A Method for Automatically Generating the Emotional Vectors of Emoticons Using Weblog Articles

SHO AOKI
Graduate School of Engineering
Tokai University
4-1-1 Kitakaname, Hiratsuka, Kanagawa 259-1292
JAPAN
9adrm002@mail.tokai-u.jp

OSAMU UCHIDA
Dept. of Human and Information Science
Tokai University
4-1-1 Kitakaname, Hiratsuka, Kanagawa 259-1292
JAPAN
o-uchida@tokai.ac.jp

Abstract: - In recent years, reputation analysis services using weblogs, message boards, and community web sites have been developed. To improve the accuracy of the reputation analysis, we have to extract emotions or reactions of writers of documents accurately. Emoticons (*Emoji*s) are often used in weblogs, and in many cases these emoticons have the role of modalities of writers of weblog articles. That is, to estimate emotions represented by emoticons is important for reputation analysis. In this study, we propose a method to create emotional vectors of emoticons automatically using the collocation relationship between emotional words and emoticons which is derived from many weblog articles.

Key-Words: - Weblog, Modality analysis, Emoticon, Emoji, Emotional word, Emotional vector

1 Introduction

In recent years, there has been a rapid spread of media such as weblogs, message boards, and social network services (SNSs), which make it possible for individuals to more easily generate information. These are collectively referred to as "consumer-generated media," and the numbers of such media users are growing at an explosive rate. For example, Facebook [1], a leading SNS, has more than 500 million users worldwide. Concurrent with the growth of SNSs, has been efforts to develop technology that could analyze such user-generated content for useful applications. One such tool is reputation analysis, which is a Web mining technology used to analyze sentences posted to media (such as weblogs and SNSs) to determine if the sentence writer has positive or negative impressions of particular a product. To accomplish this, reputation analysis utilizes natural language processing, and typically analyzes the emotional words that appear in the text of such online media.

In this study, we focus on the emoticons ("*emoji*s" in Japanese) that appear in sentences posted to weblogs and SNSs. In Japan, emoticons have long been used with mobile phone email services and have spread widely into other media formats. Because emoticons effectively communicate the emotions (modality) of the writer, many users have become accustomed to their use when posting daily writings on weblogs and other SNSs from their personal computers. By determining the writer's emotions, as expressed by the emoticons that exist in sentences posted to weblogs and SNSs, it is possible to improve the accuracy of Web mining technologies such as reputation analysis. Accordingly, we propose a methodology that automatically generates the emotional vector of emoticons. To accomplish this, we collected a large volume of weblog articles containing emoticons and analyzed the co-occurrences of emoticons and emotional words to determine how the respective emoticons are used to facilitate emotional

expression. In this study, we set 14 emotional vector dimensions, using Plutchik's emotion model [2] as a reference.

2 Existing study

2.1 Study on emoticons

In Japan, especially among younger users, emoticons (*emoji*s) provide a way to express emotions that cannot be adequately communicated in words. For example, the sweat-drop emoticon shown at the left in Fig. 1 can be used to express a wide range of emotions including embarrassment, indignation, quandary, and shock. Meanwhile, emoticons that express happiness include the heart symbol and face symbols that show happy expressions.

There have been a number of studies on emoticons in recent years. For example, Yamamoto et al. [3] proposed a method for emoticon disambiguation. Hagiwara and Mizuno [4] proposed an information retrieval method for mobile phones that enables the use of emoticons in search queries. Yamashita et al. [5] discussed applications for using the emotional vector of emoticons to deduce the mind state of email users, and for use in music retrieval systems.

2.2 Analysis of emotions and impressions from text

There has also been a great deal of study on extracting emotions and impressions from text for purposes such as information retrieval and reputation analysis. Shimizu and Hagiwara [6] proposed a method for estimating impressions based on the frequency of joint word co-occurrences in texts published on the World Wide Web. Kumamoto and Tanaka [7] proposed a method for extracting the impressions people receive from reading articles in newspapers. Emura et al. [8] proposed a method for extracting writer emotions based on the assumption that the emoticons added at the end of sentences express emotions. The kizasi.jp site [9] evaluates weblog users' emotions toward keywords that are topical among weblogs, and appends emotional word tags to those keywords.
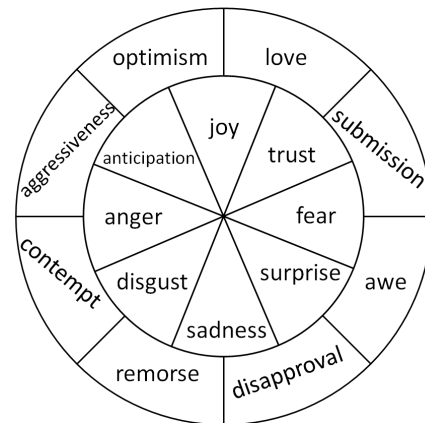

Fig. 1 Emoticon (*Emoji*) examples

Fig. 2 Plutchik's model of emotions

3 Purpose of study

Determining the emotions expressed via emoticons is important for reputation analysis efforts that utilize reviews from weblogs, message boards, and user review sites. In this study, we attempt to evaluate the emotions that are expressed via emoticons using 14 emotional vector dimensions. Yamashita et al. [5] discussed the use of emoticon emotional vectors when estimating writer mind states, but because the emotional vectors were made by questionnairing conducted by small people, questions were raised regarding the accuracy of the emotional vectors. In this study, we propose a method that can automatically generate emotional vectors of emoticons using a large volume of weblog articles. We set 14 dimensions of emotional vectors using Plutchik's emotion model (Fig. 2) [2] as a reference, and selected 288 emotional words from two Japanese dictionaries [10],[11], which we then divided into 14 basic emotions.

4 Proposed method

In this study, we used 14 dimensions of emotional vectors, using Plutchik's emotion model [2] as a reference, to express the emotions expressed in emoticons. The 14 dimensions comprise eight basic emotions (joy, trust, fear, surprise, sadness, disgust, anger, and anticipation) and six of the eight mixed emotions (love, awe, disapproval, remorse, contempt, and optimism). In this study, we assumed that emoticons used by writers in weblogs and the emotional words that appeared in the same sentence, were equivalent emotional expressions. We then prepared the emoticon emotional vectors by examining the frequency of co-occurrences between

the emoticons and emotional words. We selected a total of 288 emotional words from two Japanese dictionaries [10],[11] and subjectively categorized the words into 14 basic emotions (Table 1).

The concrete methodology used to prepare the emotional vectors is as follows. First, we collected a large volume of weblog articles and extracted only those sentences with emoticons. We then determined which sentences with emoticons also contained emotional words, and extracted those sentences. Then, we counted the co-occurrence of emoticons and emotional words in the extracted sentences. For example, in the sentence "I'm mad at myself 😡," the emotional word "mad" corresponds to the emotion "anger," so we increase the frequency of the emotion "anger" for the emoticon 😡.Using the same process, we tabulated the frequency for all of the extracted sentences. Finally, by normalizing the component values of the vectors so that they added up to a value of one, we prepared 14 dimensional emotional vectors.

5 Experiment

In order to validate the effectiveness of the proposed method, we conducted an experiment to designate emotional vectors using actual weblog articles. In this experiment, we collected weblog articles from Ameba Blog [12], which has the largest number of active users in Japan. We collected articles from 3,700 weblog and then extracted sentences that contain both emotional words and emoticons. This provided us with 45,256 sentences. Table 2 shows examples of the extracted sentences. While we can see from the table that some sentences were unsuitable for categorization, most sentences could be suitably categorized by the emotions expressed. We then prepared emoticon emotional vectors based on the extracted sentences. Table 3 shows examples of the prepared emotional vectors. We can see from the table that there are certain emoticons with large component values for specific emotions, and other emoticons for which the component values are widely distributed among various emotions. In particular, the sweat-drop emoticon has certain components that are in opposition to Plutchik's emotion model, which is to say that it is used to express components of widely contrasting emotions. Therefore, it is clear that it is not appropriate to uniquely attach specific emotions to that emoticon.

Table 1 Examples of corresponding emotional words for 14 emotions

Emotion	Emotional words
joy	Happy, fun, interesting, unexpected delight, amusing, happiness, cherish, thankful, humbled, satisfied
trust	Inevitable, unavoidable, inescapable
fear	Terrifying, scary, dangerous, eerie, scary, fear
surprise	Amazing, stupendous, fierce, splendid, incredible, amazed, surprised, astonished, shocked, stunned
sadness	Sad, cold, disconsolate, sadness, sorrow, heartrending, sentimental, pessimistic, sorrow, crying, tears, compassion, pitiful
disgust	Annoying, creepy, obnoxious, dislike, persistent, dislike, loathing, hatred, hateful, resentment, dissatisfied, offensive, disappointment, regret
anger	Audacious, irritating, chagrin, pitiable, maddening, rage, angry, infuriated, enraged, outraged, unpleasant
anticipation	Unexpected, futile
love	Lovely, cute, longing, benevolent, gentle, like, friendship, affection, loved, romance, enamored
awe	Impressive, indebted, cool, precious
disapproval	Difficult, unbearable, heartrending, disappointed, despair, pitiable
remorse	Shame, guilty, frustration, regret, jealousy
contempt	Unimaginable, silly, ridiculous, contemptible, despise scorn, condescension, ridicule
optimism	Profound, rich

6 Summary and future tasks

In this study, we proposed a methodology that can be used to automatically generate the emotional vector of various emoticons using the emotional words that co-occur with such emoticons. Furthermore, we conducted an experiment collecting and examining a large volume of weblog articles, and showed the effectiveness of the proposed method.

In our study, we set 14 dimensions for emotional vectors and subjectively divided 288 emotional words into 14 emotions. In the future, we will attempt to validate the number of dimension, and consider its application in areas such as for developing systems

that perform accurate reputation analysis and for information retrieval systems. This might be accomplished by combining our new method with existing methods for estimating emotions and impressions.

Table 2 Examples of sentences with co-occurrence of emoticons and emotional words

Emotion	Sample extracted sentences
joy	I am thankful to those who came. 💗 I'm cherishing the aquarium visit, as I haven't been for awhile. 🎶
trust	I guess it's inescapable, since everyone tired. 👋 It doesn't help to be impatient. 🙇
fear	It's scary, like a severed head. 🌀 I guess you have to get rid of the fear. 💡
surprise	It was shockingly super-tasty, something I haven't experienced in awhile. 🤯 I think it's amazing. 🎶
sadness	I held back my tears. 😥 I can hear this heartrending voice. ⬇
disgust	It's one of the top three things I most dislike in the world. ※ He's a sly fellow. 🙆
anger	I'm mad at myself. 🙆 I find that guy quite unpleasant. ※
anticipation	It's an unexpected opportunity. 💐 Well, I can't expect too much. 🌀
love	I'm starting to like him. 🍀 Well, I'm enamored. 🎶
awe	The end result is stylish and cool. 💐 He's cool because he's serious. 👍
disapproval	I was disappointed with you all. 🙆 It's warm and heartrending, and full of humanity. 🌸
remorse	I'm frustrated. 💥 You can't move forward if you worry about it all the time. ⬆
contempt	My blog is sure silly. 🌀 I was ridiculed in a loud voice. 🙆
optimism	It's rich in variation. 👍 I'm looking forward to the next Olympics. 🙆

Table 3 Examples of emoticon emotional vectors

Emoticon	Emotional vector
✨	joy: 0.62, surprise: 0.03, sadness: 0.03, disgust: 0.03, anticipation: 0.07, love: 0.17, optimism: 0.03
🎶	joy: 0.71, anticipation: 0.07, love: 0.21
⬆	joy: 0.75, surprise: 0.03, sadness: 0.02, disgust: 0.03, anger: 0.01, anticipation: 0.04, love: 0.12, remorse: 0.01, optimism: 0.01
🌀	joy: 0.13, fear: 0.07, surprise: 0.13, sadness: 0.13, disgust: 0.20, love: 0.27, remorse: 0.07
※	surprise: 0.21, disgust: 0.27, anger: 0.51

References:
[1] Facebook, http:// www.facebook.com/
[2] R. Plutchik, The Multifactor-Analytic Theory of Emotion, *The Journal of Psychology*, Vol.50, 1960, pp. 154-171.
[3] C. Yamamoto, K. Bessho, T. Uchiyama, and T. Uchiyama, A Method for Processing A Text with Emoji, *Proc.16th Annual Conf. of the Association for Natural Language Processing*, 2010, pp. 567-570 (in Japanese).
[4] M. Hagiwara and T. Mizuno, Semantic Analysis of Emoticons for Mobile Information Retrieval System, *Proc. JSAI SIG-KBS*, Vol.89, 2010, pp. 47-53 (in Japanese).
[5] R. Yamashita, K. Tani, and K. Takami, A method of deducing the user's state of mind by analyzing pictographic characters and emoticons in mobile phone emails and a method of selecting music suitable for the user's state of mind, *Proc. 2008 IEEE Region 10 Conference*, 2008.
[6] K. Shimizu and M. Hagiwara, Image Estimation of Words on Adjective Co-occurrences, *IEICE Trans. on Information and Systems*, Vol.J89-D, No.11, 2006, pp. 2483-2490 (in Japanese).
[7] T. Kumamoto and K. Tanaka, Proposal of Impression Mining from News Articles, *Lecture Notes in Computer Science*, Vol.3681, 2005, pp. 901-910.
[8] K. Emura, M. Yasugi, S. Miyazaki, T. Kuboyama, T. Aoki, and, H. Yasuda, SVM-based Emotion Extraction from Moblog Text, *Technical report of IEICE*, Vol.106, No.473, (KBSE-65), 2007, pp. 61-66 (in Japanese).
[9] kizasi.jp, http://kizasi.jp/

[10] A. Nakamura, *Dictionary of Emotive Expressions*, Tokyodo Publishing, 1993 (in Japanese).

[11] Y. Tobita and H. Asada, *Dictionary of Modern Adjective Usage*, Tokyodo Publishing, 1991 (in Japanese).

[12] Ameba BLOG, http://ameblo.jp/