

Microblog Emotion Analysis Based on SVM and CRF Multi-Feature Combination

Jing Zhang^{1, a}, Lin Zhang¹

¹SMU College of information engineering, 201306, China.

^a741293972@qq.com

Abstract

In recent years, the emotional analysis of natural language has been a hot research object in the natural language of the Internet. Currently, the most popular social communication platform for short essays in China is weibo. The discourse in weibo usually expresses the user's full emotions and has strong subjectivity. Through the use of SVM and CRF as the basis, the short text in micro-blog analysis of its effect is the most specific and reliable. Therefore, this paper will analyze its working connotation and working principle, and study the construction of emotional dictionary and relevant algorithms from multiple perspectives.

Keywords

SVM; CRF. Multi-feature combination; Weibo emotion analysis.

1. Significance and Working Principle of Weibo Emotion Analysis

1.1. Significance of Microblog Emotion Analysis

Weibo, as the most popular social network platform in China, has a growing number of users. In the process of using weibo, users can not only learn about current news, but also express their opinions in it. The emotional analysis mentioned in this paper is to explore the views of weibo users, so as to make statistics and analysis of their emotional tendency. It is the increasing use of micro-blogs that makes the significance of this work more obvious.

1.2. Working Principle of Weibo Emotion Analysis

The working principle of micro-blog emotion analysis can be classified by its working method, which can be specifically divided into two methods: analyzing micro-blog emotion by using emotional knowledge and analyzing micro-blog emotion by using feature classification. Feature classification can be divided into supervised machine learning method and unsupervised machine learning method to analyze micro-blog emotions. Firstly, among the methods based on emotional knowledge, lexicon that can express emotional direction should be taken as the basis. Delicate; Good, etc., and related reverse words include: ugly; Coarse; Evil, etc. In the process of judging the short text, the number of words in the text should be counted to judge the overall emotional direction [1]. Although such an approach is relatively direct, it also has great limitations in the current era of endless Internet vocabulary. Therefore, such an approach may be outdated.

The supervised machine learning method among the feature-based analysis methods can construct a relatively complete analysis model and analyze it based on the model. Firstly, the microblog short text is preprocessed, and useless information is deleted in the preprocessing stage. Experiments are carried out according to different parts of speech, word features and word order in sentences. At the same time, common expressions in microblog can be added into the analysis to make the results more accurate. However, in the unsupervised machine learning method, its principle actually analyzes the point mutual information and selects short sentences

in the short text of microblog, so as to define the emotional tendency of the full text. For example, "rich" and "frequent" are selected first, and the subsequent nouns are marked, respectively, to analyze the correlation degree between the positive and negative aspects. The accuracy of these two words is also affected by some factors, including emotional secondary and predictive corpus [2].

2. Emotion Dictionary Construction

2.1. Pretreatment

Preprocessing is the first step in the construction of an emotion dictionary. The significance of preprocessing is to exclude the parts irrelevant to emotion analysis in microblog phrases and sentences. Since these irrelevant contents will have adverse effects on the process of emotional construction, contents such as # topic # need to be excluded in advance, and then their sentences should be marked with word segmentation tools.

2.2. Emotional Lexicon Construction

After preprocessing, the lexicon should be constructed. In the process of lexicon construction, the lexical meanings which are not obvious enough in praise and criticism should be filtered, and the lexical contents should be modified according to the real-time trend in the network. In addition to phrases, there are also emojis as the key basis for emotional expression in weibo. The corresponding praise value of phrases and emojis should be set to depreciate for the convenience of statistics.

2.3. Construction of Derogatory Word List

In the construction of derogatory word list, common derogatory and negative words should be included. The following paragraphs will be shown as examples in the form of charts, as shown in figure 1 for details.

non	no	without
be not allowed	don't	impossibility
not yet	never	definitely not

Figure 1. Negation words table

In the above negative word list, the order is gradually strengthened from left to right and from top to bottom.

2.4. Construction of Degree Adverb List

Besides derogatory words with strong representative meanings, degree adverbs are also one of the important influencing factors. The adverbs and the degree of their emotion are shown in graph form below, as shown in figure 2.

Adverbs of degree	The degree of
slightly	The lighter
compare	light
very	The heavier
The most	heavy

Figure 2. Degree adverb degree table

3. Algorithm Design

3.1. SVM Emotion Analysis

In SVM emotion analysis, a complete SVM model is required. In order to achieve the maximum completeness of the model, five types of features will be selected for analysis. The following paragraphs will analyze the specific meanings in the form of charts, as shown in figure 3 for details.

Characteristics of words	meaning
Nature of the words	The number of words in each part of speech
Emotional vocabulary	Number of words with emotion
Derogatory term	Negative words before emotional words
Adverbs of degree	Adverbs of degree that appear before emotional words
symbol	Emoji and question marks, exclamation marks

Figure 3. Vocabulary feature table in SVM model

After classifying them according to the characteristic techniques in the above table, the frequency of occurrence of characteristic words will be counted, and the final results will be calculated by means of digital technology [3]. In the model, negative words and degree adverbs can only be included in the statistical results when they are prefixed as affective words. After many experiments, the specific accuracy of the conclusion is shown in the figure below, as shown in figure 4.

Characteristics of the combination	Positive accuracy	Negative accuracy	Synthetic accuracy
The part of speech	53.2	53.2	53.2
Solitary affective word	78.23	79.25	78.53
Affective word + part of speech	79.66	80.33	79.99
Emotional words + part of speech + derogatory words	88.66	89.98	89.03
Emotion word + part of speech + derogatory word + degree adverb	87.36	87.86	87.52

Figure 4. Experimental results of different feature combinations of SVM model

According to the statistical results, emotional words play the most important role, while derogatory words also play a certain role in the analysis results, while the special symbols and recorded degree adverbs that are not included in the table play a small role in the SVM model system.

3.2. CRF Emotion Analysis

CRF emotion analysis focuses on the labeling of sequence, and its specific principle is to conduct emotion analysis of the whole sentence, that is to say, it needs to conduct sentence polarity analysis of each short sentence in the weibo short text book. Therefore, the overall polarity will have an acute impact on the whole microblog short text. In the actual verification process, it was found that the results were still not accurate after adding polar factors of part of speech

and affective words into the model. However, when the polarity of emotional words was added separately, their accuracy was greatly improved, and then derogatory words and degree adverbs were added successively [4]. After the special symbol and other factors, it is found that it has little influence on accuracy. Compared with SVM method, its accuracy in positive microblog is higher than that in SVM emotion analysis (91.02%), and in negative microblog is lower than that in SVM emotion analysis (88.35%). The comprehensive accuracy of SVM emotion analysis was 89.71% and CRF was 90.57%.

4. Endnotes

Through the discussion of this article for SVM and CRF respectively of the combined characteristics of weibo, the essay discusses the principle and the significance of sentiment analysis, known in the weibo sentiment analysis can play a role in the process, but the word library also need to be further perfect, to improve the accuracy, hope to offer some help for future work in this paper.

References

- [1] tingting Li , donghong ji. Analysis of micro-blog emotion based on SVM and CRF multi-feature combination [J]. Computer application research, 2015, 32(04):978-981.
- [2] yuanyuan li . Emotional analysis and research on hot events facing topic-based micro-blogs [D]. Anhui university, 2016.
- [3] chun liao. Research on fine-grained affective analysis of topics based on syntax and semantics [D]. Beijing institute of technology, 2016.
- [4] xele liu , yanxiang he . Research on micro-blog emotion analysis based on multiple features [J]. Computer engineering, 2017, 43(12):160-164+172.