

Preliminary Analysis of Weibo Emojis for Sentiment Analysis of Chinese Social Media

Da Li

Rafal Rzepka

Kenji Araki

Graduate School of Information Science and Technology, Hokkaido University

In this paper, we introduce our preliminary analysis of emoticons used on Weibo, Chinese microblog. By performing a polarity annotation with a new “humorous type” added, we have confirmed that 23 emoticons can be considered more as humorous than positive or negative. We also discussed some possible related problems which might occur during any social media sentiment analysis performed on Chinese language. We believe that combining machine learning, sentiment lexicons and paralinguistic elements as emoticons can be expected to achieve better performance in the future.

1. Introduction

People have become increasingly accustomed to expressing their opinions online, especially on social media such as Twitter, Facebook and Weibo - China's biggest social media network that was launched in 2009. Such microblog data contain a vast amount of valuable sentiment information not only for the commercial use, but also psychology, cognitive linguistics or political science. Sentiment analysis of microblogs becomes the research hotspot in the field of Natural Language Processing. Study of sentiment in microblogs in English language has undergone major developments in recent years [Peng 17]. Chinese sentiment analysis research, on the other hand, is still at early stage especially in the domain of lexicons and emoticons. Emoticons convey a significant emotional information and play an important role in the sentiment analysis of social media [Novak, 15] [Guibon 16]. In this paper we focus on the emoticons used on Weibo; we analyze the characteristics of this particular set of emoticons, report on their evaluation while dividing into three categories: positive, negative and humorous. We list 23 emoticons which were annotated as “humorous” more often than “positive” or “negative”. With these novel insights into Chinese emoticons we want to help sentiment analysis researchers to consider implementing emoticons in their algorithms and improve their efficiency. We also discuss possibilities and difficulties in using these emoticons in Chinese social media sentiment analysis and applying them to other domains in the future.

2. Related Research

At present, the sentiment analysis technology generally can be divided into two categories: rule-based relying on sentiment lexicons, and machine learning-based ones relying on annotated data. Tan and Zhang [Tan 08] conducted an empirical study of sentiment categorization on Chinese documents. They tested four features - mutual information, information gain, chi-square, and document frequency; and five learning algorithms: centroid classifier, k-nearest neighbor, winnow classifier, naïve Bayes (NB) and support vector machine (SVM). Their results showed that the information gain and SVM features provided the best performances for sentiment categorization coupled with domain

or topic dependent classifiers. There are also researchers who have combined the machine learning approach with the lexicon-based approach. Chen et al. proposed a novel sentiment classification method which incorporated existing Chinese sentiment lexicon and convolutional neural network [Chen 15]. The results showed that their approach outperforms the convolutional neural network (CNN) model only with word embedding features [Kim 14]. However, existing approaches do not consider emoticons. Recently, a powerful system utilizing emoji in Twitter sentiment analysis model called DeepMoji was proposed [Felbo 17]. Its creators trained 1246 million tweets containing one of 64 common emoticons by bidirectional long short-term memory (Bi-LSTM) model and applied it to interpret the meaning behind the online messages. DeepMoji is also the most advanced sarcasm-detecting model, with an accuracy rate of 82.4 percent even outperforming human detectors who managed to acquire 76.1 percent accuracy rate.

3. Emoticons in Chinese Social Media

In the real-life (offline) dialogue between human beings, besides tone changes, we usually express emotions with body language. In social networks, this can partially achieve by using emoticons [Aldunate 16].

In Chinese social media (Weibo in our example), we found that in addition to the expression of positive and negative emotions, people tend to express a kind of humorous emotion, which is not sarcasm. Sarcasm reverses the emotion of the literal text, therefore sarcasm-detecting capability can play a significant role in sentiment analysis, especially in case of social media, however sarcasm and irony tend to convey negative emotions in general. We have noticed above mentioned humorous emotion of Weibo users when they post microblog entries which contain emoticons but are often difficult to be classified as positive or negative. It seems that some emoticons are used just for fun, self-mockery or jocosity which expresses a kind of implicit humor characteristic in Chinese culture. Emoticons seem to play an important role in expressing this kind of emotion. There is a high possibility that this phenomenon can cause a significant difficulty in sentiment detecting task, therefore we decided to analyze Weibo emoticons before adding them to our system for classifying emotions in Chinese microblogs.

Contact: Da Li, Graduate School of Information Science and Technology, Hokkaido University, lida@ist.hokudai.ac.jp

4. Analysis of emoticons used on Weibo

First we chose 67 most common emoticons out of 144 emoticons used in Weibo (see Figure 1). Only the emoticons with facial expression were chosen.



Figure 1: Most common emoticons of Weibo

We asked 12 Chinese native speakers to label these 67 emoticons by applying one of three following categories: positive, negative and humorous. The evaluators were 6 males and 6 females, each of them being a Weibo user for more than five years.

Table 1: Twenty-three emoticons conveying humor typical in Chinese culture

NUMBER	HUM (%)	NEG (%)	POS (%)
A1	41.7	25.0	33.3
G1	58.3	0.0	41.7
L1	66.7	33.3	0.0
E2	91.7	8.3	0.0
G2	58.3	0.0	41.7
M2	83.3	0.0	16.7
N2	58.3	25.0	16.7
O2	66.7	8.3	25.0
G3	66.7	0.0	33.3
H3	41.7	33.3	25.0
I3	75.0	25.0	0.0
J3	58.3	41.7	0.0
M3	50.0	25.0	25.0
O3	66.7	33.3	0.0
A4	50.0	50.0	0.0
E4	58.3	41.7	0.0
H4	50.0	50.0	0.0
I4	50.0	50.0	0.0
A5	50.0	33.3	16.7
C5	41.7	41.7	16.7
D5	75.0	8.3	16.7
F5	58.3	33.3	8.3
G5	75.0	0.0	25.0

The annotation results showed that there are 44 emoticons with obvious positive or negative emotion, and the remaining 23 emoticons are considered more as humorous than positive or negative. The results are summarized in Table 1.

5. Discussion

Our hypothesis is that the emoticons which are considered to be humorous expression in more than 50% of cases (see Table 1) usually convey humorous load. Moreover, the sentiment polarity of emoticons such as 😊, 😏, 😜 and other cases show rather clearly their ambiguity, and they should be treated with special care when we add emoticon recognition capability to our system in the future. A1 was originally a smile emoji, but more and more people believe that this emoji is used for expressing artificial smile, ridicule or self-mockery. On the other hand, when dealing with such emoticons on Weibo, we should be careful with jumping to conclusion and straightforwardly label it as a humorous, but instead we need to use a significant number of labeled examples containing these emoticons to train an appropriate model, for example by means of deep neural network. Besides, special attention should be given to emoticons such as 🐱, 🐶, 🐼, 🐾 and 🐾 because they tend to be considered to represent humorous meaning, and they became the most frequently used emoticons in Weibo in recent years (see Figure 2) ^{*1}.

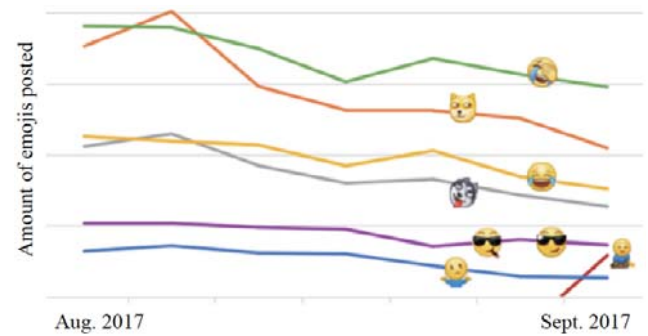


Figure 2: Top seven most frequently used emoticons in Weibo in 2017

Usually, when we process NLP tasks in social media, researchers tend to filter out images and videos as noise and concentrate on textual analysis. This often brings obvious difficulties for the sentiment analysis task. For example, as shown in Figure 3, we know that the user wants to share some gourmet food he ate during his journey, but if we only consider the text, "Morning! Third day in Sapporo, what should I eat for breakfast?" would be normally considered as neutral, but because the tweet contains 🍴 (100% positive in our annotation), a system could correctly mark the entry as positive. Based on this observation, we can notice that the sentiment analysis considering emoticons has could improve performance while analyzing short texts or the posts with images without any image processing.

The hypothesis that human facial expressions are universal [Darwin 72] is likely wrong [Gendron 14]. [Ekman 92] pointed out that the human beings of all races in the world have six basic feelings in common. However, as [Jack 14] study suggest, there

^{*1} <http://www.civiw.com/webby/20170920175205>

might be only four basic feelings. There are many unknown factors in constantly changing moods of human beings, but communication with emoticons has become a global phenomenon. On the other hand, because of different ethnic and cultural differences, misunderstandings in using facial emoticons are possible [Rzepka 17]. We noted that in the Facebook emoticons survey 2017^{*2} the emoji 🤔 ranked first in the list of most used emoticons top 10. This emoji is similar to the most used emoticons in 2017 on Weibo: 🤔 according to their survey^{*3}. It looks like a combination of 🤔 (Person Facepalming) and 😄 (Face with Tears of Joy). Both of them look like a face laughing through tears, but the expression of feelings is different when “facepalm” hand is added causing 🤔 express the resignation nuance more humorous. It also shows that we need to pay attention to the cultural background of different races in the emotional processing of multiple languages.



Figure 3: Weibo post: “Morning! Third day in Sapporo, what should I eat for breakfast?”

6. Conclusions and Future Work

In this paper, we introduced our preliminary analysis of emoticons used on Weibo, Chinese microblog. By performing a polarity annotation with a new “humorous type” added, we have confirmed that 23 emoticons can be considered more as humorous than positive or negative. We also discussed some possible problems which might occur during any social media sentiment analysis performed on Chinese language. We believe that combining machine learning, sentiment lexicons and paralinguistic elements as emoticons can be expected to achieve better performance in the future. In terms of applications, tourism information detection, tourism attraction or food recommendation

systems can also be expected to have better accuracy by implementing emoji processing. We would also like to draw attention of the sentiment researchers to the phenomenon of emoticons that can change their meaning in time. This can be an important clue suggesting that updating data for machine learning-based approaches might be crucial.

References

- [Aldunate 16] Aldunate N, González-Ibáñez R. An Integrated Review of Emoticons in Computer-Mediated Communication. *Frontiers in Psychology* (2016)
- [Chen 15] Chen, Z., Xu, R., Gui, L., Lu, Q.: Combining convolution neural network and word sentiment sequence features for Chinese text sentiment analysis. *Journal of Chinese Information Processing*. 2015;29(6). (2015)
- [Darwin 72] Darwin, C.: The expression of the emotions in man and animals John Murray (1872)
- [Ekman 92] Ekman, P.: An argument for basic emotions, *Cognition & emotion*, Vol. 6, No. 3-4, pp. 169–200 (1992)
- [Felbo 17] Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, Sune Lehmann.: Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *Association for Computational Linguistics*, pp 1616-1626, Oct 2017. (2017)
- [Gendron 14] Gendron, M., Roberson, D., Vyver, van der J. M., and Barrett, L. F.: Perceptions of emotion from facial expressions are not culturally universal: evidence from a remote culture., *Emotion*, Vol. 14, No. 2, p. 251 (2014)
- [Guibon 16] Guibon, G., Ochs, M., Bellot, P.: From Emojis to Sentiment Analysis: WACAI 2016, Jun 2016, Brest, France. (2016)
- [Jack 14] Jack, R. E., Garrod, O.G., and Schyns, P.G.: Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time, *Current biology*, Vol. 24, No. 2, pp. 187–192 (2014)
- [Kim 14] Yoon Kim. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar. (2014)
- [Novak 15] Kralj Novak P, Smailović J, Sluban B, Mozetič I, Sentiment of Emojis. *PLoS ONE* 10(12): e0144296 (2015)
- [Peng 17] Peng, H., Cambria, E., and Hussain, A.: A review of sentiment analysis research in Chinese language. *Cognitive Computation*. (2017)
- [Rzepka 17] Rzepka, R., Okumura, N., Ptaszynski, M.: Worlds Linking Faces — Meaning and Possibilities of Contemporary Pictograms: *Journal of the Japanese Society for Artificial Intelligence* 32(3), 350-355, 2017-05 (2017)
- [Tan 08] Tan, S., Zhang, J.: An empirical study of sentiment analysis for Chinese documents. *Expert Syst Appl*. 2008;34(4):2622–9 (2008)

*2 <http://www.theverge.com/2017/7/17/15984204/facebooks-most-used-emoji-hearts-tears>

*3 <http://weibo.com/ttarticle/p/show?id=2309351000124181543877781277>