Send Orders for Reprints to reprints@benthamscience.ae

The Open Cybernetics & Systemics Journal, 2015, *9*, 601-607

New Method for Sentiment Classification for Short Text

Hao Fu*

School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract: With the rapid development of the Internet, the microblog platform, BBS, e-Commerce etc. gathered a lot of short messages/text, which contained subjective sentences. These sentences often had obvious inclination which reflected the sentiment of the author. By mining the author's sentiment, such as like, angry, indignation, averseness, etc., we can analyze people's opinion for some policy, people's preferences for some commodities. So, in this paper, we proposed a new method for sentiment classification by combination of several machine learning algorithms, which included feature extraction and ontologies. The further optimization was also given in this paper. We evaluated our method on several datasets and achieved good results.

Keywords: Machine learning, sentiment classification, short message.

1. INTRODUCTION

With the rapid development of the internet, the users of it have more channels to express their views which has also developed into an important information of the text on the internet. Especially in the development process of a wired network, the users express their own views and emotions and are affected by the views on the web at the same time. In the past decade, unstructured data, especially the emotional analysis of text Web, has received increasing attention [1].

This article focuses on the emotion classification for text message of Microblogging and the language resource using emotional words and synonyms cilin body. For the division of emotion, the article uses seven classes of emotions in the emotional vocabulary, namely: anger, disgust, fear, happiness, preferences, sadness, surprise, which is on the basis of the influential Ekman's six categories of emotion [2]. "Synonyms forest" were codified by Mei Jiaju [3] in 1983, which do not only include the synonyms of a word, but also contains a number of similar words-the generalized correlation words.

In the related research of the analysis of Microblogging sentiment, Tsytsarau [1], who compared the algorithms of sentiment analysis in the past decade and compared the performance of the algorithm and finally divided it into four categories: dictionaries, statistics, semantics and machine learning. Among them, dictionaries and machine learning take larger proportion and a number of studies turned to the microblogging platform Twitter in recent years. In fact, we need to use the methods based on the dictionary or rules to get the emotional value when we use machine learning for sentiment classification. Pang [4] puts forward and evaluates three supervised classification methods: Naive Bayes, maximum entropy and support vector machine (SVM) which achieved the best performance.

For sentiment analysis of wired network, most of the methods were used to analyse the emotional polarity and strength of the text, which has a high degree of accuracy [1]. The main methods include SVM, polynomial Naive Bayes and PMI, which were also applied in the analysis of micro-Bo emotion. Sentiment analysis also includes analysis of the degree of emotional polarity, but multi-class sentiment analysis methods are penurious and have less accuracy. Secondly, the short text feature of microblogging brings enormous challenges to sentiment classification, thus, making the text feature cannot fully express the emotions Microblogging [5]. For microblogging features, we use synonyms for the transformation of the word in microblogging and we also use emotional vocabulary words to mark the emotional value of the body.

2. RELATED METHODS

This article uses a feedforward neural network and SVM as supervised learning models. Among them, the single hidden layer of the neural network not only has a strong learning ability to approach complex nonlinear functions, but also can solve the problem which the traditional parameters method cannot solve [8]. Comparison with the neural network, SVM does not need to adjust the complex models, but by solving the multi-class classification tasks needs to train a number of different models [9].

2.1. Feedforward Neural Networks

The basic model used herein feedforward neural networks was showed in Fig. (**1**).

Wherein the arrows indicate the direction of the information transfer, the error is transmitted reverse; X represents the neural network input, including 2118 features and an error adjustment node x_0, namely: $x \in \mathbb{R}$-*2119×1*.; Z represents hidden nodes, which number is d + 1, making adjustment by cross validation; $y \in \mathbb{R}$-*8×1*., containing seven categories of emotion and no emotion None; θ represents a pa-

1874-110X/15 2015 Bentham Open

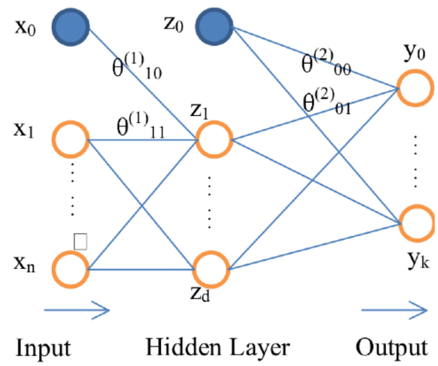


Fig (1). Two layer feedforward neural network.

rameter needed to be optimized, wherein θ-*(1)*.∈ℝ-*d*×*n*+*1*., ,θ-*(2)*.∈ℝ-*d*×*k*+*1*., $\grave{e}_{ij}^{(c)}$ denotes the connection from the node i in the layer of c+1 to the node j in the layer of c. The feedforward neural networks in this paper does not have the cross-layer connection, which means there is no connection between the input node and the output node.

According to the learning model of Fig. (**1**), the output of the K-th output node y_k is as follows:

$$y_k(x,\theta)=\sigma(\sum_{j=1}^{d}\grave{e}_{kj}^{(2)}h(\sum_{i=1}^{n}\grave{e}_{ji}^{(1)}x_i+\grave{e}_{j0}^{(1)})+\grave{e}_{k0}^{(2)}) \quad (1)$$

Among them, $\sigma(a)=1/(1+e^{-a})$, scilicet logistic function, $h(\cdot)$ can select the logistic function or tanh function, this paper chose the former as the hidden layer node transfer function. About the cost function and error propagation of feedforward neural network, you can refer to the literature.

2.2. Support Vector Machine

SVM model is one of the optimal boundary classifier (Maximum Margin Classifiers, MMC) and is different from feedforward neural networks. The objective function is as follows:

$$\mathbf{w},\mathrm{b}=\arg\max_{w,b}\left\{\frac{C}{\|w\|}\min_{n}[\mathrm{t_n}(\mathbf{w}^{\mathbf{T}}\phi(\mathbf{x_n})+b)]+\frac{1}{2}\mathbf{w}^{\mathbf{T}}\mathbf{w}\right\} \quad (2)$$

Among them, w, b are the parameters of SVM model and C is the regularization parameter with the same function of the countdown $1/\lambda$ of λ in neural Networks. $t_n\in\{-1,1\}$, which is the label value of nth training data, is the affine function of the parameter w for the predicted value for a new data. In practical applications, you can use the existing framework of SVM for model training. We can use the framework of Libsvm and make a model for each emotion category, which is combined with the method of one-vs.-the-rest for multi-class emotion classification, and finally output final result to calculate the probability. For new data x, its value is equivalent with the input of the output layer of the neural network according to the model calculation $y_k(\mathbf{x})$. The final predicted value of x is:

$$k=\arg\max_{k}(\mathrm{y_k}(\mathbf{x})) \quad (3)$$

3. PROCEDURE OF SENTIMENT CLASSIFICATION

This article used a feedforward neural network and SVM as supervised learning models. To solve the problem, the single hidden layer neural network not only has a strong learning ability to approach complex nonlinear functions, but also can solve the problem which the traditional parameters method cannot [6]. Comparison with the neural network, SVM does not need to adjust to complex models, but solving the multi-class classification tasks need to train a number of different models [7].

3.1. Text Pre-Processing

Firstly, the article will make segmentation and POS tagging for microblogging, using Stanford University segmentation and POS tagging tools respectively. Chinese word segmentation model is based on conditional random sequence and the word accuracy rate is about 95% [8]. POS tagging uses logarithmic model, combined with a variety of features and POS of areas of vocabulary [9]. POS tagging tool can directly label Chinese sentences, but it has low accuracy in comparison with combining Chinese word segmentation.

This article uses *s* as a training set of microblogging, $w=[\mathrm{w}_1,w_2,...,\mathrm{w}_d]$ represents a sequence of lexical items after word segmentation of s, $t=[\mathrm{t}_1,t_2,...,t_d]$ represents speech sequence s, d represents the number of lexical items Word segmentation after s which includes punctuation marks and special symbols.

The *s*peeches used in the messages are of 34 kinds, which have another URL label than the multi-label sets of original label [10]. In fact, more tags can provide better information, but will increase the tagging error rate.

3.2. Feature Extraction

Short test feature selection combines synonyms Cilin [3] and emotional vocabulary ontology [2], the former on the microblogging lexical items, which were conceptualized learning algorithms for suppressing over-fitting, the latter is used to obtain the value of emotional vocabulary. Synonyms Cilin make vocabulary word synonyms for three coding and the examples are as Table **1**.

Ac01 = tall dwarf giant ......; Ac03 = beautiful ladies ......

Layer 1 is represented with capital letters; layer 2 is represented with lowercase letters; layer 3 is represented with a two-digit decimal integer. In addition, HIT extended synonym Cilin, increasing the word group and atomic layers Code: Section 4 layer is represented with capital letters; 5th floor is represented with two decimal integers [11]. For example:

Aa01A05 = every man individual; Aa01C01 = crowd everyone people.

This paper uses three-synonymous coding Cilin, using Syno (w) to represent a synonymous coding for the word w, such as: Syno (Giants) = Ac01.

The emotional vocabulary words are divided into 21 main categories of emotions, such as happiness (PA), respect (PD), anger (NA). In addition to the emotional type, emo-

Table 1. Example of emotional vocabulary ontology.

Number	Vocabulary	Emotion Classification	Strength	Polarity
i	sw	Emotion (sw)	Strength (sw)	Polarity (sw)
1	dirty	NN	7	2
7	War	ND	5	2
7	War	NC	5	2
11	Qingying	PH	5	1
214	Worship	PD	5	0

Table 2. The express of feature matrix X.

	F1		F2		F3		F4	F5
w	F1PA	F1PD	F2RARE	F2Ac02	F3RARE	F3Ac02	F4NN	F51
W(1)	1	0	0	2	1	0	2	2
W(2)	0	1	1	0	0	0	4	5
W(3)	3	1	2	0	1	0	5	3

tional vocabulary body also includes emotional polarity and strength of the two dimensions. In addition, there are some secondary emotion classifications in the emotional vocabulary. This article does not distinguish between the master and slave emotional vocabulary but aims to merge, examples are:

Among them, sw represents an emotional word, $Emotion \in \{PA, PE, ..., NL, PC\}$, emotion(sw) represented emotion class lexical items sw; $Strength \in [1,9]$, which represented the emotional intensity. Among them, 1 represents the smallest emotion, 9 represents the maximum intensity $Polarity \in \{0,1,2,3\}$, represents neutral, compliment, derogatory and both appraise to justice. In the emotional Vocabulary body, the largest number is commendatory vocabulary and the minimal number is the dual nature word, with a number of 78.

Extracting a set of features 5 herein, the following example of feature matrix X is in Table **2**.

1) Feature set F1 represents the number of Vocabulary in different microblogging w emotional categories, the calculation formula is as:

$$F1_{em}^{(w)} = \sum_{i=1}^{d} I(Emotion(w_i) = em) \,;\, em \in \ Emotion \tag{4}$$

Among them, d represents the number of lexical items of microblogging w, I (dis) represents the indicator function, when the discriminant dis is true, I (dis) equals to 1, and 0 otherwise. Feature set F1 has 21 features, corresponding 21 small classes of emotional Vocabulary ontology

2) Feature set F2 represents different Vocabulary word Lin microblogging number:

$$F2_{syn}^{(w)} = \sum_{i=1}^{n} I(Syno(w_i) = syn) \,;\, syn \in \ Syno \tag{5}$$

Among them Syno (w) represents a synonym code for the word lexical items w. Syno represents Cilin coding set F2 synonyms used, is a collection of coding syn which meets the following criteria:

$$\sum_{w \in trainset} \sum_{i=1}^{d} (I(w_i) = syn) \geq \alpha \tag{6}$$

Among them, α represents a threshold value used for the feature selection and will be replaced by RARE if its frequency of occurrence in the training set is less than α synonyms Cilin alternative encoding such as the feature set of the first column F2 and F3 in Table **3**. α can effectively inhibit the training process of over-fitting, while improving the efficiency of the algorithm, Its set value can be obtained by cross-validation or manual settings and α is set to 5 in this paper.

3) Feature group F3 represents the number of emotional vocabulary words in microblogging different forest:

$$F3_{syn}^{(sw)} = \sum_{i=1}^{d} I(Syno(sw_i) = syn) \\ \& \& I(Emotion(sw_i) \in \ Emotion) \tag{7}$$

Among them, sw represents emotional vocabulary, namely the emotional Vocabulary ontology vocabulary. F3 can also set the feature selection threshold, which is the same with F2, using the RARE emotional words to replace the forest code of low frequency words. This article will manually set the threshold α of F3 as 2.

4) Feature set F4 represents the number of different lexical items in microblogging, using speech tagging results,

speech set [10], for example, $F4_{NN}^{(w)} = 2$ represents microblogging w having two common nouns, $F4_{IJ}^{(w)} = 1$ indicating that there is an interjection in w. In order to make the program better for robustness, we also join the RARE features in F4 to prevent other parts of speech due to the other parts of speech have abnormalities in pre-treatment.

5) F5 represents the number of different polarity of lexical items in microblogging, such as: $F_1^{(w)} = 6$ representing w has six neutral words.

5 sets of features characteristic matrix X contains respectively 21,1348,710,35, 4 feature, the total number of features is 2118.

3.3. Model Optimization

The first step in using a supervised learning emotional classification is to make a pre-treatment for an annotation data, getting the word sequence W and POS sequence T in microblogging. Then use the above feature extraction method to extract different characteristics from the W and T and merge into the feature vector X. For X-normalized, this article features group granularity and different characteristics of the group are individually normalized; Then X split to form the training set X_{train} and validation set X_{val}. The label value is converted to the vector form, such as $t = 2$ converts to $t = [0,1,0,0,0,0,0]$. The form of X and t are as follows:

X=$X1,X2,\ldots XnT,X\in$ ℝ$n\times m$.

$$t=t(1),t(2),\ldots,t(n)\ \mathrm{T},t\in \mathbb{R}n\times m \tag{8}$$

Among them, n is the number of marked centralized microblogging, m is the characteristic dimension, K is the number of categories. After divided into a training set, select the learning model for optimization. For feedforward neural networks, the process is as follows:

1) Select the parameter *d* and *iter* to train the neural network and set the interval of the model parameters according to the forecast results, selecting the optimal parameters. In addition to the parameters given in the code, there is also the cost function of the regularization parameter λ, which is used to determine the error and parameter values at a high cost.

In addition to the parameters given in the code, there are also the cost functions of the regularization parameter λ, which are used to determine the error right at the cost of heavy. This article is not to get the parameter λ by cross-validation, but to set λ to 1 manually. Besides, the paper chose the smaller parameter d and iter because the feature matrix X is a sparse matrix, while fewer hidden layer nodes can improve the efficiency of learning to a great extent. In addition, because of a limited set of an annotation scale, more hidden layer nodes are likely to fit the training process.

2) For each set of parameters, we need to run round full training and prediction to get cross-validation error corresponding to this set of parameters. When training the neural network, the first step is that random initialization parameters need to be optimized, è $\in [-0.1, 0.1]$ this article does not conduct random initialization times to get global optimal solution. After initialization θ, run two iterations to train the neural network, which are the two iterations to run the neural network is trained, which are the training times and each feature vector respectively. After training, we get the output of neural network *yout*∈ℝ*0.2*$\times m \times 8$, and then calculate

$$y_{d,iter} = \arg\max_{i\in[1,8]}(y_{out}^{(i)}) \tag{9}$$

In the cross-validation phase, the accuracy of forecasting can take the correct proportion of predictions and you can also take F_β value by emotion classification.

3) After selecting parameters d and iter, use an annotation set X to train the neural network and get the optimized parameter $\theta^{(1)}$ and $\theta^{(2)}$. We do the same data pre-processing and feature selection towards the test data, using neural network classification and the output of the model parameters are used to predict the results of y, the end of the algorithm.

According to the above description, model optimization algorithm for feedforward neural networks is as:

Inputs:	Training set **S,t**
Initial:	Randomly initial **è**
Define:	$[\mathbf{X_{train}}, \mathbf{X_{val}}] = feature(\mathbf{S})$
Algorithm:	**for** *d* in {25, 50, 80, …}, *iter* in {50, 100, 150,…}: **For** *i* in [1,*iter*], ***x*** in $\mathbf{X}_{\mathrm{train}}$ **è** = *adjust*(**x**,**è**,**t**,*d*) $\mathbf{y_{d,iter}}$ = predict($\mathbf{X_{val}}$,**è**) **end for** **end for** $[d,\ iter] = \arg\min_{d,iter} error(\mathbf{y_{d,iter}}, \mathbf{y_{gold}})$ **è** = *train*(**X**,**è**,**t**, *d*, *iter*) *y* = *predict*($\mathbf{X_{test}}$,$\mathbf{è^{(1)}}$,$\mathbf{è^{(2)}}$)
Output:	**è**, *d*, **y**

Among them, S represents marked microblogging corpus, $\mathbf{X_{train}}, \mathbf{X_{val}}$ represents the training set and the cross-validation set respectively; $t \in [1,8]$ represents artificial label microblogging emotions, Among them, 1-7 represents seven Class emotions, 8 represents no emotion None; d represents hidden layer nodes and iter represents the number of iterations; adjust represents the adjustment process of parameters *θ,* which is the error back transfer process; train represents a training process.

For SVM model, the optimized process is relatively straightforward. Firstly, the target value of the processing target set of training data, the target data ($t = k$) is set as t=1 and the remaining set is $t = -1/(k-1)$ when K is in training class model. Then set the kernel function and the probability of output for each category model optimization. In the prediction process of the new data, input the new data to the model of each category to obtain $y_k(x)$ and calculate an emotion classification.

4. EXPERIMENTS

Data used in this experiment from the NLP & CC 2013 evaluation tasks: the recognition of Chinese micro-blog sentiment is the overall feeling of microblogging identification. The subject is 4000 marked microblogs which has 371,697 words in total. The extracted feature matrix $X\in\mathbb{R}$-4000×2118, the label resulting matrix $t\in\mathbb{R}$-4000×1, the chosen parameters d and the different combinations of iter are $d\in\{25,50,80\}$ and $iter\in\{10,20,...,150\}$ respectively.

In this paper, the prediction accuracy of cross-validation and test data using the macro and micro average value Fβ, β is an important factor for regulating the right precision and the recall rate of weight, macro average means the same weight of each emotion classification, while the micro does the same weight of each microblogging.

The macro average is calculated as follows:

$$P\_macro=\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n}\frac{I(\mathbf{t}^{(j)})\,\&\&\,I(\mathbf{y}^{(j)}=j)}{I(\mathbf{y}^{(j)}=j)+c} \tag{10}$$

$$R\_macro=\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n}\frac{I(\mathbf{y}^{(j)})\,\&\&\,I(\mathbf{t}^{(j)}=j)}{I(\mathbf{t}^{(j)}=j)+c} \tag{11}$$

$$F_{\beta}\_macro=\frac{(\beta^2+1)\times P\_macro\times R\_macro}{\beta^2\times P\_macro+R\_macro} \tag{12}$$

Among these, t is the dimension values, y is the predicted value; n is the size of the prediction data, K is the number of categories, where the categories are not included None; c is the smoothing factor, take the smaller positive value used for the frequency of emotion for Appearing zero exception, such as fear and surprise. Running cross-validation using macro average value to obtained F1 learning curve shown in Fig. (**2**).

The horizontal axis represents the times of iterations, the vertical axis represents the macro average F1_macro, t25, v25, respectively, the training set when the number of hidden layer nodes d = 25, cross-validation set F1_macro value. As can be seen from Fig. (**2**), there are different degrees of overfitting in neural network of the different hidden layer nodes, mainly due to: 1) the number of features microblogging close to the scale of the mark set, leading to the rapid fitting training sets of neural network. 2) The emotional words, parts of speech and other performance characteristics selected by experiment cannot fully express microblogging emotion, especially the emotion with a frequency of smaller classes. 3) The cost function of the regularization parameter λ is smaller than the regular one. Moreover, when the data and characteristics are not sufficient, only adjusting λ is not enough to improve the F1 value of the cross-validation.

In cross-validation, besides the macro average, Fig. (**3**) shows the value F1 of each emotion classification, where the horizontal axis represents the emotional classification and the vertical axis represents F1 values. As we can see from the figure, the value F1 of the training set is significantly higher than the cross-validation set and the value F1 of cross-validation has a larger fluctuation, which is due to the lower frequency of part (some) emotion which has an impact on the predicted recall rate, thereby reducing the value of F1.

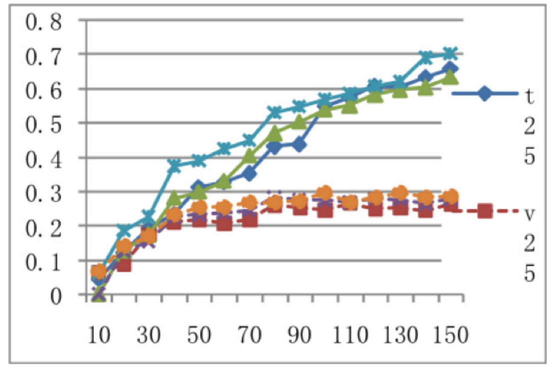
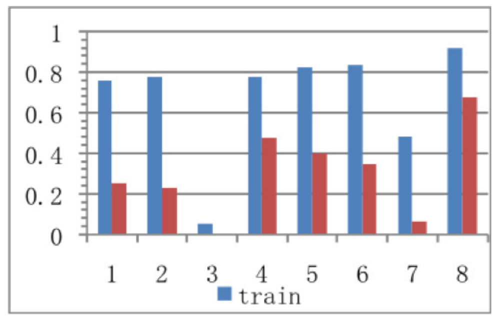
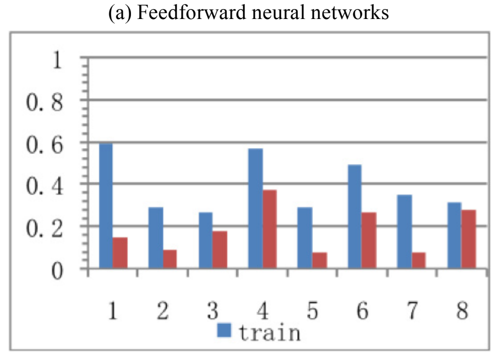


Fig. (2). Neural network model of learning curve.



(a) Feedforward neural networks



(b) SVM

Fig. (3). F1 value of each class of emotion.

In addition, the performance of feedforward neural network used in this article used on microblogging sentiment classification is superior to SVM. Comparing the accuracy of different types can be found, SVM in predicting the performance of low-frequency emotions is better, the fit is lower than the neural network. After the completion of cross-validation, using the entire set of labels to train the neural network and the prediction results of the test set is shown in Table **3**.

As can be seen from Table **3**, the results of feed forward neural networks to predict the correct rate was significantly higher than that of the recall; Emotionally judged sentences have some errors, which directly affects the recall of mood classification because an emotional sentence has been identified as no emotion None; Micro average of emotional categories is greater than the average value in the classification of F1. This is because the calculation of the average gives the macro for each type of emotion value of the same weight, therefore, fluctuations in the value of each emotion class F1

Table 3. Test set predictions.

Reviews Object	Correct Rate	Recall Rate	F1 Value
Subjective judgment sentence	0.7479	0.6335	0.6866
Average macro sentiment classification	0.2653	0.2201	0.2406
Micro-average sentiment classification	0.3655	0.309	0.3349

Table 4. Micro-average results compare.

Resource Name	Method	Number of Categories	Accuracy
Facebook Status	Maximum Entropy	4	67%
Chinese comments	ANN	6	43.6%
Experience project	Bayesian Network	5	54.5%
NLP&CC2013	Out Model	7	34.12%

will reduce the value of the macro average F1, which is showed in Fig. (**3**) that the cross-validation of emotion F1 = 0; On the other hand, smaller emotional class of F1 value appear lower in frequency, and therefore it slightly lowers the average weights, making the micro-average F1 sentiment classification to increase.

Table **4** shows the accuracy of several existing sentiment classification methods, respectively, for Facebook, Chinese comments, Experience project and microblogging text, etc. As can be seen from the table, comparing to the relative polarity analysis, multi-class emotional classification accuracy is low and inversely proportional to the number of categories.

5. CONCLUSION

In this paper, microblogging sentiment was classified by supervised learning algorithms, and it achieved certain results. However, compared to voice, microblogging text is not with the important feature of emotional expression intonations and speeds, etc., A lot of rhetoric in the text brings big challenges to analyse emotional at the same time. Therefore, the application of supervised learning algorithms on sentiment classification still requires a lot more label sets and deeper emotional features. For micro-blog sentiment analysis, the intrinsic characteristics of its text are also extremely important, in addition, the characteristics of microblogging users, microblogging theme, affected the sentiment classification of microblogging text in varying degrees.

From the experimental results, there are some over fitting phenomenon about the learning model in this paper and microblogging sentiment classification accuracy is lower. Therefore, in future research, it will not only study the discovery of emotional language features and emotional label sets, but also pay attention to the use of semi-supervised learning method to extract text microblogging of emotional characteristics, such as the micro-blog themes and context, in order to improve the accuracy of the emotion classification.

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by Natural Science Foundation of Shandong Province (Project No. 2009ZRB019PF).

REFERENCES

[1] M. Tsytsarau, and T. Palpanas, "Survey on mining subjective data on the web," *Data Mining Knowl. Discov*, vol. 24, no. 3, pp. 478-514, 2012.

[2] B. Pang, and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, no.1-2, pp.1-135, 2008.

[3] A. Go, R. Bhayani, and L. Huang, "*Twitter Sentiment Classification Using Distant Supervision*," CS224N Project Report, Stanford, pp.1-12, 2009.

[4] A. Pak, and P. Paroubek, "*Twitter as a Corpus for Sentiment Analysis and Opinion Mining*," *LREC*, 2010.

[5] C. M. Bishop, "*Pattern Recognition and Machine Learning (Information Science and Statistics)*," Springer-Verlag: New York, Inc., 2006.

[6] C. Chang, and C. Lin, "LIBSVM: A Library for Support Vector Ma-chines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 21-27, 2011.

[7] H. Tseng, "A conditional random field word segmenter," In: *Fourth SIGHAN Workshop on Chinese Language Processing*, 2005.

[8] J. Ilonen, J. Kamarainen, J. Lampinen, "Differential evolution train-ing algorithm for feed-forward neural networks," *Neural. Process Lett.*, vol. 17, no.1, pp. 93-105, 2003.

[9] I. W. Tsang, J. T. Kwok, P. Cheung, "Core vector machines: Fast SVM training on very large data sets," *J. Machine Learn. Res.*, vol. 6, pp. 363-392, 2005.

[10] S. Aixin, and L. Ee-Peng, "Hierarchical text classification and evaluation: Data Mining", In: *ICDM Proceedings IEEE International Conference*, 2001.

[11] J. K. Ahkter, and S. Soria, "*Sentiment Analysis: Facebook Status messag-es*," Technical report, Stanford University. Final Project CS224N, 2010.

Received: September 16, 2014 Revised: December 23, 2014 Accepted: December 31, 2014