word excitement is small, and the ratio of strain and anxiety are big in Q of the graph, but the ratio of excitement is big and the ratio of strain and anxiety are small in G of the graph. In short, this suggests that there are differences in the words that co-occur, even when an emoticon falls in the same emotional category.

IV. EXPERIMENTS

In Chapter III, we hypothesized that the emotion of the emoticon did not merely express the same emotion as the sentence, that it might express a different emotion relative to its context. From there we performed an analysis based on sections C and D.

It was observed that the emoticon did not necessarily express one emotion and that the emotion of the emoticon changed contextually. Thus, it is conceivable that plural emotions are expressed with one emoticon. On the other hand, someone may emphasize an emotion by using the emoticon in addition to the co-occurring emotional words or may relax the emotion by using an emoticon expressing a different emotion than the words. In the latter case, how the user wants the emotion to be understood can become a problem.

TABLE III. Cases of Emoticon Collocation with Emotional Words

Emoticon to collocate with one Emotional Word	Emoticon to collocate with two Emotional Words	Emoticon to collocate with three Emotional Words	Emoticon to collocate with four Emotional Words	Emoticon to collocate with five Emotional Words
(^▽^)	(*^o^*)	(´ω｀)	(ToT)	(^-^)
(ノдＣ。)	(￣◇￣;)	(;o;)	(; ´Д｀)	(;_;)
(￣^￣)	(￣—￣)	(^_^;)	(-_-;)	(^o^)
(◎_◎;)	(*´д｀*)	(^ω^)	(´ ; ω ; ｀)	(>_<)
(^◇^;)	(´△｀)	(^_^)	(・∀・)	(^q^)

TABLE IV. Cases of Emoticon Collocation with Two Emotional Words

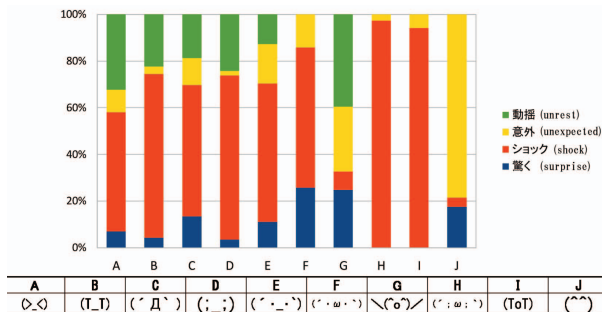| 「Joy」and 「Excitement」	(*´ω｀*)	(*´▽｀*)	(*^o^*)	(≧∇≦)	(*^_^*)
	(//∇//)	(´∀｀)	(*´д｀*)	(´ ‖ ｀)	
「Sadness」and 「Surprise」	(/_;)	(T^T)	(´△｀)	(´°ω°｀)	(ノД｀)
	(/_;)	(;_;)	(T．T)	(TT)	

264

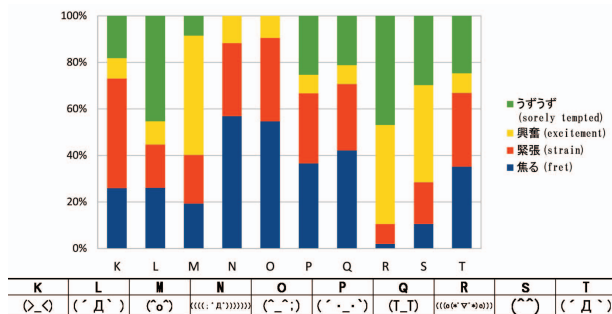Figure 3. Ratio of emoticons known to collocate with words for surprise

Figure 4. Ratio of emoticons known to collocate with words for excitement

V. CONCLUSION AND FUTURE WORK

This study aimed to understand emotional words and emoticons used in tweets. We confirmed that they could contextually collocate with different emotions. We also confirmed that the emoticons that collocate with emotional words might change into either positive or negative emotions, particularly emoticons that collocate with words in the surprise and excitement categories.

For future work, we will further investigate the emotional gap between emoticons and emotional words. Thereafter, we will try to develop an algorithm to recognize a user's intended emotion from his/her tweet, and then recommend an alternative emoticon more appropriate to the tweet, to reduce miscommunication.

REFERENCES

[1] Cui, A., Zhang, M., Liu, Y., & Ma, S.: Emotion Tokens: Bridging the Gap Among Multilingual Twitter Sentiment Analysis (Proc. 7th Asia Information Retrieval Societies Conference), pp. 238-249 (2011).

[2] Emura Y., & Seki Y.: Facemark Recommendation based on Emotion, Communication, and Motion Type Estimation in Text, The Information Processing Society of Japan, Vol. 2012, No. 1, pp. 1-7 (2012).

[3] Kato, Y., Kato, S., Sugimura, K., & Akahori, K.: The influence of affective traits on emotional aspects of message receivers in text-based communication: Examination by the experiment using e-mail communication, Educational Technology Research, Vol. 31, No. 1, pp. 85-95 (2008).

[4] Kawakami, M.: Database about emotion and the emphasis that an emoticon expresses (in Japanese), Osaka Shoin Women's University, The Human Science Research Bulletin, Vol. 7, pp. 67-82 (2008).

[5] Murakami T.: Study of the communication in the virtual community by the investigation into emoticon (in Japanese), Japanese Society for the Science of Design, pp. 62-63 (2006).

[6] Nakamura A. (Ed.): Dictionary of Emotional Expression (in Japanese), Tokyodo Publishers (1979).

[7] Paul Ekman, P., & W. V. Friesen, W. V.: Unmasking the face: A guide to recognizing emotions from facial clues. Englewood Cliffs, NJ: Prentice Hall. (1975).