

Using Simple Algorithms for Social Media Sentiment Analysis: Analyzing User Emotions on Social Platforms

Yichen Sun

Qingdao Jiaozhou Yingzi Private School Qingdao Shandong, China

EchenSun729@gmail.com

Abstract. Under the background of the vigorous development of social media, it is very important for business decision-making and social governance to accurately capture and analyze users' emotions. This article aims to explore a simple and effective way to analyze users' emotions on social media by using the basic NLP (Natural Language Processing) technology, especially the word frequency statistics method. Taking Weibo as the data source, this article constructs an emotional analysis model. By collecting data, preprocessing, constructing an emotion dictionary and calculating the emotion score, the model realizes the emotion tendency recognition of Weibo's posts. The experimental results show that the model can accurately identify the dominant emotional tendency in Weibo's posts, and perform well in capturing emotional trends. The research in this article not only provides a feasible solution for the emotional analysis of social media, but also provides reference for the follow-up research.

Keywords: social media; Emotional analysis; Word frequency statistics; Emotional dictionary; Natural language processing.

1. Introduction

In today's era of rapid development of information and digitalization, social media has become an indispensable part of people's daily life [1]. It not only changed our communication mode, but also profoundly affected the mode of information dissemination and reception [2]. Every day, hundreds of millions of users share their lives, express their opinions and exchange their emotions on various social platforms. These massive data contain rich user emotional information [3]. For enterprises, accurately capturing and analyzing these emotions can help them better understand consumer needs, optimize products and services, and formulate effective market strategies [4]. From a social perspective, sentiment analysis can assist government departments to monitor public sentiment, discover and deal with potential social problems in time, and maintain social stability [5]. Therefore, the development of efficient and accurate sentiment analysis technology is of great significance for promoting the scientificity of business decision-making and improving the accuracy of social governance [6]. The purpose of this study is to explore a simple and effective way to analyze users' emotions on social media by using basic NLP technology, especially word frequency statistics, and to provide valuable reference for related fields.

As an important branch of NLP, sentiment analysis has received extensive attention and research in recent years. Early research mostly relied on rule-based methods, and with the development of machine learning, especially deep learning technology, the accuracy and efficiency of sentiment analysis have been significantly improved [7]. Nevertheless, sentiment analysis based on simple algorithms such as word frequency statistics still has certain application value because of its easy implementation and low calculation cost. Although complex models perform well in emotional analysis tasks, the effect of basic methods such as word frequency statistics in specific scenarios cannot be ignored [8]. Based on the previous research results, this study will explore the specific application of word frequency statistics in social media sentiment analysis, and try to put forward suggestions for improvement, in order to contribute to the development of this field.

This study focuses on specific social media platforms, such as Weibo and Twitter, which are regarded as ideal data sources for sentiment analysis because of their large user base and high activity. The research will select public posts within a certain time range as the analysis object, covering a

wide range of topics to ensure the diversity and representativeness of the data. In terms of methods, this article adopts NLP technology, which is based on word frequency statistics, as the core means. Through the frequency statistics of words in posts, combined with the emotional dictionary, the user's emotional tendency is identified and quantified.

2. Theoretical basis and technical framework

2.1. NLP foundation

NLP is an interdisciplinary field of computer science, artificial intelligence and linguistics. It aims to enable computers to understand, interpret and generate human natural language [9]. The core tasks of NLP include text classification, information extraction, named entity recognition, syntactic analysis, semantic understanding and emotional analysis. Before sentiment analysis, the original text data must be preprocessed, which involves word segmentation, stop words removal, stem extraction and word shape reduction. These preprocessing operations are helpful to reduce data noise and improve the accuracy of subsequent analysis. NLP also relies on various language models and algorithms to capture the relevance and contextual information between words, so as to understand the meaning and emotional tendency of the text more accurately.

2.2. Construction of emotional analysis model

Word frequency statistics is one of the most basic and important technologies in NLP, which reveals the theme, style and emotional tendency of the text by analyzing the frequency of each word in the text. In affective analysis, word frequency statistics are often combined with affective dictionaries. Emotion dictionary is a set of words marked with emotion tendency in advance. By matching the words in the post with the emotion dictionary, the overall emotion score of the post can be calculated.

In order to reflect the importance of vocabulary in a specific context more accurately, this article uses TF-IDF (word frequency-inverse document frequency) and other weight calculation methods to weight the word frequency. Although the word frequency statistical method is simple, its efficiency and effect are often satisfactory when dealing with large-scale text data, especially in the scenes with limited resources or requirements for model complexity.

The algorithm formula is as follows:

$$TF-IDF(t, d, D) = TF(t, d) \times IDF(t, D) \quad (1)$$

Where: $TF(t, d)$ is the word frequency, indicating the frequency with which the word t appears in the document d . $IDF(t, D)$ is the inverse document frequency, which indicates the discrimination of the word t in the whole document set D .

In sentiment analysis, combined with the sentiment dictionary, the $TF-IDF$ score can be multiplied by the sentiment tendency value of each word in the sentiment dictionary, so as to get the sentiment score of the document:

$$\text{Sentiment Score}(d) = \sum_{t \in d} TF-IDF(t, d, D) \times \text{Sentiment Value}(t) \quad (2)$$

Where: $\text{Sentiment Value}(t)$ is the emotional tendency value of the word t in the emotional dictionary, which is usually a number between positive and negative, indicating the positive or negative emotional strength of the word t .

After the data preprocessing is completed, the emotional analysis model is set up. Considering that the original intention of the study is to use simple and effective algorithms, the word frequency statistics method is chosen as the core. Firstly, an emotional dictionary is constructed, which contains a large number of positive and negative emotional words, and each word is given a corresponding

emotional weight. The construction of emotional dictionary refers to a number of public emotional vocabulary resources, and makes appropriate adjustments and supplements in combination with the characteristics of Weibo language. Then, traverse the preprocessed post text, count the frequency of each word in the text, and calculate the emotional score of each post according to the emotional dictionary. In order to reflect the emotional tendency more accurately, TF-IDF method is also used to weight each word to highlight the contribution of important words to the emotional score. Finally, the emotional score of each post and the distribution of positive, negative and neutral emotions in the whole data set are obtained.

2.3. Technical framework and tools

Based on the above theoretical basis, this section designs a concise and efficient technical framework to support the whole process of social media sentiment analysis. The framework mainly includes data acquisition module, preprocessing module, emotion analysis module and result display module. The data acquisition module is responsible for grabbing post data from the selected social media platform, which is realized by API interface calling technology. The preprocessing module uses NLP tool library to clean, segment and remove stop words from the original text. The emotional analysis module is the core, which uses the word frequency statistics method and the emotional dictionary to calculate the emotional score of each post and further analyze the overall emotional trend. The results display module intuitively displays the distribution and dynamic changes of emotions through charts and visualization tools, which is convenient for users to understand and make decisions.

When implementing this technical framework, Python is chosen as the main programming language, because it has rich NLP libraries and powerful data processing capabilities. In order to improve the development efficiency and the maintainability of the code, this article also adopts modular design, each module encapsulates specific functions and can be tested and optimized independently. Through this technical framework, it aims to realize a fast, accurate and easy-to-deploy social media sentiment analysis system, which provides strong support for research and application in related fields.

3. Experimental design and implementation

In order to test the reliability of sentiment analysis model, it is necessary to collect a large number of real and representative social media data. This article chooses Weibo as the data source, because it not only has a huge user base, but also covers a wide range of topics and emotional expressions. The data collection process is carried out through the API interface provided by Weibo government, and we set a specific time range and keyword filtering conditions to ensure the relevance and timeliness of the data. After obtaining the original data, this article has done a series of preprocessing work, including removing duplicate posts, cleaning up irrelevant information, unifying text format and so on. Next, the NLTK library is used to segment the text, remove the stop words, and extract the stem, so as to reduce the data dimension and noise and lay a solid foundation for the subsequent emotional analysis. Through the statistics and analysis of the emotional scores output by the model, this article obtains rich experimental results.

On the whole, the distribution of emotions shows a certain regularity (as shown in Table 1). This verifies the effectiveness of the model in capturing emotional trends.

Table 1 Proportion of Emotional Distribution During Specific Festivals or Social Events

Time Period	Proportion of Positive Emotions	Proportion of Neutral Emotions	Proportion of Negative Emotions
During Spring Festival	60%	30%	10%
During National Day	55%	35%	10%
During a Certain Social Event	45%	30%	25%
Normal Periods	40%	40%	20%

As can be seen from Table 1, during certain festivals (Spring Festival, National Day) or social events, the proportion of positive emotions increased significantly. Among them, the proportion of positive emotions reached 60% during the Spring Festival and 55% during the National Day, which was significantly improved compared with 40% in normal times. This verifies the effectiveness of the model in capturing emotional trends, and also reflects that people are more inclined to express positive emotions during festivals or specific social events.

In this section, several specific topics are selected for in-depth analysis, and it is found that the model can accurately identify the dominant emotions in the topics (as shown in Table 2).

Table 2 Dominant Emotional Distribution for Specific Topics

Topic	Proportion of Positive Emotions	Proportion of Neutral Emotions	Proportion of Negative Emotions
Launch of New Product A	70%	20%	10%
Launch of New Product B	30%	20%	50%
Discussion on Hot Event C	45%	35%	20%
Policy Change D	25%	40%	35%

In the topic of new product release, A, the proportion of positive emotion is as high as 70%, which shows that users generally hold a positive attitude towards the product. In the topic of new product release B, the proportion of negative emotions reached 50%, reflecting that user have more dissatisfaction or criticism of the product. This different emotional distribution provides valuable user feedback for businesses and helps them better understand market demand and user preferences.

This article also compares the performance of the model in different time periods (as shown in Table 3), and finds that the model is highly sensitive to emotional changes and can capture the user's emotional fluctuations in time.

Table 3 Performance of the Model in Emotion Analysis Across Different Time Periods

Time Period	Accuracy	Recall	F1 Score
Week 1	80%	75%	0.78
Week 2	85%	80%	0.82
Week 3	78%	76%	0.77
Week 4	82%	79%	0.81

The accuracy of the first week was 80%, and the accuracy of the fourth week was 82%, with little change. This shows that the model is highly sensitive to emotional changes, can capture the user's emotional fluctuations in time, and maintain stable performance in different time periods. All the above results show that although the model is based on simple word frequency statistics method, it shows good performance and stability in practical application.

4. Conclusions

By using the basic NLP technology, especially the word frequency statistics method, this study successfully constructed an emotional analysis model for social media posts. By collecting, preprocessing and analyzing a large amount of data on Weibo platform, this article verifies the effectiveness of the model in identifying users' emotional tendencies and capturing emotional trends. The experimental results show that although the model design is relatively simple, it shows good performance and stability in practical application, which can provide valuable reference for the research and application in related fields. This study also reveals some characteristics of social media users' emotional distribution, such as timeliness of emotional tendency, topic relevance, etc. These findings help us to understand user behavior and emotional expression on social media more deeply.

With the continuous development and application of social media sentiment analysis technology, its social impact and ethical issues have become increasingly prominent. Therefore, while promoting the development of sentiment analysis technology, we must attach great importance to its ethical norms and legal constraints to ensure the legal and compliant use of technology. This includes measures such as strengthening data protection, improving algorithm transparency, and establishing user feedback mechanism to protect users' rights and interests and social public interests. At the same time, we need to strengthen interdisciplinary cooperation and public education, improve the society's cognition and understanding of affective analysis technology, and promote its healthy and sustainable development.

References

- [1] Yang Tengfei, Xie Jibo, Yan Dongchuan, et al. Emotional information extraction from social media based on deep learning and its application in disaster analysis [J]. *Geography and Geographic Information Science*, 2020,36(02):62-68.
- [2] Li Shuxing, Hu Huijun, Liu Maofu. Emotional analysis of social media images based on linguistic consistency [J]. *China Science and Technology Paper*, 2023, 18(3):322-329.
- [3] Zhang Weisheng, Wang Zhongqing, Li Shoushan, et al. Classification of emotions and intentions based on dialogue structure and joint learning [J]. *journal of chinese information*, 2020, 34(8):105-112.
- [4] Yuan Jingling, Ding Yuanyuan, Sheng Deming, et al. Emotional analysis model of image text based on visual attention [J]. *Computer Science*, 2022,49(01):219-224.
- [5] Dai Hongliang, Zhong Guojin, You Zhiming, et al. Analysis and integration method of public opinion and emotion big data based on Spark [J]. *Computer Science*, 2021,48(09):118-124.
- [6] Tan Chunhui, Chen Xiaoqi, Liang Yuanliang, et al. Emotional Analysis of Social Media Onlookers in Privacy Leaks [J]. *Information Science*, 2023,41(03):8-18.
- [7] Wang Xiao, Dong Didi, Chen Sijing, et al. Study on the theme distribution characteristics of social media and its influence on emotional tendency [J]. *Information Science*, 2023, 41(11):62-71.
- [8] Yang Changzheng. Research on the influence of social media users' stickiness and emotional load on information symbiosis behavior in emergencies [J]. *journal of the china society for scientific and technical information*, 2021, 40(6):640-655.
- [9] Feng Zeqi, Peng Xia, Wu Yachao. Tourists' Emotional Perception Based on Social Media Data Mining [J]. *Geography and Geographic Information Science*, 2022, 38(1):31-36.