

# Chinese micro-blog sentiment classification through a novel hybrid learning model

LI Fang-fang(李芳芳)<sup>1,2</sup>, WANG Huan-ting(王欢婷)<sup>1</sup>, ZHAO Rong-chang(赵荣昌)<sup>1</sup>,  
LIU Xi-yao(刘熙尧)<sup>1</sup>, WANG Yan-zhen(王彦臻)<sup>3</sup>, ZOU Bei-ji(邹北骥)<sup>1</sup>

1. School of Information Science and Engineering, Central South University, Changsha 410083, China;
2. “Mobile Health” Ministry of Education–China Mobile Joint Laboratory, Changsha 410083, China;
3. College of Computer, National University of Defense Technology, Changsha 410073, China

© Central South University Press and Springer-Verlag GmbH Germany 2017

**Abstract:** With the rising and spreading of micro-blog, the sentiment classification of short texts has become a research hotspot. Some methods have been developed in the past decade. However, since the Chinese and English are different in language syntax, semantics and pragmatics, sentiment classification methods that are effective for English twitter may fail on Chinese micro-blog. In addition, the colloquialism and conciseness of short Chinese texts introduces additional challenges to sentiment classification. In this work, a novel hybrid learning model was proposed for sentiment classification of Chinese micro-blogs, which included two stages. In the first stage, emotional scores were calculated over the whole dataset by utilizing an improved Chinese-oriented sentiment dictionary classification method. Data with extremely high or low scores were directly labeled. In the second stage, the remaining data were labeled by using an integrated classification method based on sentiment dictionary, support vector machine (SVM) and *k*-nearest neighbor (KNN). An improved feature selection method was adopted to enhance the discriminative power of the selected features. The two-stage hybrid framework made the proposed method effective for sentiment classification of Chinese micro-blogs. Experiments on the COAE2014 (Chinese Opinion Analysis Evaluation 2014) dataset show that the proposed method outperforms other schemes.

**Key words:** Chinese micro-blog; short text; hybrid learning; sentiment classification

## 1 Introduction

Micro-blog plays an important role in modern society for information sharing, exchanging and dissemination. The micro-blog consists of short texts and updates instantly, which makes building and updating personal communities much easier and more efficient. According to the 36th China Internet Development Statistics Report [1] released by the China Internet Network Information Center, by June 2015, the number of Chinese micro-blog users has reached 204 million. Huge amount of blog texts are produced by micro-blog users every moment. It is meaningful to mine useful information from these texts, such as relationship analysis, hidden pattern detection, sentiment classification. Sentiment classification aims at getting people’s subjective opinions and feelings on specific entities, personages, and events [2]. It is a core component in automatic text summarizing, questioning and answering system [3]. Nevertheless, sentiment

classification of micro-blog still remains an open issue [4, 5].

Traditional algorithms for sentiment classification can be divided into two categories. One is based on sentiment dictionary. KIM and HOVY [6] utilized sentiment dictionary for emotional tendency estimation of sentences and chapters with the weighted sum of evaluation words and phrases. LI et al [7] developed a Chinese-English bilingual information and label propagation algorithm to build sentiment dictionary. HAN et al [8] adopted an automata construction method for sentiment classification. PANG et al [9] proposed a text classification method based on emotion tendentious phrase. SUN et al [10] developed a method to sentiment classification based on conditional random field and sentiment dictionary. LUO et al [11] constructed an emotional dictionary with two-levels, and the words for different levels will get different enhancement. Sentiment dictionary based method could get high accuracy rate for dataset which has obvious emotional tendency. However, it would become very inefficient for

**Foundation item:** Projects(61573380, 61303185) supported by the National Natural Science Foundation of China; Project(13BTQ052) supported by the National Social Science Foundation of China; Project(2016M592450) supported by the China Postdoctoral Science Foundation; Project(2016JJ4119) supported by the Hunan Provincial Natural Science Foundation of China.

**Received date:** 2016–03–23; **Accepted date:** 2016–11–28

**Corresponding author:** ZHAO Rong-chang, Lecture, PhD; Tel: +86–13647449808; E-mail: byrons.zhao@gmail.com

dataset vague in emotional tendency. Moreover, it does not work well in fields with field-specific feature words, because it is virtually impossible to generate a dictionary applicable to all fields.

The other category is based on machine learning. PANG et al [12] employed Naïve Bayes, maximum entropy classification and support vector machine to achieve sentiment classification of chapters. SUN et al [13] utilized the hybrid model based on themes and emotions to achieve unsupervised sentiment classification. TAN et al [14], SOCHER et al [15] and LIU et al [16] presented semi-supervised algorithms to build sentiment classification models. BARBOSA and FENG [5] adopted support vector machine as the classifier. PARK and PAROUBEK [17] built an emotional classifier based on naive Bayes, support vector machine and conditional random fields. DAVIDOV et al [18] designed an emotional classifier based on K-nearest neighbor method, and utilized twitter's unique attributes and usage characteristics (e.g., tags and expressions) as text features. RUSTAMOV and CLEMENTS [19] utilized neural inference to obtain sentence discrimination. REN et al [20] developed a graph-based semi-supervised learning model for sentiment classification in under-resourced languages. MAO et al [21] proposed a user-related sentiment classification of Chinese micro-blogs in terms of statistical and semantic characteristics. Methods based on machine learning techniques do not need a dictionary. However, it requires a huge collection of manually labeled training datasets, which introduces too much cost in practice.

In practice, processing Chinese micro-blogs would need much more effort due to the following two characteristics. Firstly, each Chinese micro-blog cannot exceed the 140-character constraint, and contains several sentences [22]. The sentences in the same micro-blog may even tend to express opposite emotions. Secondly, Chinese micro-blog is lack of training corpus. Although English training corpus about sentiment classification has been well established based on twitter, the establishment of Chinese corpus has just begun.

To address the above problems, a novel hybrid learning model for sentiment classification of Chinese micro-blog was proposed in this work. Key contributions in this paper were as follows.

1) A novel hybrid sentiment classification model for Chinese micro-blog was proposed. The model is specifically designed for the Chinese micro-blog. In the first stage, sentiment dictionary was utilized to classify a part micro-blog dataset with clear emotional tendencies. In the second stage, an integrated classification model was presented to determine the sentiment tendencies of the remaining data that were emotionally vague. The strength of sentiment dictionary, SVM and KNN were

combined and their disadvantages were avoided in the hybrid model. Therefore, the final sentiment classification results could get higher accuracy.

2) A comprehensive processing that considered the emotion of each short sentence in micro-blog text was designed. The processing considered various situations of emotional words and their affections, such as negative word processing, degree adverb processing, dynamic emotional word processing, adversative conjunction processing and the comprehensive processing of them appeared simultaneously.

3) A novel sentiment dictionary that included the HowNet sentiment dictionary [23], the simplified Chinese emotional dictionary NTUSD [24] and some manually collected hot words in micro-blog was established. In this manner, the Chinese micro-blog corpus utilized in this paper is expanded.

## 2 Preprocessing works

Data preprocessing in this paper mainly included three steps: data cleaning, word segmentation, stop word removal and place name removal. Data cleaning was mainly utilized to remove some noise in micro-blog, such as title names, topic names, reply data, content between parenthesis and content after @. Since these contents were of no practical significance for sentiment classification, they needed to be removed. Additionally, as the emotional tendency needed to be analyzed was the data released by bloggers rather than the reply data, it should also be removed. Regular expressions were utilized in this step for effective cleaning, as shown in Table 1.

The next preprocessing step was word segmentation. Segmentation tool utilized in our work was the NLPPIR2014 proposed by ZHANG [25]. NLPPIR2014 has upgraded its system kernel 10 times and has a large dictionary scale. Moreover, it added new function of micro-blog words segmentation and added many new network hot words that appear recently in micro-blogs. After this step, words in the original micro-blog were divided into nouns, verbs, adjectives, stop words and place names, etc. Finally, since stop words and place names were meaningless for sentiment analysis, such as personal pronouns you, me, him etc., they were removed from the results of word segmentation.

## 3 Hybrid learning model

The workflow of the presented model was described as follows. The sentiment dictionary was firstly established. Then different types of words such as emotional words, negative words, degree adverbs were

**Table 1** Main cleaning rules

Type of removal	Before cleaning	Regular expression	After cleaning
Title names removal	【2014 Forbes Chinese rich list: Jack Ma, Robin Li, Pony Ma dominate the top three】Forbes Chinese network announced 2014 Chinese Rich List	【[\s\S]*?【】 \\$】	Forbes Chinese network announced 2014 Chinese Rich List
Topic names removal	#Elephants encounter beheaded and taken teeth#Just want to say that we need new legal proceedings to protect these animals	#[\s\S]*?#[# \\$]	Just want to say that we need new legal proceedings to protect these animals
Reply data Removal	No matter how short of many, only use one credit card and timely repayment//@social unspoken rule: financial tips	//@	No matter how short of many, only use one credit card and timely repayment
Content in ( ) removal	Samsung smart phone was traced loopholes exist, move your phone yet? (Share from Newsweek International)	[([\s\S]*?) \\$]	Samsung smart phone was traced loopholes exist, move your phone yet?
Content after @ removal	Save place so much, wallet can hold one million.@Gu Yue Wan@Feng Tian Lao	@[\s\S]*?[\s \\$]	Save place so much, wallet can hold one million.

taken into consideration when computing emotional scores. Next, emotional scores for each data in the whole dataset were calculated by an improved sentiment dictionary method.

For the data with extremely high or low emotional scores, their emotional tendencies could be directly determined. For the remaining data, the second stage came into play. In this stage, Vector Space Model was utilized to represent the data, and term frequency-inverse document frequency was utilized to calculate the weight of each feature. An improved feature selection method IDTC was adopted to enhance the discriminative power of the selected features. After the above mentioned steps, hybrid learning was carried out on the resulting dataset. For the data whose three classification results, namely, emotional score, KNN, SVM were all positive or all negative, their positive or negative emotional tendencies were determined.

In summary, the first stage calculated emotional scores for each data in the whole dataset, and given emotional tendencies for those data which get very high positive (or very low negative) emotional scores. In this manner, a part of data can be removed and only the remaining data with undetermined emotional tendencies need to be further determined. For these data, the second stage method was adopted. The flowchart was shown as Fig. 1.

### 3.1 Sentiment dictionary based classification

#### 3.1.1 Sentiment dictionary establishing

In the first stage, the HowNet sentiment dictionary [23] and the NTUSD [24] released by the “National Taiwan University” were utilized as the initial dictionary. Since micro-blog texts usually contain many new network hot words, the initial dictionary may not be sufficient. In this work, these hot words were manually collected according to the latest hot word list of micro-blog. Then, these words were combined with the initial dictionary. Finally, 34808 terms were contained in

the established dictionary.

#### 3.1.2 Sentiment dictionary based classification

In the process of Chinese micro-blog, the situation was very complicated. Chinese micro-blog may contain different types of words, and different words affect the emotional tendency of the sentence to a certain extent. For example, emotional words would reflect the tendency, while negative words would reverse the tendency, with a summary shown in Table 2.

Given the above situations, this paper proposed an improved sentiment dictionary based classification method. Based on the established dictionaries for negative words, degree adverbs and adversative conjunctions, processing of these words together with emotional words were fully considered. Further more, the comprehensive processing of these types of words with each other were considered. The terms utilized in the following chapters were explained in Table 3.

##### Step 1: Negative word processing

After emotional word  $w$  in a micro-blog text was identified by the sentiment dictionary, its emotional score was recorded as  $SO(w)$ . In front of  $w$ , a negative words detection window was set and  $k$  negative words were assumed to be appeared in the window. The emotional score of each phrase was recorded as  $SO(phrase)$  in the text. Taking into account that there may existed double negative words in the sentence,  $SO(phrase)$  was calculated as Eq. (1).

$$SO(phrase) = \begin{cases} SO(w), & \text{if } (k = 2n) \\ -SO(w), & \text{if } (k = 2n + 1) \end{cases}, \quad n = 0 \text{ or } 1 \quad (1)$$

##### Step 2: Degree adverb processing

Based on the 219 degree adverbs provided by HowNet [23],  $Strength(da)$  was defined as their polarity degrees and they were divided into six grades: extreme, very, relative, slight, faint and little. Each grade was assigned a value between  $-0.5$  to  $2.5$ , as shown in Table 4. The processing procedure of degree adverbs was similar with that of negative words. Suppose that there

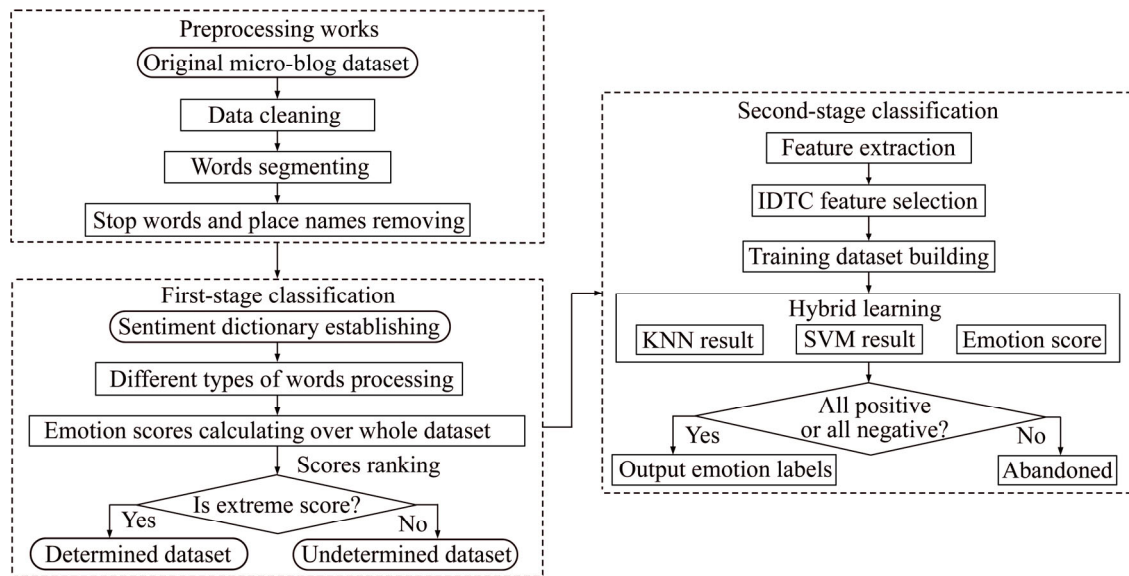


Fig. 1 Overall flowchart of hybrid learning model

Table 2 Type of words and their affections

Type	Emotional tendency
Emotional words	Reflect the tendency
Negative words	Reverse the tendency
Degree adverbs	Emphasize or weaken the tendency
Dynamic emotional words	With different collocations has different effect on the tendency
Adversative conjunctions	Emphasize or weaken the tendency of the phrase following it

Table 3 Terms and their interpretations

Term	Interpretation
$w$	Emotional word
$SO(w)$	Emotional score of $w$
$SO(phrase)$	Emotional score of phrase
$da$	Degree adverb
$Strength(da)$	Polarity score of $da$
$dew$	Dynamic emotional word
$F_{dew}$	Forward dynamic emotional word
$B_{dew}$	Backward dynamic emotional word
$cw$	Collocation word
$F_{cw}$	Forward collocation word
$B_{cw}$	Backward collocation word
$t_i$	Feature lemma of the training dataset
$C_j, C_k$	The negative and the positive category
$P(t_i)$	Frequency of the document contains $t_i$ in the training dataset
$IDTC(t_i, C_j, C_k)$	Category discriminative degree of $t_i$ to $C_j$ and $C_k$
$IDTC(t_i)$	Category discriminative degree of $t_i$ to the whole training dataset
$P(t_i C_j)$	Frequency of the document contains $t_i$ in $C_j$
$P(t_i C_k)$	Frequency of the document contains $t_i$ in $C_k$

Table 4 Six grades of polarity degrees

Grade	Examples of degree adverb	Score
Extreme	Extremely, most	2.5
Very	Extraordinarily, highly	2
Relative	Relatively, furthermore	1.5
Slight	Slightly, a little	0.8
Faint	Faintly, mildly	0.5
Little	Rarely, trivially	-0.5

was a degree adverbs detection window in front of  $w$ , if degree adverb  $da$  appeared in the window, then its degree of polarity which named  $Strength(da)$  will increase or decrease the polarity of the phrase. The calculation formula was shown as Eq. (2).

$$SO(phrase) = Strength(da) \times SO(w) \tag{2}$$

**Step 3: Dynamic emotional word processing**

Dynamic emotional words have different emotional tendencies when modifying different target words. Such as the following sentence shows. “This car is of high quality, but also of high fuel consumption”. The sentence has a dynamic emotional word “high”, and its tendency could not be gotten if only analyze the word “high”.

Therefore, the modifying target words needed to be taken into account. Such as, when analyzing the pair <quality, high>, the tendency for the phrase is positive. In the contrast, when analyzing the pair <fuel consumption, high>, the tendency is negative.

To deal with this situation, firstly a dynamic emotional words dictionary was established. Dynamic emotional words (*dew*) were divided into forward word set  $F_{dew}$ , which included words like “high”, “many”, and backward word set  $B_{dew}$ , which included words like “low”, “less”. Then a collocation word dictionary was established. Collocation words (*cw*) were divided into forward word set  $F_{cw}$ , which included words like “quality”, “function”, and backward word set  $B_{cw}$ , which included words like “fuel consumption”, “noise”. Then, *dew* in each sentence could be identified according to the dynamic emotional word dictionary, and in front and back of *dew* a noun detection window was set. If *cw* appeared in the window, then the emotional tendency of the phrase could be calculated according to its pair of case in dynamic emotional word dictionary and collocation word dictionary, i.e.,  $F_{dew}$  and  $B_{dew}$ ,  $F_{cw}$  and  $B_{cw}$ , as shown in Eq. (3).

$$\begin{cases} C(dew) = \begin{cases} 1, & \text{if } dew \in F_{dew} \\ -1, & \text{if } dew \in B_{dew} \end{cases} \\ C(cw) = \begin{cases} 1, & \text{if } cw \in F_{cw} \\ -1, & \text{if } cw \in B_{cw} \end{cases} \\ SO(phrase) = C(dew) \times C(cw) \end{cases} \quad (3)$$

#### Step 4: Adversative conjunction processing

A micro-blog text is often composed of multiple sentences, which means that the emotional tendency of the text is affected by more than one sentences. The appearance of adversative conjunction in different positions causes the focus of emotional tendency shifting among different sentences. Adversative conjunctions affect emotional tendency of the phrase mainly in two ways:

1) Emphasizing the polarity of the phrase following it. Such as the following sentence shows. “This phone looks nice, but the performance is so poor”. The sentence has a positive emotional word “nice” and a negative emotional word “poor”, and the word “poor” is following the adversative conjunction “but”. Therefore, the polarity of the whole sentence focuses on the latter part and is negative.

2) Weakening the polarity of the phrase following it. Taking adversative conjunction “although” as an example, the polarity of the phrase following “although” is weakened, and the polarity of the whole sentence focuses on the other part.

To deal with the above situation, firstly a dictionary

of adversative conjunction words was established, which included polarity-emphasizing conjunctions and polarity-weakening conjunctions. If an adversative conjunction appeared in the sentence, the polarity of the whole sentence was primarily determined by the effect of the adversative conjunction as discussed above, rather than simply summing up scores of all the phrases.

#### Step 5: Comprehensive processing

If a negative word and a degree adverb appeared simultaneously around an emotional word  $w$ , the comprehensive tendency of the emotional word and phrase were calculated by Eq. (4). If the emotional word was a dynamic emotional word, the word and its collocation were firstly treated as a whole for calculation of emotional tendency. Then, processing steps for negative words and degree adverbs were carried out.

In summary, the emotional tendency and degree of the micro-blog text was computed as follows. Firstly, the text with the above-mentioned sentiment dictionaries was matched and emotional scores of each phrase in the text was computed according to the matched result. Then, scores of all the phrases were summed up to obtain the total score of each sentence. Then, scores of all the sentences were summed up to obtain the score of the entire text. Next, the resulting score was normalized by dividing it by the number of emotional words in the text, in order to eliminate the influence of text length. Finally, the emotional tendency and degree of the text were determined by the normalized score.

$$\begin{cases} C(w) = \begin{cases} Strength(da), & \text{if } (k = 2n) \\ -Strength(da), & \text{if } (k = 2n + 1) \end{cases}, n = 0 \text{ or } 1 \\ SO(phrase) = C(w) \times SO(w) \end{cases} \quad (4)$$

### 3.2 Integrated method based classification

Since method based on sentiment dictionary is able to obtain high accuracy for datasets with obvious emotional tendencies, it was utilized to determine tendencies of text with extremely high or low scores in the first stage. On the other hand, machine learning based method works better on datasets with ambiguous emotional tendencies. Therefore, an integrated classification method which integrates sentiment dictionary classification and machine learning classification was adopted in the second stage.

The integrated classification method was composed of four steps, namely, text representation, feature selection, training dataset processing and testing dataset classification. Firstly, vector space model (VSM) was utilized to represent the micro-blog texts. Then the training dataset was manually divided into two subsets, namely, the negative dataset and the positive dataset. An

improved feature selection method, information difference between two categories (IDTC) was utilized to obtain text features of each category. Then, the above-mentioned operations were carried out on the testing dataset again. Finally, integrated classification method was utilized for the testing dataset classification.

### 3.2.1 Feature selection

Commonly utilized feature selection methods include document frequency (DF), information gain (IG), mutual information (MI) and Chi-square statistic (CHI). In this work, an improved feature selection method, IDTC, which used both occurrence frequency and category discriminative degree of the feature to represent feature information for the category, was adopted. TF-IDF (Term Frequency-Inverse Document Frequency) was utilized to calculate the weight of each feature. For a binary classification problem, category discriminative degree in IDTC was represented by MI differences, as shown in Eq. (5).

$$IDTC(t_i, C_j, C_k) = P(t_i) \times |MI(t_i, C_j) - MI(t_i, C_k)| \quad (5)$$

where  $|MI(t_i, C_j) - MI(t_i, C_k)|$  represented the MI differences of feature  $t_i$  between the two categories. Category discriminative degree of  $t_i$  to the whole training dataset was calculated according to Eq. (6).

$$\begin{aligned} IDTC(t_i) &= \sum_{j=1}^{|C|-1} \sum_{k=j}^{|C|} IDTC(t_i, C_j, C_k) \\ &= \sum_{j=1}^{|C|-1} \sum_{k=j}^{|C|} P(t_i) \times |MI(t_i, C_j) - MI(t_i, C_k)| \\ &= \sum_{j=1}^{|C|-1} \sum_{k=j}^{|C|} P(t_i) \times \left| \log \frac{P(t_i | C_j)}{P(t_i)} - \right. \\ &\quad \left. \lg \frac{P(t_i | C_k)}{P(t_i)} \right| \end{aligned} \quad (6)$$

During the feature selection, in order to reduce the computational complexity, some words that below a certain threshold from the original feature space were removed.

### 3.2.2 Integrated method based classification

In integrated method based classification, sentiment dictionary, SVM and KNN were utilized as the text classifiers. Since SVM has unique advantages in binary classification, it was utilized to determine the remaining dataset with low scores, which may have ambiguous emotional tendencies. Specifically, LIBSVM [26] software package was utilized to carry out SVM classification in this work. Note that all the default parameters provided by LIBSVM were remained unchanged during the experiment, and radial basis kernel function was selected as the kernel function. The reason for selecting this function was to confine the deviation of the classification to a relatively low level, in order to obtain a tradeoff between precision and recall rates even without parameter tuning. Since KNN is a traditional

pattern recognition method, which has been widely utilized in automatic text classification and exhibits outstanding performance on precision and recall rate, it was also utilized to classify the remaining dataset.

During the process of classification, since emotional scores had been calculated on the entire preprocessed dataset in the first stage, each text in the remaining dataset had three values, namely, emotional score, SVM classification result and KNN classification result. Emotional tendencies could be directly determined for the data with consistent positive or negative values. By utilizing the hybrid learning model, the uncertainty caused by one certain method could be minimized.

## 4 Experimental results and analysis

### 4.1 Dataset

The Chinese Information Processing Society, jointly with some well-known colleges and research institutes, launched the Chinese Opinion Analysis Evaluation (COAE) in 2008 [27]. COAE promoted the theoretical research and technical application of Chinese sentiment analysis, and established the basic dataset and the evaluation standard for Chinese sentiment analysis. The micro-blog dataset utilized in this work was provided by COAE2014, which includes 40000 texts, out of which approximately 7000 texts were labeled with sentiment tendencies. The Organizing Committees required competition teams to classify 40000 texts into positive or negative categories, and ultimately submitted the results of 10000 texts to participate evaluation.

### 4.2 Experimental result

Firstly, how many features can be chosen to achieve the best classification result need to be tested. Secondly, based on the chosen features, comparison between our proposed method and other traditional methods such as sentiment dictionary, KNN and SVM was conducted. Finally, based on the same COAE dataset, comparison between our proposed method and other methods participated in the COAE2014 was conducted. Therefore, three groups of experiments were conducted, as shown in Table 5.

Results of Experiment A suggested that the number of feature words had significant effect on the classification accuracy. Too small number of feature words will lead to insufficiently simple classification model and under-fitting problem, while too large number will lead to overly complex classification model and over-fitting problem. In Experiment A, a large number of latest Sina micro-blog texts were collected as samples, 5176 of which were utilized as training dataset and 4000 as testing dataset. Positive and negative samples in the training dataset were proportionate. Experimental results

**Table 5** Three groups of experiments

Experiment	Setting	Purpose
A	KNN algorithm, under different number of feature words	Figure out the best chose of number of feature words in the following experiments
B	Comparison between four methods, namely, sentiment dictionary, KNN, SVM and the proposed method, under the above best chosen number	Verify the effect of the proposed method
C	Comparison between the proposed method and other methods on COAE dataset	Figure out which method could achieve higher precision

of KNN algorithm under different number of feature words was shown in Table 6. When the number of feature word was 4000, the algorithm reached the highest precision rate. Therefore, 4000 was chosen as the number of feature words throughout the following experiments.

**Table 6** Precision rate under different number of feature words

Number of feature words	Precision rate
1000	0.853
2000	0.887
3000	0.891
4000	0.894
5000	0.891
6000	0.888
7000	0.887

In Experiment B, comparison among the following four methods were conducted, namely, sentiment dictionary, KNN, SVM and the proposed hybrid model. Evaluated indicators included precision rate, recall rate, F1 measure. Indicators were calculated by Eq. (7), where *System.correct* was the matching number of classification results with the manually annotated results, *System.output* was the number of classification results and *Human.Labeled* was the manual annotation results. As shown in Table 7, the hybrid model outperformed the other three methods, in terms of both precision rate and F1 measure.

$$\left\{ \begin{aligned}
 precision &= \frac{System.Correct}{System.Output} \\
 recall &= \frac{System.Correct}{Human.Labeled} \\
 F1 &= \frac{2 \times precision \times recall}{precision + recall}
 \end{aligned} \right. \quad (7)$$

**Table 7** Experimental results of different methods

Indicator	Sentiment dictionary	KNN	SVM	Proposed method
Precision	0.790	0.894	0.854	0.900
Recall	0.625	0.900	0.946	0.939
F1	0.700	0.895	0.895	0.920

In Experiment C, the proposed method was carried out on the micro-blog dataset provided by COAE2014. Evaluated indicators utilized in this paper included the above-mentioned precision, recall, F1, together with Macro\_P, Macro\_R, Macro\_F1 and Micro\_P, Micro\_R, Micro\_F1, which were the macro-average and micro-average based on the above three indicators respectively. Macro-average weights were defined in a class-wise manner, regardless of how many documents belong to it, while micro-average weights were defined in a document-wise manner. Terms and their interpretations were shown in Table 8.

**Table 8** Terms and their interpretations

Term	Interpretation	Term	Interpretation
Pos_P	Precision of the positive category	Macro_P	Macro-average of precision
Pos_R	Recall of the positive category	Macro_R	Macro-average of recall
Pos_F1	F1 of the positive category	Macro_F1	Macro-average of F1
Neg_P	Precision of the negative category	Micro_P	Micro-average of precision
Neg_R	Recall of the negative category	Micro_R	Micro-average of recall
Neg_F1	F1 of the negative category	Micro_F1	Micro-average of F1

30 well-known colleges from China submitted 50 results in COAE2014. Some best results were shown in Table 9. Values in **Best** column were selected from the best value of each item in the 50 submitted results, rather than through a certain method. **Mean** was the average value of each item in the 50 results. As shown in Table 9, the results of our proposed method are significantly higher than the mean value, with some items even better than the best value, such as Neg\_P, Macro\_P, Macro\_R, Macro\_F1, Micro\_P and Micro\_F1. The other items of the hybrid model are very close to the best value, such as Pos\_R, and Pos\_F1. Besides, the hybrid model outperforms other methods in precision rate, such as Pos\_P and Neg\_P, which indicates that for the most majority of evaluated texts, the proposed method has made the right judgment. F1 measure of our method is also among the top ranks, which further demonstrates the effectiveness of the hybrid model in sentiment

**Table 9** Part of best evaluation results

Item	Ref. [9]	Ref. [10]	Ref. [11]	Best	Mean	Proposed
Pos_P	0.954	0.964	0.964	0.977	0.891	0.977
Pos_R	0.429	0.302	0.211	0.603	0.299	0.520
Pos_F1	0.592	0.460	0.347	0.715	0.445	0.679
Neg_P	0.885	0.791	0.789	0.971	0.850	0.979
Neg_R	0.630	0.234	0.766	0.766	0.281	0.597
Neg_F1	0.736	0.361	0.778	0.778	0.428	0.742
Macro_P	0.919	0.877	0.877	0.962	0.877	0.978
Macro_R	0.530	0.268	0.489	0.543	0.309	0.559
Macro_F1	0.672	0.411	0.628	0.677	0.443	0.709
Micro_P	0.914	0.887	0.826	0.962	0.857	0.978
Micro_R	0.522	0.271	0.467	0.547	0.305	0.556
Micro_F1	0.664	0.415	0.597	0.681	0.450	0.709

classification.

## 5 Conclusions and future work

A novel hybrid learning model for sentiment classification of Chinese micro-blog was proposed. The model has following merits. Firstly, by analyzing different types of words in micro-blog that may affect the emotional tendency, a comprehensive processing is designed. Therefore, the sentiment analysis of micro-blog text can be better solved. Secondly, by combining the HowNet and NTUSD sentiment dictionary, as well as some manually collected internet hot words, the Chinese micro-blog corpus utilized in this paper is expanded. Thirdly, by utilizing the two-stage hybrid learning model, the advantages of sentiment dictionary and machine learning can be combined. Especially in the second stage, the remaining dataset with ambiguous tendencies can be better determined. Therefore, the total sentiment classification accuracy is improved.

However, the recall rates of our approach are slightly lower. For evaluated texts that are not covered by the extracted features, the hybrid model failed to make right judgments. Therefore, it remains a potential future work for us to develop more effective evaluation method for this situation.

## References

- [1] CNNIC. Statistical reports from CNNIC [EB/OL]. [2015–12–30]. <http://www.cnnic.net.cn/hlwfzyj/hlwzxbg/hlwjtjbg/>.
- [2] ZHAO Yan-yan, QIN Bing, LIU Ting. Sentiment analysis [J]. Journal of Software, 2010, 21(8): 1834–1848. (in Chinese)
- [3] XIE Li-xing, ZHOU Ming, SUN Mao-song. Hierarchical structure based hybrid approach to sentiment analysis of chinese micro blog and its feature extraction [J]. Journal of Chinese Information Processing, 2012, 26(1): 73–83. (in Chinese)
- [4] BAKLIWAL A, FOSTER J, PUIL J, O'BRIEN R, TOUNSI L, HUGHES M. Sentiment analysis of political tweets: Towards an accurate classifier [C]// Proceedings of the Workshop on Language in Social Media (LASM 2013). Atlanta: ACL, 2013: 49–58.
- [5] BARBOSA L, FENG J. Robust sentiment detection on twitter from biased and noisy data [C]// International Conference on Computational Linguistics (ICCL 2010). Beijing: CIPS, 2010: 36–44.
- [6] KIM SM, HOVY E. Automatic detection of opinion bearing words and sentences [C]// International Joint Conference on Natural Language Processing (IJCNLP 2005). Jeju Island: Springer, 2005: 61–66.
- [7] LI Shou-shan, SOPHIA Y, HUANG Chu-ren, SU Yan. Construction of Chinese sentiment lexicon using bilingual information and label propagation algorithm [J]. Journal of Chinese Information Processing, 2013, 27(6): 75–81. (in Chinese)
- [8] HAN Zhong-ming, ZHANG Yu-sha, ZHANG Hui, WAN Yue-liang, HUANG Jin-hui. On effective short text tendency classification algorithm for chinese micro blogging [J]. Computer Applications and Software, 2012, 29(10): 89–93. (in Chinese)
- [9] PANG Zhen-jun, GAO Li-bo, YAO Tian-fang. Web text tendency classification based on sentiment phrase [C]// Chinese Opinion Analysis Evaluation (COAE 2014). Kunming: CIPS, 2014: 179–186. (in Chinese)
- [10] SUN Song-tao, HE Yan-xiang, CAI Rui, LI Fei, HE Fei-yan. LEO\_WHU's report on COAE2014 [C]// Chinese Opinion Analysis Evaluation (COAE 2014). Kunming: CIPS, 2014: 27–34. (in Chinese)
- [11] LUO Yi, LI Li, TAN Song-bo, CHEN Xue-qi. Sentiment analysis on Chinese micro-Blog corpus [C]// Chinese Opinion Analysis Evaluation (COAE 2014). Kunming: CIPS, 2014: 123–130. (in Chinese)
- [12] PANG Bo, LEE L, VAITHYANATHAN S. Thumbs up? Sentiment classification using machine learning techniques [C]// Conference on Empirical Methods in Natural Language Processing (EMNLP 2002). Philadelphia, PA: ACM, 2002: 79–86.
- [13] SUN Yan, ZHOU Xue-guang, FU Wei. Unsupervised topic and sentiment unification model for sentiment analysis [J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2013, 49(1): 102–108. (in Chinese)
- [14] TAN Chen-hao, LEE L, TANG Jie, JIANG Long, ZHOU Ming, LI Ping. User-level sentiment analysis incorporating social networks [C]// International Conference on Knowledge Discovery and Data



- Mining (KDD 2011). San Diego, CA: ACM, 2011: 1397–1405.
- [15] SOCHER R, PENNINGTON J, HUANG E H, NG A Y, MANNING C D. Semi-supervised recursive auto-encoders for predicting sentiment distributions [C]// Conference on Empirical Methods in Natural Language Processing (EMNLP 2011). Edinburgh, UK: ACM, 2011: 151–161.
- [16] LIU Zhi-guang, DONG Xi-shuang, GUAN Yi, YANG Jin-feng. Reserved self-training: a semi-supervised sentiment classification method for Chinese Microblogs [C]// International Joint Conference on Natural Language Processing (IJCNLP 2013). Nagoya, Japan: ACL, 2013: 455–462.
- [17] PARK A, PAROUBEK P. Twitter as a corpus for sentiment analysis and opinion mining [C]// International Conference on Language Resources and Evaluation (LREC 2010). Valletta, Malta: DBLP, 2010: 1320–1326.
- [18] DAVIDOV D, TSUR O, RAPPOPORT A. Enhanced sentiment learning using twitter hashtags and smileys [C]// International Conference on Computational Linguistics (ICCL 2010). Beijing: ACM, 2010: 241–249.
- [19] RUSTAMOV S, CLEMENTS M A. Sentence-level subjectivity detection using neuro-fuzzy models [C]// Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2013). Atlanta: ACL, 2013: 108–114.
- [20] REN Yong, KAJI N, YOSHINAGA N, KITSUREGAWA M. Sentiment classification in under-resourced languages using graph-based semi-supervised learning methods [J]. IEICE Transactions on Information and Systems, 2014, 97(4): 790–797.
- [21] MAO Xia, JIANG Lin, XUE Yu-li. Affect computation of chinese short text [J]. IEICE Transactions on Information and Systems, 2012, 95(11): 2741–2744.
- [22] BAIKE. The definition of Chinese micro-blog [EB/OL]. [2015–12–30]. <http://www.baikewiki.com/wiki/%E5%BE%AE%E5%8D%A>.
- [23] HOWNET. The latest hownet news [EB/OL]. [2015–12–30]. [http://www.keenage.com/html/e\\_index.html](http://www.keenage.com/html/e_index.html).
- [24] SHUJUTANG. NTUSD released by the National Taiwan University [EB/OL]. [2015–12–30]. <http://www.datatang.com/data/11837>.
- [25] NLPiR. ICTCLAS 2015 [EB/OL]. [2015–12–30]. <http://ictclas.nlpir.org/>.
- [26] LIBSVM. A library for support vector machines [EB/OL]. [2015–12–30]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- [27] TAN Song-bo, WANG Su-ge, LIAO Xiang-wen, LIU Kang. Fifth Chinese opinion analysis evaluation report [C]// Chinese Opinion Analysis Evaluation (COAE 2013). Shanxi: CIPS, 2013: 5–33. (in Chinese)

(Edited by HE Yun-bin)

**Cite this article as:** LI Fang-fang, WANG Huan-ting, ZHAO Rong-chang, LIU Xi-yao, WANG Yan-zhen, ZOU Bei-ji. Chinese Micro-blog sentiment classification through a novel hybrid learning model [J]. Journal of Central South University, 2017, 24(10): 2322–2330. DOI: <https://doi.org/10.1007/s11771-017-3644-0>.