

Review of Comment Sentiment Analysis Based on Emoticons

Hongyu Qian

College of business, Shanghai university of finance and economics, Shanghai, China, 200127
Qianhy2001@163.com

ABSTRACT

Nowadays, social media is becoming more mature, and its comment data form is becoming more diversified. Emoticons are already an element that cannot be ignored in comment sentiment analysis. This article reviews the research of simple text sentiment analysis, emoticons sentiment analysis and text sentiment analysis incorporating emoticons, finding that simple text sentiment analysis is divided in two types: sentiment analysis based on sentiment dictionary method and based on machine learning method; emoticons sentiment analysis can be divided into traditional analysis methods, expression based on middle-level semantics of pictures and deep learning technology. Text sentiment analysis incorporating emoticons is establishing models to calculate scores separately and then getting the emotions, and combining the emoticons as features with text to do the sentiment analysis.

Keywords: *comment, sentiment analysis, text, emoticons, social media*

1. INTRODUCTION

Nowadays, with the rapid development of social media, the data on media platforms are not only increasing in quantity, but also becoming more diversified in types. From the text data of the previous unit to the combination of text emoticons. This kind of comment data containing both text and emoticons provides a new perspective for the study of sentiment analysis. Mining the sentiment of such comments will also become the mainstream direction of sentiment analysis in the future. The purpose of this study is to sort out the existing research situation from simple text sentiment analysis, emoticons sentiment analysis and text sentiment analysis incorporating emoticons, and to pave the way for future research on comment sentiment analysis.

2. TEXT SENTIMENT ANALYSIS

The mainstream text sentiment analysis is divided into two kinds: one is based on sentiment dictionary method, the other is based on machine learning method.

The emotional dictionary first decomposes the text into emotional words, negative words and degree adverbs, and then divides the negative words or degree adverbs that exist before the emotional words into a group. If they are negative words, the emotional weight

of the emotional words is multiplied by -1 , which is similarly multiplied by the degree value of the degree adverbs. Finally, the scores of all groups are summed, and the score is positively attributed to positive emotions, and vice versa. Yang [1] applied HowNet and NTUSD resources to expand the existing emotional dictionary and established a new emotional dictionary with tendency. Liu Ping et al. [2] proposed a new emotional weight calculation method based on the similarity calculation of Chinese words, and expanded the HOWNET emotional words set. Cai et al. [3] proposed a three-level emotional dictionary, which links emotional words with their corresponding entities, reduces the occurrence of false emotional tendencies, and adopts the method of mixed classification model, combines SVM with GBDT, and achieves the results better than those of single classification model.

The traditional sentiment analysis method based on machine learning is the bag of words model — the text is regarded as a collection of some words, and then it is brought into the algorithm for classification and clustering. But the disadvantage of the bag of words model is the loss of sequence information and semantic information between words. Le et al. [4] proposed a paragraph vector model for the defects of the bag of words model, which realized an unsupervised algorithm to automatically learn the fixed length feature representation from texts of different lengths. Li et al. [5]

proposed a RNN language model based on long and short-term memory. Compared with the traditional RNN language model, LSTM performs better in analyzing the emotion of long sentences, and can effectively obtain complete sequence information.

3. EMOTICONS SENTIMENT ANALYSIS

Emoticons as an image, so we can analyze the emotion of emoticons from the method of image emotion analysis. Image sentiment analysis can be divided into traditional analysis methods, expression based on middle-level semantics of pictures and deep learning technology.

Early image sentiment analysis mainly focuses on the underlying objective visual features of images, and uses machine learning technology to classify image sentiment. Among them, the underlying features mainly include color features, texture features and shape contour features. Machajdik et al. [6] proposed some methods to extract and combine the underlying features representing the emotional content of an image. Signs are used for image sentiment classification. These methods are mainly based on psychology and Theory and Art Theory to Extract Art Works with Emotional Expression Tamura texture feature and gray level co-occurrence matrix are applied to emotion classification based on the unique image features. Jia et al. [7] measured the basic properties of color, such as saturation, tone, color, brightness, using the basic characteristics of the image, using semi-supervised learning framework to automatically predict the emotion of art works. Wang [8] extract the color and shape of an image as its perceptual features and use RBF neural network, mapping from feature space to emotion space, to realize perceptual search of image in emotion space.

Since there is a large distance between the underlying features of the picture and the high-level emotional semantics, so many researchers began to try to construct middle semantics to express image emotion. Borth et al. [9] proposed Adjective Noun Pairs (ANP) as the middle semantic of emotion. At the same time, the article also proposed a new visual concept detector library SentiBank, which uses Color Histogram and Bureau Local Binary Pattern (LBP) descriptor, scene feature descriptor for visual emotion classification, can be used to detect 1 200 ANPs of the images; experiments show that the method acquires great improvement in accuracy and precision in detection.

With the development and progress of deep learning technology, image classification and recognition have made significant progress. In 2006, Hinton [10] proposed Deep Belief Network (DBN), using multiple hidden layer neural networks to train raw data, which transforms the high-dimensional data into understandable low-dimensional data. The data features can be better

understood and applied to digital recognition and classification tasks.

Krizhevsky et al. [11] used CNN to process image data, which has also achieved good experimental results. CNN uses reversed Back Propagation (BP) algorithm, capable of automatically learning task-related features in the data from a large number of strong tags, which is more robust than manually marking features

4. TEXT SENTIMENT ANALYSIS INCORPORATING EMOTICONS

On the past social media platforms, emoticons have not yet been popularized, so when commenting on sentiment analysis, they tend to ignore emoticons and only analyze text comments. Nowadays, the enthusiasm of the new generation of the main users of social media for emoticons makes the emotional analysis of emoticons particularly important. Jia et al. [12] used the sentiment dictionary method to analyze the sentiment polarity. When calculating the sentiment value of the text, she used the subjective method to define the value of the emoticons, and added the influence of the emoticons on the text sentiment. Chen et al. [13] divided 150 emoticons of Sina Weibo into seven categories: happiness, hobbies, anger, sadness, fear, disgust and surprise, and established the emotional symbol library. At the same time, the emotional analysis dictionary and the network terminology dictionary were established. These three parts were used as the explicit emotional characteristics of Weibo, and then the FIHC algorithm combined with frequent item set and semantic clustering was used to analyze the implicit emotional characteristics of Weibo - the part of speech and semantics of Weibo content. Huang et al. [14] proposed a microblog topic sentiment mining model TSMMF based on multi-feature fusion. The model fuses emoticons and microblog users' emotional characteristics on the basis of graph model LDA to realize microblog sentiment analysis. Ma et al. [15] considering the use of a large number of emoticons, word networking and colloquial expression on social media, modified the traditional short text emotional dictionary, and proposed a cross-media multi-emotion emotional dictionary construction method based on special emotional symbols in social networks. Combining pictures with short text content, the emotional dictionary based on social network is screened by calculating the mutual information between expression and text words. Wu [16] modeled the similarity between the expression vector and the description vector, and obtained the vectorization form of the expression symbol by using the gradient descent algorithm. Then, the expression vector and the word vector of the microblog text were analyzed based on the convolutional neural network, and finally the emotional polarity of the text was obtained. Tan et al. [17] used a bidirectional recurrent neural network model to learn text features, and

used the attention mechanism of expression symbols to obtain new features combined with text and expression symbols to analyze emotional polarity. The results show that the method can effectively identify the emotional

polarity of Weibo comments, and the accuracy and F value are beyond the method of adding emoticons as features only to the neural network.

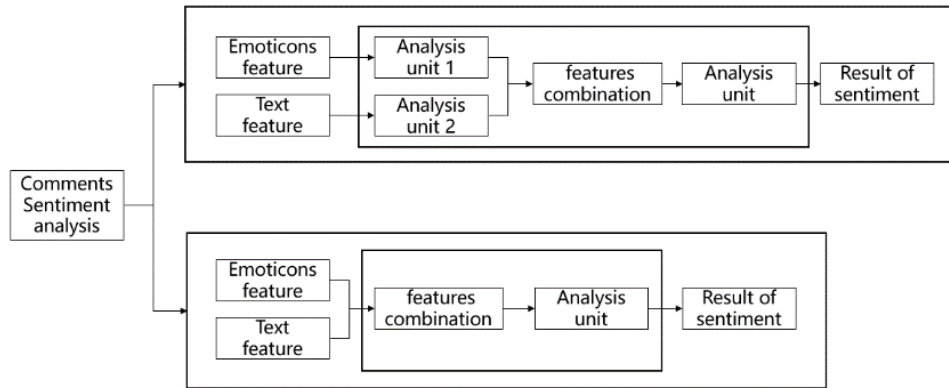


Figure 1 major coment sentiment analysis structure

5. CONCLUSION

The existing text sentiment analysis incorporating emoticons is mainly divided into two directions. The first one regards the emoticons and the text as two parallel information sources. After calculating the emotional scores with different models, it is combined to obtain the comment emotion. The second is to integrate the emoticons as a feature into the emotional analysis of text features. The article retrieves the research survey of image emotion analysis, but does not deeply study the way of using image emotion analysis in text sentiment analysis incorporating emoticons.

REFERENCES

[1] Yang Chao. Web public opinion orientation analysis based on technique of sentiment lexicon extension [D]. Northeastern University, 2009.

[2] Liu W P, Zhu Y H, Jia C L, Xiang H Z, Wen Z Q. Research on building Chinese basic semantic lexicon [J]. Computer Application, 2009,29(10):2875-2877.

[3]Yi Cai et al. "A hybrid model for opinion mining based on domain sentiment dictionary". International Journal of Machine Learning and Cybernetics 10.8(2019): 2131-2142.

[4] Le Q, Mikolov T. Distributed representations of sentences and documents. In: Proceedings of the 31st International Conference on Machine Learning (ICML014). Beijing,China: ACM Press, 2014. 1188;1196

[5] D. Li and J. Qian, "Text sentiment analysis based on long short-term memory," 2016 First IEEE International Conference on Computer Communication and the Internet (ICCCI), 2016, pp. 471-475, doi: 10.1109/CCCI.2016.7778967.

[6] Jana Machajdik and Allan Hanbury. Affective image classification using features inspired by psychology and art theory[C]., 2010.

[7] JIA J, WU S, WANG X, et al. Can we understand van Gogh’s mood? Learning to infer affects from images in social networks [C] // Proceedings of the 20th ACM International Conference on Multimedia. New York : ACM, 2012 : 857-860.

[8] WANG S F, CHENE H, WANG S H, et al. A Kansei image retrieval based on emotion model [J] . Journal of Circuits and Systems, 2003, 8 (6) : 48-52.

[9] BORTH D, JI R, CHEN T, et al. Large-scale visual sentiment ontology and detectors using adjective noun pairs [C] // Proceedings of the 21st ACM International Conference on Multimedia. New York : ACM, 2013 : 223-232.

[10] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets [J] .

Neural Computation, 2006, 18 (7) : 1527-1554.

- [11] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Image Net classification with deep convolutional neural networks, Proceedings of the 25th International Conference on Neural Information Processing Systems. Red Hook, NY, Curran Associates Inc., 2012, pp.1097-1105.
- [12] Jia S S, Di S L, Fan T R. Text emotion analysis model based on emoticons and emotional words [J]. Journal of the Hebei Academy of Sciences, 2013,30(02):11-15.
- [13] Chen T M, Miao R Y, Wang X H. Chinesemicroblog Sentiment Analysis using Both Explicit and Implicit Text Features [J]. Journal of Chinese Information Processing,2016,30(04):184-192.
- [14] Huang F L, Feng S, Wang D L, Yu G. Mining Topic Sentiment in Microblogging Based on Multi-feature Fusion[J].Chinese Journal of Computers,2017,40(04):872-888.
- [15] Ma B N, Huang Y F, Deng B X. Generating sentiment lexicon of online social network based on emotions [J]. Computer Engineering and Design,2016,37(05):1129-1133.
- [16] Wu C Q. Sentiment Analysis of Weibo Texts Combined with Emoticons[D]. Huaqiao University,2018.
- [17] Tan H, Deng S W, Qian T, Ji D H. Emoji-attentional neural network for microblog sentiment analysis[J]. Application Research of Computers, 2019, 36 (09): 2647-2650.