A Novel Calibrated Label Ranking Based Method for Multiple Emotions Detection in Chinese Microblogs

Mingqiang Wang[1], Mengting Liu[1], Shi Feng[1,2], Daling Wang[1,2], and Yifei Zhang[1,2]

[1] School of Information Science and Engineering, Northeastern University
[2] Key Laboratory of Medical Image Computing (Northeastern University), Ministry of Education, Shenyang 110819, P.R. China
{neuwangmingqiang,neuliumengting}@126.com,
{fengshi,wangdaling,zhangyifei}@ise.neu.edu.cn

Abstract. The microblogging services become increasingly popular for people to exchange their feelings and opinions. Extracting and analyzing the sentiments in microblogs have drawn extensive attentions from both academia researchers and commercial companies. The previous literature usually focused on classifying the microblogs into positive or negative categories. However, people's sentiments are much more complex, and multiple fine-grained emotions may coexist in just one short microblog text. In this paper, we regard the emotion analysis as a multi-label learning problem and propose a novel calibrated label ranking based framework for detecting the multiple fine-grained emotions in the Chinese microblogs. We combine the learning-based method and lexicon-based method to build unified emotion classifiers, which alleviate the sparsity of the training microblog dataset. Experiment results using NLPCC 2014 evaluation dataset show that our proposed algorithm has achieved the best performance and significantly outperforms other participators' methods.

Keywords: Microblog, Sentiment Analysis, Calibrated Label Ranking.

1 Introduction

With the advent of Web 2.0 technology, the mushrooming social media with user-generated-content have drawn extensive attentions from all over the world. Nowadays, people are willing to express their feelings and emotions via the microblogging services, such as Twitter and Weibo, because of easy accessibility and convenience. Therefore, the microblog has aggregated huge amount of tweets that contain people's rich sentiments. Extracting and analyzing the sentiments in microblogs has become a hot research topic for both academic communities and commercial companies.

A lot of papers have been published for analyzing the sentiments in blogs, reviews and news articles. However, there are several critical new challenges for detecting the emotions in microblogs.

(1) **Short text**. The microblog usually has a length limitation of 140 characters, which leads to extremely sparse vectors for the learning algorithms. The free and informal writing styles of users also set obstacles for the emotion detection in microblogs.

C. Zong et al. (Eds.): NLPCC 2014, CCIS 496, pp. 238–250, 2014.

(2) **Fine-grained emotions**. Different from the traditional binary classification problem, namely classifying the text into *positive* or *negative* categories, the task of emotion detection needs to identify people's fine-grained sentiments in microblogs, such as *happiness*, *sadness*, *like*, *anger*, *disgust*, *fear*, and *surprise*.

(3) **Mixed emotions**. Although the text is short, multiple emotions may be coexisting in just one microblog. Take the following tweet as an example, the twitterer expresses a *like* emotion as well as an *anger* emotion simultaneously in one tweet. Therefore, for the emotion detection task, each microblog can be associated with multiple emotion labels, which is rarely studied in the previous literature.

"@User: *What a fantastic movie! But the dinner sucks!!!!!!*"

To tackle these challenges, in this paper we regard the multiple emotions detection in microblog as a multi-label learning problem and propose a novel calibrated label ranking based method for classifying the fine-grained sentiments in the Chinese microblogs. Firstly, we transform the multi-label training dataset into single-label dataset. Then we propose a two-stage method that combining emotion lexicon and SVM to identify whether an instance (microblog or sentence) has emotion. Thirdly we construct Naïve Bayes (NB) and SVM multi-class classifiers to learn the confidence ranking value of the k emotion labels. Fourthly, we build k binary classifiers to update the ranking value and determine the threshold t. Finally we use the emotion lexicon to update the ranking value to get the final label order. For the ranking value of each label, if it is bigger than the threshold t, we put it into the relevant label set. Otherwise we put it into the irrelevant label set. To summarize, the main contributions of this paper are as follows.

(1) We propose a novel calibrated label ranking based framework for detecting the multiple fine-grained emotions in the Chinese microblogs.

(2) We combine the learning-based method and lexicon-based method to build unified emotion classifiers, which alleviate the sparsity of training microblog dataset and successfully improve the performance of the emotion classification results.

(3) Experiment results using NLPCC 2014 evaluation dataset show that our proposed algorithm significantly outperforms other team's methods, and achieve the best performance in both the "Weibo Emotion Classification" and "Emotion Sentence Identification and Classification" subtasks.

The rest of the paper is organized as follows: Section 2 introduces the related work. In Section 3, we propose our calibrated label ranking based method for multiple emotions detection in Chinese microblogs. In Section 4, we introduce the experiment setup and results. Finally we conclude the paper and give the future work in Section 5.

2 Related Work

Our work is related to two directions: multi-label learning and sentiment analysis. Multi-label learning methods originally focused on the polysemy problem of text categorization [1-2]. After years of development, the multi-label learning algorithms have been widely used in the information retrieval [3-5], personalized recommendation

[6], Web mining [7], and etc. It is a hot research topic of the scholars. Zhang et al. gave a comprehensive review of the multi-label learning algorithms in [8]. Johannes et al. proposed a problem transformation algorithm "calibrated label ranking" that could be viewed as a combination of pairwise preference learning and the conventional relevance classification technique, where a separate classifier was trained to predict whether a label was relevant or not [9]. Another representative problem transformation algorithm "binary relevance" proposed by Matthew et al. transformed the multi-label classification problem to the binary classification problem [10]. Some other algorithm adaption methods were also widely used in the multi-label learning area. This kind of methods improved the traditional supervised machine learning algorithms to solve the multi-label problem. Zhang M-L et al. proposed the MLKNN improved by the lazy learning algorithm KNN to deal with the multi-label problem [11]. Elisseeff et al. improved the kernel learning algorithm SVM to fit the algorithm to the multi-label data [12].

The sentiment analysis researches can be traced back to early 2000's. Pang and Lee [13] verified the effectiveness of applying machine learning techniques to sentiment classification, Costa [14] proposed a framework to integrate mining algorithms and software agent for building blog based sentiment applications. Zhang [15] provided a synthetic method to extract sentiment features from product reviews as product weakness finder. Huang [16] utilized the semantic information to construct a semantic sentiment space model to facilitate sentiment classification task.

Although a lot of papers have been published for multi-label learning and sentiment analysis, little work is done for detecting the fine-grained multi-label emotions in the microblog short text. In this paper, we propose a calibrated label ranking based method combined with emotion lexicon for the multiple emotions detection in Chinese Microblog.

3 Calibrated Label Ranking Based Emotion Classification

In the above section, we introduced the related work of multi-label learning and sentiment analysis. In this section, we propose a novel calibrated label ranking based method for sentence and microblog multiple emotions classification. Suppose $Y=\{y_1,y_2,\ldots,y_k\}$ represents the emotion label set {*happiness*, *sadness*, *like*, *anger*, *disgust*, *fear*, *surprise*} and $k=7$. Give the training dataset $D=\{(x_i,Y_i)|1\leq i\leq n\}$ where n is the number of instances in D and $Y_i\subset Y$, our task is to build a multi-label classifier that can predict the emotion label set $h(x)\subseteq Y$ of a new instance x. The overall framework of our proposed approach is shown in Figure 1.

In Figure 1, the proposed calibrated label ranking method has the following main steps. (1) We transform the multi-emotion label training microblog dataset into single-label dataset. (2) We build an emotional/non-emotional classifier to identify the emotional sentence in the testing dataset. (3) We propose a novel calibrated label ranking based method that integrating SVM and NB classifiers to get the ranking value of each emotion label of the sentence. (4) We construct k binary classifiers to update the ranking value and get the threshold. (5) We leverage the emotion lexicon to further update the ranking value and finally get the relevant label set by comparing the ranking value of each label with the threshold. In the following sections, we will discuss the steps of our proposed method in detail.

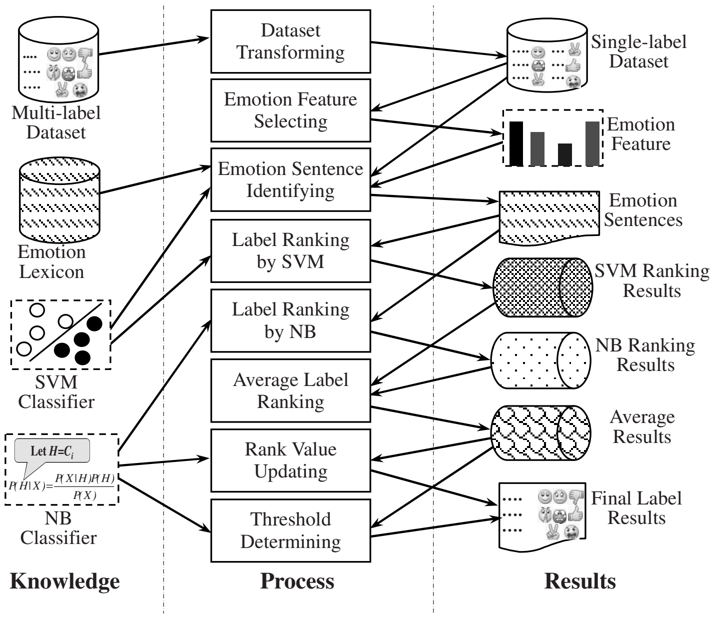

Fig. 1. The overall framework for calibrated label ranking based emotion classification

3.1 Training Dataset Transformation and Feature Selection

One popular strategy of solving the multi-label learning problem is to transform it into other well-established learning scenarios. For the given training dataset, each sentence in the microblog may contain one or two relevant emotion labels, therefore we need to transform the multi-label training dataset into the single-label train dataset so that we can use the traditional supervised machine learning algorithm to address the multi-label learning problem. The dataset transformation method is shown in Algorithm 1.

Algorithm 1. Training dataset transformation

Input: The multi-label training microblog dataset D;
Output: The single-label training dataset D' after transformed;
Description:

1. FOR every microblog m in the multi-label training dataset
2. FOR every sentence x_i in microblog m, where Y_i is the relevant emotion label set of the sentence x_i and $D=\{(x_i,Y_i)|1\leq i\leq n\}$, where n is the number of sentences in mocroblog m
3. FOR every label y_j in the label set Y, where $Y=\{$*like*, *happiness*, *sadness*, *disgust*, *anger*, *fear*, *surprise*$\}$
4. If $y_j \in Y_i$, put the training instance (x_i,y_j) into D'

In Algorithm 1, we associate each training item with only one emotion label, which transforms the multi-label dataset into a single-label dataset. Note that a sentence *x* in *D* may appear multiple times in the dataset *D*'.

Feature selection is the critical step for the text classification problem. In this paper, we select the words whose frequency is bigger than 2, which could alleviate the dimension disaster. That is to say, we ignore the words that appear less than twice in all the sentences. In addition, we select the emotion features listed in Table 1, which are potential good emotion indicators. Here we choose the punctuations, emoticons, emotion words extracted from the emotion ontology of Dalian University of Technology (DLUT) and the cyber words. Generally speaking, a sentence contains continuous punctuation always has an emotion. The emotion words and the emoticon can obviously imply that a sentence contains emotions. Some cyber words are also good symbols for people's emotions. We add these features into the user dictionary of Chinese segmentation tool, so that the tool can recognize these words and symbols.

Table 1. The Chinese Microblog Features

Feature Name	Feature Description and Examples
Punctuation	! , ? , ???, !!!, ?????!!!!!
Emoticon	
Emotion word	Extracted from the DLUT emotion ontology
Cyber word	*给力* (awesome), *稀饭* (like)

3.2 Multi-label Emotion Analysis of Weibo Sentence

To detect the multiple emotions in the microblog sentences, firstly we need to identify whether the candidate sentence is emotional or not. Then we need to detect the top 2 emotions of an emotional sentence. The target fine-grained emotion label set *Y*={*like*, *happiness*, *sadness*, *fear*, *surprise*, *disgust*, *anger*}, and if an emotional sentence contains only one emotion, then the second emotion is set to be *none*.

3.2.1 A Two-Stage Method for Emotional Sentence Identification

In this section, we proposed a two-stage algorithm that combining the lexicon based and learning based methods for emotional sentence identification. The overall procedure of the proposed identification algorithm is shown in Figure 2.

We can see from Figure 2 that in the first stage, we utilize emotion lexicon to identify the emotional sentences in the testing dataset. If at least one word of the sentence is matched in the lexicon, we regard the sentence as emotional. In the second stage, we construct a SVM based classifier to indentify the emotional sentences in the resting dataset that has no obvious emotion words. The aim of this two-stage approach is that we improve the Recall of the identification as well as keeping the Precision. Note that in the close setting, we use the DLUT emotion lexicon and in the open setting, we integrate emoticons, cyber words and extracted emotion words with DLUT lexicon.

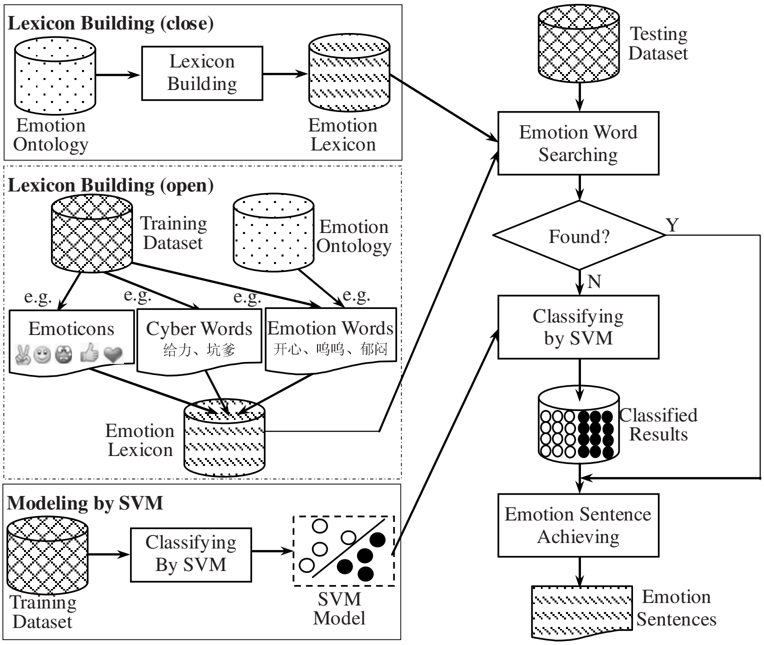

Fig. 2. The two-stage algorithm for emotional sentence identification

3.2.2 Calibrated Label Ranking

We leverage the transformed single-label training dataset to build a classifier $f(\cdot)$, by which we can get the confidence that an emotion is the proper label with a sentence. In this paper, $f(x_i,y_j)$ represents the confidence that sentence x_i has the emotion label y_j. Here $f(x_i,y_j)$ can be transformed to a rank function $rank_f(x_i,y_j)$. This rank function reflects all the real value output $f(x_i,y_j)$, where $y_j \subset Y$ and $1 \le j \le q$ (q=7). We use the Formula 1 to get the rank value.

$$rank_f(x_i, y_j) = \sum_{k=1,\, j\neq k}^{7} [\![f(x_i, y_j) > f(x_i, y_k)]\!] \tag{1}$$

where $[\![\pi]\!]$=1 if the predicate π holds, and 0 otherwise.

In this paper, we utilize the SVM and NB respectively to realize the label ranking. NB classifier can directly get the confidence of the k emotions. But SVM is a binary classifier, so here we use the 1-v-1 method to do the multi-class classification. In our experiment we adopt the average rank value $rank_f^{avg}(x_i, y_j)$ of NB and SVM.

3.2.3 Threshold Calibration

After we get the rank values of all the emotion labels, the key issue is to determine a threshold to separate the rank value list into relevant label set and irrelevant label set. In this paper we utilize the “binary relevance” algorithm to determine the threshold. The processing procedure is as follows:

(1) Transform the multi-label training dataset into k (=7) separate datasets and each dataset is associated with a unique emotion label.
(2) Construct k binary classifiers respectively corresponding with the k emotion labels using the training datasets.
(3) Employ the k binary classifiers to update the threshold and the rank value for every unseen sentence.

Binary Classifier Construction. For every binary classifier $g_j(x_i)$, we firstly construct the train dataset D_j , $j\in\{1,2,\ldots,k\}$, where j denotes the emotion label and x_i represents a sentence. The original training dataset is denoted by $D=\{(x_i,Y_i|1\leq i\leq m)\}$, where m is the number of sentences in D and Y_i is the relevant label set of the sentence x_i. We use the Formula 2 to construct the training set of the ith classifier.

$$D_j = \{(x_i, \phi(Y_i, y_j)) \mid 1 \le i \le m\}$$
$$\phi(Y_i, y_j) = \begin{cases} +1, & y_j \in Y_i \\ -1, & \text{otherwise} \end{cases} \tag{2}$$

The separate training datasets D_j ($1\leq j\leq k$) are utilized to train k binary classifiers $g_j(x_i)$ ($1\leq j\leq k$) to classify every sentence in the testing dataset. For $g_j(x_i)$, if x_i is corresponding with y_j, then $g_j(x_i)$=+1, otherwise $g_j(x_i)$=-1. This is obviously a binary classification problem.

Rank Value Updating and Threshold Calibration. We use the following rule to update the rank value of label y_j $rank_f(x_i,y_j)$ and the threshold $rank_f^*(x_i, y_V)$ where y_V is a virtual label to separate the rank list.

$$rank_f^*(x_i, y_j) = rank_f^{avg}(x_i, y_j) + [\![g_j(x_i) > 0]\!] \tag{3}$$

$$rank_f^*(x_i, y_V) = \sum_{j=1}^{7} [\![g_j(x_i) < 0]\!] \tag{4}$$

In the following section, we will further update the rank value using the emotion lexicon, and build the final multi-label classifier $h(x)$.

3.2.4 Updating Ranking Value by Emotion Lexicon

The emotion words are good indicators for emotion classification. If a sentence contains an emotion word of the label y_j, we can infer that this sentence has high probability to be associated with label y_j. Suppose E represents an emotion lexicon that has k (=7) categories, and each of the categories is associated with a label in {*happiness*, *sadness*, *like*, *anger*, *disgust*, *fear*, *surprise*}. For sentence x_i, if a word of x_i is matched in the lexicon E with label y_j, we utilize the Formula 5 to update the ranking value of y_j for x_i.

$$rank_f^*(x_i, y_j) = rank_f^*(x_i, y_j) + 1 \tag{5}$$

Finally based on the above formulas, we get the multi-label emotion classifier $h(x_i)$:

$$h(x_i) = \{y_j \mid rank_f^*(x_i, y_j) > rank_f^*(x_i, y_V),\ y_j \in Y\} \tag{6}$$

In Formula 6, the size of the relevant label set $h(x_i)$ may be bigger than two. In this situation we just select the top two emotions to fit with the evaluation task of the testing dataset. And another exception is that the relevant label set $h(x_i)$ may be empty, in this case we just select the top one emotion in the ranking list.

3.3 Multi-label Emotion Analysis of Weibo Text

The Weibo text usually consists of a group of sentences. The sentence level multi-label emotion analysis method can be seen as the basis of Weibo text level multiple emotion detection. So we can use the sentence level multi-label classification result in Section 3.2 to achieve the emotion labels at Weibo text level. Let x_i represent the sentence in Weibo m, where $1 \le i \le n$ and n is the number of sentences in m. We set the dominating emotion of the sentence the rank value 2, and 1 for the secondary emotion. Calculate the rank value $rank_m(m,y_j)$ of all the labels in Weibo m by the relevant label set of all the sentences in the Weibo.

$$rank_m(m, y_j) = \sum_{i=1}^{n} rank_m^{emo}(x_i, y_j) \tag{7}$$

where $rank_m^{emo}(x_i, y_j)$ is defined as follows:

$$rank_m^{emo}(x_i, y_j) = \begin{cases} 0, & y_i \text{ is the irrelevant label} \\ 1, & y_i \text{ is the secondary label} \\ 2, & y_i \text{ is the dominant label} \end{cases} \tag{8}$$

Similar to the sentence level multi-label emotion analysis, we employ the emotion lexicon to update the rank value of each label to get the final rank. After that we need to determine a threshold to separate the rank value list into relevant label set and the irrelevant label set. Here we defined $rank_m(m,y_V)$ as the threshold, where y_V is a virtual label. The threshold is calculated by counting the "*none*" labels of each sentence as follows.

$$rank_m(m, y_V) = \sum_{i=1}^{n} rank_m^{none}(x_i, y_V) \tag{9}$$

where y_V means the "*none*" label, and

$$rank_m^{none}(x_i, y_V) = \begin{cases} 0, & x_i \text{ has two emotions} \\ 1, & x_i \text{ has only one emotions} \\ 3, & x_i \text{ is not an emotional sentence} \end{cases} \tag{10}$$

Most sentences only contain one dominant emotion, which can easily cause the threshold $rank_m(m,y_V)$ bigger than the rank value $rank_m(m,y_j)$. We set a parameter α to balance the $rank_m(m,y_V)$. So the relevant label set $h(m)$ of Weibo m is calculated as:

$$h(m) = \{ y_j \mid rank_m(m, y_j) \geq rank_m(m, y_V) - \alpha,\ y_j \in Y \} \tag{11}$$

If the rank value of a label is bigger than the threshold, we put the label into the relevant label set. Otherwise we put it into the irrelevant label set. As the same as the sentence level emotion analysis, if there are more than two emotion labels in the relevant label set, we just select the top two emotions. And if the relevant label set is empty, we regard the Weibo text as a non-emotional. In this paper, we empirically set α=2 in the following experiment section. The tuning of parameter α is omitted due to length limitation.

Besides using the Formula 11, we can also consider the whole Weibo text as a long sentence and utilize the classification method proposed in Section 3.2 to detect the multiple emotions in the microblogs.

4 Experiment

4.1 Experiment Setup

We conduct our experiment on the real-world dataset that provided by NLPCC 2014 Emotion Analysis in Chinese Weibo Texts (EACWT) task[1]. The EACWT training dataset contains 14,000 Weibo and 45,421 sentences. There are 6,000 Weibo and 15,693 sentences in the testing dataset of EACWT. We conduct experiments using a PC with Inter Core i7, 8 GB memory and Windows 7 as the operating system.

For the sentence level emotion analysis, we firstly transform the multi-label training dataset into single-label dataset using Algorithm 1. And then we identify whether a sentence is emotional or not using the two-stage method. We employ the transformed training dataset to construct the 7-classes classifier separately by NB and SVM to get the average ranking value of each label of an emotional sentence. Next we construct 7 binary NB classifiers to determine the threshold and update the rank value. At last we use the emotion lexicon to update the rank value of each label. For Weibo text level emotion analysis, we learn the Formula 11 to detect the multiple emotions. Another alternative strategy is that we can regard the whole Weibo text as a long sentence and employ the algorithm in Section 3.2 to classify the multiple emotions in the Weibo text.

4.2 Experiment Results

There were 7 teams participated the EACWT task of NLPCC 2014. We compare our results with other participators using the Average Precision (AP).

[1] http://tcci.ccf.org.cn/conference/2014/pages/page04_eva.html

4.2.1 Emotion Sentence Identification and Classification

In the close evaluation of the *Emotion Sentence Identification and Classification*, we strictly use the given training dataset and construct the emotion lexicon by the DLUT emotion ontology. In the open evaluation, we combine the dataset of NLPCC 2013, the sample and training dataset of NLPCC 2014 to form the new training dataset. We also manually add more emotion words into the emotion lexicon.

Emotion Sentence Identification. Firstly, we evaluate our proposed two-stage method for emotion sentence identification, i.e. classifying the sentences into emotional and non-emotional categories. The results are shown in Table 2.

In Table 2, all the three methods have similar F-Measures. However, the two-stage method (Lexicon+SVM) can achieve an extreme high Recall value. That is to say, most of the emotional sentences are included in the identification results, which pave the way for the further multi-label emotion classification steps.

Table 2. The comparisons for emotion sentence identification (open)

	Precision	Recall	F-Measure
SVM	**49.24**	57.73	53.15
Lexicon	40.49	77.14	53.11
Lexicon+SVM	38.41	**87.21**	**53.40**

Emotion Sentence Classification. Our approach combines the learning based method NB and SVM with emotion lexicon to get the ranking value of each label. To evaluate the effectiveness of our proposed method, we apply NB, SVM, NB+SVM, NB+Lexicon, SVM+Lexicon respectively to get the rank value of each label. In Table 3, CLR represents the Calibrated Label Ranking based method proposed in this paper. NEUDM-1 and NEUDM-2 denote the submit versions for the evaluation that are slightly different in feature selection methods. We can see from Table 3 that the combined approach CLR achieves the best performance compared with other methods. This validate that the proposed ensemble algorithm that combining NB, SVM and emotion lexicon is effective for sentence level multiple emotion detection.

Table 3. The comparisons for sentence level emotion classification (Average Precision, Strict)

	Result ID	NB	SVM	NB+ SVM	NB+ Lexicon	SVM+ Lexicon	CLR
Close	NEUDM-1	0.4479	0.4660	0.4778	0.4704	0.4883	**0.5042**
	NEUDM-2	0.4421	0.4580	0.4700	0.4578	0.4786	**0.4911**
Open	NEUDM-1	0.4772	0.4850	0.5044	0.4963	0.5128	**0.5330**
	NEUDM-2	0.4773	0.4850	0.5044	0.4953	0.5133	**0.5317**

In Table 4, we compare the sentence level emotion classification results with other participators of NLPCC 2014 EACWT task. The AVG and MAX represent the average and max value of all the participators. The Sec-Best denotes the second best value of all the participators. We can see that our proposed method significantly outperforms other participators' methods in all the evaluation settings.

Table 4. The comparison with other participators for sentence level classification

Result ID	Close		Open	
	AP(loose)	AP(strict)	AP(loose)	AP(strict)
NEUDM-1	**0.5502**	**0.5042**	**0.5799**	**0.5330**
NEUDM-2	0.5381	0.4911	0.5785	0.5317
AVG	0.3981	0.3666	0.5158	0.4791
Sec-Best	0.3032	0.2831	0.5489	0.5175
MAX	**0.5502**	**0.5042**	**0.5799**	**0.5330**

4.2.2 Weibo Emotion Classification

In the close evaluation of Weibo emotion classification, we strictly use the given training dataset and construct the emotion lexicon by the DLUT emotion ontology. We regard the whole Weibo text as a long sentence and adopt the calibrate label ranking algorithm proposed in Section 3.2 to detect the multiple emotions in Weibo text. In the open evaluation, we make use of the open classification result of the emotion sentence classification and employ the Formula 11 to realize the Weibo text level emotion classification. In Table 5, we compare our results with the valid results of other participators of NLPCC2014 EACWT task.

Table 5. The comparison with other participators for Weibo level classification

Result ID	Close		Open	
	AP(loose)	AP(strict)	AP(loose)	AP(strict)
NEUDM-1	**0.6033**	**0.5115**	**0.5851**	**0.4960**
AVG	0.4306	0.3678	0.5145	0.4406
Sec-Best	0.5736	0.4756	0.5309	0.4668
MAX	**0.6033**	**0.5115**	**0.5851**	**0.4960**

In Table 5, our proposed method has achieved the best performance in all the participators. Note that in NEUDM-1, the AP values of the close setting are better than the AP values of the open setting, namely regarding the Weibo text as separate sentences does not lead to a better performance than regarding the whole Weibo text as a long sentence. This may be because of the accumulation of errors at the sentence level emotion classification. The Formula 11 is counting on the classification result of each sentence in the Weibo text. Therefore, combining each sentence' top ranked two labels may accumulate more errors. We can tackle this problem by regarding the whole Weibo text as a long sentence for emotion analysis.

5 Conclusions and Future Work

Recently, emotion analysis in microblogs has become a hot research topic, which aims to classify the fine-grained sentiments in the short text. Different from the traditional sentiment analysis problem, the fine-grained sentiments co-exist with each other in the short text and the emotion detection in microblog can be regarded as a typical multi-label learning problem. In this paper, we proposed a novel calibrated label ranking based multi-label emotion analysis approach for Chinese microblogs. Our proposed model combined multi-label learning algorithm and emotion lexicon together, which provided a comprehensive understanding of the embedded emotions in the short text. Experiment results based on NLPCC 2014 EACWT evaluation task showed that our proposed method significantly outperformed the methods of other participators and our proposed ensemble learning framework achieve better performance than other single classifiers or other combined approaches.

Our future work includes further improving the performance the emotional sentence identification task, which is also a critical step for the evaluation task. We intend to automatically find more effective emotion features and integrate more classifiers. We also want to propose new multi-label learning methods for the multiple emotions detection task. We will study on measuring the fine-grained emotion similarity between the Chinese microblogs and leverage the MLKNN based lazy learning method for multiple emotions detection.

Acknowledgements. This work is supported by the State Key Development Program for Basic Research of China (Grant No. 2011CB302200-G), State Key Program of National Natural Science of China (Grant No. 61033007), National Natural Science Foundation of China (Grant No. 61100026, 61370074, 61402091), and Fundamental Research Funds for the Central Universities (N120404007).

References

1. Tsoumakas, G., Katakis, I.: Multi-label classification: An overview. International Journal of Data Warehousing and Mining, 1–13 (2007)
2. Tsoumakas, G., Zhang, M., Zhou, Z.: Learning from multi-label data. Tutorial at the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2009), Bled, Slovenia (2009)
3. Schapire, R., Singer, Y.: BoosTexter: A boosting-based system for text categorization. Machine Learning 39(2-3), 135–168 (2000)
4. McCallum, A.: Multi-label text classification with a mixture model trained by EM. In: Working Notes of the AAAI 1999 Workshop on Text Learning, Orlando, FL (1999)
5. Ueda, N., Saito, K.: Parametric mixture models for multi-label text. In: Becker, S., Thrun, S., Obermayer, K. (eds.) Advances in Neural Information Processing Systems 15 (NIPS 2002), pp. 721–728. MIT Press, Cambridge (2003)
6. Song, Y., Zhang, L., Giles, L.: A sparse Gaussian processes classification framework for fast tag suggestions. In: Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM 2008), pp. 293–102. Napa Valley, CA (2008)
7. Tang, L., Rajan, S., Narayanan, V.: Large scale multi-label classification via metalabeler. In: Proceedings of the 19th International Conference on World Wide Web (WWW 2009), Madrid, Spain, pp. 211–220 (2009)
8. Zhang, M., Zhou, Z.: A Review on Multi-Label Learning Algorithms. IEEE Transactions on Knowledge and Data Engineering 26(8), 1819–1837 (2014)

9. Johannes, F.: Multi-label classification via calibrated label ranking. Machine Learning 73(2), 133–153 (2008)
10. Boutell, M.R., Luo, J., Shen, X., Brown, C.M.: Learning multi-label scene classification. Pattern Recognition 37(9), 1757–1771 (2004)
11. Zhang, M., Zhou, Z.: ML-kNN: A lazy learning approach to multi-label learning. Pattern Recognition 40(7), 2038–2048 (2007)
12. Elisseeff, A., Weston, J.: A kernel method for multi-labelled classification. In: Dietterich, T.G., Becker, S., Ghahramani, Z. (eds.) Advances in Neural Information Processing Systems 14 (NIPS 2001), pp. 681–687. MIT Press, Cambridge (2002)
13. Pang, B., Lee, L.: Thumbs up? Sentiment Classification using Machine Learning Techniques. In: Proc. 2002 Conf. on Empirical Methods in Natural Language Processing (EMNLP), pp. 79–86 (2002)
14. Costa, E., Ferreira, F., Brito, P., et al.: A framework for building web mining applications in the world of blogs: A case study in product sentiment analysis. Expert Systems with Applications 39(4), 4813–4834 (2012)
15. Zhang, W., Xu, H., Wan, W.: Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis. Expert Systems with Applications 39(9), 10283–10291 (2012)
16. Sui, H., You, J., Zhang, J., Zhang, H., Wei, Z.: Sentiment Analysis of Chinese Micro-blog Using Semantic Sentiment Space Model. In: Proceedings of 2nd International Conference on Computer Science and Network Technology, ICCSNT 2012, pp. 1443–1447 (2012)