

## Research Article

# Tibetan Microblog Emotional Analysis Based on Sequential Model in Online Social Platforms

Lirong Qiu, Huili Zhang, Zhen Zhang, and Qiumei Pu

*School of Information Engineering, Minzu University of China, Beijing, China*

Correspondence should be addressed to Lirong Qiu; [qiu.lirong@126.com](mailto:qiu.lirong@126.com)

Received 18 April 2017; Accepted 5 June 2017; Published 2 August 2017

Academic Editor: Junhu Ruan

Copyright © 2017 Lirong Qiu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of microblogs, selling and buying appear in online social platforms such as Sina Weibo and Wechat. Besides Mandarin, Tibetan language is also used to describe products and customers' opinions. In this paper, we are interested in analyzing the emotions of Tibetan microblogs, which are helpful to understand opinions and product reviews for Tibetan customers. It is challenging since existing studies paid little attention to Tibetan language. Our key idea is to express Tibetan microblogs as vectors and then classify them. To express microblogs more fully, we select two kinds of features, which are sequential features and semantic features. In addition, our experimental results on the Sina Weibo dataset clearly demonstrate the effectiveness of feature selection and the efficiency of our classification method.

## 1. Introduction

Microblogs are widely used in people's daily life. Since this kind of tool is easy to use, it attracts more and more people. Similar to Twitter, Sina Weibo has become the most important microblogging tool in China. In the first quarter of 2016, there were 261 million active monthly users and 120 million active daily users in Sina Weibo. Weibo provides such a platform, where users can not only share information, but also express their emotions by emoticons, voices, and videos [1].

With the explosive growth of users, more and more sellers have realized that the microblog platforms can be used for e-commerce and begin to advertise. Besides users who write microblogs in Mandarin, Tibetan users have the same necessity. For example, Tibetan users would like to advertise for some products. They describe such products in Tibetan and then sell them. This necessity increases, while the Tibetan users increase. According to the statistics of user groups of Sina Weibo [2], the daily updating of Tibetan microblogs is about 2,000.

Customers tend to read product reviews before they buy something online. And merchants prefer hot sellers. If there exist some solutions, which one can help them make decisions? Undoubtedly, emotion analysis can help to understand

opinions and product reviews. Different from Mandarin, Tibetan language has its own characteristics. Tibetan words and phrases are quite different from Mandarin words and phrases. Thus, it is hard to make a simple mapping. It is a challenge to deal with Tibetan microblogs. Another challenge is lack of basic knowledge. Since Tibetan nationality is a kind of minority, few researchers pay attention to Tibetan microblog analysis. However, Tibetan microblog analysis, especially emotion analysis of Tibetan microblogs, is quite useful.

Different from the traditional analysis of emotions, we divide Tibetan sentences into seven categories, instead of positive and negative tendency classification. The tendency classification can achieve a good classification effect whether it is based on the dictionary method or statistical learning method [3]. To tackle the novel and challenging problem of analyzing the emotions of Tibetan microblogs, we consider different kinds of features and classify microblogs by using a multiclassification model. We make several contributions.

First, we analyze the emotions of Tibetan microblogs by using sequential features. Given a set of text, we construct sequences by vectorization. In every sequence, we treat emotional words, conjunctions, negative words, and other emotional characteristics as items. Potential sequential features can be obtained by sequential pattern mining under the

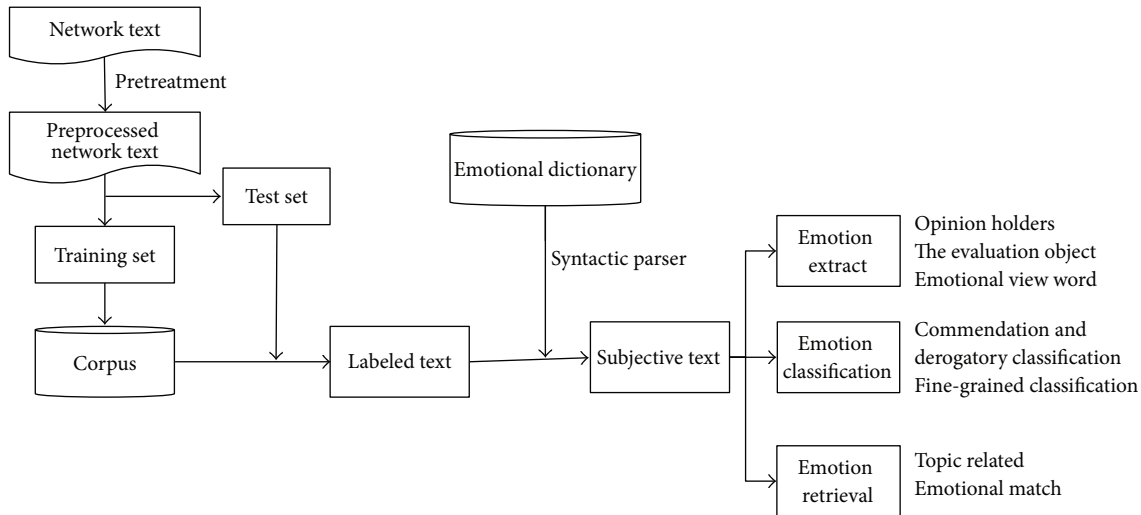


FIGURE 1: Basic process of emotion analysis.

control of the minimum support or the minimal confidence. The main reason to select sequential features is that sequential pattern mining can find latent information.

Second, we continue emotion analysis of Tibetan microblogs by adding semantic features. Semantic features include the emotional words, degree adverbs, and negative words. We construct the space vectors by using part-of-speech features (N-POS) model and syntactic dependencies.

Third, we verify our feature selection and multiclassification on Sina Weibo dataset. The experimental results confirm the effectiveness of feature selection and the efficiency of multiclassification.

The rest of the paper is organized as follows. We review the related work in Section 2. In Section 3, we analyze the emotions of Tibetan microblogs by using sequential features. In Section 4, we continue our analysis by adding semantic features. We report the empirical evaluation results in Section 5 and conclude the paper in Section 6.

## 2. Related Work and Technology

Tibetan microblog emotional analysis on sequential model is related to sequential pattern mining and sentiment analysis.

*2.1. Emotional Analysis.* The systematic study of sentiment analysis began in the early 21st Century. The basic process of emotion analysis is shown in Figure 1, which includes the acquisition of network text resources, text preprocessing, corpus construction, emotional dictionary construction, and the related processing of emotional analysis.

In 2002, Pang et al. firstly applied the method of supervised learning to carry on the emotion tendency classification of movie review text [4]. In the same year, Turney proposed the unsupervised emotion classification method based on semantic tendency [5]. Two kinds of emotional analysis methods were derived: the emotion analysis methods based

on supervised learning and the emotion analysis methods based on unsupervised learning.

At present, using machine learning to analyze the sentiment was still the mainstream. Sentiment analysis methods based on supervised learning were also called the methods based on machine learning. The methods of emotion analysis based on unsupervised learning can be divided into the dictionary-based analysis methods and the rule-based analysis methods.

The specific methods of dictionary-based emotion analysis were to construct the emotional dictionary firstly, in which each emotional word must be given an emotional polarity and an emotional value. Then according to the polarities and the values of the emotional words in the sentence, we computed the emotional score for each sentence. Putting forward an analysis model of fuzzy sets and constructing the membership function, we measured the distance between the emotional tendencies for the sentence in various levels (positive, negative, and neutral). The shortest emotional tendency was determined as the emotional tendency of the whole sentence [6].

Based on the emotional dictionary, the rule-based analysis methods introduced the logic relationship between components in the statement. This kind of methods was an optimization. Rule-based methods took full account of the grammatical relations and sentences. They treated a sentence as a whole of grammatical structure, rather than a stack of words. Statistical similarity between sentences has been studied without sentence parsing [7], which uses the words in sentence as features.

In [8], Liu et al. pointed out that the Tibetan text extracted from pages with natural tag information can be used to construct original corpus, word and phrase corpus, text categorization corpus, and so on; “emotion” was defined as a positive opinion and a negative opinion by Das and Chen in 2001 [9]. In [10], Liu analyzed users’ views, attitudes, and feelings in the topics and defined emotional analysis (opinion

mining) as subjective information in 2012. Text analysis and mining are closely related to a multilingual environment, and multilingual corpus construction is an important task for opinion mining in multilingual context [11].

Aoki and Uchida argued that the estimated emotions expressed by the emoticons were important for the reputation analysis and a method of automatically creating the emotion vectors of emoticons was presented, which used the relationship between emotional words and emoticons from many blog posts [12]. A lot of research works study the feasibility of using deep learning methods in emotional analysis according to the relationship between adjacent words in a sentence. Gao et al. proposed a new model of emotional polarity conversion model to strengthen textual association [13]. Zhu et al. used HowNet on the Chinese word semantics to carry out the emotional tendencies calculation [14]. And Decheng had carried on the emotion analysis of Chinese sentence semantics by using syntactic structure and dependency relationship [15].

The above studies were used for English or Chinese emotion analysis, and there was no effective way to analyze the emotions of Tibetan microblogs.

**2.2. Sequential Pattern Mining.** Sequential pattern mining model [16] was also used for weighting data, where the pattern was considered as a feature and the support of the pattern was considered as the weight. Sequential pattern mining with broad applications, including analysis of customer shopping patterns, web access patterns, and DNA sequences, is a well-studied topic in data mining. It was first introduced by Agrawal and Srikant in [17]. Given a set of sequences, they want to find all frequent subsequences whose occurrence frequency in the set of sequences is no less than a user-specified `min_support`. Many efficient algorithms were proposed to mine sequential patterns. A priori-based algorithms, such as GSP [18], tried to reduce search space based on the apriori property of sequential property. And pattern growth-based algorithms, like FreeSpan and PrefixSpan [19], mined sequential patterns by projecting the sequence database based on subsequences and then growing subsequences. However, because of the difficulty to select an appropriate threshold in practice, the task to mine top- $k$  sequential patterns was proposed. Tzvetkov et al. proposed an algorithm called TSP [20], which aimed to mine top- $k$  frequent closed sequential patterns of length no less than a given minimum length. They thought closed sequential patterns were compact representations of frequent patterns. To reduce time and space complexity, many fast methods were proposed. Temporal relations in text are sequential relations to formulate an event network for information analysis including opinion mining [21].

### 3. Emotion Analysis Based on Sequential Model

In this section, we first introduce our sequential model and then use the model to classify Tibetan text.

**3.1. Sequential Model.** Let  $I = \{i_1, i_2, \dots, i_n\}$  be a set of items. An item set  $X$  is a subset of  $I$ : that is,  $X \subseteq I$ . A sequence  $s$  is an

TABLE 1: A toy database.

SID	Sequence
1	$\langle\langle CD \rangle\rangle\langle\langle ABC \rangle\rangle\langle\langle ABF \rangle\rangle\langle\langle ACDF \rangle\rangle$
2	$\langle\langle ABF \rangle\rangle\langle\langle E \rangle\rangle$
3	$\langle\langle ABF \rangle\rangle$
4	$\langle\langle DHG \rangle\rangle\langle\langle BF \rangle\rangle\langle\langle AGH \rangle\rangle$
Length	Pattern
1	$\langle\langle A \rangle\rangle, \langle\langle B \rangle\rangle, \langle\langle D \rangle\rangle, \langle\langle F \rangle\rangle$
2	$\langle\langle AB \rangle\rangle, \langle\langle AF \rangle\rangle, \langle\langle B(A) \rangle\rangle, \langle\langle BF \rangle\rangle, \langle\langle D(A) \rangle\rangle, \langle\langle D(B) \rangle\rangle, \langle\langle D(F) \rangle\rangle, \langle\langle F(A) \rangle\rangle$
3	$\langle\langle ABF \rangle\rangle, \langle\langle AB(A) \rangle\rangle, \langle\langle D(BF) \rangle\rangle, \langle\langle D(B)(A) \rangle\rangle, \langle\langle D(F)(A) \rangle\rangle$
4	$\langle\langle D(BF)(A) \rangle\rangle$

ordered list of item sets, denoted by  $s = \langle s_1 s_2 \dots s_l \rangle$ , where  $s_j$  is an item set. A sequence with length  $l$  is called an  $l$ -sequence. A sequence  $\alpha = \langle a_1 a_2 \dots a_n \rangle$  is called a subsequence of another sequence  $\beta = \langle b_1 b_2 \dots b_m \rangle$  and  $\beta$  a supersequence of  $\alpha$ , denoted as  $\alpha \subseteq \beta$ , if there exist integers  $1 \leq j_1 < j_2 < \dots < j_n \leq m$  such that  $a_1 \subseteq b_{j_1}, a_2 \subseteq b_{j_2}, \dots, a_n \subseteq b_{j_n}$ .

A sequence database  $S$  is a set of tuples  $\langle \text{sid}, s \rangle$ , where  $\text{sid}$  is a `sequence_id` and  $s$  a sequence. A tuple  $\langle \text{sid}, s \rangle$  is said to contain a sequence  $s$ , if  $\alpha$  is a subsequence of  $s$ . The support of a sequence  $\alpha$  in a sequence database  $S$  is the number of tuples in the database containing  $\alpha$ : that is,  $\text{sup}_S(\alpha) = |\{ \langle \text{sid}, s \rangle \mid \langle \text{sid}, s \rangle \in S \cap (\alpha \subseteq s) \}|$ .

The frequent patterns are often used to generate association rules. Consider the rule  $X \Rightarrow Y$ , where  $X$  and  $Y$  are sets of items. The confidence of the rule  $X \Rightarrow Y$  is equal to the ratio of the support of  $X \cup Y$  to the support of  $X$ .

We then give an example to illustrate the above concepts, as shown in Table 1. We denote the marked sequence dataset  $S$  and the minimum support `min_sup` = 2. The dataset contains 4 sequences and 8 items. The aim is to find all of the frequent  $k$ -sequences with the support greater than or equal to 50%. We find that if  $\langle BF \rangle$  exists in a sequence, there is a 75% probability that  $\langle A \rangle$  appears with  $\langle BF \rangle$ . In other words, the rule  $\langle BF \rangle \Rightarrow \langle ABF \rangle$  has a confidence of 75%. Another example is that the rule  $\langle D(BF) \rangle \Rightarrow \langle D(BF)A \rangle$  has the confidence of 100%.

**3.2. Sequence Classification of Tibetan Text.** This subsection introduces the classification processing of Tibetan microblog text. In the sequence, each sentence is considered as an item set (here, the single hammer symbol is considered as the sentence division), where conjunctions are also separately expressed as items. Then the text is denoted as a sequence, which contains multiple items. We give an example in Table 2.

The text contains three sentences. Taking into account the negative word (“མེ”) and the turning conjunctive (“འོན་ཀྱང”), we can obtain the sentence sequences which are shown as follows:

$\langle\langle \text{none} \rangle\rangle\langle\langle \text{neg, like} \rangle\rangle\langle\langle \text{but} \rangle\rangle\langle\langle \text{happiness} \rangle\rangle$

TABLE 2: Tibetan text.

Text
དེ་མིང་རྒྱུགས་ལེན་གྱི་མཚུགས་འབྲས་ཐོན་པ་དང་། (the test results came out today)
གྲུབ་འབྲས་མི་བཟང་། (the result is not high)
འོན་ཀྱང་ལ་ཡར་ཐོན་ཆེན་པོ། (but the progress is great)

Conjunctions can be used to reflect the interrelationship between words (such as the undertaking relationship, the turning relationship, and the causal relationship). And the relationship between the words has great influence on the emotions of the whole text. Thus, it is important to add a conjunction as a sequence feature when the sentence is serialized.

The steps to build a sequence database in Tibetan text are shown as follows:

- (1) Each sentence of the Tibetan text is used to determine the sentence's emotions based on dictionary method.
- (2) We combine the emotional tags of each sentence with the conjunctions of sentences head to convert the text into a sequence.
- (3) In the training set, we express the emotional tags of each microblog pointing to a class. For example, the emotional tag of the text in the above example is "happy," and the following input sequence can be obtained as follows:

((none)(neg, like)(but)(happiness), happiness)

Based on the above method, we can construct a sequence database. By serializing the Tibetan sentences in the training set, we can use the sequential pattern mining algorithm to find the frequent sequence rules which satisfy the minimum support or the minimum confidence. These sequential rules represent the specific patterns of different emotional categories.

#### 4. The Emotional Classification Based on Semantic Features of Tibetan

In this section, we select the features used to classify Tibetan microblogs and analyze the reasons why we select them.

*4.1. Semantic Features Selection.* Emotional classification associates the given text with one or more emotional categories according to the emotional characteristics of the text. The key is to select emotional characteristics properly. The effective features are able to significantly improve the performance of the classifiers. We still use the classical vector space model to express microblogs. Although the traditional feature extraction algorithms can achieve a good effect in the subject text classification, due to the neglect of the semantic features, such algorithms have poor effect to apply on the emotional classification directly. Based on the characteristics of microblogs, this paper explores the extraction of emotional features of microblogs from different angles. Microblogs'

emotions hide a lot of category features, which can be extracted from the grammar structure and the semantic level. To this end, this paper presents the following four types of semantic features as the candidate feature vectors that constitute microblogs' emotional vector space model.

(1) *The Characteristics of Emotional Words.* Emotional words contain emotional colors and express people's inner feelings, such as pleasure, anger, and disgust. This kind of words has an important reference value. In a real microblog language environment, network vocabularies, phrases, short sentences, and emoticons have emotional tendencies. In this paper, emotional words are extracted based on the emotional dictionary of Tibetan microblogs.

(2) *Affective Factors.* The so-called affective factors are negative words, degree words, and related words. The appearance of such words often affects the emotional changes or emotional strength in a sentence: for example, “དེ་མ་ཐག་ཏུ་མི་ཡོང་མོད། འོན་ཀྱང་བརྗོད་པ་བསྐྱེད་ན་རིམ་གྱིས་འགྲུབ་ཡོང་།” (although we cannot succeed immediately, but if you work hard, you will eventually complete) contains the emotional impact factors “མི (not),” “འོན (but),” and so on. These emotional words can determine the user's emotional trends.

(3) *Statistics of Words.* According to the statistics and research, we found that a single part of speech or several consecutive parts of the combination contain subjective information and objective information. The N-POS (Part-of-Speech) model is a corpus-based statistical natural language model. When  $N$  is 3, the three consecutive words are combined into a pattern. This paper treats the sequences of three consecutive parts as emotional characteristics.

Examples are as follows:  
 མོགས་བཟང་ལོ་ལྷོད་དང་འཕམ་པས་དགའ་བོ་བྱུང་། (my best friend, nice to meet you).

Parts of speech are as follows: མོགས་/ng: noun, བཟང་ལོ་/a: adjective, ལྷོད་/rh: pronoun, དང་/ks: lattice, འཕམ་/vt: verb, པས་/h: name markup, ལོ་/ki: lattice, དགའ་བོ་/ng: noun, བྱུང་/vi verb, །/xp: symbol.

The three-POS features of the sentence are as follows: nouns, adjectives, and pronouns; adjectives, pronouns, and lattice; pronouns, lattice, and verb; lattice, verb, and name markup; verb, name markup, and lattice; name markup, lattice, and noun; lattice, noun, and verb; noun, verb, and symbol.

(4) *The Characteristics of Semantic Dependency.* We use the semantic relation between words and words to reveal the syntactic structure of sentences. The semantic dependency is the main element of the syntactic structure of the grammatical grammar. It refers to the binary relation of the word pair in the sentence, one of which is called the central word and the other is called the subsidiary word. The dependency expresses a semantic dependency between the central word and the subsidiary word. By exploring the interdependence between the central words in the sentence and the subsidiary



TABLE 3: Tagging scheme of Tibetan syntactic analyzer.

Name	Symbol
Normal name	NN
Buddhist Sanskrit	NF
Neutral adjective	AX
Object preposition	P2
Prepositional phrases	PP
Implement prepositions	P3
Positive adjective	AZ
Judgment verb	VJ
Verb phrases	VP
Name	NR
Negative adjective	AF
Verb of existence	VE
Noun phrases	NP
Adverb	AD
Interjection	IJ
Personal pronouns	RR
Sentence	IP
—	—

words of the central word, we can obtain effective emotional characteristics and help the emotional classification achieve better results.

The analysis of this paper adopts the Tibetan syntactic parser developed by Long Congjun and other people who come from the Chinese Academy of Social Sciences, which is a highly optimized probability of context-free grammar and lexical dependency analyzer. The parser constructs the syntactic tree library in square brackets (“and”) and uses the detailed label scheme of Table 3 to label it. With the help of the binary dependencies, we take the bottom-up method to build the binary tree of the sentences.

Example sentences are as follows:  
 དེ་རིང་ནི་ཨ་པའི་རྒྱལ་ཆེན་ལ་བཀའ་ཤིས་བདེ་ལེགས་ལྷུ། (today is Father’s Day. Good luck!)

The syntactic parsing result of this sentence is as follows:  
 [[[IP [NPU [NP [ng དེ་རིང་ (Today)]] [U [up ཅི་ (particle of pause)]]] [VP [KP [NP [KP [NP [ng ཨ་པ་ (father)]] [K [kg འི་ (genitive)]]] [N [ng རྒྱལ་ཆེན་ (holiday)]]] [K [kx ལ་ (allative)]]] [VP [NP [ng བཀའ་ཤིས་བདེ་ལེགས་ (good luck)]] [V [vt ལྷུ།]]]]] [PU [xp |]]].

We use the syntactic tree, which was first introduced by Lucien Tesniere [22], to denote the sentence structure dependencies. The generated syntactic tree is shown in Figure 2.

The corresponding sentence structure dependencies are as follows: NPU (NP, U), KP (NP, K), NP (KP, N), KP (NP, K), VP (NP, V), VP (KP, VP), IP (NPU, VP), and ROOT (IP, PU).

From Figure 2, we can find that the subject of the microblog is རྒྱལ་ཆེན་ (festivals), and the predicate is བཀའ་ཤིས་བདེ་ལེགས་ལྷུ།. The microblog grammar structure is obvious.

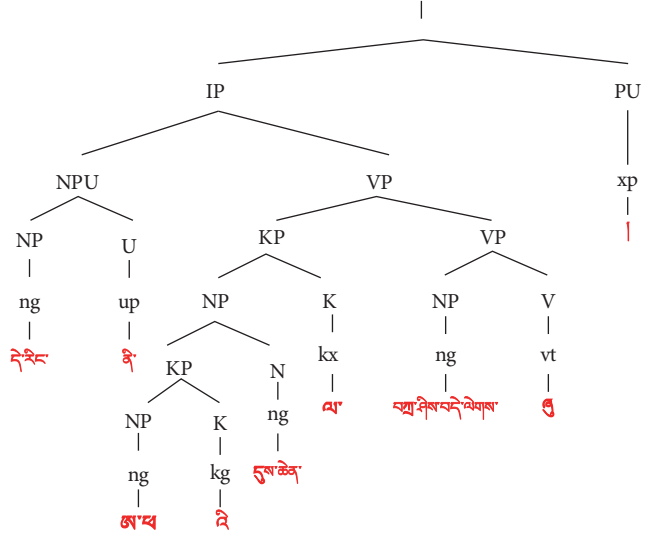


FIGURE 2: Syntactic tree structure of Tibetan microblogging.

4.2. *Emotional Vector Space Model of Microblogs.* In this paper, we use the vector space model to express the text of microblogging short text. The set of microblogs is  $D = \{d_1, d_2, \dots, d_n\}$ . We can use feature set  $T = \{t_1, t_2, \dots, t_m\}$  to vectorize the text space, for example, the vectorization of text  $d_i$  is  $d_i = \{(t_1, w_{i1}), (t_2, w_{i2}), \dots, (t_m, w_{im})\}$ , where  $t_i$  ( $1 \leq i, j \leq m$ ) is the characteristic, and  $w_{ij}$  is the characteristic weight corresponding to the characteristic term  $t_i$ . In order to solve the data sparse problem of microblogs, we select sequential rules, semantic features, and facial features.

The semantic features include emotional word features, affective factors, part-of-speech features, and syntactic dependent features. The emotional dimension of emotional words is fixed by 7 and eigenvalues are the number of emotional words in each category. Affective factors mainly have conjunctions, negative words, and the adverbs of degree adverbs, where the conjunctions and the negative words use the word frequencies as the characteristic weights; the degree adverbs use the weight coefficients in the table as the weights. There are more lexical features and more syntactic dependent features. We use the  $\chi^2$  statistical method to select the effective feature items and use TF-IDF values as weights.

An emotional symbol is the unique emotional feature of a microblog, which can be classified by “good,” “happy,” “sad,” “hate,” “scared,” “angry,” and “fear.” These seven kinds of emoticons are selected as the feature items, and the feature weights are expressed by the conditional probabilities of the expression symbols in the corresponding categories.

## 5. Experiment and Analysis

5.1. *Experimental Data.* The experimental data of this paper came from Sina Weibo. The main reason to select this data set is that Sina Weibo is the most widely used platform and can provide abundant training corpus. It is easy to build sequential features since the microblogs are temporal and

TABLE 4: Details of the experimental data.

Category	Good	Happy	Sad	Hate	Scared	Angry	Fear
Quantity	9600	4800	2400	1200	600	300	300

follow the power law distribution. We collect microblogs with the following steps.

First, we built a microblog crawl seed set (mainly Tibetan microblogging users) manually and selected 126 users who had high frequencies of Tibetan microblogging as the initial seeds.

Second, we randomly selected user seeds for crawling and, at the same time, we filtered crawled text. If we planned to save the Tibetan microblogs, we would obtain the related comments and the other users who had participated in this microblogging activity. If these users had not been visited, they would be added to the seed collection for the following visit.

Finally, we stored the crawl Tibetan microblogs and related comments according to the microblogs' IDs.

We fixed the time interval by February to July in 2016. At last, we obtained 300,000 pieces of Tibetan microblogs. After preparing Tibetan microblog text and labeling the data by hand, we used xml tag library to store the microblogging information. The storage fields included the id, the user name, the original text, the Tibetan text, the theme, Tibetan word tags, Tibetan syntactic tree, and emotional logos, as shown in Figure 3.

By labeling and filtering the crawled microblogs, we selected 19200 pieces of microblog text. The details of each emotional category are shown in Table 4. 50% of the corpus are used to train the model and 50% are used for the classification test.

**5.2. Evaluation Indicators.** Accuracy rate, recall rate, and  $F$  value are generally used as the evaluation indicators. The accuracy refers to the ratio of microblog text and the accuracy is consistent with the results of manual labeling for the classification. The recall refers to the proportion of the correctly predicted samples represented in all samples belonging to the class. In order to deal with the multicategory problem, macroaverage and microaverage are used as evaluation criteria. The accuracy, recall rate, and  $F$  value of macroaverage and microaverage are calculated as follows:

Macro Accuracy Rate

$$= \frac{1}{7} \sum_i \frac{|\text{correctly classified microblogs in category } i|}{|\text{microblogs in category } i|};$$

Macro Recall Rate

$$= \frac{1}{7} \sum_i \frac{|\text{correctly classified microblogs in category } i|}{|\text{microblogs in manual labeling category } i|};$$

Macro  $F$  Rate

$$= \frac{2 \times \text{Macro Accuracy Rate} \times \text{Macro Recall Rate}}{\text{Macro Accuracy Rate} + \text{Macro Recall Rate}};$$

Micro Accuracy Rate

$$= \frac{\sum_i |\text{correctly classified microblogs in category } i|}{\sum_i |\text{microblogs in category } i|};$$

Micro Recall Rate

$$= \frac{\sum_i |\text{correctly classified microblogs in category } i|}{\sum_i |\text{microblogs in manual labeling category } i|};$$

Micro  $F$  Rate

$$= \frac{2 \times \text{Micro Accuracy Rate} \times \text{Micro Recall Rate}}{\text{Micro Accuracy Rate} + \text{Micro Recall Rate}}.$$

(1)

With the sequential rule mining method introduced in Section 3.1, we use the training set to construct the sequential rules. Each training corpus is represented by a vector. The sequential pattern  $X$  in each rule  $\{X \rightarrow y\}$  is used as a feature. If the corresponding sequence of a microblog contains  $X$ , its corresponding feature value is set by 1, otherwise set by 0.

Figure 4 shows how the support threshold affects the emotional classification. With the increasing of the support, the feature dimension begins to decrease. When the support degree increases to 0.035, the effect achieves the best, indicating that the sequence feature can be more effective for emotional recognition. However, the reduction of the feature dimension leads to the relative deterioration of the classification effect.

Figure 5 shows the performance changes for how the minimum confidence threshold affects the effect of emotion classification. In this figure, the minimum confidence varies from 0.005 to 0.5, and we can see that the micro  $F$  value is stable over a large range, and the  $F$  value decreased significantly with the increasing of minimum confidence value. This shows that the influence of the confidence on this method is not particularly large.

## 6. Conclusion

This paper analyzes the emotions of Tibetan microblogs, which is helpful to understand opinions and product reviews for Tibetan customers. First, we use sequential rules to classify Tibetan microblogs. To improve the effectiveness, we add semantic features, which include the characteristics of emotional words, affective factors, statistics of words, and semantic dependency. We also propose to analyze Tibetan microblogs by building syntactic tree and define emotion calculation processing by this tree structure. The experimental results confirm the effectiveness of feature selection and the efficiency of multiclassification. As future work, we will

```

<?xml version="1.0" encoding="UTF-8"?>
<elements>
  <element wid="M_8UyioB">.....</element>
  <element wid="M_A00QCaplp">.....</element>
  <element wid="M_A00TH31xe">.....</element>
  <element wid="M_A00IJ38II">.....</element>
  <element wid="M_A01Tn0QXX">.....</element>
  <element wid="M_A01wp3M0n">.....</element>
  <element wid="M_A02d7eT0A">.....</element>
  <element wid="M_A02kagdKU">.....</element>
  <element wid="M_A07P33kDn">.....</element>
  <user>Tibetvictor</user>
  <text>ངའི་ཉེ་མོན་ལྟགས་ཅན་གྱི་ལུ་བ་གཟུགས་བརྒྱན་མགྲིན་ཚབ་བྱས་པ། ལྟུབ་བསྐྱོར་གནང་རོགས། [smile]</text>
  <fine>ངའི་ཉེ་མོན་ལྟགས་ཅན་གྱི་ལུ་བ་གཟུགས་བརྒྱན་མགྲིན་ཚབ་བྱས་པ། ལྟུབ་བསྐྱོར་གནང་རོགས། </fine>
  <seg>ང/rh འི/kg ཉ/n རེ/n མོན/v ་/n ལྟགས/v/n ཅན/vf གྱི/kg ལུ/v བ/nf གཟུགས/v/n བརྒྱན/v མགྲིན/v ཚབ/v/n བྱས/v ལ/h |/xp ་/n ལྟུབ/v བསྐྱོར/v གནང/v རོགས/v |/
xp</seg>
  <topic></topic>
  <emo>[smile]&amp;</emo>
  <syn>[ [ [IP [IP [VP [NP [KP [NP [NP [KP [NP [rh འ:]] [K [kg འི:]]] [N [ng ཉེ་མོན:]]] [N [ng ལྟགས་ཅན:]]] [K [kg གྱི:]]] [NP [N [ng ལུ་བ:]]] [N [ng གཟུགས་བརྒྱན:]]] ] [VP [D [d མགྲིན་ཚབ:]] [V [vt བྱས:]]] [I [T [h བ:]]] ] [PU [xp |]]] [IP [VP [D [a ལྟུབ:]]] [VP [NP [ng བསྐྱོར་གནང:]]] [V [vt རོགས:]]] ] [PU [xp |]]] ]</syn>
  <type>like</type>
</element>
<element wid="M_A07bSnB1G">.....</element>
<element wid="M_A07dXeqai">.....</element>
<element wid="M_A07fYdaEg">.....</element>

```

FIGURE 3: Tibetan microblog corpus.

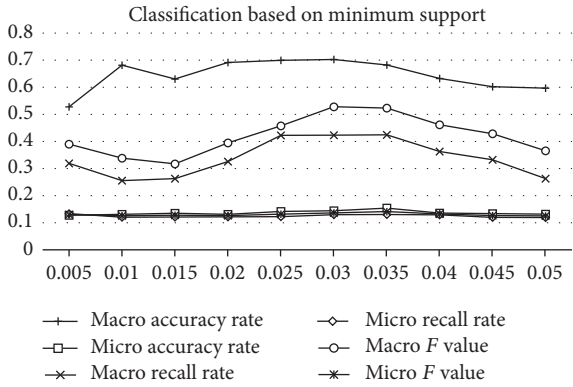


FIGURE 4: Experimental comparison of minimum support thresholds.

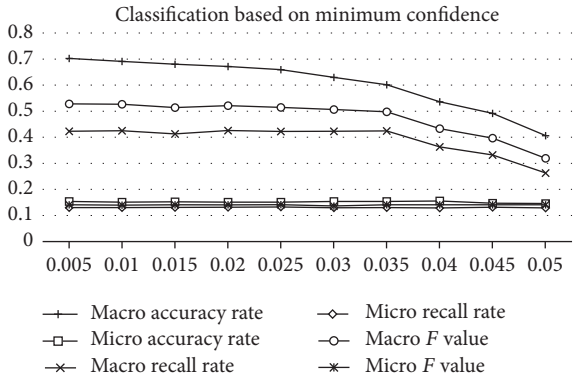


FIGURE 5: Comparison of minimum confidence thresholds.

analyze emotion of Tibetan microblogs without emotional words.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper or the funding.

## Acknowledgments

This work was supported by the National Nature Science Foundation of China (no. 61672553) and the (Ministry of Education in China) Project of Humanities and Social Sciences (Project no. 16YJCZH076).

## References

- [1] W. Ling, L. s. Marujo, C. Dyer, A. W. Black, and I. Trancoso, "Mining parallel corpora from Sina Weibo and Twitter," *Computational Linguistics*, vol. 42, no. 2, pp. 307–343, 2016.
- [2] 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, 2016.
- [3] Y.-Y. Zhao, B. Qin, and T. Liu, "Sentiment analysis," *Journal of Software*, vol. 21, no. 8, pp. 1834–1848, 2010.
- [4] B. Pang, L. Lee, and S. Vaithyanathan, "Sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing—Volume 10 (EMNLP '02)*, pp. 79–86, Association for Computational Linguistics, Stroudsburg, Pa, USA, July 2002.
- [5] P. D. Turney and M. L. Littman, "Measuring praise and criticism: inference of semantic orientation from association," *ACM Transactions on Information Systems*, vol. 21, no. 4, pp. 315–346, 2003.
- [6] K. Lun-Wei, W. Tung-Ho, L. Li-Ying et al., "construction of an evaluation corpus for opinion extraction," in *Proceedings of the NTCIR-5*, pp. 513–520, Japan, 2005.
- [7] J. Zhang, Y. Sun, H. Wang, and Y. He, "Calculating statistical similarity between sentences," *Journal of Convergence Information Technology*, vol. 6, no. 2, pp. 22–34, 2011.
- [8] H. Liu D, H. Nuo M, and L. Ma, "Mining Tibetan web text resources and its application," *Journal of Chinese Information Processing*, vol. 29, no. 1, pp. 170–177, 2015.
- [9] S. Das and M. Chen, "For Amazon: Extracting market sentiment from stock message boards," in *Proceedings of the Asia Pacific finance association annual conference (APFA)*, pp. 35–43, 2001.
- [10] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1–167, 2012.
- [11] J. Zhang, Y. Sun, and A. J. Jara, "Towards semantically linked multilingual corpus," *International Journal of Information Management*, vol. 35, no. 3, article 1387, pp. 387–395, 2015.
- [12] S. Aoki and O. Uchida, "A Method for Automatically Generating the Emotional Vectors of Emoticons using Blog Articles," *International Journal of Computers*, vol. 5, pp. 346–353, 2011.
- [13] K. Gao, Y. S. Li, and Y. D. Ruan, "A Micro-blog sentiment analysis approach," *Journal of Chinese Information Processing*, vol. 29, no. 4, pp. 40–49, 2015.
- [14] Y. Zhu L, J. Min, and Y. Zhou, "Semantic orientation computing based on HowNet," *Journal of Chinese Information Processing*, vol. 20, no. 1, pp. 14–20, 2006.
- [15] L. Decheng, "Tianfang Yao. Semantic polarity analysis and opinion mining on Chinese review sentence," *Computer application*, vol. 11, 2006.
- [16] . Jian Pei, . Jiawei Han, B. Mortazavi-Asl et al., "PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth," in *Proceedings of the 17th IEEE International Conference on Data Engineering*, pp. 215–224, IEEE Computer Society, Los Alamitos, Calif, USA, 2001.
- [17] R. Agrawal and R. Srikant, "Mining sequential patterns," in *Proceedings of the IEEE 11th International Conference on Data Engineering*, pp. 3–14, San Jose, Calif, USA, March 1995.
- [18] R. Srikant and R. Agrawal, "Mining sequential patterns: Generalizations and performance improvements," in *Proceedings of the 5th International Conference on Extending Database Technology*, pp. 3–17, 1996.
- [19] J. Han, J. Pei, B. Mortazavi-Asl, Q. Chen, U. Dayal, and M.-C. Hsu, "FreeSpan: Frequent pattern-projected sequential pattern mining," in *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery in Databases*, pp. 355–359, Boston, MA, USA, August 2000.
- [20] P. Tzvetkov, X. Yan, and J. Han, "TSP: mining top-k closed sequential patterns," *Knowledge and Information Systems*, vol. 7, no. 4, pp. 438–457, 2005.
- [21] J. Zhang, C. Yao, Y. Sun, and Z. Fang, "Building text-based temporally linked event network for scientific big data analytics," *Ubiquitous Computing*, vol. 20, no. 5, p. 13, 2016.
- [22] B. Plank and A. Moschitti, "Embedding semantic similarity in tree kernels for domain adaptation of relation extraction," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, ACL 2013*, pp. 1498–1507, Sofia, Bulgaria, August 2013.





# Hindawi

Submit your manuscripts at  
<https://www.hindawi.com>

