JCSE Journal of Computing Science and Engineering

Regular Paper
Journal of Computing Science and Engineering,
Vol. 10, No. 4, December 2016, pp. 103-110

Use of Word Clustering to Improve Emotion Recognition from Short Text

Shuai Yuan
National Engineering Research Center for E-Learning, Central China Normal University, Wuhan, China
yuanshuai.china@163.com

Huan Huang*
School of Education, South Central University for Nationalities, Wuhan, China
huanghuan@mail.scuec.edu.cn

Linjing Wu
School of Educational Information Technology, Central China Normal University, Wuhan, China
wlj_sz@126.com

Abstract
Emotion recognition is an important component of affective computing, and is significant in the implementation of natural and friendly human-computer interaction. An effective approach to recognizing emotion from text is based on a machine learning technique, which deals with emotion recognition as a classification problem. However, in emotion recognition, the texts involved are usually very short, leaving a very large, sparse feature space, which decreases the performance of emotion classification. This paper proposes to resolve the problem of feature sparseness, and largely improve the emotion recognition performance from short texts by doing the following: representing short texts with word cluster features, offering a novel word clustering algorithm, and using a new feature weighting scheme. Emotion classification experiments were performed with different features and weighting schemes on a publicly available dataset. The experimental results suggest that the word cluster features and the proposed weighting scheme can partly resolve problems with feature sparseness and emotion recognition performance.

Category: Smart and intelligent computing

Keywords: Emotion recognition; Affective computing; Word clustering

I. INTRODUCTION

Since Minsky proposed that intelligent machines should have affective capability, making a machine that understands and expresses emotions has become a hot research topic in the artificial intelligence community. Recently, Picard proposed a concept called "affective computing", and defined it as computing that relates to, arises from, or influences emotions [1]. Emotion recognition is one of the most important components of affective computing. It has been implemented in many kinds of media, and includes facial expression, speech characteris-

Open Access http://dx.doi.org/10.5626/JCSE.2016.10.4.103 http://jcse.kiise.org

Received 13 July 2016; **Revised** 21 November 2016; **Accepted** 24 November 2016
*Corresponding Author

 pISSN: 1976-4677 eISSN: 2093-8020

tics, and physiological signals [2]. However, text is still the most popular communication medium at present, and for many applications it is still important to recognize emotions from texts. For example, if an intelligent dialogue tutor can recognize emotions from a user's dialogue, it can give more adaptive and human-like feedback. In text-to-speech synthesis, if a system can recognize emotions from text, the system can generate speech that sounds more natural and engaging. In short, recognizing emotions from texts is significant for the implementation of natural and engaging human-computer interaction.

Currently, the approaches that recognize emotions from text can be categorized into two groups: knowledge-based approach and machine learning-based approach. Knowledge-based approach usually begins by constructing an affective lexicon, and then combining some syntactic and semantic rules to recognize emotions from text. These methods perform well in some special domains. However, their performances are highly dependent on the quality of their affective lexicon and syntactic rules. Furthermore, it is expensive and complex to build high quality affective lexicons and syntactic rules. The machine learning-based approach considers emotion recognition as a classification problem. It extracts features from emotion-labeled texts, and represents each text as a feature vector. Usually, the features are thousands of words or n-grams that occur in the labeled texts. Machine learning algorithms, such as Support Vector Machine (SVM), Naïve Bayes, and Maximum Entropy, are used to train an emotion classifier. Although these methods require building an emotion-labeled corpus beforehand, it is easier than manually constructing an affective lexicon and syntactic rules. Because the emotion classifier has been previously trained, these approaches are more efficient when recognizing emotions from a new text. There are many efficient classification algorithms available, and machine learning-based approach appears to be very simple. However, existing works have suggested that these algorithms do not perform as well as they do for topic text classification [3-5]. One important reason is that the processed texts in emotion recognition are usually very short, leading to large, sparse feature space [6, 7].

In order to improve the performance of emotion recognition in short texts, this paper focuses on the feature sparseness problem. To resolve this problem, we proposed to use word clusters instead of words that were usually used as features. Word clusters or concepts have been used as features in text classification. However, as far as we know, they have not been used in emotion recognition. Furthermore, we proposed a detailed word clustering algorithm along with a novel weighting scheme. The weighting scheme can measure the feature values more accurately. More importantly, our experiments suggest that this approach can largely improve the effectiveness of emotion recognition from short text.

II. RELATED WORK

Currently, there are a lot of literatures related to detecting sentiment or emotion in text. According to their research perspectives, these literatures can be divided into two categories. Most literatures focus from a commercial perspective, on sentiment detection in reviews from users, such as product reviews and movie reviews [8-10]. They attempt to identify the polarity of sentiment in text: positive, negative, or neutral. We generally name these researches as sentiment analysis or opinion mining. Recently increasing literatures are focused on detecting multiple emotions in social web text, such as the messages in social network sites, blogs and discussion forms, from the perspective of emotional psychology [5, 11, 12]. They attempt to identify various emotions in the text, usually focusing on the six so-called basic emotions of anger, disgust, fear, joy, sadness, and surprise [13]. We generally call these researches as fine-grained sentiment detection or emotion recognition. In this paper, we focus on detecting emotions in short text. Here, we will only review literatures targeting multiple emotion recognition.

Approaches to recognizing emotions in text can be classified into two groups. One group of emotion recognition approach is based on affective knowledge, such as emotional keywords, syntactic rules, and semantic rules. Elliott's affective reasoner used 198 unambiguous emotional words with some modifiers, clue phrases, and heuristic rules to recognize the emotions expressed in text [14]. Subasic and Huettner [11] built a fuzzy affective lexicon, and further proposed a novel, convenient fusion of natural-language processing and fuzzy logic techniques for analyzing affect content in free text. Boucouvalas and Zhe [15] used a tagged dictionary to extract the emotional words in a sentence, and applied a parser to identify the associated objects, following which he employed syntactic rules to recognize emotions in real-time dialogues. Liu et al. [16] proposed an approach to learning a small society of linguistic affect models from a large-scale real-world generic knowledge base, the Open Mind Common Sense (OMCS). Neviarouskaya et al. [12] created an affect database, and based on this database, they proposed a rule-based affect analysis model to recognize emotions expressed by text messages. Although approaches based on affective knowledge perform well in some specialized domains, they hugely depend on a previously created affective lexicon and rules. The generalization capability of these approaches is not very good, because it is difficult and expensive to create an emotion lexicon and related rules that are appropriate for all domains.

Another group of emotion recognition approach is based on statistical machine learning techniques, including supervised learning methods and unsupervised learning methods. These approaches deal with emotion recognition as a classification problem. Alm et al. [3] proposed a supervised learning method to classify the sentences of fairy tales into three categories: positive, negative, and neutral. In [17], she further refined and improved the feature set, and presented the experimental results of fine-grained emotion classification. Aman and Szpakowicz [5] constructed a corpus of blog posts. Based on the corpus, they used the obvious emotion words present in the sentence as features, and applied Naïve Bayes and SVM to classify sentences from blog posts into emotional or non-emotional categories. They also utilized unigrams and emotional words as features to classify the sentences into six basic emotions. To recognize emotions of news headlines, Katz et al. [18] used a supervised learning method based on a unigram model, and Strapparava and Mihalcea [19] proposed several semantic methods using the Latent Semantic Analysis (LSA) technique. Machine learning-based approaches are more efficient and more generalized, since they avoid the difficult task of creating emotion lexicons and related rules. However, their commonly used text categorization methods encounter significant performance degradation when applied to emotion recognition. This is because of the intrinsic nature of text characteristics, and because the statistical machine learning algorithms require longer input for reasonable accuracy.

III. EMOTION RECOGNITION BASED ON WORD CLUSTERING

The framework for emotion recognition based on word clustering is shown in Fig. 1.

This framework consists of three basic phases: the word clustering phase, the training phase, and the classification phase. In the word clustering phase, all the effective words in the emotion-labeled corpus are clustered by their semantic similarity. In the training phase, each text is represented as a feature vector, and the SVM learning

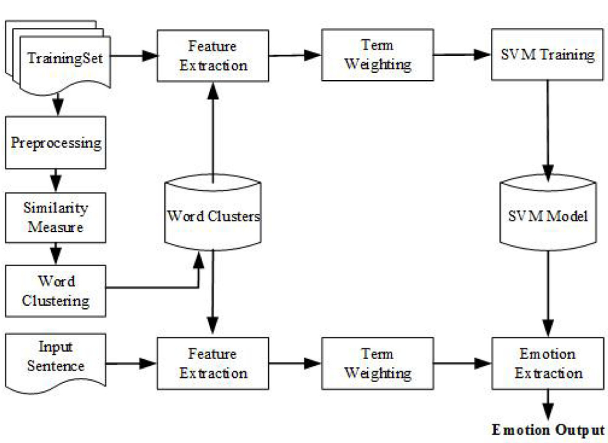


Fig. 1. The framework of emotion recognition based on word clustering.

algorithm is applied to train an emotion classifier. In the classification phase, an unlabeled sentence is also represented as a feature vector and is classified into a basic emotion by the emotion classifier. In the following subsection, we will illustrate key steps of the framework.

A. Semantic Similarity Measure

Semantic similarity measure is important in the process of word clustering. Researchers have proposed many different approaches to determining the conceptual similarity in taxonomy. These can be categorized as information content approaches and conceptual distance approaches.

Our semantic similarity measure is based on the combined approach proposed by Jiang and Conrath [20], which was derived from the conceptual distance approach by adding information content as a decision factor. We used WordNet as the taxonomy to find out the shortest path that links two concepts. When calculating the weighting value of the edge link between two adjacent concepts, only the link strength is considered. The link strength is defined as the difference between a child concept and its parent concept.

In light of the above description, assume that there are two words w_1 and w_2, and $Dist(w_1, w_2)$ represents the semantic similarity/distance between w_1 and w_2. The semantic similarity/distance can be calculated as follows:

$$Dist(w_1, w_2) = \min(IC(c_1) + IC(c_2) - 2 \times IC(LSuper(c_1, c_2))) \quad (1)$$

where c_1, c_2 denote a possible sense of word w_1 and w_2; $LSuper(c_1, c_2)$ is the lowest super-ordinate of c_1 and c_2; $IC(c_i)$ is the information content of sense/concept c_i.

Following the notation in information theory, the information content of sense/concept c can be quantified as follows:

$$IC(c) = -log\left(\frac{\sum_{w \in words(c)} freq(w)}{N}\right) \quad (2)$$

where $words(c)$ is the set of words representing concept c, $freq(w)$ denotes the frequency of word w in corpus, and N denotes the total frequency of the corpus. The frequencies of concepts are estimated using the frequencies from a universal semantic concordance, SemCor [21], a semantically tagged corpus consisting of 100 passages from the Brown Corpus.

B. Word Clustering Algorithm

Though there are many word clustering algorithms, some are more difficult than others to use in emotion recognition. First, our semantic similarity measure is based on the hierarchical structure in WordNet, in which only nouns and verbs are organized hierarchically. This means our introduced semantic similarity measure can only be applied to noun pairs and verb pairs. Second, because emotion words are important, it is more effective to group them by their part-of-speech and emotion category. Therefore, we categorize the words as content words and emotion words, and use different methods to cluster them.

For the emotion words, we use the emotion word lists [22] to group them into 24 clusters. The words in each cluster have the same part-of-speech and the same emotion category. For the content words, we propose an agglomerative clustering algorithm to group similar words into a cluster. This algorithm groups each word into a separate cluster, and successively merges clusters until a stopping criterion is satisfied. The proposed word clustering algorithm includes two main processes. First, it calculates the semantic similarity of each pair of words. If their similarity is more than the predefined threshold, α, they are the synonyms for each other. The word and its similar words are then initialized as a separate cluster. Second, it discovers the two nearest word clusters and predicts whether their similarity is more than the predefined threshold, β. If it is, the two nearest word clusters are merged, otherwise the procedure is over. The similarity measure of word clusters is quantified as follows:

$$sim(wc_i, wc_j) = \frac{|wc_i \cap wc_j|}{min(|wc_i|, |wc_j|)} \tag{3}$$

where $sim(wc_i, wc_j)$ denotes the similarity of word clusters wc_i and wc_j; $wc_i \cap wc_j$ is the intersection of word clusters wc_i and wc_j; $|wc_i|$ is the number of words in cluster wc_i.

C. Text Representation based on Word Clusters

A text must be represented in computational form before it can be processed and analyzed. The vector space model is one of the most commonly employed models in text mining. It selects certain features to build a high dimensional vector space, and every text then becomes a vector in this space. Usually, the features are words or *n*-grams that occur in the corpus, and the feature values are calculated by their Presence, Absolute Frequencies, Relative Frequencies, or TFIDF (term frequency inverse document frequency) values. These methods cannot be applied directly when the features are word clusters. We propose a new weighting scheme for word clusters, which is based on the discrimination degree of clusters and the representation degree of words.

The representation degree of a word is the extent to which a word represents a word cluster. A word is closer to the center of a word cluster whose representation degree is stronger. If a word is closer to the center of a cluster, the variance of the semantic distances between the word and other words will be smaller. According to this notion, we assume that a word cluster c is represented as $c = (w_1, w_2, \ldots, w_i, \ldots, w_n)$, and w_i denotes one of the words belonging to cluster c. The representation degree of the word w for cluster c can be quantified as follows:

$$r = \frac{1}{STDEV(Dist(w, w_i))} \tag{4}$$

where $Dist(w, w_i)$ denotes the semantic similarity/distance between word w and word w_i, and $STDEV(Dist(w, w_i))$ denotes the standard variance of $Dist(w, w_i)$.

The discrimination degree of a word cluster is its capability to discriminate between different kinds of emotions. We assume that the distribution of a word in different emotions is imbalanced, which makes the discrimination degree stronger. According to the notation of information theory, the discrimination degree of a word can be quantified as the reciprocal of its information entropy. Assume that a word cluster c including n words is represented as $c = (w_1, w_2, \ldots, w_i, \ldots, w_n)$, and the m emotion categories is represented as $e = (e_1, e_2, \ldots, e_j, \ldots, e_m)$. The discrimination degree of a word cluster can be quantified as follows:

$$d = \frac{1}{n}\Sigma_{i=1}^{n}\frac{1}{\Sigma_{j=1}^{m} - log(P(w_i|e_j))} \tag{5}$$

where d is the discrimination degree of the word cluster c; $P(w_i|e_j)$ denotes the conditional probability of encountering the word w_i given the texts belonging to emotion e_j.

Based on the representation degree of words and the discrimination degree of word clusters, the weighting value of the word cluster wc_i in text t can be quantified as follows:

$$weight_i = d_i \times \Sigma_{i=1}^{n} k \times r_{ij} \tag{6}$$

where d_i denotes the discrimination degree of word cluster wc_i; r_{ij} denotes the representation degree of a word w_j in the cluster wc_i; k is the number of occurrences of word w_j in the text t; n is the total of words in cluster wc_i.

D. Classifier based on SVM

The SVM was introduced by Vapnik and colleagues, and has been very successful in text categorization and many other areas of application. In text categorization, it has been proved that the SVM has better performance than other statistical and machine learning approaches [13]. Furthermore, Dumais et al. [23] have reported that linear SVM performs better than nonlinear SVM in text categorization.

In this paper, we use the linear SVM implemented in the LIBLINEAR package [24] to train a multi-class emotion classifier. The multi-class emotion classifier is implemented in the one-versus-all strategy. If we suppose there are K (K>2) classes, this strategy would then construct K binary SVMs. The original training data are separated into two classes. The first is the original class of K classes; all others are separated into the second class. Each binary SVM is then trained on the dataset. A new text sample is classified to a class whose binary SVM has the largest values.

IV. EXPERIMENTS AND EVALUATION

We conducted two group experiments on a publicly available dataset. In the first group, we studied the impact of varying parameters in the word clustering algorithm. In the second group, we compared the performance of emotion recognition when we used different features and weighting schemes.

A. Dataset

Our experiments were conducted on the publicly available corpus, News Headlines. It was developed by Strapparava and Mihalcea [22] for the SemEval 2007 task on "Affective Text". The task focused on the emotion classification of news headlines, including the emotion annotation subtask and the valence labeling subtask. For the emotion annotation subtask, each news headline was annotated with six integers between 0 and 100, which indicate the different degrees of six basic emotions proposed by Ekman [25]. For the valence labeling subtask, the interval for the valence annotation was set to [-100, 100]. The value 0 indicates the news headline is neutral, 100 indicates the news headline is highly positive, and -100 indicates the news headline is highly negative. The News Headlines corpus consisted of two datasets: a development dataset, including 250 annotated news headlines, and a test dataset with 1,000 annotated news headlines. In our experiment, we combined the two datasets and classified each news headline to the emotion category with the highest degree. The distribution of emotion classes is presented in Table 1.

B. Evaluation Criteria

In this paper, we consider emotion recognition as a multi-class classification problem. Thus, we use precision, recall, and F1 measure to evaluate the performance of the emotion classifier for each emotion category, and use the macro-average and micro-average to evaluate the performance of the emotion classifier for all classes.

The macro-average equally weighs all classes, regardless of how many texts belong to them, which can be quantified as follows:

$$MacroAVG = \frac{X_1 + X_2 + \cdots + X_N}{N} \tag{7}$$

where X_i is the precision, recall, or F1 value for an emotion category.

The micro-average equally weighs all the documents, thus favoring the performance on common classes. Since we suppose a text belongs to only one emotion category, hence the micro-averages of precision, recall and F1 measure are equal, and they are also equal in overall accuracy. They can be qualified as follows:

$$MicroAVG = \frac{C_1 + C_2 + \cdots + C_N}{M} \tag{8}$$

where C_i is the number of texts that are correctly classified into the emotion category by the classifier. N is the number of emotion classes, and M is the total of texts.

Table 1. Distribution of emotion classes in the dataset

Angry	Disgusted	Fearful	Happy	Sad	Surprised
91	42	194	441	265	219

C. Results and Analysis

According to the procedures of emotion recognition, we conducted the following two group experiments on the above dataset. All the experiments are performed using 5-fold cross validation, with 80% train and 20% test data. When the LIBLINEAR package is used to train the emotion classifier, the parameters are set to s = 1, c = 1.6. Besides, we also compared our results with those obtained by existing systems.

1) Parameter Determining

When word clusters are used as features, the results of word clustering are important for the performance of emotion recognition. As mentioned in Section III-A, we categorize the words as emotion words and content words. We grouped the emotion words into 24 clusters based on their part-of-speech and emotion category. Content words were grouped using the algorithm introduced before to cluster them. There are two parameters, α and β, which determine the results of word clustering, and need to be pre-assigned.

To discover the optimal values for emotion classification, we conducted an experiment with the weighting scheme Presence. When the Presence scheme is applied to word cluster features, a feature's weighting value is 1 if one of the words in the cluster is presented in the text; otherwise the weighting value is 0. In the experiment, we first set β to 0.45 by experience, and varied α from 0.45 to 0.70 to find the optimal α value. Next, we set α to the optimal value, and varied β from 0.45 to 0.70 to find out the optimal β value. The experimental results for the macro-average of F1 values and micro-average of F1 values conducted are shown in following tables.

Table 2. The results with different α on News Headlines Corpus

β = 0.45	α					
	0.45	**0.50**	**0.55**	**0.60**	**0.65**	**0.70**
Mic-F1	0.4552	0.4400	0.4680	0.4497	0.4544	0.4504
Mac-F1	0.3784	0.3571	0.3858	0.3632	0.3717	0.3750

Mic-F1: micro-average of F1 values, Mac-F1: macro-average of F1 values.

Table 3. The results with different β on News Headlines Corpus

α = 0.55	β					
	0.45	**0.50**	**0.55**	**0.60**	**0.65**	**0.70**
Mic-F1	0.4680	0.4704	0.4552	0.4760	0.4576	0.4744
Mac-F1	0.3858	0.3888	0.3711	0.3954	0.3689	0.3906

Mic-F1: micro-average of F1 values, Mac-F1: macro-average of F1 values.

According to Tables 2 and 3, we find that when α = 0.55 and β = 0.60, the emotion classification performance is best on the News Headlines Corpus. Therefore, in the next experiment, when the word clusters are used as features, we use these optimal values.

2) Comparison of Different Features and Weighting Schemes

Features and weighting schemes are important for the performance of emotion classifiers. To test the validity of using word clusters as features, we compared the results from using emotion words, unigrams and word clusters as features with the Presence scheme. To validate the proposed weighting scheme for word clusters, we also compared our weighting scheme to the Presence scheme. Wang et al. [26] suggested that the Presence scheme has the best classification results in sentiment classification. Therefore, we decided not to compare our weighting scheme with Absolute Frequency, Relative Frequency, and TFIDF.

In Table 4, E/P represents emotion words as features, and applies the Presence scheme. Analogously, U/P and WC/P represents unigrams and word clusters as features and applies the Presence scheme. WC/D&R represents word clusters as features and applies our weighting scheme, which is based on the discrimination degree of the word cluster and the representation degree of the word.

From Table 4, we see that when we used unigrams instead of emotion words as features, the micro-average and macro-average of precisions, recalls and F1 values are largely improved, which suggest that some non-emotional words are also important for emotion recognition. When we used word clusters instead of unigrams as features, the micro-averages of precisions, recalls and F1 values are enhanced more than 3%, which suggests that the overall accuracy of the emotion classifier was improved. All the macro-averages were also enhanced more than 1%, which suggests that the word cluster feature can improve the generalization capability of the emotion classifier. When we used the proposed weighting scheme instead of the Presence scheme, the micro-average and the macro-average of precisions, recalls and F1 values were further improved slightly. Based on these improvements, we conclude that the word cluster feature and the proposed weighting scheme enhances the overall performance of emotion recognition.

Table 4. The micro-average and the macro-average of precisions, recalls, and F1 values (unit: %)

	Features and weighting schemes			
	E/P	**U/P**	**WC/P**	**WC/D&R**
Mic-P/R/F1	26.57	44.56	47.60	48.57
Mac-P	24.95	39.09	40.64	40.83
Mac-R	18.78	38.20	39.23	39.50
Mac-F1	10.99	38.11	39.54	39.70

Mic: micro-average of precisions, recall and F1 value, Mac-P: macro-average of precisions, Mac-R: macro-average of recalls, Mac-F1: macro-average of F1 values.

Although the overall accuracy of our approach is below 0.5, the existing researches also had a limited success. Table 5 shows the overall results of three systems participating in emotion annotation task of the SemEval-2007: SWAT, UA and UPAR7, and five systems proposed by Strapparava and Mihalcea [19]: WN-affect presence, LSA single word, LSA emotion synset, LSA all emotion words and NB trained on blogs. The results suggest that the task of emotion recognition is difficult. This is probably because the texts analyzed are typically short and some related emotions can exist in one text.

Table 5. The macro-average of precisions, recalls, and F1 values of each system (unit: %)

	Mac-P	**Mac-R**	**Mac-F1**
WN-affect presence	38.32	1.54	4.00
LSA single word	9.88	66.72	16.37
LSA emotion synset	9.20	77.71	13.38
LSA all emotion words	9.77	90.22	17.57
NB trained on blogs	12.04	18.01	13.22
SWAT	19.46	8.61	11.57
UA	17.94	11.26	9.51
UPAR7	27.60	5.68	8.71

Mac-P: macro-average of precisions, Mac-R: macro-average of recalls, Mac-F1: macro-average of F1 values.

V. CONCLUSION AND FUTURE WORKS

In this paper, we proposed a novel emotion recognition approach based on word clustering for short texts. Based on the semantic similarity measure of words, we first proposed a word clustering algorithm. Next, we used word clusters as features, and further proposed a new weighting scheme based on the discrimination degree of word clusters and the representation degree of words. In our experiments, we examined emotion recognition performance on a publicly available dataset when different features and weighting schemes were used. The experimental results showed that 1) using word clusters as features can largely reduce the dimension of feature space; 2) when short texts are represented by word cluster features with the proposed weighting scheme, the emotion classifier performs better for most of the specific emotions; and 3) using word cluster features and the proposed weighting scheme can also improve the whole performance of emotion recognition.

This study also suggests some interesting problems for further exploration. We see from the experimental results that the word cluster features and the proposed weighting scheme are not effective for the emotion 'disgusted'. This may suggest that different features are adaptive for different emotions. Besides, considering the intensity of emotion words may improve emotion recognition performance. In future work, we will attempt to further improve and refine our method and deal with the above problems.

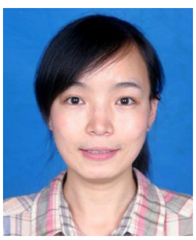
AKNOWLEDGMENTS

This work is supported by the National Nature Science Foundation of China (No. 61272205), and supported by the Educational Science Planning in Hubei Province (No. 2015GB025), and supported by the Fundamental Research Funds for the Central Universities, South-Central University of Nationalities.

REFERENCES

1. R. W. Picard, *Affective Computing*, 1th ed., Cambridge: MIT Press, 1997.
2. R. A. Calvo and S. D'Mello, "Affect detection: an interdisciplinary review of models, methods, and their applications," *IEEE Transactions on Affective Computing*, vol. 1, no. 1, pp. 18-37, 2010.
3. C. O. Alm, D. Roth, and R. Sproat, "Emotions from text: machine learning for text-based emotion prediction," in *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, Vancouver, Canada, 2005, pp. 579-586.
4. S. Aman and S. Szpakowicz, "Identifying expressions of emotion in text," in *Proceedings of the 10th International Conference on Text, Speech and Dialogue*, Plzen, Czech Republic, 2007, pp. 196-205.
5. S. Aman and S. Szpakowicz, "Using Roget's thesaurus for fine-grained emotion recognition," in *Proceedings of the 3rd International Joint Conference on Natural Language Processing*, Hyderabad, India, 2008, pp. 312-318.
6. T. Xu and Q. Peng, "Extended information inference model for unsupervised categorization of web short texts," *Journal of Information Science*, vol. 38, no. 6, pp. 512-531, 2012.
7. L. Gaudette and N. Japkowicz, "Compact features for sentiment analysis," in *Advances in Artificial Intelligence*, *Lecture Notes in Computer Science vol. 6657*, Heidelberg: Springer, pp. 146-157, 2011.
8. B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1-2, pp. 1-135, 2008.
9. B. Liu, *Sentiment Analysis and Opinion Mining*, San Rafael, CA: Morgan & Claypool Publishers, 2012.
10. R. Feldman, "Techniques and applications for sentiment analysis," *Communications of the ACM*, vol. 56, no. 4, pp. 82-89, 2013.
11. P. Subasic and A. Huettner, "Affect analysis of text using fuzzy semantic typing," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 4, pp. 483-496, 2001.
12. A. Neviarouskaya, H. Prendinger, and M. Ishizuka, "Affect analysis model: novel rule-based approach to affect sensing from text," *Natural Language Engineering*, vol. 17, no. 1, pp. 95-135, 2011.
13. T. Joachims, "Text categorization with support vector machines: learning with many relevant features," in *Proceedings of the 10th European Conference on Machine Learning*, Chemnitz, Germany, 1998, pp. 137-142.
14. C. Elliott, "The affective reasoner: a process model of emotions in multi-agent system," Ph.D. dissertation, Northwestern University, Chicago, IL, 1992.
15. A. C. Boucouvalas and X. Zhe, "Text-to-emotion engine for real time internet communication," in *Proceedings of the 3rd International Symposium on Communication Systems, Networks and Digital Signal Processing,* Stafford, VA, 2002, pp. 164-168.
16. H. Liu, H. Lieberman, and T. Selker, "A model of textual affect sensing using real-world knowledge," in *Proceedings of the 8th International Conference on Intelligent User Interfaces,* Miami, FL, 2003, pp. 125-132.
17. C. O. Alm, "Affect in text and speech," Ph.D. dissertation, University of Illinois at Urbana-Champaign, IL, 2008.
18. P. Katz, M. Singleton, and R. Wicentowski, "SWAT-MP: the SemEval-2007 systems for task 5 and task 14," in *Proceedings of the 4th International Workshop on Semantic Evaluations*, Prague, Czech Republic, 2007, pp. 308-313.
19. C. Strapparava and R. Mihalcea, "Learning to identify emotions in text," in *Proceedings of the 23rd Annual ACM Symposium on Applied Computing*, Fortaleza, Brazil, 2008, pp. 1556-1560.
20. J. J. Jiang and D. W. Conrath, "Semantic similarity based on corpus statistics and lexical taxonomy," in *Proceedings of the International Conference on Research in Computational Linguistics,* Taiwan, 1997, pp. 19-33.
21. G. A. Miller, C. Leacock, R. Tengi, and R. T. Bunker, "A

semantic concordance," in *Proceedings of the ARPA Workshop on Human Language Technology,* Stroudsburg, PA, 1993, pp. 303-308.
22. C. Strapparava and R. Mihalcea, "SemEval-2007 task 14: affective text," in *Proceedings of the 4th International Workshop on Semantic Evaluations*, Prague, Czech Republic, 2007, pp. 70-74.
23. S. Dumais, J. Platt, D. Heckerman, and M. Sahami, "Inductive learning algorithms and representations for text categorization," in *Proceedings of the 7th International Conference on Information and Knowledge Management*, Bethesda, MD, 1998, pp. 148-155.
24. R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang, and C. J. Lin, "LIBLINEAR: a library for large linear classification," *Journal of Machine Learning Research*, vol. 9, pp. 1871-1874, 2008.
25. P. Ekman, "Facial expression and emotion," *American Psychologist*, vol. 48, no. 4, pp. 384-392, 1993.
26. S. Wang, D. Li, L. Zhao, and J. Zhang, "Sampling cutting method for imbalanced text sentiment classification based on BRC," *Knowledge-Based Systems*, vol. 37, no. 2, pp. 451-461, 2013.



Shuai Yuan

Shuai Yuan received the Master degree in educational technology from Central China Normal University (CCNU), Wuhan, China, in 2011. She is currently a Ph.D. student in National Engineering Research Center for E-Learning, Central China Normal University. Her major is Management Science and Engineering, and her research interests include social network analysis and automated essay scoring.

Huan Huang

Huan Huang received the Ph.D. degree in educational technology from Central China Normal University (CCNU), Wuhan, China, in 2014. He is currently a lecturer in the School of education, South-Central University for Nationalities. His current research interests include educational data mining and emotion detection.

Linjing Wu

Linjing Wu received the Ph.D. degree in educational technology from Central China Normal University (CCNU), Wuhan, China, in 2013. She is currently a lecturer in the School of Educational Information Technology, Central China Normal University. Her main research interests include knowledge representation, adaptive learning and educational data mining.