Research on the Majority Decision Algorithm Based on WeChat Sentiment Classification

Shengtai Zhang$^{1,a}$, Feifei Wang$^{1,b}$, Juliang Zhang$^{1,c}$, Fan Duo$^{2,d}$

$^1$School of economics and management, Beijing University of Posts and Telecommunications, Beijing 100876, China; $^2$School of information and communication engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China.

Abstract: Sentiment analysis mainly studies the emotional tendencies of texts from grammar, semantic rules and other aspects. The texts from social network are characterized by less words, irregular grammar, data noise and so on, which have increased the difficulty of emotion analysis. In order to improve the performance of machine learning in sentiment analysis, this study proposed the Majority Decision Algorithm to classify the emotional tendentious of the text in WeChat, combined the characteristics of five classifiers and integrated the classification results of five classifiers, eventually the text can be classified in WeChat. Firstly, this study utilized the BlueStacks to crawl the cache of WeChat Moment developed by Tencent company. Secondly, the cache was processed by Python to get the WeChat dataset. After the Chinese word segmentation, data cleaning and segmentation, the sentiment classification experiment were carried out using different classifiers. Finally, a Majority Decision Algorithm composed of five classifiers was established. It included, Naive Bayes (sklearn), Naive Bayes (SnowNLP), SVM (linear), SVM (RBF) and SGD. Then, the comparison was carried out between the performance of the algorithm and the five classifiers. Results show that the precision rates of the five classifiers are 0.8598, 0.8154, 0.8511, 0.8739 and 0.8678; the recall rates are 0.8544, 0.8482, 0.9380, 0.9226 and 0.9349; F1 scores are 0.8571, 0.8315, 0.8924, 0.8975 and 0.9001, respectively. The algorithm of the Precision rate, Recall rate and F1 score were 0.8804, 0.9349 and 0.9069, respectively, indicating that algorithm in current study significantly improved the performance, which can be effectively applied into the new text form of WeChat Moment. The conclusions of the study can provide theoretical reference for sentiment classification of Chinese text based on machine learning.

Keywords: WeChat; Sentiment Classification; Majority Decision Algorithm (MDA)

1. Introduction

With the advent of mobile Internet era, a variety of approaches are employed to obtain information. The applications of smart phone have brought convenience as well as mass of information to modern generation. By the end of 2016, the number of users on mobile social website has reached to 695 million(CNNIC, 2016). The most popular social media in China (i.e. WeChat) was Founded in 2011, WeChat has quickly accumulated users relying on QQ users and mobile directory, and become the China’s mainstream communication tool, well-known for the highly convenience, interactivity and innovativeness. As one of the new media platforms, WeChat has a set of social, sharing and recreational functions. It has brought revolutionary changes to people’s life, in which people can use to record their moments of life, interact with friends, express their views and make comments on hot social events. At present, the main research of
sentiment analysis is based on Twitter (Tellez et al., 2017), Facebook (Srividya et al., 2017), microblog (Ficamos et al., 2017) and other social networks. As the mainstream online social networking tool, WeChat expresses all the feelings and emotions of users. However, the classification of these feelings has some difficulties. For example, some users put a “smile” expression in the latter WeChat text while showing negative emotion in former sentence. Although the “smile” expression is a commendatory term, it shows different emotions in different contexts. Therefore, it is necessary to deeply analyze the emotion in WeChat text through algorithm to enhance the performance of text sentiment classification.

Sentiment classification is a special method of text classification. As the number of subjective texts increasing rapidly in social media, a new research field, sentiment analysis, appeared in the beginning of year 2000. Nowadays, the research of sentiment analysis mainly focuses on the following two types of methods: one is the Lexical-based methods (A kind of sentiment analysis that is composed of some words and phrases which are able to express positive or negative emotions). The emotional lexicons-based approach can preliminarily determine the emotional tendencies of the text (Xianghua Fu, 2017). However, it can’t be applied into all cases, because emotional lexicons cannot contain all evaluated words, and some emotional words have different emotional polarity in different contexts. Since the emotions of these words in different situations may be the opposite, thus it is not enough to solely use emotional lexicons for sentiment classification.

The other method is based on machine learning (also known as corpus-based method). The text is evaluated as words, and convert into several features. Then use the machine learning techniques, such as Naïve Bayes, Support Vector Machines, Maximum Entropy are applied to predict sentiment tendency. Ye Q (2009) used supervised machine learning algorithm to carry on sentiment classification of the comment in tourism blog. The experimental results illustrated that the SVM and N-gram algorithm were superior to Naive Bayes, but the drawback was that he just made a comparison of several algorithms by the performance, there was no proposed solution for improving the performance. Catal C (2016) et al. developed a sentiment classification model that used the voting algorithms to improve the performance of machine learning. But this combination research was limited to emotion analysis of microblogging in Turkish. Therefore, on the basis of Chinese text in WeChat Moment, we established an algorithm to optimize the performance of five classifiers after the process of Chinese word segmentation, data cleaning and data segmentation, in order to improve the performance of machine learning in Chinese emotion classification.

2. Related work

The text sentiment analysis, also known as opinion mining, simply, is defined as the process of analysis, process, induction and reasoning for subjectivity with emotion. The initial sentiment analysis stems from the predecessors' analysis of words with emotion, for example, "excellent" is a word with a positive attitude, and "poor" is a word with a negative attitude (Zhao Yanyan et.al, 2010). At present, while using machine learning for sentiment analysis, we mainly Support Vector Machine, Naive Bayes, Maximum Entropy, KNN and other algorithms for sentiment classification after marking the training corpus and test corpus (Anjaria et. al, 2014). The WeChat Moment is characterized by large amount of information, fast updating and rich contents. However, there is little research of WeChat because it has semi-closed characteristics, and without the open API interface, besides, it cannot be displayed on the webpage. Thus, the data acquisition on WeChat is
a new challenge.

At present, the research of machine learning in sentiment analysis is mainly concerned with the selection of the best algorithm and the improvement of performance. This kind of promotion is divided into two kinds: one is to improve the performance of single classifiers by optimizing the algorithm. The other one is using multi-classifier method to combine the advantages of each algorithm to improve the performance.

The method of using algorithm optimization to improve the performance of the classifier is mainly reflected by the comparison with the classification results of several algorithms. For example, L Jiang et.al (2016), used the deep feature weighting (DFW) to optimize the performance of Naive Bayes, and estimated the conditional probability of Naive Bayes by deeply calculating the characteristic weighting frequency from the training data, so as to improve the classification performance. Mullen et.al (2004), used hybrid method to optimize the Support Vector Machine and analyzed the syntactic relations and characteristics of the text in order to realize the emotional division and the results showed a better classification performance. Wijksgatan O (2013) used Stochastic Gradient Descent and text feature methods to analyze emotions in short texts. Based on the above research, we found that the classification performance of Naive Bayes and Support Vector Machine is better, while SGD can do a quick classification, and it can not only predict the sample category, but also calculate the probability of classification. However, there is a limit to the optimization of algorithm to optimize the performance of classifiers, so we consider the use of multi-classifiers method for emotion classification.

The study of multi-classifiers is based on the combining use of several classifiers of the same algorithm or several classifiers of different algorithms. For example, ChalothomT et.al (2015) combined four kinds of classifiers that including the Support Vector Machine, Naive Bayes, SentiStrength and Stacking on the emotion analysis of Twitter, and achieved good classification results. Fersini E (2014) proposed an ensemble method that was based on the multi-Bayesian classifier, which considered the uncertainty and reliability of each individual classifier. Used multi-classifiers model to make sentiment classification for social media Twitter. Ouyang Chunping et.al (2014) proved that the integration method based on NB, SVM, KNN was superior to a single algorithm, and verified that the "NB + SVM" method was superior to "NB + KNN" in the multi-strategy integration method. S Li (2007) proposed a multiple classifiers method to judge the emotional polarity of the subjectivity texts by using several combining rules to design the multi-classifiers model. The experimental results indicated that the performance of each combination method was better than any individual classifier, and the performance of simple addition rule was the best, which was 2.56% higher than the best single classifier. Based on the above research, the use of classifier combination method can effectively improve the classifiers' performance in text classification. Therefore, we established the Majority Decision Algorithm to carry out emotion classification of WeChat dataset, which combined the advantages of five classifiers, and the results show that it could improve the performance of individual classifiers on sentiment classification of WeChat datasets.

The remaining parts are organized as follows. Section 3 introduces the methodology, followed by section 4 which presents the experiment results, and section 5 is the conclusions and future work.

3.Methodology
In the experiment, we used two machine learning libraries, sklearn(scikit-learn) and SnowNLP, to analyze the sentiment of texts. At the same time, we chose three algorithms based on the domestic and foreign research, then considered the characteristics of Chinese text. It includes: Naive Bayes(NB), the performance is excellent in text classification; Support Vector Machine(SVM), has advantages over small samples; and Stochastic Gradient Descent(SGD), which is based on logistic regression analysis, and the system requires little overhead while using SGD, it is especially advantageous in the case of massive text.

3.1 Feature Conversion

After the Chinese word segmentation, a sentence will be divided into several separate phrases by spaces. We used weight calculation method, the TF-IDF, to transform texts into eigenvectors. In the experiment, we compared three methods of calculating weights, presence, TF and TF-IDF. Results show that the performance of TF-IDF combined with classifiers is the best. TF-IDF (term frequency–inverse document frequency) is a commonly used weighting technique for information retrieval and information discovery. It is used to evaluate the importance of a word to a file set or a file in a corpus.

In this study, we use TF-IDF to calculate the feature’s weight, and the calculating formula is:

$$Tf-idf(i, j) = \frac{n_{i,j}}{\sum n_{j} \times \log \frac{|D|}{1+|[j : t_i \in d_j]|}}$$

(1)

$n_{i,j}$ denotes the j-th characteristic emotion word in the i-th document, $|D|$ denotes the number of documents.

The importance of the word is proportionally increased to the number of times it appears in the document, but it is proportionally decreased to the frequency it appears in the corpus. The forms of TF-IDF are often used by search engines, as a measure or rating of the correlation between a file and a user's query. In the experiment, we used the CountVectorizer method and the TF-IDF transformer method in the sklearn library to generate the eigenvectors for each sentence.

3.2 Naive Bayes Algorithm

In the experiment, we used Bayesian classifiers which are in the sklearn and SnowNLP of Python machine learning library to analyze the sentiment of Chinese text. Naive Bayes is a supervised learning algorithm, which is used to solve many machine learning problems. Compared with other complex machine learning algorithms, Naive Bayes is a simple algorithm based on probability theory. It predicts the individuals with these characteristics by using the prior probabilities of each category and the conditional probability of occurrence of specific characteristics in each category. The Naive Bayes hypothesis is quoted when calculating the conditional probability, that is, the probability of occurrence of each feature is independent. The reason why we introduce the Naive Bayes assumption is to avoid too many combinations of features, which may lead to the sparse of datasets. The formula of Naive Bayes is:

$$P(c_j | d) = \frac{P(c_j)P(d \mid c_j)}{P(d)} = \frac{P(c_j)\prod_{i=1}^{m}P(f_i \mid c_j)}{P(d)}$$

(2)

For our study, the denominator of (2) is fixed, as we only focus on the relative size, so we only calculate the numerator. We give the category of highest probability of text. After several
training, we get the estimation of $P(c_j)$ and $P(f_i | c_j)$:

$$
\hat{P}(c_j) = \frac{N_j}{N} \quad (3)
$$

$$
\hat{P}(f_i | c_j) = \frac{\sum_{k=1}^{N} I\{c = c_j \text{ and } f = f_i\}}{\sum_{k=1}^{N} I\{c = c_j\}} \quad (4)
$$

The most fundamental feature of the Naive Bayes classifier is the introduction of Naive Bayes hypothesis, that is, the words in the document are conditional independence, which is a very strong assumption. Although it is not very common, the result is very good. Besides, in practical use, there is a "zero probability" problem in Naive Bayes. The so-called zero probability problem is that the final probability will be zero if a component does not appear in the training set in the process of calculating a new instance. This is unreasonable, we can’t consider that a certain event is impossible to happen just because we don’t observe it. To solve this problem, we introduce Laplace smoothing technology. That is, we add some number to the numerator and denominator when calculate the possibility, to realize that the probability will not be zero, while all the probability value is still added up to one. Laplace smoothing formula is as follows, supposed that there are $H$ kinds of categories:

$$
\hat{P}(f_i | c_j) = \frac{\sum_{k=1}^{N} I\{c = c_j \text{ and } f = f_i\} + 1}{\sum_{k=1}^{N} I\{c = c_j\} + H} \quad (5)
$$

### 3.3 SVM

Support Vector Machine (SVM) is a new classification method developed in recent years. It is a learning system using linear function hypothesis in high dimensional feature space, and has good performance in classification. The SVM model is proposed by Cortes C et. al (1995), which has advantages in solving small samples, nonlinear and high dimensional pattern recognition problems and can be applied to other machine learning problems such as function fitting. SVM has been widely used in the theory research of pattern recognition, knowledge discovery and in the related technological research of computer vision and image recognition, bioinformatics and natural language processing. In natural language processing, SVM is widely used in phrase recognition, word disambiguation, text automatic classification, information filtering and so on.

SVM is developed from optimal separating plane of linear separable cases. The basic idea can be explained by Figure 1: First, in the case of linear separable cases, we find two optimal separating planes $H1$, $H2$ ($H$ is the classification hyperplane between them) of two kinds of samples from the original space. In the case of linear inseparable, the slack variable is added for analysis. By using the non-linear mapping, transform the samples in the low-dimensional input space into the high-dimensional attribute space to make it linear, making it possible that analyze the nonlinear property of samples using linear algorithm from high-dimensional attribute space, and find the optimal separating plane in this feature space. Secondly, it constructs the optimal separating plane in the attribute space by using the principle of structural risk
minimization, so that the classifier is globally optimal, and reaches the upper bound with certain probability of the expected risk in the whole sample space.

Figure 1. Optimal Separating Plane

Suppose that the linear separable sample set is \((X_i, Y_i), i = 1, ..., n, x \in \mathbb{R}^d, y \in \{+1, -1\}\) are category symbols. The general form of the linear discriminant function in the d-dimensional space is the category symbol. The general form of the linear discriminant function in the d-dimensional space is: \(g(x) = w \cdot x + b\), and the classification linear equation is \(w \cdot x + b = 0\). Normalizing the discriminant function, so that all two samples meet \(|g(x)| = 1\), which makes the samples which is closest to the separating plane meet \(|g(x)| = 1\). And the sorting interval is equal to \(2/\|w\|\). Therefore, to maximize the interval is equivalent to minimize the \(\|w\|\) (or \(\|w\|^2\)). To correctly classify all samples by the classification interface, then it should meet:

\[
y_i[(w \cdot x) + b] \geq 1, i = 1, 2, ..., n
\]

(6)

The optimal separating plane is one which can meet the condition (6) and minimize \(\|w\|^2\). The Lagrangian optimization method can be used to convert the optimal separating plane problem into a simple dual problem, and obtain the optimal separating function:

\[
f(x) = \text{sgn}\{(w^* \cdot x) + b^*\} = \text{sgn}\{\sum_{i=1}^{n} \alpha_i^* y_i (x \cdot x_i) + b^*\}
\]

(7)

In this study, we use the two kernel functions (linear and radial basis) of SVM to judge the emotional tendencies of sentences and identify emotional sentences.

1. Linear kernel function

\[
K(x, y) = x^T y + c
\]

(8)

2. Radial basis function (RBF)

\[
K(x, x_i) = \exp\{-\frac{|x - x_i|^2}{\sigma^2}\}
\]

(9)
3.4 SGD

While applying the machine learning algorithm, we usually use gradient descent method to train the used algorithm. The SGD classifier does text classification by logical regression. The classifier is very efficient if it is based on logical regression. It can not only predict the category of the sample, but also can calculate the probability of classification. The principle is:

The hypothesis function of the general linear regression function is:

$$h(\theta) = \sum_{j=0}^{n} \theta_j x_j$$  \hspace{1cm} \text{(10)}

Its energy function (loss function) is:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (y^i - h_{\theta}(x^i))^2$$  \hspace{1cm} \text{(11)}

Since the batch gradient method requires all training samples when updating each parameter, the training process becomes unusually slow as the number of samples increases. The Stochastic Gradient Descent (SGD) is proposed to solve the drawbacks of the bulk gradient descent.

The energy function (11) can be changed as follows:

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{2} (y^i - h_{\theta}(x^i))^2 = \frac{1}{m} \sum_{i=1}^{m} \cos t(\theta, (x^i, y^i))$$  \hspace{1cm} \text{(12)}

$$\cos t(\theta, (x^i, y^i)) = \frac{1}{2} (y^i - h_{\theta}(x^i))^2$$  \hspace{1cm} \text{(13)}

The corresponding gradient can be obtained by taking the derivative of $\theta$ using the loss function of each sample and to update $\theta$:

$$\hat{\theta} = \theta_j + (y^i - h_{\theta}(x^i))x^i_j$$  \hspace{1cm} \text{(14)}

The SGD is updated over each sample. If there is a large number of samples (for example, hundreds of thousands), then it may be iterated to the optimal solution only using tens of thousands or thousands of samples. But SGD may bring some noise, cause that we cannot always follow the overall optimization direction every iteration in SGD. Although it has a high training speed, compared with other algorithms, it has a lower accuracy.

3.5 Majority Decision Algorithm

3.5.1 Model design ideas

The MDA model is a combination algorithm that performs decisions by applying a combination of several classifiers. Catalan et.al. (2016) proposed that the performance can be enhanced by a variety of rules in the design of voting algorithm, including majority voting, minimum probability, maximum probability, probability multiplication and probability averaging. The voting classifier uses the technique of majority vote to combine the machine learning methods, that is, get the classification results through "voting" by three classifiers. The classifiers of the model are less, and it is applied to Turkish. On this basis, we put forward a new model, selected five classifiers with different advantages, combined the results of five classifiers and apply the model to WeChat Chinese text.

Figure 2 shows the ideas of our model, suppose that we have a classification problem: five classifiers and two classes. If the prediction results of five classifiers are: classifier 1 $\rightarrow$ class 2,
classifier 2 → class 2, classifier 3 → class 1, classifier 4 → class 2, classifier 5 → class 1, then the final result is going to be class 2, based on the majority decision rule. In the method of probability averaging, the classification is determined based on the maximum value of the average of predicted probabilities. Our algorithm is a majority choice system, and the premise of this algorithm is that the classification results of each algorithm for each text can be determined and known. The precision of the MDA can be determined by the classification results of each text. This algorithm requires that the performance of the five classifiers also be good, so we selected five classifiers with different advantages to establish our MDA model.

Figure 2. WeChat Sentiment Classification Model.

3.5.2 The theoretical derivation of the model

In the multiple classifiers system, assume that the accuracy of the classifier with the worst performance is a, and that the accuracy of all other classifiers are all a. While combing the five classifiers to classify the samples, the accuracy of multiple classifiers system is:(P denotes accuracy)

\[ P = C_5^3 \cdot a^3 (1-a)^2 + C_5^2 a^3 (1-a)^1 + C_5^4 a^5 (1-a)^0 \]  
\[ P = a^3 (6a^2 - 15a + 10) \]  
\[ y = P - a = a^3 (6a^2 - 15a + 10) - a \]

As shown in equation (17), when a> 0.5, then P> a, shows that the accuracy of the multiple classifiers system is improved and higher than the accuracy of the classifier with the worst performance.
In Figure 3, the abscissa represents a, and the ordinate represents y. By further analysis of Figure 3, we can know that when the difference between the maximum and minimum values of the accuracy of each classifier is less than 15%, the accuracy of the multiple classifiers system can be higher than the accuracy of the classifier with the highest performance. This conclusion proves that the classification results by multiple classifiers system wouldn’t be the worst if we cannot know the performance of each classifier. When the accuracy of each classifier is known, and the difference between the maximum and minimum values of the accuracy of each classifier is less than 15%, then the performance of the multiple classifiers system can be the best.

3.6.3 The evaluation index of the model

We use the precision rate, recall rate, F1 score as measures of the performance of classifiers. For classification, the terms “true positives, true negatives, false positives, and false negatives” are concluded from the comparison of the classification results by the classifier and experts. The terms “positive and negative” indicate the classifier’s prediction (sometimes known as the expectation), and the terms “true and false” mean whether that prediction corresponds to the experts’ judgment or not (sometimes known as the observation).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Prediction Positive</th>
<th>Prediction Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition positive</td>
<td>True Positive (TP)</td>
<td>False Negative (FN)</td>
</tr>
<tr>
<td>Condition negative</td>
<td>False Positive (FP)</td>
<td>True Negative (TN)</td>
</tr>
</tbody>
</table>

Precision rate(P)is in terms of prediction results, and it shows how many of the samples predicted positive are correctly classified.

Precision rate: \( P = \frac{TP}{TP + FP} \)

Recall rate(R)is in terms of original sample, and it shows how many of the positive samples are correctly classified.

Recall rate: \( R = \frac{TP}{TP + FN} \)

F1 is the balanced value of P and R to evaluate the overall result of classification.

\( F1 = \frac{2*P*R}{P+R} \)

4. Results and Analysis

4.1 Data processing
The data processing part is mainly to deal with data, including three sub-modules: Chinese word segmentation, data cleaning and data segmentation. The module of data cleaning mainly includes: wipe off the stop words, punctuation and so on. The Chinese word segmentation module mainly divides the text into words. The data segmentation module is used to segment the training dataset and the test dataset, which is respectively used for the training and evaluating of the model. If the system is still in the testing and verifying phase, we should divide the data (which has already experienced the word segmentation) according to a certain proportion by the data segmentation module. If it is already in the normal phase of the system operation, we do not need the data segmentation module to segment the data.

4.1.1 Chinese word segmentation

The data comes from the social platform WeChat, which is a Chinese social media, so it is needed to preprocess the Chinese text. In an English text, there are some clear spaces among words. However, in Chinese texts, there is no obvious interval among Chinese words. It becomes the key of preprocessing that how to split a Chinese sentence into several correct phrases. For example, a Chinese sentence may be like this: "今天我查阅了资料", the sentence is split into several phrases: "我", "今天", "借", "了","一本","书".(Figure 4) This should be done before the following classifier training. In the experiment, we would use jieba library from Python component for Chinese word segmentation.

![Figure 4. Chinese word segmentation](image)

4.1.2 Data cleaning

After the word segmentation, there are many words without emotion in the Chinese sentences, such as "的", "一个" and so on, we call them stop words. Removing these words can improve the accuracy of classification. The main task of data cleaning is to remove some common pause words and punctuations in Chinese. We first construct a lexicon of stop words, which stores some common stop words and Chinese punctuations. And then check the results of each word segmentation in turn, remove the word if it is in the lexicon

4.1.3 Data segmentation

As a supervised learning process, the training and testing of the model are inseparable. After filtering all data, the experimental data was divided into training set and testing set in accordance with the proportion of 8:2. The selection was completely random for concrete data, there was no intersection between the positive and negative sets. Finally, we got four documents: train-pos.txt, train-neg.txt, test-pos.txt, test-neg.txt. Each file had carried on the data cleaning and Chinese word segmentation. And put the original text into a line, the specific example is shown in Figure 5:

(1) Training sets for classifiers: train-pos.txt contains 4000 positive texts, train-neg.txt
contains 4000 negative texts, for training data of classification model;

(2) Testing sets for classifiers: test-pos.txt contains 1000 positive texts, test-neg.txt contains 1000 negative texts, for testing data of classification model.

Figure 5. Data processing

4.2 Experimental Design

Using BlueStacks to get nearly 20,000 data from WeChat Moment. (BlueStacks is a software that can simulate Android system in computer, it can get the highest authority of Android system, so we can get the cache data.) The process of crawling data is shown in Figure 6:

First, installed WeChat in BlueStacks, then let volunteer login WeChat, opened WeChat Moment, sliding to the bottom of the Moment, then extracted cache from the system, and finally used python to process the cache and got the dataset. After preprocessing the data, there were 12966 data available, and we asked experts to judge the emotional tendencies of these selected corpus. After that, we selected the equal number of positive and negative texts, each 3000 as our experimental datasets. We used 80% of these texts to train five basic classifiers, used the remaining 20% to test our model’s performance. Figure 7 shows the process of our experiment:
4.3 Experiment results and related analysis

4.3.1 Classification results and analysis of basic classifiers

Table 2 shows the classification results of WeChat texts by the five basic classifiers. We use default parameters of classifiers.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Precision rate</th>
<th>Recall rate</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Bayes (sklearn)</td>
<td>0.8598</td>
<td>0.8544</td>
<td>0.8571</td>
</tr>
<tr>
<td>SGD</td>
<td>0.8678</td>
<td>0.9349</td>
<td>0.9001</td>
</tr>
<tr>
<td>SVM (linear)</td>
<td>0.8511</td>
<td>0.9380</td>
<td>0.8924</td>
</tr>
<tr>
<td>SVM (RBF)</td>
<td>0.8739</td>
<td>0.9226</td>
<td>0.8975</td>
</tr>
<tr>
<td>Naive Bayes (SnowNLP)</td>
<td>0.8154</td>
<td>0.8482</td>
<td>0.8315</td>
</tr>
</tbody>
</table>

The experimental results show that the accuracy of SGD and RBF-based SVM are higher than others; in terms of recall rate, SGD, linear SVM and RBF-based SVM are higher than others; for F1 score, SGD performs best. By adjusting the parameters, three indicators have increased by about 1-2%. From the experimental results we can see that no classifier can achieve the best on the three indicators, so we need MDA to improve the performance of individual classifiers.

4.3.2 Classification results and analysis of Majority Decision Algorithm Model

While using a single classifier for sentiment classification, it is easy to make classification mistakes because of some errors. If we use MDA to make emotion classification for a particular text, although every classifier may make mistakes, for the probability of classification errors, the MDA is lower than individual classifiers. Therefore, MDA improves the performance of sentiment classification on WeChat datasets. The results of MDA are shown in Figure 8:
We calculated the emotional tendencies of the remaining 20% WeChat datasets. Figure 8 shows the classification results. Comparing to the five single classifiers, our model is acceptable. We can see that the three indicators reached desired results, and the performance of the classifiers can be improved by the use of MDA.

5. Conclusions
Aiming to improve the performance of machine learning in the short text of WeChat Moment, this study discusses that the methods of classical sentiment classification, the establishment of feature vector in Chinese text and the design idea of MDA. Based on the analysis of the existing sentiment analysis methods, a Majority Decision Algorithm model, composing of five classifiers, Naive Bayes (sklearn), Naive Bayes (SnowNLP), SVM (linear), SVM (RBF) and SGD, was proposed. The following conclusions were obtained:

(1) Majority Decision Algorithms can effectively improve the performance of sentiment classification by single classifiers. Our experiment showed that MDA combined the advantages of single classifiers in the classification process, and made significant improvements in terms of accuracy rate, recall rate and F1 score.

(2) When the performance of each classifier is known, if the difference between the maximum and minimum values of the accuracy of each classifier is less than 15%, we could get the best performance by the use of majority decision algorithm.

(3) Majority decision algorithm validates the performance of multiple classifiers system, improve the performance of machine learning in emotion classification of Chinese text, providing technical support for emotion analysis of Chinese text.

The conclusions have certain reference value for the study of emotion classification based on WeChat text. However, this study is only a preliminary exploration of machine learning in the field of WeChat sentiment analysis, and mainly focuses on the analysis of sentiment polarity of the text. The future work needs to further explore the emotional analysis combing with the characteristics of social network, such as the study of emotional communication and emotional transmission of social networks, mining the evolution law of opposite emotion in groups, mining emotional communities and so on.

Acknowledgments
This work was sponsored by National Natural Science Foundation of China (NSFC) under grant number 71571022.

References