IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 9, 2018, accepted March 19, 2018, date of publication March 28, 2018, date of current version April 23, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2820163

Advanced Sentiment Classification of Tibetan Microblogs on Smart Campuses Based on Multi-Feature Fusion

LIRONG QIU, QIAO LEI, AND ZHEN ZHANG
School of Information Engineering, Minzu University of China, Beijing 100081, China

Corresponding author: Lirong Qiu (qiu_lirong@126.com)

This work was supported in part by the National Nature Science Foundation of China under Grant 61672553 and in part by the (Ministry of Education in China) Project of Humanities and Social Sciences under Project 16YJCZH076.

ABSTRACT Sentiment analysis is an important problem in natural language processing, which plays an important role in many fields, such as information forecasting, knowledge classification, and product review. Because Tibetan microblogs have their own unique form, particularly the heterogeneous features, such as the emoticons, the grammatical relations, and the speech, the existing sentiment analysis method has difficulty in analyzing the emotions that such microblogs express. In this paper, we propose a sentiment classification method for Tibetan microblogs based on multi-feature fusion. To better study the affection of affective features, this paper first determines the theme of Weibo texts and chooses smart campuses as theme of Weibo texts for analyzing the influence of each feature on the sentiment of the microblog. Then, these features are fused as a multi-feature, and the sentiment of the Tibetan microblog is classified according to the multi-feature fusion. The experimental results demonstrated that the sentiment classification algorithm based on feature fusion improved the accuracy of microblog sentiment classification.

INDEX TERMS Sentiment analysis, smart campus, Tibetan microblog, multi-feature fusion, sentiment classification.

I. INTRODUCTION

Smart campuses based on the idea of building a digital campus use new information technologies, such as cloud computing, Internet of Things and virtualization, to change the way in which school administrators, teachers, students and parents interact with campus resources. Microblogs with a smart campus theme have appeared in large numbers.

With the rapid development of the Internet, social networks, which are represented by microblogs, are gaining increasing popularity as an increasing number of users are more willing to express their views or express their feelings on social networks. At present, the most representative online social networking products are Facebook, Twitter, Tencent QQ and the Sina microblog in China. The rapid development of social networks provides a large number of valuable comments on figures, events, and products, and such information contains a variety of emotions and emotional color differences, such as joy, anger, sadness, criticism and like.

As an important channel for information exchange and news dissemination, the microblog platform has greatly influenced the lives of modern people. This platform can bring together the emotional states of a large number of users or their views on an event in a short time. Compared to traditional text, microblog text has a peculiarity, namely, the text content expression is generally fragmented and irregular. Meanwhile, the content of microblogs is quite complicated, involving subjects such as politics, culture, economics, society, and so forth.

In addition to the above common features, Tibetan microblogs also have unique features; for example, many Tibetan microblogs are mixed with Chinese, and some even contain English. Statistics have shown that most of these Chinese texts are translated from the Tibetan content to help other users better understand the content.

According to statistics, the focus of Tibetan users relates to the content of news, social economy, language culture, science and technology education, religion, culture and arts, tourism, the environment, Tibetan medicine and pharmacology. These contents not only contain common parts that are the same as other national users but also contain

Tibet's national culture characteristics. For example, such microblogs give more attention to content relevant to Tibetan Buddhism [1].

Tibetan microblogs have been increasingly concerned about technology education. Sentiment analysis has been considered to more clearly understand the sentiments of the microblog text and to determine the theme of the microblog content regarding smart campuses. Sentiment analysis technology refers to the processes through which subjective texts' sentiments are analyzed, processed, summarized and inferred. Through this technology, we can understand public opinion perception on an event or product. It plays an important role in many fields, such as information forecasting, public opinion monitoring, product review, and so forth [2], [3].

With the popularization and development of social networks, increasingly more Tibetan netizens are willing to express themselves on figures, events, products and other subjective ideas through sentiment analysis [4]. However, the traditional rules of text can hardly cover the analysis of Tibetan microblog sentiment tasks. Moreover, user groups in Tibet are the minority from the perspectives of cost, profit and other commercial considerations; thus, commercial companies haven not provided specialized services for Tibetan topics and Tibetan comments. Consequently, there are no sentiment analysis works about Tibetan social networks at present.

II. RELATED WORK

In 2001, S. Das and M. Chen defined ''emotion'' as the positive and negative opinions from a message [5]. In 2012, B. Liu defined sentiment analysis (also known as opinion mining) as the subjective information. In other words, sentiment analysis is to analyze users' opinions, attitudes, and emotions contained in topics according to the products, services, events and topics [6].

Liu *et al.* [7] noted that the Tibetan text extracted from pages with the natural tag information can be used to build a raw text classification corpus, Internet word corpus, phrase corpus and so on. Cheng *et al.* [8] used online product reviews as samples to investigate the characteristics and strategies in the attitude analysis of short texts, and a text attitude analysis system was constructed based on dictionaries and rules. Zhu *et al.* introduced a simple HowNet-based method for semantic orientation computation of Chinese words [9]. W. Du et al. proposed a novel scalable word semantic orientation computing framework, first building an undirected graph using word similarity computing technology and then dividing the word-to-word graph based on the ''minimum-cut'' approach. In this way, function optimization is adopted in this word semantic orientation computing framework and resolved by using a simulated annealing algorithm [10]. Xia *et al.* [11] proposed a dual-view co-training algorithm based on a dual-view bag-of-words representation for semi-supervised sentiment classification. Yessenalina *et al.* [12] proposed a joint two-level approach for document-level sentiment classification that simultaneously extracts useful (i.e., subjective) sentences and predicts document-level sentiment based on the extracted sentences.

S. Aoki and O. Uchida suggested that the estimated emotions represented by emoticons are important for reputation analysis, and they proposed a method to create emotional vectors of emoticons automatically using the collocation relationships between emotional words and emoticons, which are derived from many weblog articles [13]. Later, Barbosa and Feng [14] proposed an approach to automatically detect sentiments on Twitter messages (tweets) that explores some characteristics of how tweets are written and meta-information of the words that compose these messages. Tang *et al.* [15] proposed a representation learning approach to build a large-scale sentiment lexicon from Twitter. A. Agarwal et al. introduced a novel approach of adding semantics as additional features into the training set for sentiment analysis. For each extracted entity (e.g., iPhone) from tweets, its semantic concept (e.g., ''Apple product'') is added as an additional feature, and the correlation of the representative concept with negative/positive sentiment is measured [16]. Gao *et al.* [17] proposed a sentiment analysis approach based on sentiment unit and opinion target, and the extraction of sentiment unit and sentiment evaluation object was based on the co-occurrence probability. Liang *et al.* [18] explored the feasibility of performing Chinese microblog sentiment analysis by deep learning and proposed a novel sentiment polarity transition model based on the relationship between neighboring words in a sentence to strengthen the text association.

Cheng-gong [19] proposed a sentiment analysis method based on polarity, in which the modifiers and polarity words were combined into polarity phrases, and the phrases were used as the basic unit to analyze the polarity of sentences and texts. Xie *et al.* [20] compared the performances of three methods based on emoticons, the sentiment lexicon and the hybrid approach over the hierarchical structure using SVM, and the experiments showed that the SVM-based hybrid approach achieved the best performance. Zhang *et al.* [21] used the emoticons from microblogs combined with emotion words to build the Chinese sentiment corpus, constructed a Bayes classifier and used the entropy to improve the performance based on the corpus. Yuan *et al.* [22] proposed a method of constructing the semantic feature space by combining the Tibetan sentence structure and semantic feature vectors, and carrying out a Tibetan microblog emotion analysis method.

However, the aforementioned sentiment analysis methods are for English or Chinese. The differences in language uniqueness and the users' focus have made it difficult to copy social network sentiment analysis technology in English and Chinese directly to Tibetan.By reading a large amount of paper, the mechine learning algorithm was introduced into the field of Tibetan sentiment analysis to improve the accuracy of Tibetan sentiment analysis and to expand the

emotional resources of Tibetans and establishing a dictionary of Tibetan emotions can make the tagging corpus more complete.To conduct sentiment analysis and keyword extraction on Tibetan topics on social networks, it is necessary to concentrate on the characteristics of the Tibetan language and Tibetan users' interests. In this paper, we propose a sentiment classification method for Tibetan microblogs based on multi-feature fusion.

III. FUNDAMENTAL WORK

A. SENTIMENT CLASSIFICATION OF MICROBLOGS BASED ON SEQUENCE RULES

Microblog text is generally brief with a relatively straightforward emotional expression and simple grammar, and texts with similar grammatical relationship patterns generally share similar ways of expressing emotion. This section classifies the sentiment through analyzing sentiment relations of text sequences in Tibetan microblogs. First, the microblog is serialized according to its sentiment features, potential sequence features are extracted according to the training corpus, and then the sequence feature that satisfies the minimum support and minimum confidence is selected as the vector feature. Second, the sequence is vectorized according to the features of the vector, the classification model is trained, and the microblog classification is tested.

In the microblog corpus, a simple sequence combination is more frequent, but some rare combinations of sequences also exist; thus, it is not possible to select the sequence feature only through the minimum support and the minimum confidence. We propose a multi-minimum support strategy to solve this problem, in which the minimum support of a rule depends on the product of the minimum frequency of the items in the training set and the parameter. This strategy can not only sustain rare sequence features with a low support but also solve the phenomenon of over-fitting with frequent sequence features. The process of microblog sentiment classification based on the sequence features is shown in Figure 1. The specific steps of microblog sentiment classification based on the sequence features are as follows:

A) Serialize the training corpus according to the features of emotional words, negative words and conjunctions. If there is a conjunction feature in the text, then it will be used as an item alone.

B) Construct an n-dimensional feature vector according to the sequence mining pattern based on a multi-minimum support strategy.

C) Vectorize the feature of the test text; use X as a feature in every $X \rightarrow y$ sequence pattern. If the sequence of a microblog contains X, then the corresponding vector eigenvalue should be set to 1; otherwise, it should be set to 0.

D) Use the classification models trained with the corpus, which are annotated to classify the corpus to be classified.

B. THE SENTIMENT CLASSIFICATION BASED ON EMOTICONS

There are many emoticons in microblogs, which can express emotions more vividly than words, and microblog users are

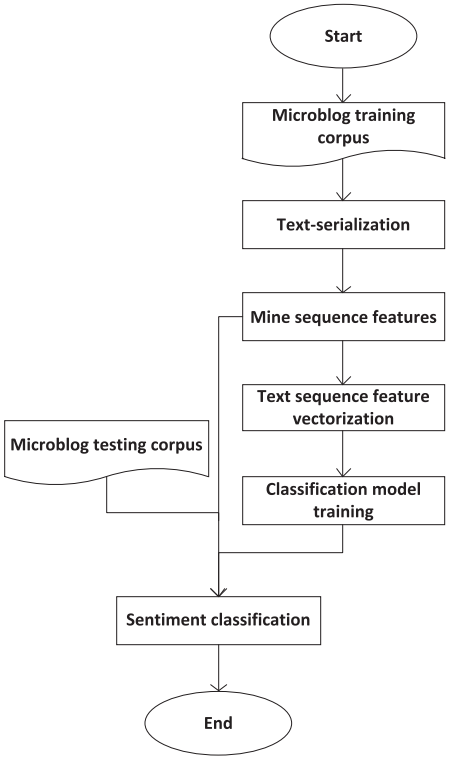


FIGURE 1. Flow chart of sentiment classification based on sequence rule features.

increasingly accustomed to using emoticons to express their emotions when publishing a microblog. Therefore, we can forecast the sentiment tendency of the text and judge the sentiment classification through analyzing the impact of emoticons on the sentiment of microblog text. As shown in Figure 2, emoticons are becoming increasingly diverse; in addition to the default emoticons, there are more diverse types of emoticons, such as ''mood'' and ''rage comic''.

A microblog has short text with fewer features, and as an important sentiment feature, the emoticon plays an important role in effectively expressing emotion. According to statistics and related studies, in the 279,931 collected public Tibetan microblogs, there are 54,655 microblogs that have been annotated, which accounts for 19.52% of microblogs. Additionally, there are 2.56 emoticons on average in each microblog. In this paper, we artificially select 94 emoticons and classify them into happiness, like, surprise, sadness, disgust, anger and fear, as shown in Table 1. The number shows the number of emoticons that belong to each category. In this section, we divide the microblog text into different sentimental categories through matching its emoticons.

To analyze the influence of emoticons on microblog text sentiment, we assume that the microblog contains emoticons (E) as T and extract the emoticon set in the microblog as $\{w_i\}$, where i indicates the position of the emoticons within the blog. The prior probabilities of every emoticon belong to the sentiment categories that can be calculated based on the naive Bayes classifier, which is necessary to predict the

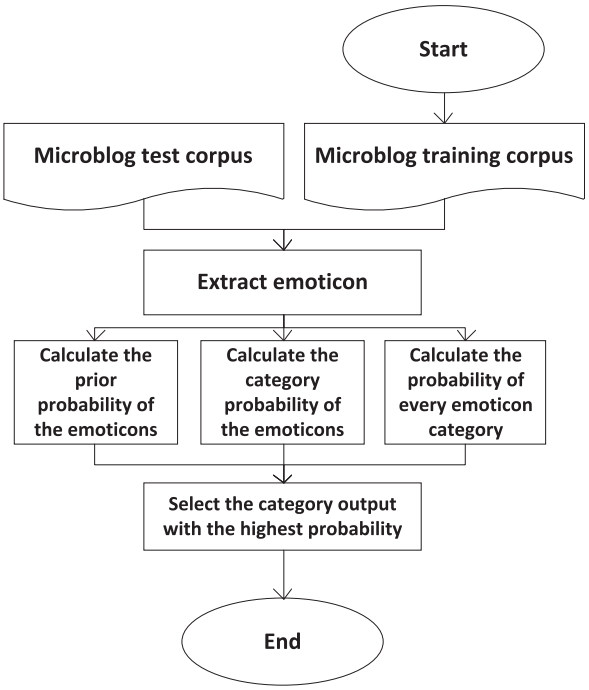


FIGURE 2. Flow chart of sentiment classification based on sequence rule features.

TABLE 1. Microblog emoticon dictionary.

Category	Emoticons	Number
like	Love, Awesome, good, Like, Hug, Shake	23
happiness	Wild Laugh, haha, laugh, Yeah, Very happy	21
disgust	Despise, Spit, doge, Sly, Thumbs Down, Nose Pick, Slight	14
fear	Fear, Thunder, Shy, Nervous, Horror, Hail	17
anger	Anger, Scold, Silent, Pooh-pooh, Fist, Scream	18
sadness	Sadness, Disappointed, Sorrow, Purr, Heart-broken, Shrunken	12
surprise	Surprise, Amazed, Huge Sweat, Sweat, Shock, Panic	10

sentiment of microblog text.

$$P(w_i|c_j) = \frac{n^{c_j}(w_i) + 1}{\sum_q (n^{c_j}(w_q) + 1)}$$

where j = 1, 2, 3, 4, 5, 6, 7, $n^{c_j}(w_i)$ shows the number that appears in the category in all training corpora.

We predict the sentiment category of the corpus to be classified based on the following formula:

$$c^*(t) = \underset{j}{\operatorname{argmax}} P(c_j) \prod_i P(w_i|c_j)$$

$P(c_j)$ is the prior probability of c_j.

Since one piece of microblog contains fewer emoticons, the training corpus is divided into five equal-sized copies when calculating the prior probability of the emoticon. One copy is used to calculate the prior probability of the emoticons according to its generated classifier, and the remaining four copies are superimposed to amend the classifier, which is to avoid a certain emoticon leading to classification deviations.

TABLE 2. The microblog corpus includes emoticons.

Category	like	happiness	sadness	disgust	surprise	anger	fear
Number	**34985**	**5827**	**2517**	**1959**	**447**	**248**	**298**

To analyze the sentiment of microblogs, the emoticons are extracted first because the emoticons in a microblog are expressed with [ ], such as "[hee hee]", "[haha]", and "[sad]". Thus, it is easy to extract the emoticons. Second, the probability that each emoticon belongs to a different sentiment category is calculated according to the prior probabilities of every emoticon. If the emoticons belong to the same category, then their category probability is added and the total category probability is obtained. Therefore, the category with the maximum probability is selected as the sentiment category of the microblog. The process is shown in figure 2.

The specific steps of forecasting sentiment classification of a microblog based on sequence rules are as follows:

a) Extract the emoticons in the microblog, marked as $\{E_i\}$;

b) Calculate the prior probability $P(e_i|c_j)$ of the emoticon e_i according to the training samples, in which $e_i \in E$, $j = 1, 2, 3, 4, 5, 6, 7$;

c) Calculate the category probability $c^*(j)$ of the emotion;

d) Count the probability of each category $\sum(c^*(j))$ and select the maximum category max $(c^*(j))$ as output.

The data of the collected Tibetan microblog corpus that expresses the sentiment through emoticons mentioned above are shown in Table 2. According to the empirical analysis based on emoticons, the prior probabilities of microblog emoticons in different sentiments can be calculated on the basis of the above process steps, as shown in Table 3.

IV. THE SENTIMENT CLASSIFICATION ALGORITHM BASED ON FEATURE FUSION

A. FEATURE SELECTION

A microblog's part of speech features, emoticon features and grammatical relation features all reflect the microblog's sentiment in one aspect, but the sentimental classification effects of a single-feature-based microblog are not obvious. To improve the accuracy of microblog sentiment classification, this section designs a sentiment classification algorithm based on feature fusion.

The features and meanings of the selected algorithms are shown in Table 4. From the aspect of part of speech, the N-dimensional features such as "d-v-v", "r-d-v", ... "r-d-a" are selected using the 3-POS pattern feature, and every dimension of the feature is the χ^2 statistic of the pattern. From the aspect of grammar dependence relationship, the N-dimensional features such as "VP(KP,VP)", "VP(KP,VP)", ... "NP(KP,N)" are selected, and the features use 3-POS pattern as weight. From the aspect of emoticons, a 7-dimensional feature of like, happiness, sadness, disgust, surprise, anger and fear is selected, and the value of each feature dimension shows the probability that the microblog belongs to this category. From the aspect of sentiment words,

TABLE 3. The prior probabilities of some emoticons.

	Number	like	happiness	sadness	disgust	surprise	anger	fear
[Haha]	2953	0.891297	0.051134	0.002709	0.027768	0.022689	0.000338	0.004063
[Happy]	411	0.768856	0.209246	0.002433	0.004866	0.009732	0.002433	0.002433
[Bow]	429	0.011655	0.960373	0.002331	0.004662	0.016317	0.002331	0.002331
[Great]	60	0.016667	0.9	0.016666	0.016666	0.016667	0.016667	0.016667
[Support]	166	0.006024	0.656626	0.006024	0.066265	0.253012	0.006024	0.006024
[Sad]	223	0.013452	0.040358	0.004484	0.919282	0.0134529	0.004484	0.004484
[Sly]	185	0.005405	0.037838	0.005405	0.005405	0.935135	0.005405	0.005405
[Nervous]	24	0.041667	0.083333	0.041667	0.083333	0.041666	0.041666	0.666666
[Sweat]	233	0.025751	0.030043	0.901287	0.012875	0.004291	0.008583	0.017167
[Shy]	143	0.020979	0.027972	0.006993	0.006993	0.923077	0.006993	0.006993
[Surprise]	355	0.008450	0.036620	0.819718	0.061971	0.042253	0.028169	0.002816

TABLE 4. Features and meanings of feature fusion algorithm.

Feature	Number	Meaning	Normalized Value
Part of Speech Feature (F_{pos})	1	d-v-v	VALUE_DVV
	2	r-d-v	VALUE_RDV
	...	...	..
	N	r-d-a	VALUE_RDA
Syntactic Dependency ($F_{grammar}$)	N+1	VP(KP,VP)	VALUE_G1
	N+2	VP(KP,VP)	VALUE_G2
	...	...	...
	N+M	NP(KP,N)	VALUE_GM
Emoticon (F_{emot})	N+M+1	happiness	VALUE_ HAPPINESS
	N+M+2	like	VALUE_LIKE
	N+M+3	anger	VALUE_ANGER
	N+M+4	sadness	VALUE_SADNESS
	N+M+5	fear	VALUE_FEAR
	N+M+6	disgust	VALUE_DISGUST
	N+M+7	surprise	VALUE_SURPRISE
Sentiment Word (F_{dict})	N+M+8	happiness	VALUE_ HAP_NUM
	N+M+9	like	VALUE_LIK_NUM
	N+M+10	anger	VALUE_ANG_NUM
	N+M+11	sadness	VALUE_SAD_NUM
	N+M+12	fear	VALUE_FEA_NUM
	N+M+13	disgust	VALUE_DIS_NUM
	N+M+14	surprise	VALUE_SUR_NUM

it also selects the 7-dimensional feature of like, happiness, sadness, disgust, surprise, anger and fear, and each feature dimension value denotes the normalized value of the number of sentiment words in microblogs.

Consider microblog feature vector (V),

$$\mathrm{V} = \left\{F_{pos}, F_{grammar}, F_{emot}, F_{dict}\right\},$$

in which F_{pos} denotes the 3-POS mode feature, corresponding to feature 1-N;

$F_{grammar}$ indicates the feature of the grammatical dependency, corresponding to feature N-N + M;

F_{emot} indicates the feature of the emoticons corresponding to feature N+M-N+M+7;

and F_{dict} indicates the feature of the sentiment words, corresponding to feature N+M+7- N+M+14.

B. VECTOR REPRESENTATION OF MICROBLOG TEXT

We vectorize the microblog text based on the feature set $\mathrm{T} = \{t_1, t_2, \ldots, t_m\}$, and the vector form of the text is shown as follows: $d = \{(t_1, w_{i1}), (t_2, w_{i2}), \ldots, (t_m, w_{im})\}$ $t_i(1 \leq i \leq m)$ represent the features, and w_{ij} is the weight of t_i. In this paper, we vectorize the microblog text according to the features selected from sequence rules, semantics and the emoticons.

We define a space vector $\mathrm{Vec} = \{F_1, F_2, F_3, F_4, F_5, F_6\}$ to describe the features of text. F_1 is the feature of emotional words. There are 7 types of emotional words that are selected as the feature, and the word frequency is selected as its weight. F_2 is the feature of emotional affect factors; the parallel conjunctions, turning words, negative words and degree adverbs are selected as the feature, and the word frequency is selected as its weight. F_3 is the feature of sequence rules, 50-dimensional features are selected according to the minimum support degree of 0.035, and the weight is the value of TF-IDF. F_4 is the feature of 3-POS, the former 200 dimensions of 3-POS are selected as the feature based on the χ^2, and the weight is the value of TF-IDF. F5 is the feature of syntax dependency, and the former 140 dimensions of 3-POS are also selected as the feature based on the χ^2. F6 is the feature of emoticons. There are 7 types of emoticons selected, and the prior probability of the emotion category is used as the characteristic weight.

C. MULTI-CLASSIFICATION STRATEGY IMPLEMENTATION BASED ON SVM

According to the statistical results of the existing training corpus, it is found that there is a serious imbalance in the proportion of the number of corpora in different sentiment categories. To avoid oversampling and undersampling in the training process, we consider the problem of the unbalanced corpus in the combination of the multi-classification strategy, and according to the details of the number of corpora, the multi-classification strategy based on a partial binary tree is adopted. The structure chart is shown in Figure 3. The data proportion of the training corpus is shown in Table 5. The actual distribution of all types of sentiments in the microblog also shows the following proportion relation. Therefore, it is suitable to adopt the partial binary tree classification strategy.

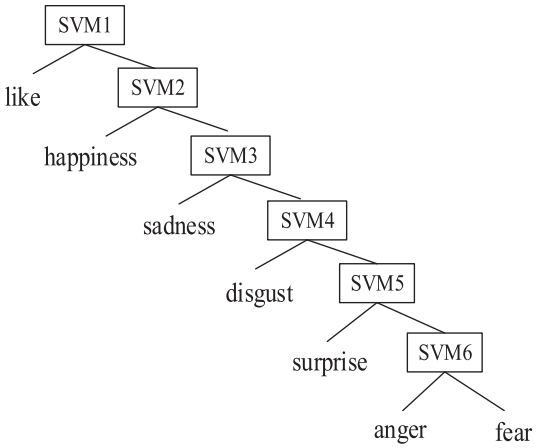


FIGURE 3. Multi-classification structure of partial tree.

TABLE 5. Training corpus data details.

Category	like	happiness	sadness	disgust	surprise	anger	fear
Number	12000	6000	3000	1500	750	375	375

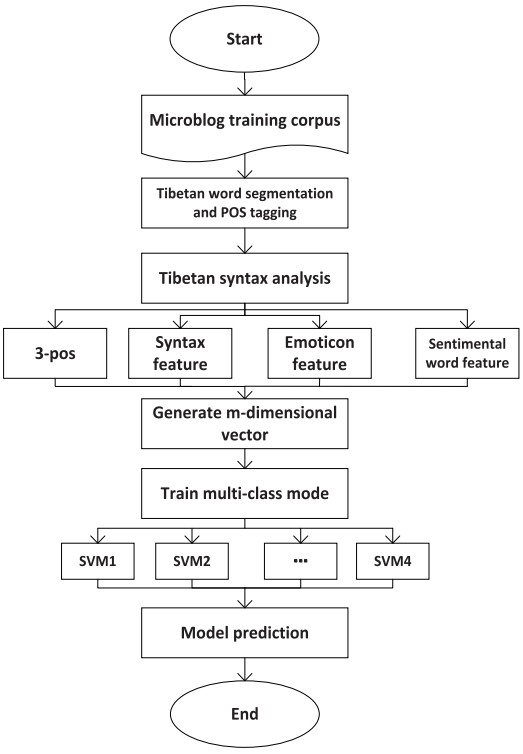


FIGURE 4. Feature fusion algorithm implementation flow.

Feature extraction is an important step. There are three main steps in extracting features with a sentimental tendency in a microblog, as shown in Figure 4. First, we use the tools of Tibetan word segmentation and syntactic analysis developed by Nong Congjun from the Chinese Academy of Social Sciences for the preparation work of word segmentation, POS tagging and syntactic dependency. Then, according to the sentiment dictionary and pre-processed microblog, we extract and calculate the features. Finally, we train each SVM model according to the generated vectors and test the experimental results of the multi-classification model with the test corpus.

V. EXPERIMENTS AND ANALYSIS

A. DATA SET

The Sina Microblog increases the number of requests on the default public API, and it cannot directly call the interface to obtain a large number of microblogging resources. In the course of the experiment, microblog crawling tools were designed mainly for a large number of Tibetan microblog resources. The specific processes are as follows: first, manually build the microblog crawling seed set with a focus on Tibetan microblog users, and select 126 Tibetan microblog users with higher frequency as the initial seed set. Second, crawl the microblog information related to smart campuses according to randomly selected user seeds and filter the microblog text after crawling every piece of information. If it is a Tibetan microblog, then save it and obtain access to related comments; while accessing other participants of the microblog, if these users have not been visited, they will be added to the seed set for subsequent visits. Finally, according to the ID of each microblog, archive the Tibetan microblog and related comments after crawling. We set the time period from February 2016 to July 2016. Through one-week of crawling with the crawler tool, a total of nearly 300,000 Tibetan microblogs are acquired, and the microblog texts are processed as follows:

1) Remove miscellaneous items: remove the microblog hyperlinks, sources, user locations, and user information;
2) Theme extraction: extract the related subjects from #theme#;
3) Emoticons: extract emoticons through regular rules;
4) Tibetan word segmentation and Tibetan syntactic analysis and processing.

After pretreatment of the Tibetan microblog text, by manual processing of markers, we use the JSON format file to store the processed microblog information, which includes microblog id, user name, original text, microblog Tibetan text, theme, Tibetan POS tagging, Tibetan syntax tree and sentiment identity, as shown in Figure 5.

B. EVALUATION INDICATORS

Accuracy, recall and F value are generally used as the evaluation indicators. Accuracy rate refers to the proportion of microblog entries in the classification results consistent with the manual annotation results in the number of such types of text. The recall rate refers to the proportion of the manually identified micro-blog text that is consistent with the manual annotation results in the test set accounting for the total number of texts. In other words, the proportion of correctly predicted samples is considered to belong to all samples of

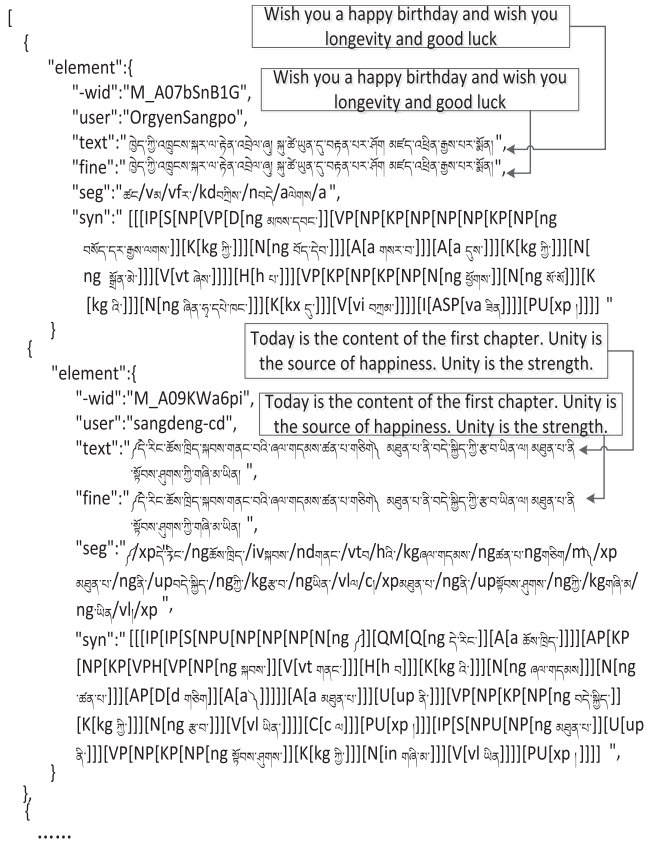


FIGURE 5. Pretreatment of Tibetan microblogging information.

that class. To solve the multi-category problem, we set the evaluation criteria as macro-average and micro-average. Qiu *et al.*, based on the calculation formula, proposed a method to calculate the accuracy rate, recall rate and F value of macro-average and micro-average [23].

C. EXPERIMENT DESIGN

The design and analysis of the proposed algorithm were introduced in the previous section. Next, we need to design comparative experiments to verify the effectiveness of the proposed multi-feature fusion of the Tibetan microblog sentiment algorithm. In section 3.1, the sequence rule model was introduced, and it classifies the microblog sentiment according to the association rules. In addition to the rule method, this paper takes into account the unique features of emoticons, sentimental words and semantics in microblogs, and relative contrast experiments are performed for combinations of different features. The contrast experiments are designed as follows:

1) Experiment I is $F1 + F2$, designed to verify the influence of multi-feature fusion based on emotion words and emotional affect factors on the accuracy of sentiment classification.
2) Experiment II is $F1 + F2 + F3$, designed to verify the influence of multi-feature fusion based on emotion words, emotional affect factors and the sequence rules.
3) Experiment III is $F1 + F2 + F4$, designed to verify the influence of multi-feature fusion based on emotion words, emotional affect factors and the 3-POS.
4) Experiment IV is $F1 + F2 + F5$, designed to verify the influence of multi-feature fusion based on emotion words, emotional affect factors and the syntax dependency.
5) Experiment V is $F1 + F2 + F6$, designed to verify the influence of multi-feature fusion based on emotion words, emotional affect factors and the emoticons.
6) Experiment VI is $F1 + F3 + F4$, designed to verify the influence of multi-feature fusion based on emotion words, the sequence rules and the 3-POS.
7) Experiment VII is $F1 + F3 + F5$, designed to verify the influence of multi-feature fusion based on emotion words, the sequence rules and the syntax dependency.
8) Experiment VIII is $F1 + F4 + F5$, designed to verify the influence of multi-feature fusion based on emotion words, the 3-POS and the syntax dependency.
9) Experiment IX is $F1+F3+F4+F5$, designed to verify the influence of multi-feature fusion based on emotion words, the sequence rules, the 3-POS and the syntax dependency.
10) Experiment X is $F1 + F2 + F3 + F4 + F5$, designed to verify the influence of multi-feature fusion based on emotion words, emotional affect factors, the sequence rules, the 3-POS and the syntax dependency.
11) Experiment XI is $F1 + F3 + F4 + F5 + F6$, designed to verify the influence of emotion words, the sequence rules, the 3-POS, the syntax dependency and the emoticons.

TABLE 6. The results of contrast experiments.

Contrast experiment	Evaluation Indicators Macro accuracy	Macro recall	Macro F value	Micro F value
Experiment I	0.3771	0.1702	0.2346	0.2753
Experiment II	0.8206	0.5094	0.6286	0.5563
Experiment III	0.5483	0.2662	0.3584	0.3515
Experiment IV	0.3501	0.2293	0.2771	0.3035
Experiment V	0.703	0.6522	0.6767	0.7381
Experiment VI	0.7752	0.5673	0.6551	0.5996
Experiment VII	0.7332	0.5581	0.6338	0.5977
Experiment VIII	0.4492	0.2844	0.3483	0.3562
Experiment IX	0.7435	0.5824	0.6532	0.6118
Experiment X	0.7517	0.5844	0.6576	0.6156
Experiment XI	0.8826	0.8786	0.8806	0.8966

D. RESULTS AND ANALYSIS

To verify the effectiveness of the proposed method based on multi-feature fusion, eleven contrast experiments are designed, which are described in detail in section 5.3. The results are shown in Table 6, Figure 6, and Figure 7.

The results presented in table 6, figure 6 and figure 7 reveal that the proposed method is effective. The results of experiments I, II, III, IV and V, which have fused the sequence rules, the 3-POS, the emoticons and the syntax dependency

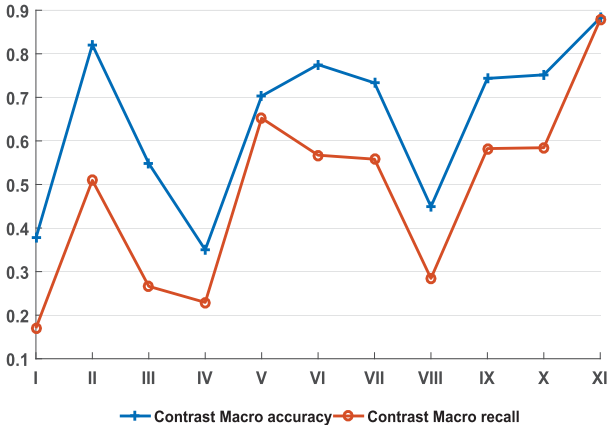


FIGURE 6. The results of contrast experiments.

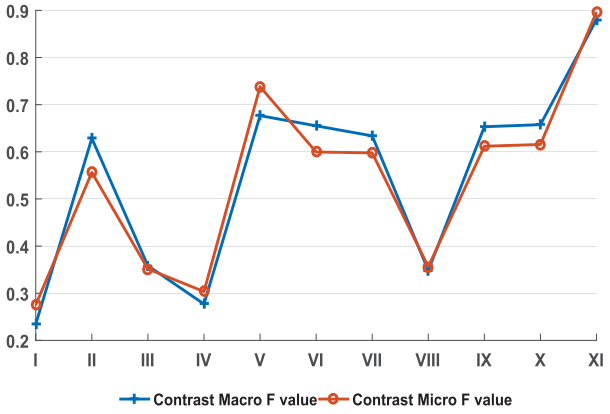


FIGURE 7. The results of contrast experiments.

on the basis of emotion words and emotional affect factors, prove that each feature has different effects on the accuracy of sentiment classification. In particular, the emoticon feature has the greatest influence on the results, which verified that the emoticon feature can clearly distinguish different sentiments. Moreover, experiment II reveals that the sequence rules also have a great influence. However, experiments VI, VII and VIII verify that the syntax dependency feature can improve the performance to a certain extent, but it is not obvious. The results of experiments IX and X reveal that the emotional affect factors have little impact on the accuracy of sentiment classification. Experiment XI, which fused the emotion words feature, the sequence rules, the 3-POS, the syntax dependency and the emoticons, verifies that the multi-feature fusion algorithm can achieve the best performance because the features selected from multiple dimensions can compensate for the sparseness of a single feature.

VI. CONCLUSION

In this paper, we propose a multi-feature fusion method to analyze the sentiment regarding Tibetan microblogs related to smart campuses. First, obtained for Tibetan microblogs with a smart campus theme, the influences of each feature on the sentiment of the microblog are analyzed, particularly the sequence feature and emoticon features, which have a great impact on the sentiment of the microblog. Then, we analyze the sentiment of the microblog according to the multi-feature fusion. To verify the effectiveness of the proposed algorithm, we design eleven contrast experiments. The experimental results confirm that the speech features, emoticon features and grammatical relation features all reflect the microblog's sentiment in one aspect and show that our sentiment classification algorithm based on feature fusion is effective, which can improve the accuracy of microblog sentiment classification.

REFERENCES

[1] C. B, "The new development of standardization and standardization of tibetan terms—A summary of the second session of the third working group on standardization of tibetan terminology," *China Tibetol.*, vol. 2, pp. 189–192, 2014.

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

[3] Y. Ai-Ming, L. Jiang-Hao, and Z. Yong-Mei, "Method on building Chinese text sentiment lexicon," *J. Frontiers Comput. Sci. Technol.*, vol. 7, no. 11, pp. 1033–1039, 2013.

[4] Z. S. Q. W. S. Zhou and S. X. S. Yunchen, "Overview on sentiment analysis of Chinese microblogging," *Comput. Appl. Softw.*, vol. 30, no. 3, p. 45, 2013.

[5] S. Das and M. Chen, "Yahoo! for Amazon: Extracting market sentiment from stock message boards," in *Proc. 8th Asia Pacific Finance Assoc. Annu. Conf.*, vol. 35. Bangkok, Thailand, Jul. 2001, p. 43.

[6] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.

[7] H. D. Liu, M. H. Nuo, and L. L. Ma, "Mining tibetan Web text resources and its application," *J. Chin. Inf. Process.*, vol. 29, no. 1, pp. 170–177, 2015.

[8] N. C. Cheng, M. Hou, and Y. L. Teng, "Short text attitude analysis based on textual characteristics," *J. Chin. Inf. Process.*, vol. 29, no. 2, pp. 163–169, 2015.

[9] Y.-L. Zhu, J. Min, Y.-Q. Zhou, X.-J. Huang, and L.-D. Wu, "Semantic orientation computing based on HowNet," *J. Chin. Inf. Process.*, vol. 20, no. 1, pp. 14–20, 2006.

[10] W. Du, S. Tan, X. Yun, and X. Cheng., "A new method to compute semantic orientation," *J. Comput. Res. Develop.*, vol. 46, no. 10, pp. 1713–1720, 2009.

[11] R. Xia, C. Wang, X.-Y. Dai, and T. Li, "Co-training for semi-supervised sentiment classification based on dual-view bags-of-words representation," in *Proc. ACL*, vol. 1. 2015, pp. 1054–1063.

[12] A. Yessenalina, Y. Yue, and C. Cardie, "Multi-level structured models for document-level sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process., Assoc. Comput. Linguistics*, 2010, pp. 1046–1056.

[13] S. Aoki and O. Uchida, "A method for automatically generating the emotional vectors of emoticons using weblog articles," in *Proc. 10th WSEAS Int. Conf. Appl. Comput. Appl. Comput. Sci.*, 2011, pp. 132–136.

[14] L. Barbosa and J. Feng, "Robust sentiment detection on twitter from biased and noisy data," in *Proc. 23rd Int. Conf. Comput. Linguistics, Posters, Assoc. Comput. Linguistics*, 2010, pp. 36–44.

[15] D. Tang, F. Wei, B. Qin, M. Zhou, and T. Liu, "Building large-scale twitter-specific sentiment lexicon: A representation learning approach," in *Proc. 25th Int. Conf. Comput. Linguistics, Techn. Papers*, 2014, pp. 172–182.

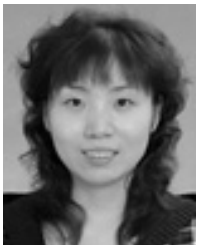
[16] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of twitter data," in *Proc. Workshop Lang. Soc. Media, Assoc. Comput. Linguistics*, 2011, pp. 30–38.

[17] K. Gao *et al.*, "A micro-blog sentiment analysis approach," *J. Chin. Inf. Process.*, vol. 29, no. 4, pp. 40–49, 2015.

[18] J. Liang *et al.*, "Deep learning for Chinese micro-blog sentiment analysis," *J. Chin. Inf. Process.*, vol. 28, no. 5, pp. 155–161, 2014.

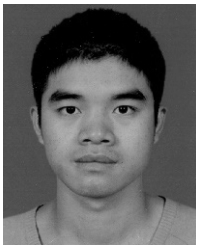
[19] Z. Cheng-Gong, "A sentiment analysis method based on a polarity lexicon," *J. Shandong Univ.*, vol. 47, no. 3, pp. 47–50, 2012.

[20] L. Xie, M. Zhou, and M. Sun, "Hierarchical structure based hybrid approach to sentiment analysis of chinese micro blog and its feature extraction," *J. Chin. Inf. Process.*, vol. 26, no. 1, pp. 73–83, 2012.
[21] S. Zhang, L. Yu, and C. Hu, "Sentiment analysis of Chinese micro-blog based on emotions and emotional words," *Comput. Sci.*, vol. 39, no. 11A, pp. 146–148, 2012.
[22] B. Yuan, T. Jiang, and H. Yu, "Emotional classification method of Tibetan micro-blog based on semantic space," *Appl. Res. Comput.*, vol. 33, no. 3, pp. 682–685, 2016.
[23] L. Qiu, H. Zhang, Z. Zhang, and Q. Pu, "Tibetan microblog emotional analysis based on sequential model in online social platforms," *Complexity*, vol. 2017, no. 1, 2017, Art. no. 5342601.



LIRONG QIU received the Ph.D. degree in computer sciences from the Chinese Academy of Science in 2007. She is currently an Associate Professor of computer sciences with the Information Engineering Department, Minzu University of China. Her current research interests include different aspects of natural language processing, artificial intelligence, and distributed systems.

QIAO LEI is currently pursuing the master's degree with the Minzu University of China. His current research interests include natural language processing, artificial intelligence, and social networking analysis.



ZHEN ZHANG is currently a Graduate Student with the School of Information Engineering, Minzu University of China. His current research interests include natural language processing, artificial intelligence, and sentiment analysis.