

Topic Sentiment Analysis in Chinese News

Ouyang Chunping, Zhou Wen⁺, Yu Ying, Liu Zhiming and Yang Xiaohua

*School of Computer Science and Technology, University of South China, Hunan
Hengyang, 421001, China*

⁺*mrwentian@foxmail.com*

Abstract

Sentiment analysis in news is different from normal text sentiment analysis. News usually have a specific topic, a focus semantic emotion, therefore, this paper, based on the principal of using Emotion Dependency Tuple (EDT) as the basic unit of news emotion analysis, resolves topic sentiment analysis in news into three progressive sub-problem, namely, topic sentence recognition, EDT extraction and topic sentiment analysis. We use an improved TF-IDF and cross entropy to extract feature set of topics. Then, based on space vector model, calculate the topic association of a sentence and extract topic sentence. Finally, we construct topic sentence based on EDT and complete clustering of news topic sentiment. This method is evaluated using COAE2014 dataset, and differential means shows that our results close to the best results. This shows that the topic based EDT could effectively improve performance of sentiment analysis in news.

Keywords: *sentiment analysis, EDT, topic sentence extraction, News context*

1. Introduction

Context sentiment analysis, also known as opinion mining, is the process of mining, analyzing and differentiating users' opinion, hobbies and emotions. It is a cross-domain scientific research field, which includes probabilistic theory, data statistics analysis, computer linguistics, natural language processing, machine learning, information retrieval, ontology and many other related fields. Context sentiment analysis has drawn the attention of many institute and researcher recently due to the important application value.

Sentiment analysis can be categorized as three levels of progressive research task, namely sentiment information extraction, sentiment information classification and sentiment information retrieval and summarization. This paper focuses on sentiment information classification. The objective is to classify context sentiment into positive or negative terms and even specific classes. Depending on the granularity of context analysis, sentiment information classification can be divided into word level, phrase level, sentence level and paragraph level. Currently, there are two main branches in sentiment classification: sentiment knowledge based and feature based. The former sums up and set weights on sentiments or polar words (or language units) based on existing sentiment knowledge base. The latter extracts meaningful class related features from context and apply machine learning method for classification. There have many researches on both sentiment research fields all around the world. Kim *etc.*, [3] uses the first idea by summing up and adding weight to English text to acquire the polar of sentences and paragraphs. Turney [4] used the idea mutual information in unsupervised learning methods and distinguish emotion of article by calculating difference in mutual information of word combination chosen by pre-defined rule and “excellent” and “poor” of seed words. Pang *etc.*, [5] started to use machine learning models to classify English text. There are many researches which are distance learning supervision based on

SVM, reinforced KNN corpus learning, Naive Bayesian (NB) feature learning, etc [6]. Due to language difference, learning method cannot be used directly in Chinese language sentiment analysis. Researchers in China have done research for sentiment analysis based on Chinese. There some researches which was based on HowNet emotion word dictionary [7]. They used semantic relation field and automaton to do emotion classification. Moreover, some researchers use CRF and information gain algorithm combined with feature selection method to achieve sentiment classification [8].

The above methods has made progress in sentiment analysis in both english and chinese, but these researches all consider sentiment analysis as a set of words [9] (bags-of-words). But in reality, people not only use emotion words to express emotions, but also use a certain level of emotion expression structure. Bag-of-words models ignore syntactic and semantic relations in emotion word thus making accuracy low. Also, some researcher present sentiment analysis based on dependency analysis. Matsumoto use dependency of a sentence as SVM features to classify sentiment. Wu [10] use dependency analysis to perform sentiment analysis in comment text. These method using dependency relations has improved the performance in emotion classification.

Dependency Grammar (DG) theory is set up by French linguistic Lucien Tesnière in 1959. He believes that sentence is a complete organized structure, where components are words. Words will generate relations with the adjacent word, which forms a framework of sentence, and express meaning [11]. Most research based on dependency grammar uses dependency structure and high accuracy machine learning analysis method, but does not achieve the transformation from syntactic level to semantic level [12,13]. This type of syntactic dependency analysis will introduce noise unrelated to emotion, but making it more efficient for syntactic analysis using dependency grammar reasonable assumption. News has different characteristics. It has a stable structure, obvious topic and emotion expression concentrate in topic sentence [14]. In news sentiment analysis, if news topic and topic words are not considered, information un-related to topic will be introduced. To avoid unnecessary noise, this paper borrows the idea of using verb in dependency tree to construct sentence framework, and propose a basic structure to describe news emotion: Emotional dependency tuple (EDT) and analyze emotion tendency based on EDT. EDT uses topic feature words as core, with other modification word are attached to core word. We presented a novel sentiment analysis model based on emotional dependency tuple, which regards the topic word of sentence as headword of the emotional dependency tuple, improves the efficiency of the sentiment analysis.

In the remainder of the introduction, we motivate our interest in the task, describe the topic sentiment analysis model for Chinese news, and discuss some examples highlighting some difficulties. In Section 2, we summaries how to accomplish sentiment analysis in Chinese news and which techniques will be applied. Trying to illustrate the key theories in our method, we tempt to refine the task in Section 3 and Section 4. At last, we describe our experiment data and results. The paper ends with a conclusion.

2. Analysis Pipeline for Topic Sentiment of Chinese News

News context data exist in portal sites, blogs and bulletin in large scale. Most of it is sentiment oriented and sentiment analysis can provide important evidence for users to grasp social dynamics and distinguish public status [15, 16]. News report is an important carrier for news events, which has standard grammar, correct syntax, and reasonable adjective modifiers, and needs to justify state clearly 6W (when, where, who, why, how, what) of news events [17]. 6W usually appears in the title, beginning and end of a report. The title of a news report

is considered as the eye of a news event. It is usually a length restricted, single lined statement and contain rich information. So, it is important to enhance the mining of titles, first sentence and first paragraph in order to mine more information.

The main source of text sentiment evidence comes from emotional words. But single sourced emotional word dictionary doesn't record new words, hot words, deformed words and potential emotion words in time, thus leading to limited coverage in emotion evidence [18, 19]. We use emotion word and evaluation word from HowNet as the foundation, and combine it with DUT emotion dictionary after removing neutral words and classifying seven emotion words into positive and negative words. Then it is united with Taiwan University's Chinese emotion word dictionary (NTUSD), Sougou word dictionary's new word potion and basic dictionary [20, 21]. Words in Chinese sentences don't have obvious separators, so lexicon processing is required before text classification. We use NLPiR tool, which is based on HMM, as our text parser. Before parsing, internet words that are domain related and words from the emotion ontology are add to the custom dictionary. First, new words extracted from the whole document are added to the parser dictionary. Then, we do parsing and POS tagging to each sentence to increase accuracy. The results of parsing is as shown: $D_i = \{S_1, S_2, S_3, \dots, S_j, \dots, S_n\}$, where S_j is the j th sentence in document D_i , and $S_j = \{W_1, W_2, \dots, W_k, \dots, W_m\}$, where W_k is the k th word in the sentence S_j . Based on the characteristic of new documents, topic features are first extracted. Then, topic sentence candidate set is then extracted from the document. Later, EDT is extracted from topic sentence to construct EDT analysis model for topic polar distinction. Finally, topic sentence subjective-objective classification is used to select subjective news topic sentiment with priority. Figure 1 shows the detailed process:

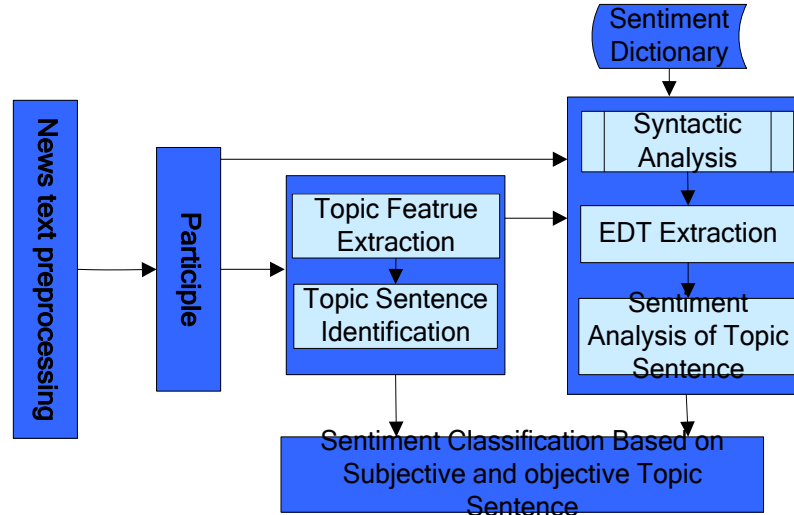


Figure 1. The News Tendency Analysis Flow Chart

News tendency (NT): there are at least two types of news emotion tendency, one is new event tendency (NET), for example, natural disasters, human casualty and property loss are examples of negative news, while technology, sports and culture advance has positive tendency. The other represents the subjective tendency of new reporters (NRT), for example, the news event “discount on high speed train ticket” draws both positive reports and negative

reports. This paper places subjective tendency prediction in higher priority. NET will be the final tendency for ones that doesn't have subjective tendency.

Sentiment topic sentence: sentiment topic sentence must be able to express section topic and general tendency. Therefore, sentiment tendency topic sentence contains two main features: essential keyword and sentiment tendency keyword. Topic keyword is used to summarize the topic of a section, sentiment tendency keyword is used to summarize tendency of section.

Emotion dependency tree (EDT): using core words (CW) as topic features, emotion words (EW) as dependent on core word, degree word (DW) and negative word (NW) series as modifier for core word and emotion word, the basic structure of the emotion expression is constructed, its matching pattern is shown as formula (1).

$$EDT=[*NW/DW][*[*NW/DW]EW]CW[*[*NW/DW]EW] \quad (1)$$

For example, for topic sentence "RouShiJi industry in China is not very good, now our company have to compensate the 45 million now.", "RouShiJi industry" is regarded as core word, "not very good" is regarded as emotion word used to modify core word, and negative word "not" and degree word "very" forms a negative degree structure used to modify emotion word. The conditions for "RouShiJi industry" or "industry" to be core words is that it must exist in topic word set.

Topic sentence of News sentiment classification: uses news event tendency and news emotion tendency as classification foundation. Extract news sentiment topic sentence from news report that is closest to topic, as subjective emotion of news. Extract news event topic sentence from news report that has highest similarity to topic as objective emotion of news.

3. Identification of the Topic Sentence of Chinese News

Directly analyzing tendency of news article, is often disturbed by emotion elements unrelated to topic, and is often difficult to distinguish news report and news event. Given the above problems, this paper presents a two-step process: first extract topic sentence from new article, and then distinguish emotions of topic sentence to suppress interference.

3.1. Feature Sets of News Topic Construction

The topic an article can be expressed by topic feature words, then topic feature set of news article can be obtained by using TF-IDF. A traditional TF-IDF method doesn't consider the frequency of a word in a document, which can have certain influence on the characterization of a word. Considering this problem, we improve TF-IDF:

$$IFIDF_{ik} = a \cdot tf_{ik} \cdot idf_k \cdot f_k = a \cdot tf_{ik} \cdot \log\left(\frac{N}{N_k}\right) \cdot \left(\frac{Num_k}{Num_{all}}\right) \quad (2)$$

Where f_k is the ratio of the frequency of feature value W_k in document D_i over the total. There are totally N documents, and N_k represents the number of documents with feature values. Num_k is the frequency of feature value W_k in document D_i . And Num_{all} is the frequency of feature value W_k in all documents.

Using formula (2), the TF-IDF value of every word can be calculated, forming a feature subset. To improve the accuracy of topic feature, we extract topic features using a method based on crossed entropy, thus getting another feature subset. then we obtain a first feature set

by using equation 3 to do normalization on topic similarity for the two features $T(\text{tfidf})$ and $T(\text{cross})$:

$$T(\text{temp}) = \alpha \cdot T(\text{tfidf})_i + \beta \cdot T(\text{cross})_i \quad (3)$$

Where α and β are weights for adjusting the weight of the two methods. Due to the fact that new headline can better characterize the topic, at first, through the analysis of the distribution of topic score in $T(\text{temp})$ features two thresholds is set to 8.5 and 5.5. Then carry out these operations in $T(\text{temp})$: sequently take word W_k from topic, if $T(\text{temp})$ exist W_k with topic value lower than 8.5, then set W_k to 8.5; else if $T(\text{temp})$ doesn't contain W_k , then add W_k to $T(\text{temp})$ and set topic value to 5.5. The complete topic feature set T is then obtained by merging topic words and $T(\text{temp})$.

3.2. Topic Sentence Extraction

Because topic feature words are centered on topic sentence of new reports, topic sentence from news article is extracted by using vector space model to calculate the cosine distance of topic feature word and news article. The detailed steps are as follows:

Step 1: Calculate cosine similarity. Use all values from topic feature set T as vector dimension space to construct the feature vector $V(T)$ and the sentence vector $V(S_j)$:

$$V(T) = \{\text{weight1}, \text{weight2} \dots \text{weightn}\} \quad (4)$$

$$V(S_j) = \{N_1 \cdot \text{weight1}, N_2 \cdot \text{weight2} \dots N_n \cdot \text{weightn}\} \quad (5)$$

Where n is the number of feature value in T . $V(T)$ uses the similarity of its feature value and topic as the weight of each dimension. $V(S_j)$ uses the product of the number of feature words and its similarity as the dimension weight. Then calculate the cosine similarity $Score(\text{cos})$ of $V(T)$ and $V(S_j)$ as the topic base value of a sentence.

Step 2: Calculate location score. Topic feature extraction has not considered the location feature of the sentence. Sentences at the beginning and ending of the article are more likely to be topic sentences, while sentences in the middle of the article it is more likely to be details. So, higher scores for sentences at the beginning and ending of the article and lower scores for sentences in the middle meets the characteristics of quadratic function. This paper uses function $a \cdot (x - 0.5)^2$ to calculate the location score of a sentence $Score(\text{loc})$.

Step 3: Extract sentence length feature. Longer sentences have higher probability of having more feature value. This paper uses quotient of the sum of TF-IDF value of every word in a sentence and the square value of the sentence length as the length value $Score(\text{len})$ to adjust the influence of sentence length.

Step 4: Calculate title similarity. Title can reflect over 90% of the topic. Considering the fact the title similarity can complement for the deficiency of feature vector method, title similarity $Score(\text{title})$ is the count of words in a sentence that also appears in the title.

Step 5: The topic sentence is extracted by weight fusion after normalization of the above scores. an article at most has four topic sentences. The fusion formula is as follows:

$$Score_{set} = \gamma_i \cdot \sum_{i=1}^4 Score_i \quad (6)$$

When i is equal to 1 to 4, $Score_i$ indicates the score of step 1 to 4, and γ_i is the weight of each item.

4. Sentiment Analysis of the Topic Sentence

Emotion expression of news sentence uses emotion dependency tuple as units. Emotion value of sentence is a comprehensive emotion representation of EDT. Extracting EDT is a process of syntactic analysis of a sentence, analysis of dependency, extracting EDT from dependency, and finally distinguish new tendency based on EDT.

4.1. Emotional Dependency Tuple Extraction

Document structured can be extracted from news article through syntactic analysis. Information extraction based on this can obtain more precise knowledge and improve performance of information extraction system. In this paper, Stanford parser is used to analyze dependency of a sentence and extract the dependency tree. Using the sentence “Reporters also found a lot of investors are very optimistic.” as an example, EDT extraction process is as follows:

(1) First convert parsed text and POS into string sequence “Reporters/NN also/AD found/VV a lot of/CD investor /NN very/AD optimistic /VA”, then perform syntactic analysis to construct syntactic tree and dependency relation;

(2) Extract topic feature word that can be used as evaluation object as the head word of the emotion tuple, for example “investor” in the sentence.

(3) Search for head word in all leaf nodes in syntactic tree, then using extraction rules in Table 1 extract modifier words for head word in all sibling nodes and sub-tree of sibling nodes of head word. Three pairs with structure <head word, modifying> can be extracted, <investor, very>, <investor, optimistic>, using rule;

Table 1. Emotional Dependency Tuple Extraction Rules

ID	Head-word	Modifying Word
1	Noun	Adjective, Verb, Noun, and JJ
2	Verb	Adjective and Adverb
3	Pronoun	Adjective, Verb, Noun, and JJ

(4) Extract the head word, negative dependency of modifying word, degree dependency advmod, nummod from the dependency tree to construct a complete EDT. The complete EDT contains a head word with its modifying words and a number (or zero) of dependency relations.

4.2. Sentiment Identify for the Topic Sentence

Construct emotion distinguishing model based on emotion atom group, $S(sub)$ represents the emotion value for head word with initial value 1. $S(dec)$ represents emotion value for modifier word with initial value 0. $S(term)$ represents emotion base value for the emotion atom group. Obtain head word and modifier word from emotion word dictionary, where positive emotion word value of 1 and negative emotion word value of -1. Then, calculate negative degree $NegW(word)$ of head word and each modifier word: obtain negative dependency of head word and each modifier word, and obtain a negative emotion $NegW(word) = -NegW(word)$ for each negative dependency. Obtain degree modifier for negative dependency, obtain degree modifier $NegW(word) = NegW(word) * W(word)$ for each degree modifier word, where $W(word)$ is the degree coefficient for degree word. So, emotion value for the whole emotion tuple will be:

$$S(term) = S(sub) * NegW(sub) * [\prod_i^n S(dec) * NegW(dec_i) + 1] \quad (7)$$

Where n is the number of modifier for head word, emotion polarity of emotion tuple is dependent on polarity of head word and number of emotion polarity of modifier word. When there is no modifier word or modifier word have no emotion, then $S(term)$ is determined by polarity of head word. Emotion value of sentence is the sum of all emotion dependency tuple, and when a sentence has no emotion tuple, emotion value of a sentence is calculated based on emotion word dictionary. The model for emotion value of a sentence is as:

(8)

Where n is the number of emotion tuple in a certain sentence. When $n = 0$, emotion value of a sentence is the summation of all emotion value of emotion word where m is the number of emotion word and emo is the emotion word. Using this model, the emotion value of topic sentence can be calculated.

4.3. Sentiment Classification of Chinese News

In the above, topic sentiment includes two types of emotion, new report and new event. News report emotion represents the subjective emotion of the author, and the news event emotion represents the objective emotion of the news event. So, classifying sentence emotion with subjective emotion will obtain subjective emotion based on news report.

Step 1: extract subject predicate dependency from dependency tree. The subject usually appears first, so the first pair of subject predicate relation extracted is subject and predicate.

Step 2: using grammatical person and POS of an object, in addition with predicate to distinguish subjective and objective state of a sentence. If the object is first person, pronoun then annotate as subject sentence. News report with specific predicate words as predicate is annotated as objective words.

Step 3: select most relevant subjective sentence to topic from candidate topic sentence as emotion key sentence to obtain emotion of news report. When no subjective sentences exist, and then use most relevant objective topic sentence as final emotion type of news article.

5. Experimental Results

5.1. Parameters Setting

We used the Sixth Chinese Opinion Analysis Evaluation corpus (COAE2014) which is released by the China Conference on Information Retrieval. This corpus has 10000 texts including news, blogs, and forums. We selected 500 News which are randomly labeled from the corpus for our experiments. For gold standard label boundaries, the datasets have been labeled by annotator, so that the key sentence and opinion of the sentiment are marked, and 15 topic words are labeled in each text. Then, we utilized a rule-based method to remove some erroneous examples. Finally, we get a labeled corpus for our task which includes 280 texts.

In Topic feature extraction, we adopt an aggregating method of TFIDF and Cross Entropy based weights. The methodology gets the weights of TFIDF and Cross Entropy during the aggregation through analyzing the similarity of topic feature with title and annotation results. Figure 2 shows the similarity result In the case of different weights. Finally, we selected 9:1 proportion between TFIDF and Cross Entropy.

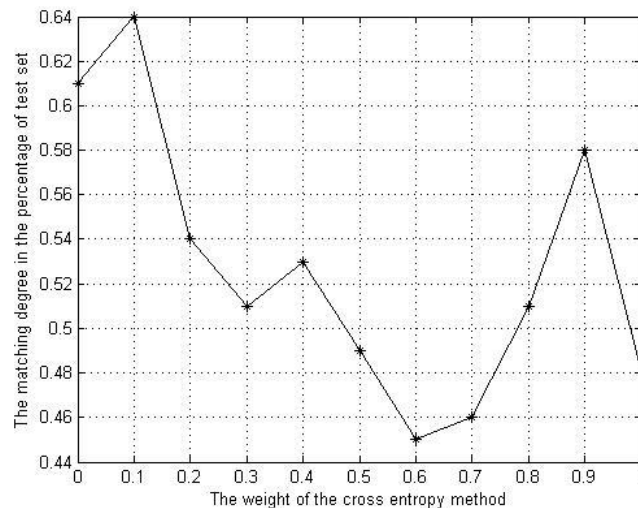


Figure 2. Topic Features Extraction Parameter Weights Contrast

In Topic feature extraction, we adopt an aggregating method of TFIDF and Cross Entropy based weights. The methodology gets the weights of TFIDF and Cross Entropy during the aggregation through analyzing the similarity of topic feature with title and annotation results. Figure 2 shows the similarity result In the case of different weights. Finally, we selected 9:1 proportion between TFIDF and Cross Entropy.

In topic sentence extraction, we used the four features which are the Cosine distance of topic feature vector, the sentence position, the sentence size, and the similarity with title. In our experiments, the evaluation of each feature was based on the number of the matching the

topic sentence labeled by human. We utilized the results of evaluation to determine the weight on each feature. The weight learning uses the same optimization objective as EM algorithm. Finally, we get the four weights: $r1:r2:r3:r4 = 0.55:0.2:0.1:0.15$

5.2. Results and Observations

We use Precision (P), Recall (R), F measure (F1) and Micro-average (Micro) to evaluate the overall performance. Referring to the evaluation standards of NLP&CC 2013 [22], the formulas of evaluation standards for News sentiment classification are listed as follows.

$$Precision = \frac{\#system_correct(emotion = Y)}{\#system_proposed(emotion = Y)} \quad (9)$$

$$Recall = \frac{\#system_correct(emotion = Y)}{\#annotated(emotion = Y)} \quad (10)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (11)$$

In above formulas, #system_correct denotes the number of the submitted results which match with the manually annotated results, #system_proposed denotes all the number of the submitted results, #annotated denotes the number of the manually annotated results.

The result shows that our method leads to a higher performance in Precision, Recall and F-measure than those approaches released by COAE2014. In Micro-average, our results are the closest to the best results. The details about the results are shown in Table 2.

Table 2. The Results of Sentiment Analysis of Chinese News

	R	P	F1	Accuracy	Micro Recall	Micro Precision	Micro F-measure
EDT-Based	0.2152	0.0683	0.1037	0.0600	0.3495	0.0932	0.1472
median	0.0676	0.0517	0.0571	0.0405	0.2201	0.0682	0.1041
max	0.2394	0.0846	0.1092	0.0652	0.3887	0.1041	0.1642

The results of our experiments are the second best in all the approaches, and very close to the best. This shows our sentiment analysis model based on emotional dependency tuple is a practical and worthy of further study methods. Our approach adopt topic feature to construct emotional dependency tuple, and can avoid the impact of non-topic sentiment. Our model can further enhance, for example, we improve the recall and precision by the augment of syntactic and semantic evidence. We can also lift performance by adding synonym dictionary, sentiment dictionary and headword dictionary. Consequently, the scalability of our sentiment analysis model based on emotional dependency tuple is very strong.

6. Conclusions and Future Work

To analyze the topic sentiment of Chinese news, our task is divided into three sub-task: identifying the topic sentence, extracting the emotional dependent tuples, and the topic sentiment analysis. Divide - and - conquer strategy is used to complete the different layer of

sub-task to ease analyzing sentiment of news as a whole. We presented a novel sentiment analysis model based on emotional dependency tuple, which improve the efficiency of the sentiment analysis. This model regards the topic words of sentence as head-words of the emotional dependency tuple, and removes the noise of non-topic sentiment. We found the augment of syntactic and semantic evidence has a great help to lift the performance of sentiment analysis through our experiments. This will be our further works. Sentiment classification in the news text is very useful to individual and corporation, such as taking effective measures to reduce the negative impact on the news media, Internet, and other media. This paper proposed the method based on emotional dependency tuples is preliminary discussed for sentiment analysis, which is a worthy of further exploration and research, has broad application prospects.

Acknowledgements

This research work is supported by Hunan Provincial Natural Science Foundation of China (No.13JJ4076), the Scientific Research Fund of Hunan Provincial Education Department for excellent talents (No.13B101), Foundation of University of South China (No.2012XQD28), the Construct Program for the Key Discipline in University of South China (No.NHxk02), and the Construct Program for Innovative Research Team in University of South China.

References

- [1] Y. Zhao, B. Qin and T. Liu, "Sentiment Analysis", *Journal of Software*, vol. 21, no. 8, (2010).
- [2] T. F. Yao, X. W. Cheng, F. Y. Xu, H. Wu and R. Wang, "A Survey of Opinion Mining for Texts", *Journal of Chinese Information Processing*, vol. 22, no. 3, (2008).
- [3] S. M. Kim and E. Hovy, "Automatic detection of opinion bearing words and sentences", *Proceedings of the Second International Joint Conference on Natural Language Processing*, (2005); Jeju Island, Korea.
- [4] B. Pang, L. Lee and S. Vaithyanathan, "Sentiment classification using machine learning techniques", *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Morristown, (2002); NJ, USA.
- [5] P. Turney, "Semantic orientation applied to unsupervised of reviews", *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, (2002); USA.
- [6] B. Pang and L. Lee, "Opinion mining and sentiment analysis", *Foundations and trends in information retrieval*, vol. 2, no. 1, (2008).
- [7] X. Jun, Y. X. Ding and X. L. Wang, "Sentiment classification for Chinese news using machine learning methods", *Journal of Chinese Information Processing*, vol. 21, no. 6, (2007).
- [8] W. Yang, J. J. Song and J. Q. Tang, "A Study on the Classification Approach for Chinese MicroBlog Subjective and Objective Sentences", *Journal of Chongqing Institute of Technology*, vol. 27, no. 1, (2013).
- [9] M. B. Salem and J. S. Salvatore, "Detecting masqueraders: A comparison of one-class bag-of-words user behavior modeling techniques", *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, vol. 1, no. 1, (2010).
- [10] Y. B. Wu, Q. Zhang and X. J. Huang, "Phrase dependency parsing for opinion mining", *Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics*, (2009); Boston, USA.
- [11] L. Y. Luo, J. L.V. Mejino Jr. and G. Q. Zhang, "An Analysis of FMA Using Structural Self-Bisimilarity", *J Biomed Inform*, vol. 46, no. 3, (2013).
- [12] J. L. Zhang, E. Liu, J. Wan, Y. Ren, M. Yue and J. Wang, "Implementing Sparse Matrix-Vector Multiplication with QCSR on GPU", *Applied Mathematics & Information Sciences*, vol. 7, no. 2, (2013).
- [13] G. Qiu, B. Liu, J. Bu and C. Chen, "Expanding Domain Sentiment Lexicon through Double Propagation", *the 21st International Joint Conference on Artificial Intelligence*, (2009); Pasadena, California, USA.

- [14] C. Zhang, D. Zeng, J. Li, F. Y. Wang and W. Zuo, "Sentiment analysis of Chinese documents: From sentence to document level", *Journal of the American Society for Information Science and Technology*, vol. 60, no. 12, (2009).
- [15] S. Feng, Y. C. Fu, F. Yang, D. Wang and D. Zhang, "Blog Sentiment Orientation Analysis Based on Dependency Parsing", *Journal of Computer Research and Development*, vol. 49, no. 11, (2012).
- [16] J. L. Zhang, Y. Chen, J. Wang, E. Y. Liu, Y. Y. Yin, Y. J. Ren and C. H. Li, , "The Implementation and Optimization of Irregular Application Task Models Based on the Cell BE Processor", *International Journal of Digital Content Technology and its Applications*, vol. 6, no. 2, (2012).
- [17] W. Wang, D. Y. Zhao and W. Zhao, "Identification of Topic Sentence about Key Event in Chinese News", *Acta Scientiarum Naturalium Universitatis Pekinensis*, vol. 47, no. 5, (2011).
- [18] J. L. Zhang, H. Ding, J. Wan, Y. Ren and J. Wang, "Designing and Applying an Education IaaS System based on OpenStack", *Applied Mathematics & Information Sciences*, vol. 7, (2013).
- [19] L. X. Xie, M. Zhou and M. S. Sun, "Hierarchical Structure Based Hybrid Approach Sentiment Analysis of Chinese Microblog and Its Feature Extraction", *Journal of Chinese Information Processing*, vol. 26, no. 1, (2012).
- [20] H. H. Wu, , A. C. R. TSAI, R. T. H. Tsai and J. Y. J. HSU, "Building a Graded Chinese Sentiment Dictionary Based on Commonsense Knowledge for Sentiment Analysis of Song Lyrics", *Journal of Information Science & Engineering*, vol. 29, no. 4, (2013).
- [21] G. Xu, X. F. Meng and H. F. Wang, "Build Chinese Emotion Lexicons Using A Graph-based Algorithm and Multiple Resources", *Proceedings of the 23rd International Conference on Computational Linguistics*, (2010); Beijing, China.
- [22] C. P. Ouyang, X. H. Yang, L. Y. Lei, Q. Xu, Y. Yu and Z. M. Liu, "Multi-strategy Approach for Fine-grained Sentiment Analysis of Chinese Microblog", *Acta Scientiarum Naturalium Universitatis Pekinensis*, vol. 50, no. 1, (2014).

Authors



Ouyang Chunping, she received her Ph.D. degree from University of Science Technology Beijing, China, in 2011. Now, she is working in University of South China, where she is an associate professor and the dean of software engineering department. Her research interests are in Semantic Web, social computing and sentiment analysis of text.



Zhou Wen, he received his B.E. degree from Department of Computer Science of University of South China, Hengyang, in 2012. Now, he is working his M.E. Degree at University of South China. His research interests include natural language processing and data mining. He is corresponding author of this paper.



Yu Ying, she received his M.E. degree from Department of Computer Science of University of South China, Hengyang, in 2012. Now, he is lecture of University of South China. His research interests include education information technology and sentiment analysis of microblog.



Liu Zhiming. He received his Ph.D. degree from National University of Defense Technology, China, in 2012. Now, he is professor of University of South China, supervisor of postgraduate. He has authored and coauthored over 20 research publications in peer-reviewed reputed journals. His research interests include information retrieval, Cloud Computing and big-data mining.



Yang Xiaohua. He received his Ph.D. degree from CAS, China, in 1999. From 2000 to 2001, he was a visiting scholar at Wollongong University. Yang is vice-president of University of South China, and professor of computer science, doctoral tutor. He has authored and coauthored over 50 research publications in peer-reviewed reputed journals. He has served as the program committee member of various international conferences and reviewer for various international journals. His research interests include natural language processing and information retrieval.