An Automatic Method to Generate the Emotional Vectors of Emoticons Using Blog Articles

Sho Aoki, Osamu Uchida

Abstract—In recent years, reputation analysis and opinion mining services using the articles written in personal blogs, message boards, and community web sites such as Facebook, MySpace, and Twitter have been developed. To improve the accuracy of the reputation analysis and the opinion mining, we have to extract emotions or reactions of writers of documents accurately. And now, graphical emoticons (emojis in Japanese) are often used in blogs and SNSs in Japan, and in many cases these emoticons have the role of modalities of writers of blog articles or SNS messages. That is, to estimate emotions represented by emoticons is important for reputation analysis and opinion mining. In this study, we propose a methodology for automatically generating the emotional vectors of graphical emoticons automatically using the collocation relationship between emotional words and emoticons which is derived from many blog articles. The experimental results show the effectiveness of the proposed method.

Keywords—Collective intelligence, Consumer-generated media, Blog, Emoticon, Emotional vector, Emotional word, Emoji, Opinion mining, Reputation analysis, User-generated content.

I. INTRODUCTION

In recent years, there has been a rapid spread of media such as blogs, wikis, message boards, customer review sites, and social networking sites (SNSs), which make it possible for individuals to more easily generate information. These are collectively referred to as “consumer-generated media” or “user-generated content,” and the numbers of such media users are growing at an explosive rate. For example, Facebook [1], a leading SNS, has more than 600 million users worldwide and was the top visited website in the United States in 2010. 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 [2] and opinion mining [3]. Both of them can be considered as kinds of Web mining technologies used to analyze sentences posted to media (such as blogs, customer review sites, and SNSs) to determine if the sentence writer has positive or negative impressions of, for example, a particular product. To accomplish this, reputation analysis and opinion mining utilizes the technology of natural language processing [4], and typically analyzes the emotional words that appear in the text of such media [5], [6].

In this study, we focus on the graphical emoticons (“emojis” in Japanese [7]) that appear in sentences posted to blogs. In Japan, emojis have long been used with mobile phone email services [7]–[10]. Especially among younger users, emojis 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. Now, emojis have spread widely into other media formats. For example, in gmail, we can use graphical emoticons easily (Fig. 2). Because emojis effectively communicate the emotions (modality) of the writer, many users have become accustomed to their use when posting daily writings on blogs and SNSs not only from mobile phone but also from their personal computers (Fig. 3 shows an example of blog article with emojis). By determining the writer’s emotions, as expressed by the emoticons that exist in sentences posted to blogs and SNSs, it is possible to improve the accuracy of Web mining technologies such as reputation analysis and opinion mining. For example, the following two texts are same but the emoticons at the end of these texts are not equal.

“The dish I ordered wasn’t good.☹”
“The dish I ordered wasn’t good.😢”

If we only use the text for estimating the emotion, it is natural that two emotions derived from these texts are equal. However, it is expected that the emotion of the writer of the second sentence is madder than the first writer’s one. It is evident from the example that to use the emoticon for analyzing writers’ emotions is very important. Because of this situation, there have been many studies on estimating the emotion by using emoticons. However, almost all of them were research for not

Fig. 1 Examples of graphical emoticons (Emojis)
graphical emoticons (emojis) but ASCII glyphs emoticons such as “:-)”.
Accordingly, in this paper we propose a methodology that automatically
Generates the emotional vector of graphical emoticons (emojis).
To accomplish this, we collected a large
volume of blog articles containing graphical 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

II. PREVIOUS STUDY

A. Study on Emoticons

There have been lots of studies on ASCII glyphs emoticons
[12]–[20]. For example, Tanaka et al. [12] proposed methods
for extracting emoticons n text and classifying them into some
emotional categories. Kato et al. [13] examined the
relationships between four emotional states – anger, joy,
sadness, and guilt – and four different emoticons in mobile
phone email communication. Ptaszynski et al. [14] proposed a
system for affect analysis of emoticons based on theory of
kinesics. Yuasa et al. [15] showed that emoticons convey
emotions without the cognition of faces by using fMRI.

On the other hand, studies on graphical emoticons (emojis)
are in an early phase. Yamamoto et al. [21] proposed a method
for emoji disambiguation. Hagiwara and Mizuno [22] proposed
an information retrieval method for mobile phones that enables
the use of emoji in search queries, and their method is embedded
into a practical mobile information retrieval system [23].
Yamashita et al. [24] discussed applications for using the
emotional vector of emoji to deduce the mind state of email
users, and for use in music retrieval systems.

B. Analysis of Emotions and Impression from Text

There has also been a great deal of study on extracting
emotions and impressions from text for purposes such as
reputation analysis and opinion mining [5], [25]–[42]. For
example, Shimizu and Hagiwara [25] proposed a method for
estimating impressions based on the frequency of joint word
Kumamoto and Tanaka [26] proposed a method for extracting
the impressions people receive from reading articles in
newspapers. Emura et al. [27] 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 [28] evaluates weblog users’ emotions toward keywords
that are topical among weblogs, and appends emotional word
tags to those keywords.

III. PURPOSE OF STUDY

Determining the emotions expressed via emoticons is
important for reputation analysis and opinion mining by
utilizing reviews from blogs, message boards, consumer review
sites, and SNSs. In this study, we attempt to evaluate the
emotions that are expressed via emoticons using not one
sentiment but 14 emotional vector dimensions, because it is
expected that some graphical emoticons mean a lot of emotions.
Yamashita et al. [24] discussed the use of emotional vectors of
graphical emoticon (emojis) when estimating writer mind states,
but because the emotional vectors were made by
questionnairesing 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 graphical emoticons (emojis) using a large
volume of blog articles. We set 14 dimensions of emotional
vectors using Plutchik’s emotion model (Fig. 4) [11] as a reference, and selected 288 emotional words from two Japanese dictionaries [43], [44], which we then divided into 14 basic emotions.

IV. METHODOLOGY

In this study, we used 14 dimensions of emotional vectors, using Plutchik’s emotion model [11] 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 emotion 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 [43], [44] 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 blog articles and extracted only those sentences with graphical emoticons (emojis). 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 “自分に腹が立つ 😝 (I’m angry at myself),” the emotional word “腹が立つ (angry)” corresponds to the emotion “anger,” so we increase the frequency of the emotion “anger” for the emoticon “舂”。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 their added up to a value of one, we prepared 14 dimensional emotional vectors.

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Emotionally expressive words</th>
</tr>
</thead>
<tbody>
<tr>
<td>joy</td>
<td>うれしい (happy), 楽しい (fun), おもしろい (interesting), 幸福 (blessed), 思いがけない (unexpected), 喜ぶ (delight), 愉快 (amusing), 慰め (comfort), 楽し (cherish), 感謝 (thankful), 満足 (satisfied)</td>
</tr>
<tr>
<td>trust</td>
<td>やむを得ない (inevitable), 仕方がない (unavoidable), しょうがない (inescapable)</td>
</tr>
<tr>
<td>fear</td>
<td>恐ろしい (terrifying), 恐い (scary), 恐怖 (fear), 危ない (dangerous), 不気味 (eerie), 恐い (scary)</td>
</tr>
<tr>
<td>surprise</td>
<td>凍える (freezing), すさまじい (stupendous), 激しい (fierce), 素晴らしい (splendid), とてもない (incredible), 驚く (amazed), びっくり (surprised), 仰天 (astonished), 衝撃 (shocked)</td>
</tr>
<tr>
<td>sadness</td>
<td>悲しい (sad), 寒い (cold), 悲観 (pessimistic), やるさい (disconsolate), 悲哀 (sorrow), 悲痛 (heartrending), 悔い (regret), 悔しい (regretful), 悔絶 (despair), 悔絶 (despair), 不満 (dissatisfied), 心外 (offensive), 残念 (disappointment), 遺憾 (regret)</td>
</tr>
<tr>
<td>disgust</td>
<td>できる (possible), 腹が立つ (angry), 腹立たしい (irritable), 騒ぐ (noisy), 意地悪 (bitter), 意地悪 (bitter), 憎悪 (hate), 憎悪 (hatred), 憎悪 (hatred), 憎悪 (hatred)</td>
</tr>
<tr>
<td>anger</td>
<td>あつがましい (audacious), 怒り (rage), いらだしい (irritating), くやしい (chagrin), けんか (fight), 腹立たしい (maddening), 腹が立つ (angry), うがい (sneezing), 激怒 (furious), 慣れる (compassion)</td>
</tr>
<tr>
<td>anticipation</td>
<td>願ってもない (unexpected), 恐れない (fear)</td>
</tr>
<tr>
<td>love</td>
<td>いとしい (loving), 可愛い (cute), 恋しい (loving), 慈悲深い (benevolent), 愛 (love), 愛 (romance), 愛 (love)</td>
</tr>
<tr>
<td>awe</td>
<td>働い (impressive), かたじけない (indebted), かっこう良い (cool), 勉強 (precious)</td>
</tr>
<tr>
<td>disapproval</td>
<td>辛い (difficult), やりきれない (unbearable), 切ない (heartrending), 失望 (disappointed), 絶望 (despair), 恨えない (unbearable)</td>
</tr>
<tr>
<td>remorse</td>
<td>うそぶかない (shame), やりましい (guilty), 悔い (regret), 恨し涙 (frustrated tears), 悔い (regret), 悔い (regret)</td>
</tr>
<tr>
<td>contempt</td>
<td>あり得ない (unimaginable), くだらない (silly), ほかばかる (ridiculous), 軽蔑 (contemptible), 感激 (gratitude), 滅び (scorn), 憎う (contemptible)</td>
</tr>
<tr>
<td>optimism</td>
<td>感嘆深い (profound), 待ち遠しい (look forward), 豊か (rich)</td>
</tr>
</tbody>
</table>
V. 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 blog articles written in Japanese from Ameba Blog [45], which has the largest number of active users in Japan. We collected articles from 3,700 blog sites 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 generated emoji emotional vectors based on the extracted sentences. Figures 5–16 show examples of the prepared emotional vectors. We can see from the figures 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, for example, the 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. The knowledge derived from the experimental results indicates that the number of kinds of emotions represented by graphical emoticons may be larger than that of ASCII glyphs emoticons, and we think that this consequence is an important outcome of this study.

VI. SUMMARY AND FUTURE WORKS

In this study, we proposed a methodology that can be used to automatically generate the emotional vector of various graphical emoticons (emoji) using the emotional words that co-occur with such emoticons. Furthermore, we conducted an experiment collecting and examining a large volume of blog 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 opinion mining. This might be accomplished by combining our new method with existing methods for estimating emotions and impressions.

ACKNOWLEDGMENT

We would like to thank Professors Shohachiro Nakanishi and Hiroaki Kikuchi of Tokai University for valuable discussions.
Table 2: Examples of sentences with co-occurrence of emoticons and emotionally expressive words

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Sample extracted sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>joy</td>
<td>来てくださった方々に感謝いたします❤️ (I am thankful to those who came.) 久々の水族館、楽しみだよ💖 (I’m cherishing the aquarium visit, as I haven’t been for awhile.)</td>
</tr>
<tr>
<td>trust</td>
<td>みんな笑われていたからしょうがないよね😊 (I guess it’s inescapable, since everyone was tired.) 焦ったってしょうがないから🤗 (It doesn’t help to be impatient.)</td>
</tr>
<tr>
<td>fear</td>
<td>生音みたいで怖いね、これ🌙 (It’s scary, like a severed head.) やはり恐怖心を取りなければなりません😭 (I guess you have to get rid of the fear.)</td>
</tr>
<tr>
<td>surprise</td>
<td>久しぶりに、衝撃的に激ウマだったぁ😍 (It was shockingly super-tasty, something I haven’t experienced in awhile.) すごいなって思います🌙 (I think it’s amazing.)</td>
</tr>
<tr>
<td>sadness</td>
<td>泣きを我慢していた😢 (I held back my tears.) 悲痛な声が聞こえてくる悲痛な声が聞こえてくる悲痛な声が聞こえてくる悲痛な声が聞こえてくる悲痛な声が聞こえてくる悲痛な声が聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞く聞いています😭 (I’m starting to like him.) いや〜惚れました?&gt;&gt; (Well, I’m enamored.)</td>
</tr>
<tr>
<td>anger</td>
<td>自分に腹が立つ(Self mad at myself.) あの男、つくづく不愉快な奴だ😡 (I find that guy quite unpleasant.)</td>
</tr>
<tr>
<td>disgust</td>
<td>ねがってもない大チャンスだぞ!(It’s an unexpected opportunity.) まぁ、期待してもしょうがないけど둬(Well, I can’t expect too much.)</td>
</tr>
<tr>
<td>love</td>
<td>好きになってきてるんです❤️ (I’m starting to like him.) いや〜惚れました😢 (Well, I’m enamored.)</td>
</tr>
<tr>
<td>awe</td>
<td>出来上がりがオシャレでかっこいい!(The end result is stylish and cool.) 真剣だからかっこいい!(He’s cool because he’s serious.)</td>
</tr>
<tr>
<td>disapproval</td>
<td>みなさんには失望しました😢 (I was disappointed with you all.) 暖かくて切ない、人間味溢れる内容です лид (It’s warm and heartrending, and full of humanity.)</td>
</tr>
<tr>
<td>remorse</td>
<td>悔しいぜ😢 (I’m frustrated.) くよくよしても、前に進めないさ↑ (You can’t move forward if you worry about it all the time.)</td>
</tr>
<tr>
<td>contempt</td>
<td>私のブログはくだらないね〜criminal. My blog is sure silly.) 大声で馬鹿にされた!(I was ridiculed in a loud voice.)</td>
</tr>
<tr>
<td>optimism</td>
<td>バリエーション豊かだよね😍 (It’s rich in variation.) 次回のオリンピックが待ち遠しいです!(I’m looking forward to the next Olympics.)</td>
</tr>
</tbody>
</table>
REFERENCES

Information Science, Tokai University. His research interests include information theory, image processing, Web computing, and natural language processing.

He is a member of IEEE, IEICE (The Institute of Electronics, Information and Communication Engineers), IPSJ (The Information Processing Society of Japan), IIEEJ (The Institute of Image Electronic Engineers of Japan), and JSAI (The Japanese Society for Artificial Intelligence).