

Analysis of Emoticons in Japanese Twitter Dataset

Rina Murakami

s1200090

Supervised by Prof. Maxim Mozgovoy

Abstract

This paper describe the relation between emoticons and emotions in text. We use Japanese tweets corpus that was marked with the emotions the tweets express on the basis of Ekman and Plutchik emotion categories. As a result, we conclude that some emoticons are strongly correlated with certain emotions, while other emoticons can correspond to a variety of different emotions.

1 Introduction

Generally, we communicate with other people only using text data like in e-mail, chat, and social networking services (SNS) in the Internet. In this case, it is difficult to explain one's determination or emotions accurately. On the other hand, the reader may get into trouble if he/she cannot understand writer's true emotions or meaning. This is because messages in plain text cannot convey key information as the face-to-face communication does, in which a listener can guess speaker's emotions from a face, voice, and gesture.

Emoticons are strings of characters and punctuation symbols comprising images of faces, used as one of the ways to convey non-verbal information in text communication. In the present research we discuss the relation between text and emoticons that help to enable smoother communication. In [1], a set of Japanese and English Twitter messages (tweets) was categorized into eight types of emotions. Each tweet was processed by three independent coders. The eight types of emotion were Ekman's six basic emotions (anger, disgust, sadness, surprise, fear, and happiness) [2], and two extra emotions — pride in achievement (associated with positive self-evaluation) and embarrassment (a negative self-conscious emotion). These two extra emotions were selected from Ekman's extended set [3]. The first six emotions overlap with the corresponding emotions from the Plutchik's wheel [4] (see Fig.1). Moreover, the corpus has a "Skipped" category, reserved for tweets containing gibberish or foreign language text.

In the present research we use the Japanese section of the resulting marked corpus, obtained in [1]. We presume that the emotions expressed in tweets are linked with the emoticons found in the text. We first collect emoticons from the corpus of tweets, and assign each emoticons the same emotion as the corresponding tweet. Then the function of emoticons



Fig. 1 Plutchik's wheel of basic emotions

in tweets is examined. In addition, the ratio of the eight emotions expressed by each emoticon is calculated, so we also can find unexpected ways of using the emoticons. The focus of this research is to find out the true relationship between the emotions expressed by tweets and those by emoticons.

2 Emoticon Analysis

2.1. Exploring Emotional Meaning of Emoticons

To establish the link between emotions and emoticons, we first manually created the list of all emoticons found in the Japanese section of the corpus [1]. We removed tweets, marked as gibberish or having no emotional content. Then we counted how many times each emoticon was marked with a particular emotion. For example if a certain emoticon is given the emotion of anger and disgust, it gets one "Anger score" and "Disgust score." When the same emoticon appear many times, its score is increased. If the same emoticon appear having another emotion, it will receive the score of that emotion's category. Next, we calculated emotion percentage. This is a ratio of emotions each emoticon express, obtained with the following formula:

$$\text{Emotion percentage} = \frac{\text{Score of single emotion}}{\text{Total emotion score}}$$

"Score of single emotion" is one of the eight emotion scores (such as "Anger score" or "Disgust score")

"Total emotion score" is the total score of each emoticon. In other words, the sum of "Anger score", "Disgust score" and so on. The emotion percentage

shows the ratio of emotions that each emoticon may have. It is useful when we discuss the links between emotions and emoticons of tweets.

2.2. Results of Corpus Analysis

General information about our corpus is provided in the Table 1.

Table 1: General corpus information

Tweets in corpus	4338
Tweets having emoticons	791(18.2%)
Total unique emoticons	399
Total emoticons	965
Tweets marked as having emotions	2215

The Table 2 shows the distribution of emotions in our corpus. “Happiness” is the most widespread emotion, comprising nearly 40% of all emotional scores.

Table 2: Emotion percentage in the corpus

Emotions	Ang	Dis	Sad	Sur	Fea	Hap	Pri	Emb
Probability	3.79	4.38	17.47	10.65	2.17	39.59	8.80	3.07

The Table 3 shows the emotion percentages of ten most popular emoticons. The most widespread emoticon in the corpus is $\backslash (^o^o) /$, occurring 40 times in the corpus. This table shows that there is no direct correspondence between emoticons and emotions. In the table, most of the emoticons are smiles, but they are also associated with “Anger”, “Sadness” and other negative emotions. In the following section, we will study this observation in more detail.

Table 3: Emotional percentages of individual emoticons

Emoticon	fre	Ang	Dis	Sad	Sur	Fea	Hap	Pri	Emb
$\backslash (^o^o) /$	40	5.1	1.3	12.7	17.7	0.0	34.2	13.9	0.0
^^	39	5.5	2.7	4.1	5.5	0.0	43.8	15.1	5.5
(^o^)	30	1.9	1.9	3.7	3.7	0.0	66.7	11.1	1.9
(' ▽ ') /	28	3.6	3.6	3.6	7.3	3.6	56.4	9.1	1.8
(^^)	23	7.6	3.8	3.8	7.6	0.0	54.7	9.4	1.9
(^O^)	21	2.2	2.2	6.7	0.0	2.2	55.6	6.7	0.0
(' ω ')	20	3.0	4.6	53.0	12.1	12.1	7.6	0.0	3.0
(^_^)	17	2.9	0.0	0.0	8.6	0.0	65.7	5.7	0.0
(' ▮ ')	16	0.0	2.7	64.9	5.4	2.7	5.4	10.8	0.0

2.3. Emoticon Ambiguity

To understand the contexts of emoticon use, we performed the following process. First, we normalized emotion percentages so the largest value becomes 1.0 for each emoticon. This is done with a simple calculation:

$$\text{Normalized percentage of emotion} = \frac{\text{Largest percentage among emotions} / \text{Percentage of emotion}}$$

Then we find a ratio between the largest percentage and the second largest one. The resulting value (“emotional confidence”) shows the degree of ambiguity of each emoticon. If the ratio is high, it is likely that the emoticon is used almost always in the same emotional context. If the ratio is low, the emoticon carries a broad emotional value, and can be used in a variety of contexts.

The Table 4 shows emotional confidence values of the same emoticons as in the Table 3. The most frequent emotion is almost always “Happiness”, but the second frequent emotion varies greatly across emoticons. The average emotion confidence in the Table 4 is 5.13.

The Table 5 shows ten emoticons with the highest emotional confidence value. The average emotional confidence in the Table 5 is 8.1. Again, most probable emotion is “Happiness”. It is curious to note that some emoticons carry emotional value opposite to the facial expression of the emoticon. For example, the crying face expresses “Happiness”, while the smile face expresses “Anger.”

Table 4: Emotional confidence values for most frequent emoticons in the corpus

Emoticon	Fre	Largest Emotion	Second Largest	Ratio
\(^o^)/	40	Happiness	Surprise (0.5)	1.9
^^	39	Happiness	Pride (0.3)	2.9
(^o^)	30	Happiness	Pride (0.2)	6.0
(^▽^)/	28	Happiness	Pride (0.2)	6.2
(^^)	23	Happiness	Pride (0.2)	5.8
(^o^)/	21	Happiness	Sadness, Pride (0.1)	8.3
(^ω^)	20	Sadness	Surprise (0.2)	4.4
(^A^)	17	Happiness	Surprise (0.1)	7.7
(^D^)	16	Happiness	Pride (0.2)	6.0
(^ω^)	15	Happiness	Surprise (0.5)	2.1

Table 5: Emoticons with the highest emotional confidence

Emoticon	Fre	Largest emotion	Second largest	Ratio
(;ω;)	9	Sadness	Happiness (0.08)	12.5
(^^)/	7	Happiness	Pride (0.09)	11.0
(^▽^)	5	Happiness	Anger (0.11)	9.0
(^o^)/	21	Happiness	Sadness, Pride (0.12)	8.3
(^A^)	17	Happiness	Surprise (0.13)	7.6
(^▽)	11	Happiness	Anger, Pride (0.13)	7.5
(^▽^*)	9	Happiness	Surprise (0.13)	7.5
(#A.#)	5	Happiness	Pride (0.14)	7.0
(^AAA)	12	Happiness	Surprise (0.15)	6.5
(^▽^)/	28	Happiness	Pride (0.16)	6.2

3 Discussion

In this research, we assigned the emoticons the same emotion as the corresponding tweets so emoticons have emotions based on emotions of tweets. If a certain emoticon a high emotional confidence value, it expresses a specific emotion. However, Table 3, 4, 5 show that some emoticons do

not correspond to one emotion. The emotion of the emoticon is often set by its use in a sentence. In other words, the emotion is that emoticon expresses varies depending on the situation.

“Happiness” is highest emotion in this research. In the Table 2, the probability of “Happiness” is 40%, larger than of any other emotion. This is maybe related to the fact that the Japanese people grow up with the perception of “be considerate of the situation of another person.” They do not tweet their negative emotions because they are afraid that their negative emotions will negatively affect others.

On the other hand, “Happiness” is a wide emotion that is also related to “Joy”, “Satisfaction”, “Love” and other positive emotions. In other words, the highest probability of “Happiness” can be explained with the fact that the range of “Happiness” is too wide.

In the Table 5, the most unambiguous emoticons expresses “Sadness”, but it can be used to express “Happiness” in certain cases. For example, “ありがと一あと2日、頑張る一(; ω ;) (Thanks! I do my best in two days).”

“Happiness” and “Sadness” are completely opposite emotions as Plutchik’s wheel shows it. The crying face here expresses a “cry for joy”, so the crying face is used as “Happiness”. “Cry for joy” is crying due to one’s excessive joy. It complements “Happiness” emotion that the plain text cannot express.

Likewise, the third emoticon in the Table 5 is a smiling face, but it may also expresses “Anger”. For example, in the phrase “キレそう (^▽^)” (I feel I am going to be rage.)” it express anger while laughing. According to [5], Japanese usually hide own negative emotion by smile. In [5], this smile is called “Japanese Smile”. The results in this research, “Japanese Smile” is used in text communication. In the context like shown in the example above, the smile emoticon feels softer or more joke-like than if the real anger emoticon is used (like in “キレそう(#`D)`”).

4 Conclusion

In this research, the relation between text and emoticons is discussed. Some emoticons have different emotional meanings when used in texts. Sometimes the emoticon may express an emotion opposite to the emoticon facial expression because of the content of text where the emoticon is used. In other words, the emotion of emoticon greatly depends on the contents of text. Emoticons do not necessarily express emotions originally associated with them. Furthermore, emoticons serve as complements of emotions that text cannot express in

plain tweets or appease an emotion directly contained in the text.

However, the result of this research is biased due to initial emotion categorization. Especially, "Happiness" takes 40% of the total emotional content. In the future, if we perform a more fine-grained classification of emotions, we will be able to get more reliable results.

Reference

[1] A. Danielewicz-Betz, H. Kaneda, M. Mozgovoy, and M. Purgina, "Creating English and Japanese Twitter corpora for emotion analysis", International

Journal of Knowledge Engineering, 2015, vol. 1(2), pp. 120-124.

[2] P. Ekman, "An argument for basic emotions", *Cognition & Emotion*, vol. 6, no. 3-4, 1992, pp. 169-200.

[3] P. Ekman, "Basic Emotions," in *Handbook of Cognition and Emotion*, T. Dalgleish and M. Power, Eds., John Wiley & Sons, 1999, pp. 45-60.

[4] R. Plutchik, "The nature of emotions," *American Scientist*, vol. 89, no. 4, 2001, pp. 344-350.

[5] M. Nakamura, "The Communication of Emotion and Moral Emotions: The psychological background of the mutualism in a symbiotic society",

宇都宮大学国際学部研究論集(Utsunomiya University Faculty of International Studies Research Paper Collection), vol.30, pp75-86, 2010, (only