IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 24, 2019, accepted December 12, 2019, date of publication December 18, 2019,
date of current version December 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2960655

Chinese Micro-Blog Sentiment Analysis Based on Multiple Sentiment Dictionaries and Semantic Rule Sets

JIESHENG WU[1], KUI LU[1], SHUZHI SU[1], AND SHIBING WANG[2]
[1]School of Computer Technology and Engineering, Anhui University of Science and Technology, Huainan 232001, China
[2]School of Computer and Engineering, Fuyang Normal University, Fuyang 236037, China

Corresponding author: Jiesheng Wu (wjs_19960317@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61806006, and in part by the Natural Science Research Project of Colleges and Universities in Anhui Province under Grant KJ2018A0083.

ABSTRACT Sentiment analysis of Chinese micro-blog based on sentiment dictionary has become a challenging research subject in the field of artificial intelligence. However, due to insufficient sentiment words, Chinese micro-blog sentiment analysis is difficult to process high accuracy. Aimed at this issue, we propose a method for constructing multiple sentiment dictionaries, which mainly constructs original sentiment dictionary, emoji dictionary, and other related dictionaries. Among them, we have innovatively constructed a Chinese micro-blog new word sentiment dictionary. Multiple sentiment dictionaries increase the coverage of sentiment words. At the same time, we further analyze semantic rule sets between Chinese micro-blog texts and take the inter-sentence analysis rules and sentence pattern analysis rules into the sentiment analysis of Chinese micro-blog, which further improves the accuracy of Chinese micro-blog sentiment analysis. Finally, based on the method of multiple sentiment dictionaries and semantic rule sets, we propose an algorithm for Chinese micro-blog sentiment calculation from complex sentences to clauses and then from clauses to words, and finally combined with the emoji. This algorithm can accurately classify Chinese micro-blog into positive Chinese micro-blog, negative Chinese micro-blog, and neutral Chinese micro-blog. The experimental results show that this method has greatly improved the sentiment analysis of Chinese micro-blog.

INDEX TERMS Sentiment analysis, sentiment dictionary, sentiment calculation, Chinese micro-blog, semantic rule sets.

I. INTRODUCTION

With the popularity of the mobile Internet in recent years, Chinese micro-blog has become a new social media platform for the public, which has become a social tool that people use every day. On Chinese micro-blog, thousands of people use it to express their opinions, suggestions, or feelings on some current hot topics. The subjective Chinese micro-blog news published by these people involves all aspects of life, which contains rich sentiment information. More and more institutions use sentiment information for analysis and detection to make decisions and judgments for the future.

However, how to fully extract valuable sentiment information from massive Chinese micro-blog data is our current research topic. Due to the flexibility, randomness or diversity of Chinese micro-blog data, it is impossible to rely on the manual collection and sorting out the sentiment information contained therein. In recent years, there are computer technologies to mine and analyze sentiment information in text data [1], [2]. Therefore, this paper proposes a method based on multiple sentiment dictionaries and semantic rule sets to analyze Chinese micro-blog. Obviously, the sentiment analysis of Chinese micro-blog can not use the sentiment dictionary developed by the English language. Therefore, this paper constructs multiple Chinese language sentiment dictionaries, including an original sentiment dictionary, an adverb dictionary, a conjunction dictionary, an emoji dictionary, a negative and double negative dictionary. At the same time, since Chinese micro-blog is a real-time social media platform, there are often some unknown new online words, so a Chinese micro-blog new word sentiment dictionary is constructed. These sentiment dictionaries increase the coverage of sentiment words and help to improve the sentiment analysis of Chinese micro-blog. Finally, combined with the semantic rule sets between the Chinese micro-blog texts to calculate the sentiment value, according to the magnitude of the sentiment

The associate editor coordinating the review of this manuscript and approving it for publication was Rashid Mehmood.

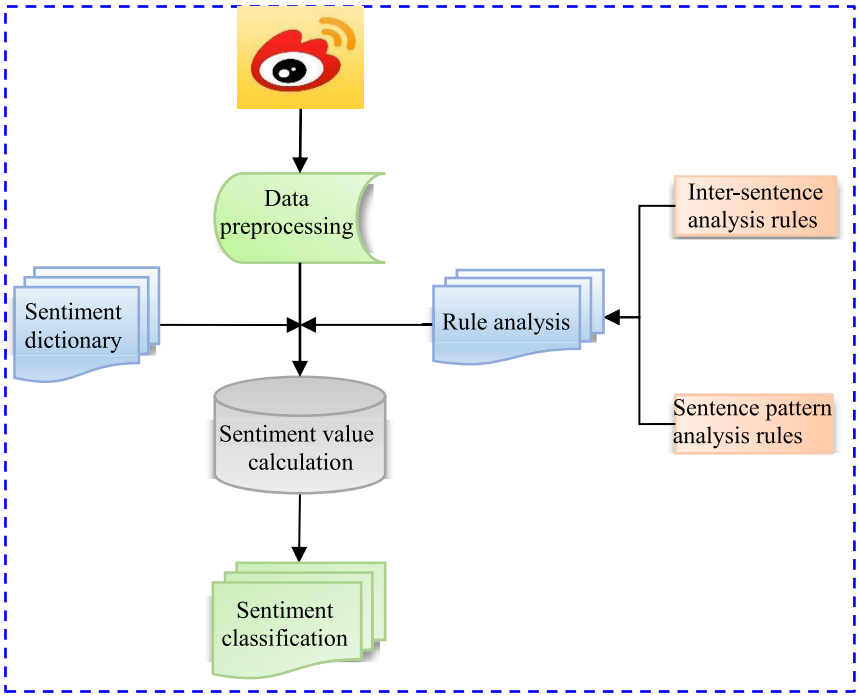


FIGURE 1. The flow chart of Chinese micro-blog sentiment analysis based on multiple sentiment dictionaries and semantic rule sets.

value, Chinese micro-blog is divided into positive Chinese micro-blog, neutral Chinese micro-blog, and negative Chinese micro-blog. The flow chart of Chinese micro-blog sentiment analysis based on multiple sentiment dictionaries and semantic rule sets is shown in Figure 1.

The contributions of this method mainly consist of the following three aspects:

A. CONSTRUCTION OF MULTIPLE SENTIMENT DICTIONARIES

Firstly, get Chinese micro-blog data. Since the obtained Chinese micro-blog data is disorderly, it is necessary to preprocess it, and then it is divided into corresponding Chinese texts, emojis and so on. Then construct multiple sentiment dictionaries, including an original sentiment dictionary, an adverb dictionary, a conjunction dictionary, an emoji dictionary, a negative and double negative dictionary, and a Chinese micro-blog new word sentiment dictionary, among which the Chinese micro-blog new word sentiment dictionary it is a key point of this method. It is necessary to identify the sentiment new words, and then construct the Chinese micro-blog new word sentiment dictionary for these sentiment new words. Finally, give each word in the sentiment dictionary a certain sentiment value.

B. ANALYSIS OF SEMANTIC RULE SETS

Chinese micro-blogs are mainly composed of Chinese texts except for a few elements such as emojis. As we all know, the text is a combination of one or more sentences. Although the sentence is composed of words, its structure is regular, so it is very important to analyze the semantic rules. In Chinese text, a text is composed of multiple complex sentences, and a complex sentence is composed of multiple single sentences. The single sentence that forms a complex sentence is also called a clause. The analysis of the semantic rules of Chinese micro-blog text mainly analyzes the sentence pattern analysis rules and inter-sentence analysis rules of each sentence in the text. The combined analysis of the two rules can more accurately analyze the sentiment of Chinese micro-blog.

C. SENTIMENT VALUE CALCULATION OF CHINESE MICRO-BLOG AND SENTIMENT CLASSIFICATION

The sentiment dictionary matching is performed on each word and emoji, and the sentiment value is calculated according to the weighted summation of the sentiment values. Then, based on this, the semantic rules between the Chinese micro-blog texts are considered. In this article, we design an algorithm of sentiment calculation, which calculates the sentiment value of Chinese micro-blog from complex sentences to clauses and then from clauses to words, and the final sentiment value of Chinese micro-blog is obtained. Finally, the Chinese micro-blog is divided into positive Chinese micro-blog, neutral Chinese micro-blog, and negative Chinese micro-blog.

In this paper, the sentiment analysis of Chinese micro-blog will be carried out in order from the above three aspects, and the argument analysis will be given in combination with the experiment.

II. RELATED WORKS

Understanding the sentiments contained in the text is an important goal of the development of artificial intelligence. The analysis of sentiments is closely related to the development of human beings. At present, many scholars have invented many methods to analyze the emotions of texts [3], [4]. This section mainly introduces the methods of sentiment analysis in recent years. Most of the methods of using sentiment dictionary for sentiment analysis are convenient, because it is easy to deal with fine-grained text sentiments, and then the method of using machine learning for sentiment analysis is mostly, its role faster and more accurate, more suitable for text in the chapter class.

A. SENTIMENT ANALYSIS USING SENTIMENT DICTIONARY

As early as the 20th century, text sentiment analysis in other languages had already started, Riloff and Shepherd conducted a dictionary study [5]. Littman and Turney used the ISA algorithm to perform sentiment analysis on the text, provided that the extended dictionary was used and the experimental results were increased to 74% [6]. In the latest sentiment analysis work, Valdivia and others extracted and compared the polarity and viewpoints of Tripadvisor's comments [7], and achieved some good results. Diamantini et al. used a dictionary-based approach to analyze social network content and introduced methods for eliminating word ambiguity and negating associations [8]. Sailunaz et al. directly analyzed the sentiment of twitter, mainly dealing with the problem of extracting sentiments and emotions [9]. In recent years, Chinese research scholars have gradually developed three Chinese sentiment dictionaries, namely the sentiment dictionary developed by the Dalian University of Technology [10], the sentiment dictionary developed by Taiwan University [11], and the sentiment dictionary developed by HowNet [12]. Taboada and Brooke et al. proposed a method for extracting features using a sentiment dictionary and used it in text sentiment analysis [13]. Dragut et al. automatically construct a domain sentiment dictionary with different vocabulary polarities in multiple sentiment dictionaries [14]. Feng et al. proposed a mutual refinement sorting model for word emoticons for learning sentiment dictionaries based on graphical expressions in data [15]. For the sentiment analysis of Chinese micro-blog, Li et al. proposed a method but performed sentiment analysis based on the sentiment dictionary of the two languages [16]. Wu et al. constructed a sentiment dictionary belonging to the field of commodity shopping, including nouns, adjectives, adverbs and so on. The words are sorted by the TF-IDF algorithm to obtain a sentiment dictionary of the field [17].

B. SENTIMENT ANALYSIS USING MACHINE LEARNING

With machine learning, you first need to manually mark part of the data to build a data training set and extract features to train the classification model. Finally, the trained classification model classifies the unknown data to achieve sentiment classification of the data [18], [19]. At present, the most mature machine learning classification methods include the maximum entropy model [20], the naive Bayesian method [21] and the support vector machine method [22]. Chaffer et al. implemented a text-based sentiment approach based on supervised learning [23]. McDonald and Tackstrom proposed a model of shallow variable structure based on the conditional random field model to perform sentiment analysis on the text [24].

In recent years, the continuous development of neural networks and deep learning, and most scholars use them for text sentiment analysis is more and more common. The deep learning proposed by Geoffrey Hinton et al. in 2006 provides new ideas for solving these issues [25], [26]. Sun et al. combined the deep belief network and the feature extraction method to obtain an extended feature extraction method to solve the problem of sparse short text extraction features [27]. Hu et al. built a keyword lexicon based on the LSTM model, which can help to mine the potential language in the text, and further improve the correctness of the text sentiment orientation judgment [28]. Giatsoglou, Vozails, and Diamantaras et al. proposed a method that uses Word2Vec to combine contextual information with a sentiment dictionary to greatly enhance the role of sentiment analysis [29]. Although the word vector embedding technique considers the context of the word but ignores the sentiment of the whole text, Tang, Wei, and Qin et al. used the contextual relationship of words to solve the sentiment analysis of text [30].

C. OTHER METHODS

Sentiment dictionary and machine learning are widely used by many scholars in studying text sentiments. In fact, there are many other methods besides these two methods. Some scholars have applied these methods to the text sentiment analysis task. Gao Ge et al. used the concept hierarchy and duality theory in the conceptual hierarchy network theory HNC and used symbols to generate some new words in symbols, which is convenient for calculating sentiment values [31]. For texts with subjective nature such as product reviews, Pandey and Rajpoot et al. proposed an algorithm based on K-values and search that can perform sentiment analysis on the text [32]. Some researchers have noticed other features that affect sentiment analysis. For example, Purver et al. studied the classification of six basic sentiments for Chinese micro-blog and confirmed it by far-supervised learning [33]. Usually, when emotionally classifying text, it is assumed that the individual who published the text is independent, but Grandi et al. believe that processing a set of data in this way may lead to inaccurate or incorrect results, thus proposing a combination of personal sentiments and preference information. Algorithms are used to provide a more accurate analysis of collective sentiments [34]. Eliacik et al. also agree that individuals are not independent. The degree of trust an individual has in a group affects the results of sentiment classification, so they take social

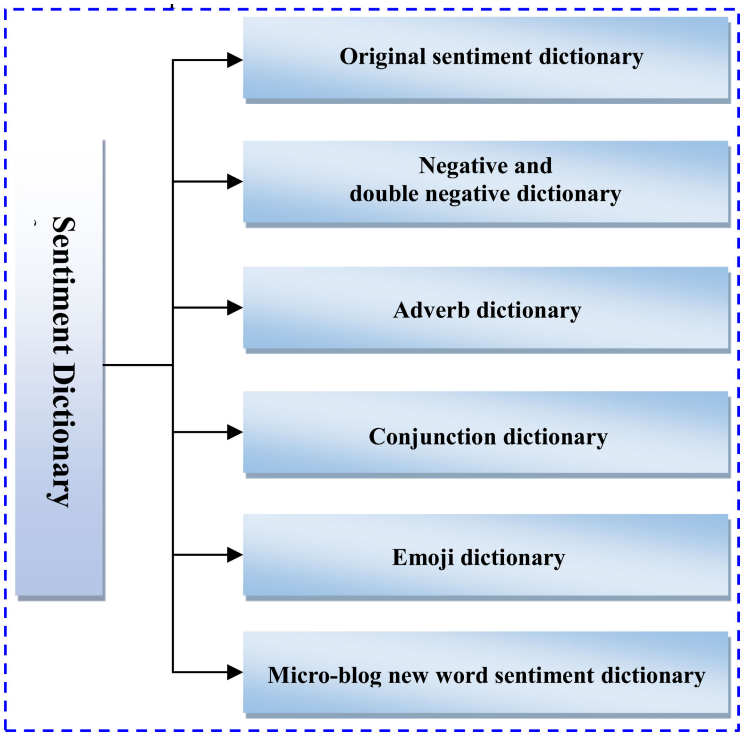


FIGURE 2. The structure of multiple sentiment dictionaries.

networks into account and focus on influential users [35]. At present, Adversarial training has become an active research topic and has been vigorously developed in the field of artificial intelligence. Therefore, Han et al. commented that in the next generation of sentiment computing, Adversarial training will be applied to sentiment analysis and introduce a series of future The direction of sentiment analysis research [36], [37].

III. CONSTRUCTION OF SENTIMENT DICTIONARY

This section introduces how to integrate and optimize the original sentiment dictionary, construct multiple sentiment dictionaries and a Chinese micro-blog new word sentiment dictionary to conduct Chinese micro-blog sentiment analysis. The structure of multiple sentiment dictionaries is shown in Figure 2.

A. CHINESE MICRO-BLOG PREPROCESSING

Chinese micro-blog has the characteristics of complexity, randomness, and colloquialism. Before the sentiment analysis, it needs to use pre-processing to filter the data. Therefore, the following steps are taken:

1) DELETING THE CHINESE MICRO-BLOG LINK, ANIMATION, PICTURE, AND VIDEO; DELETING "@USERNAME" AND "#TOPIC#" IN THE CHINESE MICRO-BLOG

These content have little to do with sentiment analysis, so delete them.

2) TRANSLATING TRADITIONAL CHINESE AND ENGLISH INTO SIMPLIFIED CHINESE CHARACTERS

This is for the convenience of follow-up work because the Chinese micro-blog text is composed of multiple languages, so the unified transformation into simplified Chinese characters is conducive to sentiment analysis.

3) RESERVED THE EMOJIS IN THE CHINESE MICRO-BLOG

Because emojis are external manifestations of states, related to sentiments, they can participate in sentiment value calculations.

4) WORD SEGMENTATION

Since Chinese micro-blog is mainly composed of text, it is segmented by the word segmentation tool.

5) DELETING STOP WORDS IN THE TEXT

Such as the auxiliary word "的", the pronoun "她,他" and the like.

After the preprocessing is completed, the Chinese micro-blog is divided into strings of words and several emojis. For example, "我校运动员小明在短跑男子 500 米决赛中夺冠 [Our school athlete Xiao Ming won the sprint men's 500m final]." It will become {我校, 运动员, 小明, 在, 短 跑, 男子,500, 500, 米, 决赛, 中, 夺冠}. At present, the Chinese sentiment dictionary does not have a complete and mature sentiment dictionary. Therefore, in addition to constructing the existing original sentiment dictionary, we also construct a negative

TABLE 1. An example of the original sentiment dictionary.

Sentiment Word	Part of Speech	Sentiment Value	Orientation
Hopeless(令人绝望的)	Adj	9	2
Beauty(美丽)	Noun	5	1
Happy(开心的)	Adj	5	1
Climb(爬)	Verb	0	0

TABLE 2. An example of the negative and double negative dictionary.

Type of Word	Word	Value
Negative word	Seldom(很少),never(从不),scarcely(几乎没有)…	-1
Rhetorical word	Can't you see(难道你看不到), Why not(为什么不), Why even bother(何必多此一举)…	-2
Double negative word	Have to(不得不), really not(并非), no not than(无不大于)…	1

and double negative dictionary, an adverb dictionary, a conjunction dictionary, and an emoji dictionary.

B. CONSTRUCTION OF MULTIPLE SENTIMENT DICTIONARIES

1) ORIGINAL SENTIMENT DICTIONARY

The original sentiment dictionary used in this paper is also the sentiment dictionary developed by the Dalian University of Technology, which contains a total of 27,476 words [10]. The words in the original sentiment dictionary are composed of three types: positive sentiment words, neutral sentiment words, and negative sentiment words. In this paper, the number "1" is used to represent a positive sentiment word, "0" is used to represent a neutral sentiment word, and "2" is used to represent a negative sentiment word. At the same time, the sentiment words are divided into five levels according to the magnitude of their sentiment values, which are respectively represented by the numbers "1, 3, 5, 7, 9", where "1" means the smallest and "9" means the maximum. The sentiment value of the neutral sentiment word is "0". An example of the original sentiment dictionary is shown in Table 1.

2) NEGATIVE AND DOUBLE NEGATIVE DICTIONARY

The negative and double negative word dictionary includes three types of words, respectively negative words, rhetorical words, and double negative words. Their roles are different. Negative words and rhetorical words change the sentiment orientation of the sentiment words they modify, but double negative words equal positive words and do not change the sentiment orientation of the modified sentiment words. This paper summarizes 25 words, which constitute a negative and double negative dictionary, and their values are represented by the numbers "−1, −2, 1". An example of the negative and double negative dictionary is shown in Table 2.

3) ADVERB DICTIONARY

The adverb dictionary is a dictionary developed by HowNet [12], with a total of 219 degree adverbs. These adverbs are divided into six levels, the levels are super, extreme, very, relatively, slightly, and insufficiently. The adverb is to enlarge the sentiment value of the sentiment word it has modified by a certain multiple. Therefore, the numbers "3, 2.5, 2, 1.5, 1, 0.5" are used to represent the expansion multiple of each level. An example of the adverb dictionary is shown in Table 3.

4) CONJUNCTION DICTIONARY

Conjunctions are to connect sentences to sentences, mainly including transition relationship conjunctions, concession relationship conjunctions, progressive relationship conjunctions, causality relationship conjunctions, and hypothetical relationships conjunctions. This paper collects some commonly used conjunctions to construct a conjunction dictionary, and assigns a certain value according to its relationship attribute, using the numbers "0.5, 1.5, 2, 1.5, 1" represents the value. An example of the conjunction dictionary is shown in Table 4.

5) EMOJI DICTIONARY

Chinese micro-blog emoji has a strong sentiment orientation in Chinese micro-blog, which has a certain effect on

TABLE 3. An example of the adverb dictionary.

Level	Adverb word	Multiple	Number
Super	Especially(特别),excessively(过度),unduly(过分)…	3	30
Extreme	Absolutely(绝对), extremely(极度), too(过于)...	2.5	69
Very	Really(的确), but(不过), only(唯一)...	2	42
Relatively	Fully(充分), big deal(大不了), completely(十分)...	1.5	37
Slightly	Slightly(稍微), less(少), a little(一点)...	1	29
Insufficiently	Hardly(毫不), weakly(弱), not very(不堪)...	0.5	12

TABLE 4. An example of the conjunction dictionary.

Relationship	Conjunction	Value	Number
Transition	But(但是),however(然而),while(而)…	0.5	10
Concession	Even if(即使), even though(尽管), although(虽然)...	1.5	10
Progressive	Even(甚至), and(并且), as well as(以及)...	2	9
Causality	So(所以), because(因为), due to(由于)...	1.5	14
Hypothetical	If(如果), in case(假使), suppose(假使结果会…)...	1	12

judging the sentiment orientation of Chinese micro-blog. In this paper, we use the partial emojis with high frequency in Chinese micro-blog to construct the emoji dictionary, a total of 217 emojis, and divide these emojis into five levels, and give each level of emojis a certain sentiment value. An example of the emoji dictionary is shown in Table 5.

C. CHINESE MICRO-BLOG NEW WORD SENTIMENT DICTIONARY

As a social media platform, Chinese micro-blog always generates some new words on the Internet. Therefore, it is crucial to identify and mine new words, and then collect these new words to construct a dictionary of new words. First, we should identify new words based on statistical methods, and then perform sentiment recognition in new words. The specific methods are as follows:

1) USING STATISTICAL METHODS TO MINING NEW WORDS

Definition 1 (String Frequency): In all possible string sets $W = \{w_1, \ldots, w_i, \ldots, w_n\}$ of the text field D, the definition string frequency $N(w_i)$ represents the number $w_i \in W$ occurrences in the text field D.

$N(w_i)$ is one of the criteria for new words. When judging whether a string constitutes a new word in the text, the first thing to consider is the frequency at which it appears in the text. When a string is repeated more than the value, it is considered to constitute a new word.

Definition 2 (Internal Coupling): The string w is divided into all possible combinations of two substrings:

$\{(w_{11}, w_{12}), \ldots, (w_{i1}, w_{i2}), \ldots, (w_{n1}, w_{n2})\}$, calculated by the following formula:

$$IC(w) = \frac{1}{n}\sum_{i=1}^{n}\frac{P(w)}{P(w_{i1}) \times P(w_{i2})} \tag{1}$$

The resulting $IC(w)$ is referred to as the internal coupling degree of the string w. Where $P(w)$ represents the probability of occurrence of the string w in the text field D, calculated by the formula:

$$P(w) = \frac{N(w)}{N_D}, \tag{2}$$

where $N(w)$ represents the number of occurrences of the w string in the text field D, and N_D represents the total number of words in the text field.

Whether a string can be worded or not is closely related to the internal tightness of the string. This degree of tightness can be measured by this degree of internal coupling.

Definition 3 (Neighbor Word Set Information Entropy): In the text field D, all sets of words $C = \{c_1, \ldots, c_i, \ldots, c_n\}$ that may appear to the left or right of the string w are called the left or right neighbor word set of w.

For $C = \{c_1, c_2, \ldots, c_i, \ldots, c_n\}$ through the following formula:

$$IE(w) = -\sum_{i=1}^{n}\frac{n_i}{n}\log\frac{n_i}{n}, \tag{3}$$

the calculated $IE(w)$ is called the left or right neighbor word set information entropy of w. Where n_i represents the number of times c_i appears as the left or right neighbor of w, and n

TABLE 5. An example of the emoji dictionary.

emoji	Sentiment Value	Number
	2	33
	1	42
	0	77
	-1	33
	-2	32

represents the sum of the number of occurrences of all words in the neighbor word sets C as the left or right neighbor of w.

Then calculate by the following formula:

$$IE_{\min}(w) = \min\{IE_{left}(w), IE_{right}(w)\} \tag{4}$$

The smaller value of the left and right neighbor word set information entropy is obtained as a criterion for judging whether the character string w can be flexibly applied in the text by one word.

The neighbor word set information entropy can well measure the external application ability of a string. The neighbor word set represents a set of single words that may appear on both sides of a word, and information entropy can introduce the possibility of an event occurring. If the possibility is greater, the greater the amount of information that needs to be understood, the higher the information entropy. If a string is indeed a potential word, its use in the text is more diverse, and the neighbor words appearing on both sides are more uncertain. This uncertainty can be calculated by information entropy.

Whether a string can be worded is related to the above three definitions. The Chinese micro-blog text is a text composed of a series of words. First, we use a long string to represent the Chinese micro-blog text and set the length of the new word to a value. This article is set to 7. And for each of the three definitions above, each definition must determine a parameter threshold. If no condition is met, that is, the threshold range is exceeded, the string is not a word. Also, if the word is not found in an existing dictionary, it will become a new word.

2) SENTIMENT ANALYSIS OF NEW WORDS USING IMPROVED PMI ALGORITHM

Through the above methods, new words can be identified and excavated, but the sentiment orientation of these words needs to be further recognized, thus constructing a Chinese micro-blog new word sentiment dictionary. We need to count and sort the new words identified by the above methods according to the word frequency, and filter them from top to bottom to filter out words with strong sentiment orientation and high frequency of words as known words. How to use these words to judge the sentiment orientation of unknown new words?

In response to this issue, we use the PMI(Point-wise Mutual Information) algorithm. The basic idea of the algorithm is to count the probability that two words appear simultaneously in the text. If the probability is larger, the correlation is tighter and the degree of relevance is higher. The calculation formula for the mutual information between the two words w_1 and w_2:

$$PMI(w_1, w_2) = log_2 \frac{p(w_1, w_2)}{p)(w_1) \times p(w_2)}, \tag{5}$$

formula (5) is to judge the degree of similarity of a pair of words. This formula can be used to judge the degree of similarity between unknown new words and known words. If the result is large, it means that the two words have a high degree of similarity and sentiment orientation. However, it is not acceptable to use this formula to judge the sentiment orientation of a word. On Chinese micro-blog, there are many unknown new words. Therefore, this paper screens out a certain number of positive and negative sentiment words to form a positive sentiment word set W_p and a negative sentiment word set W_n according to the frequency of the string, and improves the formula (5) to obtain the following formula for judging the sentiment orientation of the unknown new word:

$$Sen\_PMI(w) = \sum_{w_p \in W_P} PMI(w, w_p) - \sum_{w_n \in W_N} PMI(w, w_n), \tag{6}$$

formula (6) is to judge the sentiment orientation of the new word. If its result is larger than 0, it is a positive new word; if its result is less than 0, it is a negative new word; if its result is equal to 0, it is a neutral new word.

The above method can identify new words and sentiment orientation judgments and finally construct a Chinese micro-blog new word sentiment dictionary. This paper identifies and mines 164 new words in 2018, and divides these words into 4 levels, and give each new word a certain sentiment value. An example of the Chinese micro-blog new word sentiment dictionary is shown in Table 6.

IV. SENTIMENT ANALYSIS OF THE SEMANTIC RULE SETS

The main part of Chinese micro-blog is text, which is composed of Chinese characters, and there are certain grammatical relations and semantic rules between the texts. This section explains the addition of the semantic rule sets between texts to the sentiment analysis of Chinese micro-blog, in which the semantic rule sets consists of two parts, the first is the inter-sentence analysis rules, and the second

TABLE 6. An example of the Chinese micro-blog new word sentiment dictionary.

New word	Sentiment Value	Number
Real incense(真香), skr, koi(锦鲤)...	2	18
Official Xuan(官宣), Buddhism(佛系), confirmed eyes(确认过眼神)...	1	40
Arrange(安排), cool(凉凉), big pig hoof(大猪蹄子)...	-1	65
Pothole(坑爹), Nima(尼玛), middle-aged greasy male(中年油腻男)...	-2	41

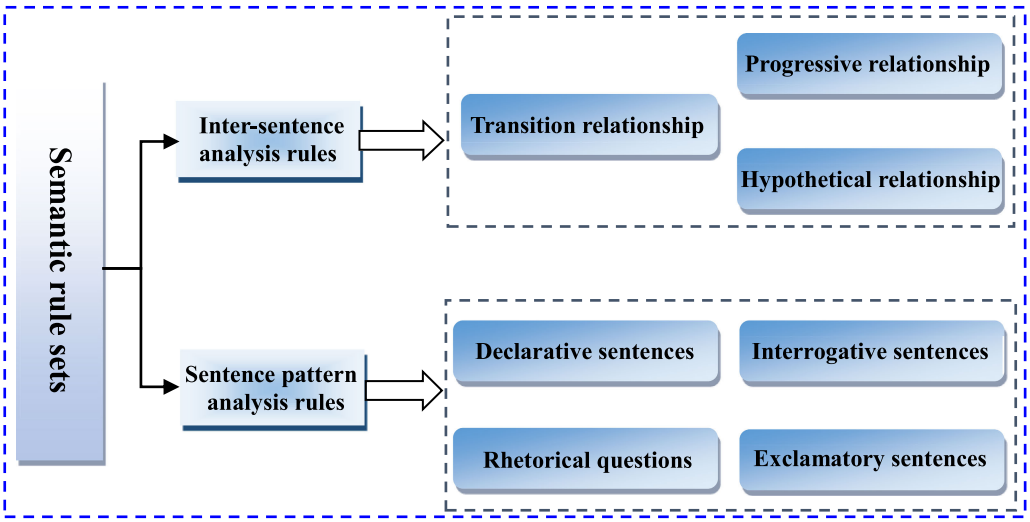


FIGURE 3. The architecture of the semantic rule sets.

is the sentence pattern analysis rules. The architecture of the semantic rule sets is shown in Figure 3.

A. INTER-SENTENCE ANALYSIS RULES

A Chinese micro-blog text can be cut into several complex sentences by punctuation. A complex sentence can be divided into several clauses. The inter-sentence analysis rule considers the relationship between clauses and clauses. There are three main types of inter-sentence relationships: the transition, progressive and hypothetical relationship. In this section, S is used to represent the entire complex sentence, and S_i is used to represent the individual clause of the complex sentence. The definition set $\{S_1, S_2, S_3, \ldots, S_i\}$ is the set of clauses that make up a single complex sentence, and R_i is used to represent the sentiment value of the inter-sentence analysis rule on the clause S_i.

1) TRANSITION RELATIONSHIP RULE

In the transition relationship, the sentiments are reversed before and after the transition. The sentiments of the clauses before the transition will be weakened, and the sentiments of the subsequent clauses will be highlighted. The latter clauses are opposite to the sentiment orientation of the previous clauses. The specific rules are defined as follows:

Case 1: If there is only a single transition word in the complex sentence S that belongs to the category "但是,可是 [But, but]", and it appears in the clause S_i, the sentiment value R_i of the clause before S_i is set to 0, and the sentiment value R_i of the clause after S_i is set to 1.

Case 2: If there is only a single transition word in the complex sentence S that belongs to the category "虽然,尽管 [Although, despite]", and it appears in the clause S_i, the sentiment value R_i of the clause before S_i is set to 1, and the sentiment value R_i of the clause after S_i is set to 0.

Case 3: If a pair of transition words belonging to the category "虽 然...但 是...[However...although...]" appears in the complex sentence S, and "但是 [but]" appears in the clause S_i, the sentiment value R_i of the clause before S_i is set to 0, and the sentiment value R_i of the clause after S_i is set to 1.

2) PROGRESSIVE RELATIONSHIP RULE

Progressive relationship, as the name suggests, in this relationship rule, each clause of the complex sentence gradually

enhances the sentiment according to the order from front to back. The rule definition is as follows:

If progressive conjunction appears in the complex sentence S, the sentiment value of each clause is:

$$R_i = 1, R_{i+1} = 1.5, \ldots, R_j = 1 + 0.5 \times (j - i).$$

3) HYPOTHETICAL RELATIONSHIP RULE

An assumption about the reality is called a hypothetical relationship. The sentiment it expresses is mainly in the first half of the complex sentence, while the sentiment in the latter half is relatively weak. For example: If A, then B. Then the complex sentence emphasizes content A. The specific rules are defined as follows:

Case 1: If there is no negative hypothetical conjunction in the complex sentence S, but hypothetical conjunction that belongs to the category "那么[Then]" appears, and it appears in the clause S_i, the sentiment value R_i of the clause before S_i is set to 1, the sentiment value R_i of the clause after S_i is set to 0.5.

Case 2: If there is a negative hypothetical conjunction in the complex sentence S, and hypothetical conjunction that belongs to the category "那么[Then]" appears, and it appears in the clause S_i, the sentiment value R_i of the clause before S_i is set to -1, the sentiment value R_i of the clause after S_i is set to -0.5.

The three inter-sentence analysis rules described above can affect the sentiment orientation of the entire Chinese micro-blog text, so they should be considered in the sentiment analysis, as for other inter-sentence analysis rules such as causality, juxtaposition, etc. The impact of sentiment analysis is negligible.

B. SENTENCE PATTERN ANALYSIS RULES

This section explains the influence of the sentence pattern of the complex sentence on the sentiment orientation of the entire text. This paper mainly discusses four common sentence patterns: interrogative complex sentences, rhetorical questions, declarative complex sentences, and exclamatory complex sentences. They often end with punctuation marks such as "?" "!" "。" "," A text is represented by D, and the text is divided into sentences, which are also called complex sentences. They are defined by a set as $\{D_1, \ldots, D_i, \ldots, D_n\}$, where a single complex sentence is represented by D_i, and T_i represents the sentiment value of the sentence pattern analysis rule on the complex sentence D_i. The specific rules are defined as follows:

Case 1: If there is a complex sentence D_i in the Chinese micro-blog text ends with an exclamation point "!", it means that the complex sentence is an exclamatory sentence and its sentiment value T_i is set to 1.5.

Case 2: If there is a complex sentence D_i in the Chinese micro-blog text ends with the question mark "?", it can be divided into two cases. The first case is that there is a rhetorical question word at the end of the complex sentence, then the complex sentence is a rhetorical question; the second case is that the complex sentence does not with a question mark "?", but there is a rhetorical question word at the end, then the complex sentence is also a rhetorical question. In both cases, the sentiment value of the complex sentence is set to -1.

Case 3: If there is a complex sentence D_i in the Chinese micro-blog text ends with the question mark "?" and there is no rhetorical question word at the end, it means that the complex sentence is an interrogative sentence and its sentiment value T_i is set to 0.

Case 4: If there is a complex sentence D_i in the Chinese micro-blog text ends with a punctuation mark "。" and other punctuation marks, it means that the complex sentence is a declarative sentence, and its sentiment value T_i is set to 1.

V. CHINESE MICRO-BLOG SENTIMENT CALCULATION

This section uses the basics of Sections III and IV, we can carry out the overall Chinese micro-blog sentiment calculation from words and emojis to sentences. The process of Chinese micro-blog sentiment calculation is shown in Figure 4.

In this section, D is used to represent the entire text, each complex sentence in the text is represented by D_i; at the same time, S is used to represent the corresponding single complex sentence, and S_i is the individual clause in the complex sentence. V represents the total sentiment value of the entire Chinese micro-blog, R_i is the sentiment value generated by the inter-sentence analysis rule of the clause, T_i is the sentiment value generated by the sentence pattern analysis rule of the complex sentence, and sen_i is the sentiment value obtained by the word matching the sentiment dictionary. The Chinese micro-blog sentiment calculation is as follows:

1) The formula for calculating the sentiment value $V(W_i)$ of the word is as follows:

$$V(W_i) = \begin{cases} N \times A \times sen_i \times 0.5 \\ A \times N \times sen_i \times 1 \end{cases} \quad (7)$$

In formula (7): k represents the number of negative words. If the relative positional relationship between negative words and degree adverbs is different, the sentiments are different. For example: "不太好看 [not too good-looking]" and "太不好看 [too bad-looking]", the expression of sentiment is completely different, it is obvious that the sentiment of the second sentence is much stronger than the sentiment of the first sentence. Therefore, if the degree adverb is in front of the negative word, multiply the formula (7) by 1; if the degree adverb is behind the negative word, multiply the formula (7) by 0.5.

N corresponds to a value of the negative and double negative dictionary, A corresponds to a multiple of the adverb dictionary, and sen_i corresponds to a sentiment value of the sentiment dictionary. W_i represents the sentiment word. A word is not only related to its own sentiment value, but also to the degree adverbs and negative words modified in front of it, so they should be taken into account the calculation.

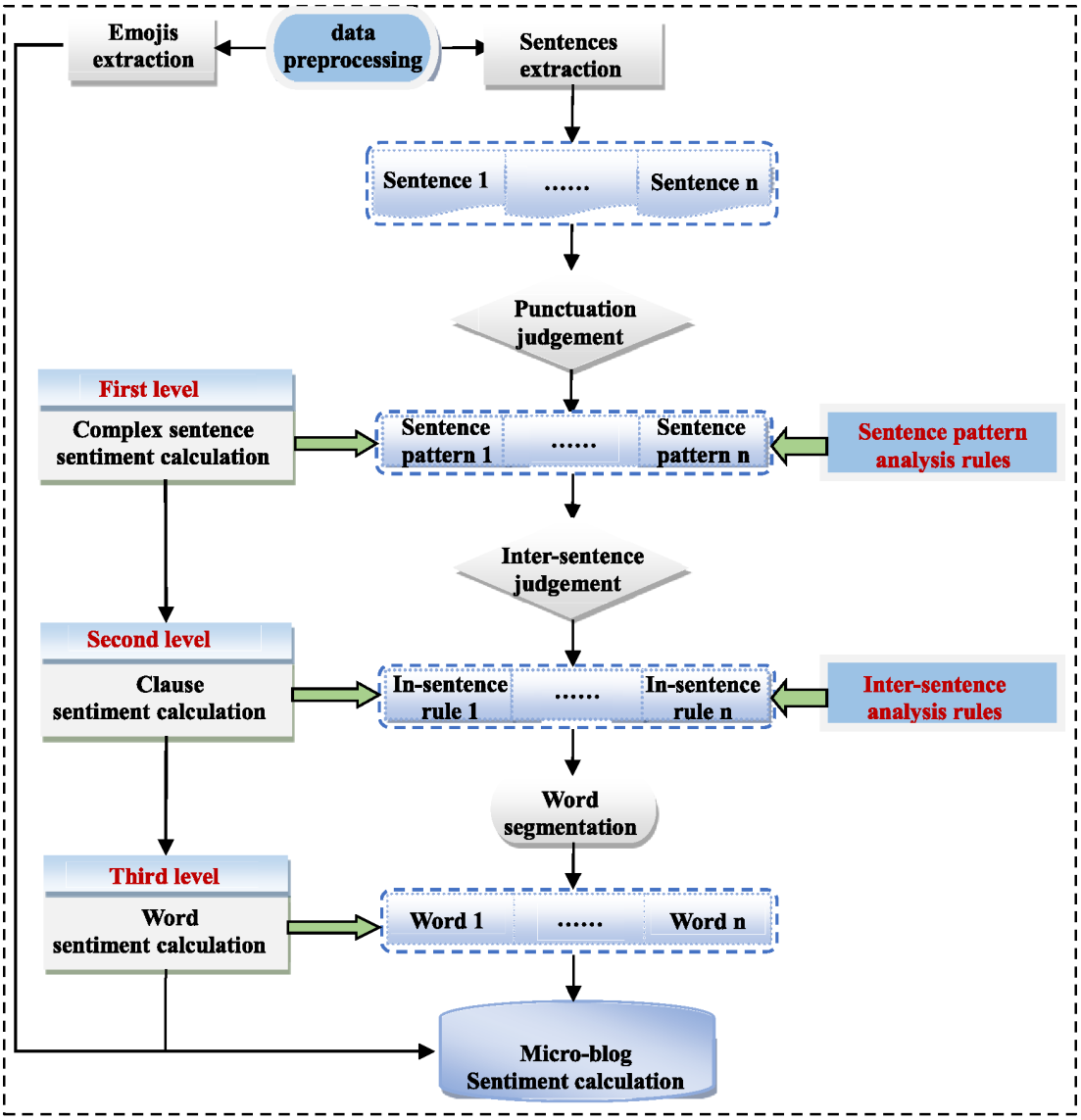


FIGURE 4. The process of Chinese micro-blog sentiment calculation.

2) The formula for calculating the sentiment value $V(S_i)$ of the clause is as follows:

$$V(S_i) = \sum_{i=1}^{n} V(W_i) \times R_i \tag{8}$$

In formula (8): $\sum_{i=1}^{n} V(W_i)$ is the sum of the sentiment value of each word in the clause, and then multiplied by the sentiment value of the inter-sentence analysis rule, and the final result is the sentiment value of the clause.

3) The formula for calculating the sentiment value $V(D_i)$ of the complex sentence is as follows:

$$V(D_i) = \sum_{i=1}^{n} V(S_i) \times T_i \tag{9}$$

In formula (9): $\sum_{i=1}^{n} V(S_i)$ is the sum of the sentiment value of each clause in the complex sentence, and then multiplied by the sentiment value of the sentence pattern analysis rule, and the final result is the sentiment value of the complex sentence.

4) The formula for calculating the sentiment value V_{text} of the Chinese micro-blog text is as follows:

$$V_{text} = \sum_{i=1}^{n} V(D_i) \tag{10}$$

In formula (10): $\sum_{i=1}^{n} V(D_i)$ is the sum of the sentiment value of each complex sentence. thereby obtaining the sentiment value of the Chinese micro-blog text.

5) The formula for calculating the sentiment value V_{emoji} of the emojis is as follows:

$$V_{emoji} = \frac{1}{n} \sum_{i=1}^{n} V(sen_i) \tag{11}$$

In formula (11), the emojis sentiment value is calculated as: first, the sentiment values obtained by matching each emoji to

the emoji dictionary are added and then divided by the total number of emojis, thereby obtaining the sentiment value of emojis.

6) The formula for calculating the sentiment value V of the Chinese micro-blog is as follows:

$$V = a \times V_{text} + b \times V_{emoji} \tag{12}$$

In formula (12), the sentiment value of the Chinese micro-blog is composed of two parts: the sentiment value of the Chinese micro-blog text and the sentiment value of the Chinese micro-blog emojis, where a represents the proportion of text in Chinese micro-blog, and b represents the proportion of emoji in Chinese micro-blog. By analyzing the calculation result V, if $V > 0$, it means that the sentiment orientation of the Chinese micro-blog is positive; if $V < 0$, it means that the sentiment orientation of the Chinese micro-blog is negative; if $V = 0$, it means that the sentiment orientation of the Chinese micro-blog is neutral.

According to the above, the algorithm of Chinese micro-blog sentiment analysis can be designed. The algorithm flow is shown in Algorithm 1:

VI. EXPERIMENTS

In this section, we will give specific experimental methods and results to verify the effectiveness of the proposed method.

A. THE EXPERIMENTAL METHOD

First, the data source crawls the two related topics on Chinese micro-blog through the crawler tool and then performs sentiment analysis on the data. The experimental steps are as follows:

1) *Get Experimental Data:* Use the crawler software to crawl the two hot topics on the Chinese micro-blog "#Short Video Rectangle#" and "#" I am not a drug god" burst red lead social hot discussion#" data.

2) *Manually Label the Sentiment Orientation of the Data:* Because the sentiment orientation of the data obtained is unknown, the sentiment orientation of the data needs to be manually labeled. First, randomly select three experimental students, and then let them alternately subjectively judge the sentiment orientation of the data, and finally the statistical results.

Algorithm 1 The Sentiment Analysis Algorithm for a Chinese Micro-Blog

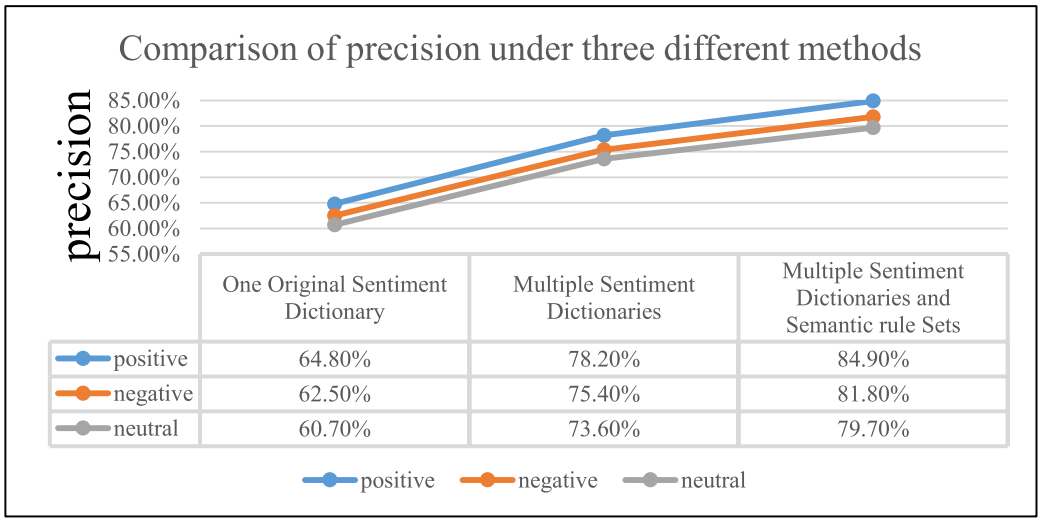
Input: a Chinese micro-blog

Output: the sentiment orientation of the Chinese micro-blog

$V = 0$;

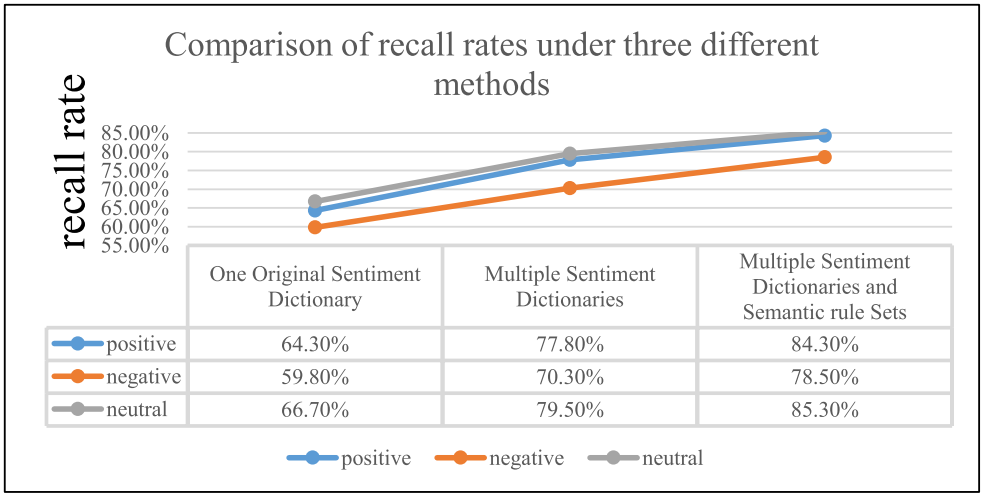
A Chinese micro-blog is divided into a Chinese micro-blog text and e emojis, in which the Chinese micro-blog text is cuted into n complex sentences, each individual complex sentence is cuted into m clauses, and each individual clause is cuted into o sentiment words;

```
for (i = 1; i++; i <= n)
  {   V_text = 0;
  V_text = V_text + V(D_i);
  for (j = 1; j++; j <= m)
    {   V(D_i) = 0;
    V(D_i) = [V(D_i) + V(S_j)]×T_i;
   for (k = 1; k++; k <= o)
    {   V(S_j) = 0;
      V(W_k) = N×A×sen_k;
      V(S_j) = [V(S_j) + V(W_k)]×R_j;
    }
  }
  }
for (l = 0; l++; l <= e)
  {V_e = 0; V_e = V_e + sen_l; }
V_emoji = V_e/e;
V = a×V_text+b×V_emoji;
if (V > 0)
  Output this is a positive Chinese micro-blog;
  else if (V < 0)
    Output this is a negative Chinese micro-blog;
      else
        Output this is a neutral Chinese micro-blog;
    end if
end if
```



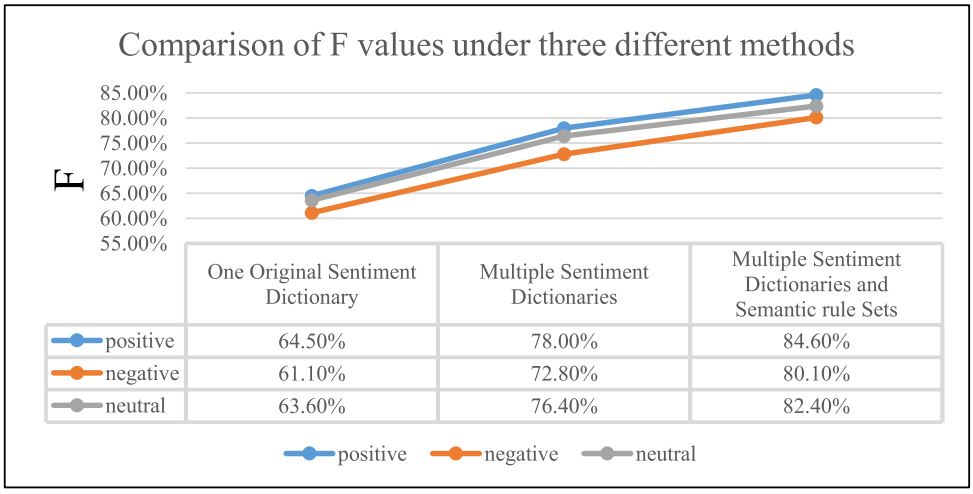
	One Original Sentiment Dictionary	Multiple Sentiment Dictionaries	Multiple Sentiment Dictionaries and Semantic rule Sets
positive	64.80%	78.20%	84.90%
negative	62.50%	75.40%	81.80%
neutral	60.70%	73.60%	79.70%

FIGURE 5. Comparison of precision under three different methods.



	One Original Sentiment Dictionary	Multiple Sentiment Dictionaries	Multiple Sentiment Dictionaries and Semantic rule Sets
positive	64.30%	77.80%	84.30%
negative	59.80%	70.30%	78.50%
neutral	66.70%	79.50%	85.30%

FIGURE 6. Comparison of recall rates under three different method.



	One Original Sentiment Dictionary	Multiple Sentiment Dictionaries	Multiple Sentiment Dictionaries and Semantic rule Sets
positive	64.50%	78.00%	84.60%
negative	61.10%	72.80%	80.10%
neutral	63.60%	76.40%	82.40%

FIGURE 7. Comparison of F values under three different methods.

3) *Construct Multiple Sentiment Dictionaries:* According to the corresponding method, the six sentiment dictionaries are constructed.

4) *Sentiment Analysis of Chinese Micro-Blog Topics:* Based on an original sentiment dictionary, based on multiple sentiment dictionaries and based on multiple sentiment

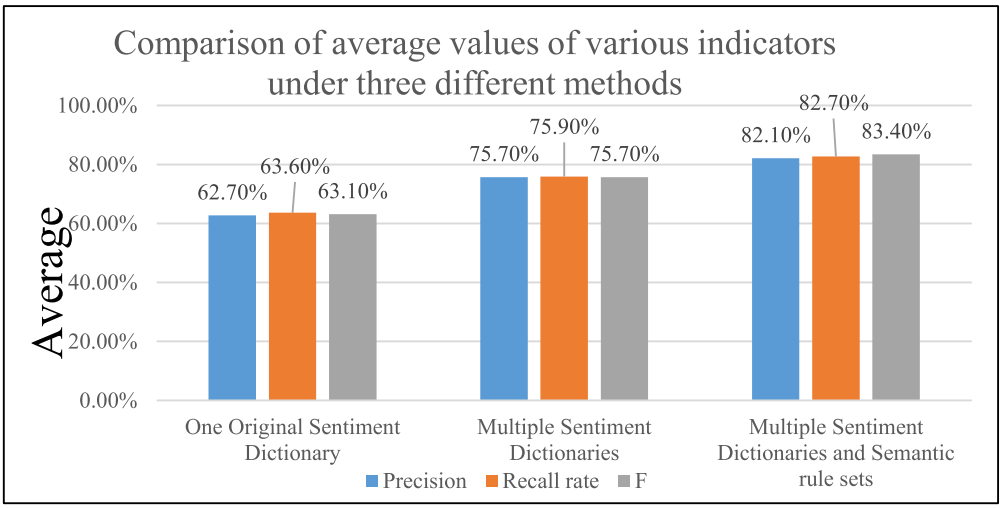
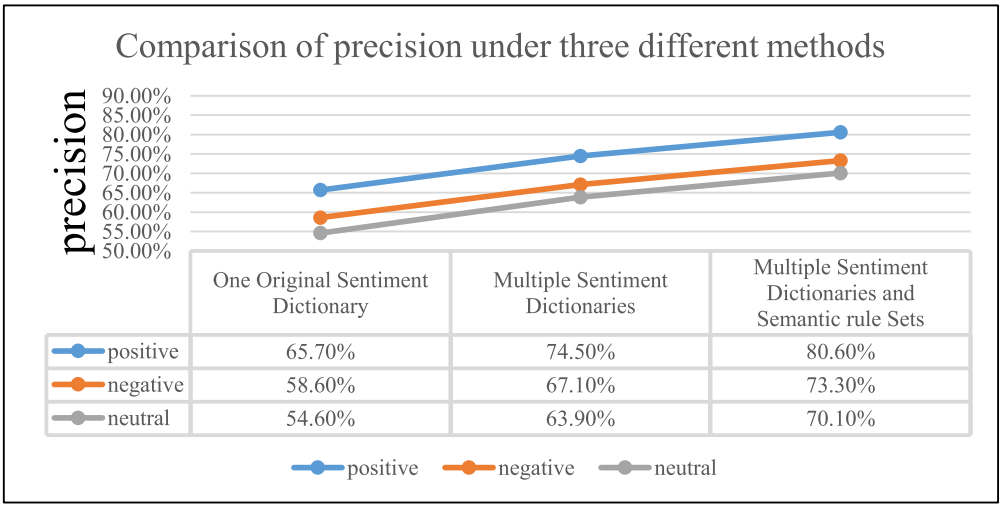


FIGURE 8. Comparison of average values of various indicators under three different methods.



	One Original Sentiment Dictionary	Multiple Sentiment Dictionaries	Multiple Sentiment Dictionaries and Semantic rule Sets
positive	65.70%	74.50%	80.60%
negative	58.60%	67.10%	73.30%
neutral	54.60%	63.90%	70.10%

FIGURE 9. Comparison of precision under three different methods.

dictionaries and semantic rule sets, three sets of experiments were carried out on these two topics respectively, and the sentiment analysis results of Chinese micro-blog were obtained.

B. THE EXPERIMENTAL DATA

In this paper, crawler software is used to crawl the datasets of these two Chinese micro-blog topics, and then the Chinese micro-blogs are labeled, and the sentiment value of each Chinese micro-blog is calculated and classified. A total of 25,720 Chinese micro-blogs belonging to the topic "#Short Video Rectangle#" were selected, including 18,634 positive data, 1,385 negative data, and 5,701 neutral data. A total of 17,695 Chinese micro-blogs belonging to the topic "#" I am not a drug god" burst red lead social hot discussion#" were selected, including 10,672 positive data, 2,856 negative data, and 4,167 neutral data.

C. EXPERIMENTAL PERFORMANCE EVALUATION INDICATOR

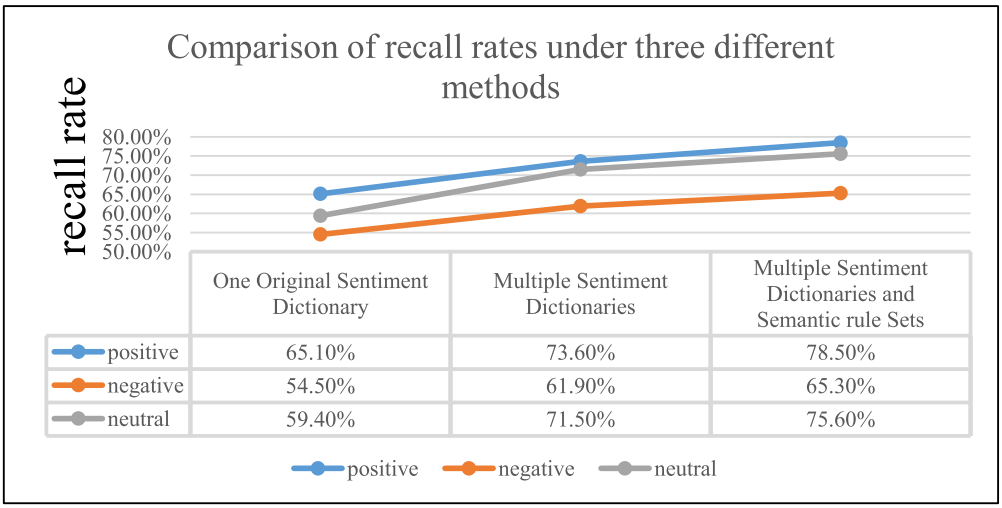
In this experiment, we mainly use the following three indicators for analysis, namely precision, recall rate, and comprehensive metric (F).

D. EXPERIMENTAL RESULTS AND ANALYSIS

We have done three sets of experiments using innovative methods and crawled data:

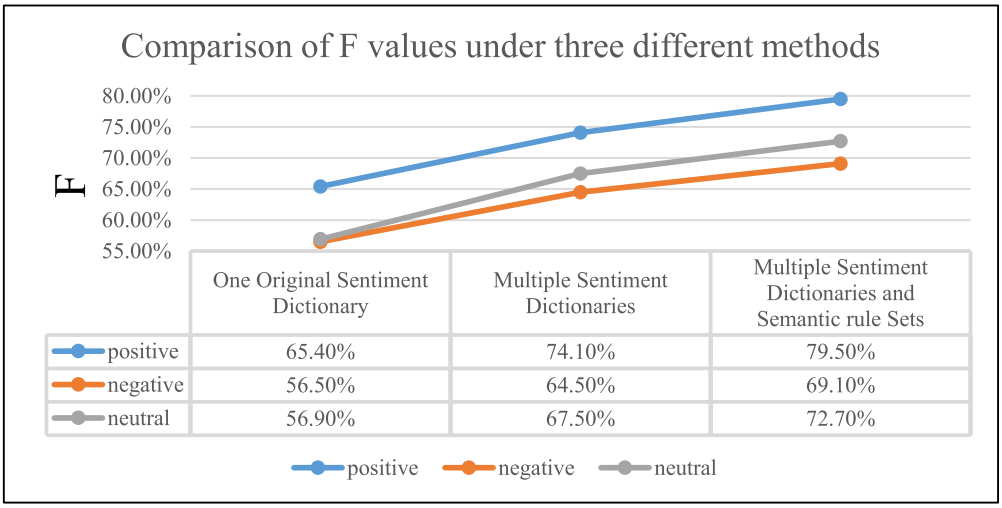
1) The first set of experiments: using a method based on one original sentiment dictionary to conduct sentiment analysis on the Chinese micro-blogs of the above two topics, and perform sentiment classification.

2) The second set of experiments: using a method based on multiple sentiment dictionaries to conduct sentiment analysis on the Chinese micro-blogs of the above two topics, and perform sentiment classification.



	One Original Sentiment Dictionary	Multiple Sentiment Dictionaries	Multiple Sentiment Dictionaries and Semantic rule Sets
positive	65.10%	73.60%	78.50%
negative	54.50%	61.90%	65.30%
neutral	59.40%	71.50%	75.60%

FIGURE 10. Comparison of recall rates under three different method.



	One Original Sentiment Dictionary	Multiple Sentiment Dictionaries	Multiple Sentiment Dictionaries and Semantic rule Sets
positive	65.40%	74.10%	79.50%
negative	56.50%	64.50%	69.10%
neutral	56.90%	67.50%	72.70%

FIGURE 11. Comparison of F values under three different methods.

3) The third set of experiments: using a method based on multiple sentiment dictionaries and semantic rule sets to conduct sentiment analysis on the Chinese micro-blogs of the above two topics, and perform sentiment classification.

Through the above three sets of experiments, the experimental results are compared and the experimental results obtained according to the performance evaluation indicators are shown in the following figures.

The analysis of the experimental results of the topic ''#Short Video Rectangle#'' is shown in Figures 5, 6, 7 and 8 below:

The analysis of the experimental results of the topic ''#'' I am not a drug god'' burst red lead social hot discussion#'' is shown in Figures 9, 10, 11 and 12 below:

Through the illustration of the above experimental results, we can draw the following relevant conclusions:

1) Using the method of this paper to improve the precision of Chinese micro-blog sentiment analysis. If it only relies on an original sentiment dictionary, the precision is lower, because Chinese micro-blog contains many features that ordinary text does not have, so it is necessary to construct multiple sentiment dictionaries on the basis of the original, improve the coverage of the dictionary, and take into account the semantic rule sets, which is more conducive to the sentiment analysis.

2) Through the analysis of the experimental results of the two topics, it can be seen that the precision of the topic ''#Short Video Rectangle#'' is higher than the precision of the topic''#''I am not a drug god'' burst red lead social hot discussion#'', Because the former has a lot of positive data, and the latter is a satirical topic. so there are many ironic Chinese micro-blogs on this topic, such as ''The drug dealers in the movie are really good, they can selling medicines to patients, they are really kind!'', which is ''good'' and ''kind'' are positive sentiment words, but in fact they are ironic, negative Chinese micro-blogs, so in the follow-up Chinese micro-blog's sentiment analysis can continue to improve the semantic rules.

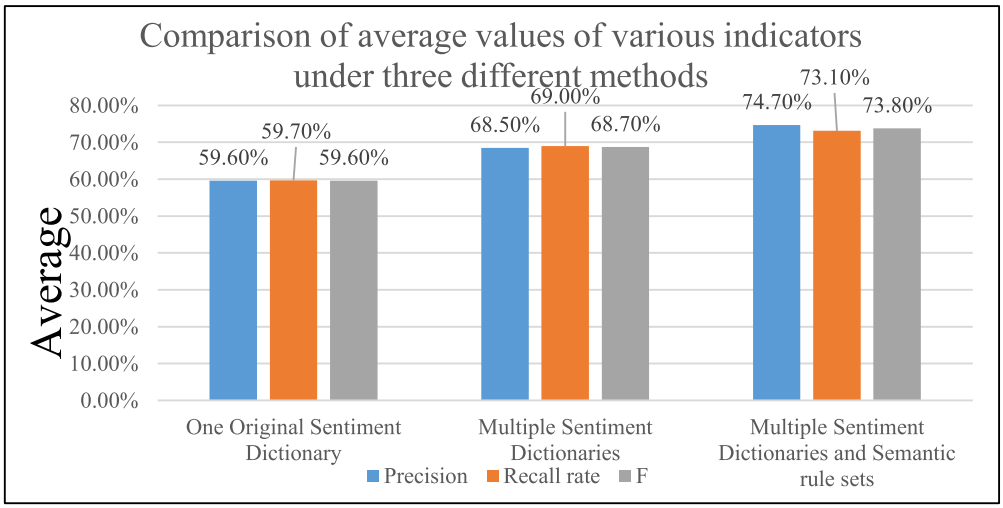


FIGURE 12. Comparison of average values of various indicators under three different methods.

3) Through the data in the figures and tables, it is found that the recall rate and F value of the positive Chinese micro-blogs are high. This is because these two topics are hot topics in society, and many netizens have a supportive attitude, so this leads to much positive Chinese micro-blog data.

4) By comparing the F values, it can be found that after constructing six sentiment dictionaries, the F value is greatly improved. This is because on the basis of the six sentiment dictionaries, especially the Chinese micro-blog new word sentiment dictionary is added, and the matching Chinese micro-blog is wider.

VII. CONCLUSION

In this paper, we propose a Chinese micro-blog sentiment analysis method based on multiple sentiment dictionaries and semantic rule sets. It mainly constructs multiple sentiment dictionaries. And the semantic rule set mainly includes inter-sentence analysis rules and sentence pattern analysis rules. The contribution of this paper is also to propose a new algorithm for Chinese micro-blog sentiment calculation, which calculates the sentiment value of Chinese micro-blog and classifies it.

Although this method has achieved certain effects, there are still some issues to consider. For example, weakly supervised learning technology can be used to further improve the correct rate of sentiment analysis, because Chinese micro-blog is a social media platform, which contains many users' likes. We will use weak annotation information to further study the above work in the future.

REFERENCES

[1] H. Ma, M. Jia, D. Zhang, and X. Lin, "Combining tag correlation and user social relation for microblog recommendation," *Inf. Sci.*, vol. 19, no. 4, pp. 325–337, 2017.

[2] M. Dragoni and G. Petrucci, "A neural word embeddings approach for multi-domain sentiment analysis," *IEEE Trans. Affect. Comput.*, vol. 8, no. 4, pp. 457–470, Oct./Dec. 2017.

[3] H. Peng, E. Cambria, and A. Hussain, "A review of sentiment analysis research in Chinese language," *Cognit. Comput.*, vol. 9, no. 4, pp. 423–435, 2017.

[4] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intell. Syst.*, vol. 31, no. 2, pp. 102–107, Mar./Apr. 2016.

[5] E. Riloff and J. Shepherd, "A corpus-based approach for building semantic lexicons," in *Proc. 2th Conf. Empirical Methods Natural Lang. Process.*, 1997, pp. 117–124.

[6] P. D. Turney and M. L. Littman, "Measuring praise and critism," *ACM Transl. Inf. Syst.*, vol. 21, no. 4, pp. 315–346, 2003.

[7] A. Valdivia, E. Hrabova, and L. Chaturvedi, "Inconsistencies on TripAdvisor reviews: A unified index between users and sentiment analysis methods," *Neurocomputing*, vol. 353, no. 8, pp. 3–16, 2019.

[8] C. Diamantini, A. Mircoli, and D. Potena, "Social information discovery enhanced by sentiment analysis techniques," *Future Gener. Comput. Syst.*, vol. 95, no. 6, pp. 816–828, 2019.

[9] S. Kashfia and A. Reda, "Emotion and sentiment analysis from Twitter text," *J. Comput. Sci.*, vol. 36, no. 9, pp. 1–42, 2019.

[10] J. M. Chen, H. F. Lin, and Z. H. Yang, "Automatic acquisition of emotional vocabulary based on syntax," *CAAI Trans. Intell. Syst.*, vol. 4, no. 2, pp. 100–106, 2009.

[11] P. Li, P. Di, and L. G. Duan, "Document sentiment orientation analysis based on sentence weighted algorithm," *J. Chin. Comput. Syst.*, vol. 36, no. 10, pp. 2252–2256, 2015.

[12] Y.-Y. Zhao, B. Qin, and T. Liu, "Sentiment analysis," *J. Softw.*, vol. 21, no. 8, pp. 1834–1848, 2010.

[13] M. Taboada, J. Brooke, and M. Tofiloski, "Lexi-con-based methods for sentiment analysis," *Comput. Linguistics*, vol. 37, no. 2, pp. 267–307, 2011.

[14] E. C. Dragut, H. Wang, and P. Sistla, "Polarity consistency checking for domain independent sentiment dictionaries," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 3, pp. 838–851, Mar. 2015.

[15] S. Feng, K. Song, D. Wang, and G. Yu, "A word-emoticon mutual reinforcement ranking model for building sentiment lexicon from massive collection of microblogs," *World Wide Web*, vol. 18, no. 4, pp. 949–967, Jul. 2015.

[16] Y. Li and X. Li, "A lexicon-based multi-class semantic orientation analysis for microblogs," *Acta Electronica Sinica*, vol. 44, no. 9, pp. 2068–2073, 2016.

[17] X. Wu and L. Wang, "Investigation on sentiment of reviews with shopping field dictionary construction," *Comput. Technol. Develop.*, vol. 27, no. 7, pp. 194–199, 2017.

[18] E. Boiy and M.-F. Moens, "A machine learning approach to sentiment analysis in multilingual Web texts," *Inf. Retr.*, vol. 12, no. 5, pp. 526–558, 2009.

[19] Q. Ye, Z. Q. Zhang, and R. Law, "Sentiment classification of online reviews to travel destinations by supervised sentiment machine learning approaches," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6527–6535, 2009.

[20] A. Abbasi, H. Chen, and A. Salem, "Sentiment analysis in multiple languages," *ACM Trans. Inf. Syst.*, vol. 26, no. 3, pp. 1–34, 2008.

[21] J. K. Rout, K. K. R. Choo, and A. K. Dash, "A model for sentiment and emotion analysis of unstructured social media text," *Electron. Commerce Res.*, vol. 18, no. 1, pp. 119–181, 2017.

[22] A. S. Manek, P. D. Shenoy, M. C. Mohan, and K. R. Venugopal, "Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier," *World Wide Web*, vol. 20, no. 2, pp. 135–154, 2017.

[23] S. Chaffer and D. Inkpen, "Using a heterogeneous Dataset for emotion analysis in Text," in *Proc. 24th Can. Conf. Adv. Artif. Intell.*, St. John's, NF, Canada, 2011, pp. 62–67.

[24] O. Tackstrom and R. McDonald, "Discovering fine-grained sentiment with latent variable structured prediction models," in *Proc. Eur. Conf. Adv. Inf. Retr.*, Dublin, Ireland, 2011, pp. 368–374.

[25] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[26] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[27] X. Sun, C. Li, and F. Ren, "Sentiment analysis for chinese microblog based on deep neural networks with convolutional extension features," *Neurocomputing*, vol. 210, pp. 227–236, Oct. 2016.

[28] F. Hu, L. Li, and Z. L. Zhang, "Emphasizing essential words for sentiment classification based on recurrent neural networks," *J. Comput. Sci. Technol.*, vol. 32, no. 4, pp. 785–795, 2017.

[29] M. Giatsoglou, M. G. Vozalis, K. Diamantaras, A. Vakali, G. Sarigiannidis, and K. C. Chatzisavvas, "Sentiment analysis leveraging emotions and word embeddings," *Expert Syst. Appl.*, vol. 69, pp. 214–224, Mar. 2017.

[30] D. Tang, F. Wei, N. Yang, T. Liu, M. Zhou, and B. Qin, "Sentiment embeddings with applications to sentiment analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 2, pp. 496–509, Feb. 2016.

[31] G. Gao, J. Luo, and Y. Wang, "Analyzing textual sentiment based on HNC theroy," *Data Anal. Knowl. Discovery*, vol. 1, no. 8, pp. 85–91, 2017.

[32] A. C. Pandey, M. Saraswat, and D. S. Rajpoot, "Twitter sentiment analysis using hybrid cuckoo search method," *Inf. Process. Manage.*, vol. 53, no. 4, pp. 764–779, 2017.

[33] M. Purver and S. Battersby, "Experimenting with distant supervision for emotion classification," in *Proc. 13th Conf. Eur. Chapter Assoc. Comput. Linguistics*, Avignon, France, 2012, pp. 482–491.

[34] Grandi, A. Loreggia, F. Rossi, "A Borda count for collective sentiment analysis," *Ann. Math. Artif. Intell.*, vol. 77, nos. 3–4, pp. 281–302, 2016.

[35] A. B. Eliacik and N. Erdogan, "Influential user weighted sentiment analysis on topic based microblogging community," *Expert Syst. Appl.*, vol. 92, pp. 403–418, Feb. 2018.

[36] E. Cambria, S. Poria, B. Liu, and A. Hussain, "Computational intelligence for affective computing and sentiment analysis," *IEEE Comput. Intell. Mag.*, vol. 14, no. 2, pp. 16–17, May 2019.

[37] J. Han, Z. Zhang, B. Schuller, and N. Cummins, "Adversarial training in affective computing and sentiment analysis," *IEEE Comput. Intell. Mag.*, vol. 14, no. 2, pp. 68–81, May 2019.

[38] Y. Ma, H. Peng, T. Khan, E. Cambria, and A. Hussain, "Sentic LSTM: A hybrid network for targeted aspect-based sentiment analysis," *Cogn. Comput.*, vol. 10, no. 4, pp. 639–650, 2018.

[39] Z. Yu, Z. T. Wang, and L. M. Chen, "Featuring, detecing, and visualizing human sentiment in chinese micro-blog," *ACM Trans. Knowl. Discovery Data*, vol. 10, no. 4, pp. 1–23, 2016.

[40] G. Xu, Z. Yu, and H. Yao, "Chinese text sentiment analysis based on extend sentiment dictionary," *IEEE Access*, vol. 7, pp. 43749–43762, 2019.

[41] S. Zhang, Z. Wei, Y. Wang, and T. Liao, "Sentiment analysis of chinese micro-blog text based on extended sentiment dictionary," *Future Gener. Comput. Syst.*, vol. 81, pp. 395–403, Apr. 2018.

[42] Y. Cai, K. Yang, and D. P. Huang, "A hybrid model for opinion mining based on domain sentiment dictionary," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 8, pp. 2131–2142, 2019.

[43] L. Wu, F. Morstatter, and H. Liu, "SlangSD: Building, expanding and using a sentiment dictionary of slang words for short-text sentiment classification," *Lang. Resour. Eval.*, vol. 52, no. 3, pp. 839–852, 2019.

JIESHENG WU received the bachelor's degree in computer science and technology from the Anhui University of Science and Technology, China, in 2017, where he is currently pursuing the master's degree. His current research interests involve natural language processing, sentiment analysis, and machine learning.

KUI LU received the Ph.D. degree in control theory and control engineering from the China University of Mining and Technology, Beijing, China, in 2004. He is currently a Professor with the School of Computer Science and Engineering, Anhui University of Science and Technology. His current research interests include computer network and communication, and big data mining.

SHUZHI SU received the Ph.D. degree from the School of Internet of Things Engineering, Jiangnan University. He is currently a Full Lecturer with the College of Computer Science and Engineering, Anhui University of Science and Technology, China. His research interests include sentiment analysis, information fusion, and multimodal pattern recognition.

SHIBING WANG received the Ph.D. degree in circuits and systems from Anhui University, China, in 2010. He is currently a Postdoctoral Researcher with the Dalian University of Technology and also a Professor with the School of Computer and Information Engineering, Fuyang Normal University, China. His research interests include nonlinear circuits and systems, chaos control and synchronization, chaotic signal, and information processing.