

Research on Methods of Microblogging Sentiment Feature Extraction Based on Dependency Parsing

Li Yonggan^{1,a}, Zhou Xueguang^{2,b}, Guo Wei^{2,c}, Zhang Huanguo^{1,d}

¹School of Computer, Wuhan University, Wuhan 430072, China

²Department of Information Security, Navy University of Engineering, Wuhan 430033, China

^alyglcn@163.com, ^bzxcg196610@hotmail.com, ^cdaguoguo945@163.com, ^dliss@whu.edu.cn

Keywords: Microblogging; Dependency Parsing; Opinion Word Expansion; Sentiment Feature Extraction

Abstract. Microblogging emotional elements extraction recall rate is extremely low without opinion words expansion. Aiming at the feature extraction of Microblogging sentiment, a sentiment feature extraction method based on dependency parsing for the Microblogging emotional has been proposed in this paper, abbreviated as MSFE_DP. The specific steps are as follows: First, the existing opinion words were expanded to build a special emotional thesaurus. Second, according to the six dependencies modes of emotional words and evaluation objects, evaluation units were identified and extracted automatically. Finally, a complete evaluation object was obtained in combination with the evaluation object merging algorithm. Through participating in NLP&CC Chinese Microblogging emotional public evaluation, MSFE_DP took the third place among the 22 approaches in the evaluation tasks, which indicates that our method is effective.

Introduction

In the Web 2.0 environment, a large number of network media sharing platforms have sprung up, and users are increasingly inclined to the Internet to express their views on a variety of products or events. By the end of 2014 China Microblogging coverage has reached 54.5%, and through updating short message within 140 characters, Microblogging made instant sharing with multiple tools come true, which has become the mainstream interactive platform and Internet mass media center for China's Internet users, seen in [1]. Microblogging users continue to grow and become mobile gradually, resulting in large amount of data, rapid produce speed, and serious content fragmentation of Microblogging. Those characteristics determine that it is not feasible to manually monitor and process micro-blog content. How to automatically detect Microblogging and control emotions Microblogging topic from the mass Microblogging is one of the hot spots for network content security researches according to the reference [2].

There are a large number of evaluations of information generating which users are involved in and are used to evaluate commodities, events or person performed on the Microblogging, which often contain people's emotions and tendencies, so Microblogging sentiment analysis technology came into being. Microblogging sentiment analysis, also known as Microblogging opinion mining, refers to the analyzing, processing, summarizing and reasoning process of the subjective Microblogging with emotional colors. Microblogging sentiment feature extraction is the underlying task of the Microblogging sentiment analysis, a both important and difficult assignment, which is designed to extract meaningful information unit in Microblogging emotional evaluation, make non-structure Microblogging into structured text, and provide data for subsequent analysis of emotion, seen in [3].

It is not enough to just get the emotional words for emotional information extraction. Some emotional words have obvious emotion, but the emotional tendency is based on the evaluation object, such as "big, small, light and heavy." As in the sentence "The room was very large and spacious", the emotional tendency of "large" is compliment; while in "the noise is very large outside the room", the emotional tendency of "large" is derogatory. This phenomenon in reference [4] is an important

basis for our research. Therefore, we use the emotional extension based on dependency association analysis in this paper to improve the recall rate.

According to the reference [2], emotional information extraction mainly includes the extraction of the evaluation words, the evaluation objects, opinion holders and the combination of evaluation units. The result of independently using the above method to extract the Microblogging emotional factor is not ideal. Based on the above conclusions, through the dependency pattern analysis of the Microblogging emotional words and evaluation object, this paper achieves the Microblogging extraction of emotional elements.

The paper is organized as follows: Section 1 introduces the background and significance of the paper. Section 2 describes related work and motivation, and Section 3 describes the emotional factor extraction method based on dependency parsing. Section 4 is openly involved in NLP&CC2012 evaluation experimental results and analysis. Finally it summarizes the research work and proposes future research directions.

Related Work and Motivation

A. Researches and hot spots for Microblogging

Traditional text mining theme has been extensively studied, but because of micro-blog with some structural aspects of the social network information, the using in its ordinary text mining algorithms modeling is not accurate enough. Mainstream related study abroad includes Twitter analysis by H. Kwak et al., seen in [5], and TwitterRank thought proposed by J Weng et al. in [6]; and Domestic Zhang Chenyi et al. proposed a model based on the LDA generated Microblogging, and Microblogging theme auxiliary mining which can be seen in [7]; Li Si proposed the LQE model and Bowen averaging algorithm to extract related blog, seen in [8]. The above results of Microblogging researches are from multiple angles, which provide an important reference to deal with Microblogging in this paper.

B. Dependency Analysis

Dependency parsing is first proposed by L.Tesniere in [9], the core of which is represented by coupling the relationship between words. In the 1970s, Robinson made four axioms about dependency grammar dependency seen in [10], and due to the special nature of the Chinese language structure, there is a fifth axiom seen in [11]. Penn Chinese tree is a Chinese Structure Grammar Library appeared earlier and studied by more people, seen in [12], and ctbparsen open source toolkit can automatically obtain the dependencies between words in [13]. Currently dependency parsing method has been widely used, including: hierarchical dependency parsing in [14], based on analysis of emotional words associated with the expansion of dependency and analysis, which can be seen in [15], and other content-based active learning Chinese dependency syntax. The above researches from different fields of dependency analysis were carried out on the text, which made significant contributions to conduct Microblogging grammatical structure dependency analysis in this paper.

C. Emotional Information Extraction

Currently emotional information extraction methods are widely used in many fields, and there are two methods, namely corpus-based approach and dictionary-based approach on the evaluation of the words in terms of extraction and discrimination. Hatzivassiloglou et al. extracted a lot of adjectives evaluation of the words from a large corpus of Wall Street Journal in [16]. Kobayashi et al. via the syntactic analysis (characteristics, evaluation words) obtained evaluation of object in [17]. Qiu and others use the dependencies between word emotion and evaluation of object to obtain the evaluation of object in [18]. Hou Min et al, from the Microblogging language features and sentiment analysis strategy, put forward views sentence extraction strategy based on the identification and evaluation of the emotional object dictionary and phrase semantic rules in [19]. Ms Wang Qian, He Tingting extracted text emotional elements from the phrase dependent analysis in [20]. The above findings for text emotional processing provide an important basis for dealing with the emotional elements extraction in this paper, which regards Microblogging as the processing object.

D. Related Work Introduction

In previous studies, the proposed method associated with reference [4], [18], [19] and [20].

Both reference [4] and [18] conduct emotional word expansion by the dependency parsing,

where the method of Qiu et al. can only extend the emotional words which have a direct dependency in [18], while the OWP algorithm can simultaneously extend the term with a direct emotional interdependence and indirect dependencies in [4], and the recall rate experiments of [4] is better than the method in [18]. Therefore, OWP method is one of the key technologies to achieve the goal of our paper.

Reference [19] ranked 1st and 2nd in NLPC2012 emotional elements extracted evaluation, which can be seen in [21]. The main methods can be concluded as follows: on the basis of inducing the features of Microblogging language, it constructed the emotional phrase dictionary, determined the polarity of the sentence by the phrase rules, focused on the negative form, established the topic-based strategies like OBJ form, and completed the micro-Bo sentiment analysis. Interdependent way was conducted in this paper to accomplish the automatic creation of emotional dictionary, and then we carried out the Microblogging emotional factor extraction based on the dependency parsing.

Similar to the article, reference [20] participated in the COAE2008 evaluation; however, there are the following differences: firstly, its object is a text rather than dealing with the Microblogging; secondly, the experimental data set is small, only about 3000 sentences; thirdly, it adopted the trial and error evaluation method which you can always adjust the optimization. Compared reference [20], the object in this paper is a Microblogging with processing data set of 17500 micro-Bo and 32000 sentences; the results were processed without any reference answers to the control adjustment; and evaluation results were directly submitted to the meeting and the public.

Key Technologies

Taking into account the particularity unlike ordinary text, such as short text of Microblogging, random choice of words, more negative views, strong emotional expressions, lack of rational evaluation, mainly using oral language, etc., we firstly put forward the emotional word expansion algorithm to simultaneously extend emotional words with direct or indirect dependence, and dependence relationship through pruning branches of the dependency tree and merging operations and the given expansion strategy; then built dependency evaluation unit model, and took advantage of dependency parsing rules to get the evaluation unit and the evaluation object; finally, automatically identified the evaluation units by the six summarized evaluation unit modes, and combined with the evaluation object merging algorithm to obtain the complete evaluation object.

A. Emotional Word Expansion Algorithm Based on Dependent Correlation Analysis

This paper improves the result of which in [18], and put forward a processing algorithm suitable for application in the evolution of communities prone tendencies network information (such as micro-blog), called OWP algorithm in [4]. Algorithm is as follows:

- Find new words and phrases;
- Prune Dependency tree, including pruning and merging operations;
- Expand emotional words.

Through the dependency tree pruning and normalization two words with indirect dependencies were changed into direct dependency and emotional word expansion. Then a basic emotion thesaurus was established, and the extensible object of emotional words were positioned on adjectives and verbs, while the evaluation object and extracting object were positioned on nouns and noun phrases.

In addition, it is necessary to make commendatory and derogatory emotion thesaurus and evaluation object library separated. The emotional words and evaluation object will be extended in accordance with the principle of the same polarity, while it must be negated in the processing of the polarity of negative words plus the emotional words in dependency tree merge operation.

TABLE 1 EVALUATION UNIT MODE

| Evaluation unit mode | Part of speech of emotional words | Part of speech of evaluation objects |
|---|-----------------------------------|--------------------------------------|
| $\{NP\} \xleftarrow{SUB} \{AV\}$ | VA, VV | NN, NR, NT, PN |
| $\{NP\} \xleftarrow{SUB} \{VC\} \xrightarrow{PRD} \{NP\} \xrightarrow{NMOD} \{JV\}$ | VA, JJ | NN, NR, NT, PN |
| $\{AV\} \xrightarrow{OBJ} \{NP\}$ | VA, VV | NN, NR, NT, PN |
| $\{NP\} \xleftarrow{SUB} \{VV\} \xrightarrow{VMOD} \{AV\}$ | VA, VV | NN, NR, NT, PN |
| $\{JV\} \xleftarrow{NMOD} \{NP\} \xleftarrow{SUB} \{VV\}$ | VA, JJ | NN, NR, NT, PN |
| $\{VV\} \xrightarrow{OBJ} \{NP\} \xrightarrow{NMOD} \{JV\}$ | VA, JJ | NN, NR, NT, PN |

B. Dependent Mode of Evaluation Unit

Emotional words and evaluation objects have mappings, and through the dependencies parsing of the evaluation objects and emotional words, we will make evaluation objects and emotional words grouped into six kinds of dependency model, which propose extraction method of the emotional elements based on dependency parsing. In order to overcome the noise which may be brought in the extraction of the emotional words and evaluation objects only relying on dependencies parsing, we increased restrictions on the speech of words when dependency analysis is conducted.

As it is shown in Table 1, the candidate evaluation units were filtered by the evaluation unit mode-dependent, with the evaluation unit modes from the top down and the priority level from high to low.

In the evaluation unit dependent mode column, words in bold show us emotional words and evaluation objects.

It is worth noting that the priority level of the evaluation unit dependency pattern from 1 to 6 is gradually decreased. That is to say, it is supposed to match evaluation unit dependent mode 1 firstly. If it works, the other candidate evaluation units will be screened after that; if it does not work, mode 2 will be matched, and so on.

C. Merging algorithm evaluated

In the six interdependent modes of emotional words, the evaluation object of Mode 1, Mode 2, Mode 4 and Mode 5 is all subjects, while the evaluation object of Mode 3 and Mode 6 is objects. In order to extract the evaluation object as complete and specific as possible, it is necessary to conduct a merge operation to obtain complete components of the subject or object. For example, the sentence in Chinese Pinyin: Wo xihuan ipad3 de pingmu, which could be translate in English as I like the screen of ipad3, scilicet, it should be extracted "ipad3 de pingmu", in English as "the screen of ipad3", scilicet, rather than just "pingmu" in English as "screen", scilicet.

In order to distinguish between the original evaluation object and the evaluation object obtained after merging, the evaluation of the original object is called evaluated baseline words. In the operation of merging evaluation objects, it should be started with the left neighbor word of the baseline evaluation object word, judging from right to left. If the father node is the benchmark or the right neighbor word, it will be integrated into the evaluation of the object; it will be stopped integrating when it is the beginning of a sentence or its father node is neither the benchmark nor the right neighbor word.

Merging algorithm of the evaluation object is described in Table 2 as below.

TABLE 2 EVALUATED MERGE ALGORITHM

```

Name: Appraisal Target Merge Algorithm
Input:  $T_{in}$  //the original appraisal target
Output:  $T_{out}$  //the final appraisal target
1.  $W(i) = T_{in}$ ;
2.  $j = i - 1$ ;
3. while ( $j \geq 0$ ) do
   //all the left words  $W(j)$  of  $W(i)$ 
4.   if ( $W(j).parent = i \parallel W(j).parent = j + 1$ ) do
5.      $j --; k = j; k \leq i; k ++$ 
6.   end if
7.   else do
8.     for ( $j --; k = j; k \leq i; k ++$ ) do
9.        $T_{out} = T_{out} + W(k)$ ;
10.    end for
11.    break;
12.  end else
13. end while

```

D. MSFE_DP Algorithm

The algorithm steps for Microblogging emotional elements extraction by using the method of MSFE_DP are as follows:

- a) Do emotional word extension by using the methods of emotional word extension based on the use of the dependency relation analysis [14] to the view sentence;
- b) Get candidate emotional words and candidate evaluation objects;
- c) Use evaluation unit dependent mode to choose the candidate evaluation unit;
- d) Get a complete evaluation object by merging operations to the dependency tree.

Among the sentences waiting for analysis, all the emotional words and extended emotional words will be listed as the candidate emotional words, while all the nouns or pronouns will be listed as the candidate objects.

Experimental Results and Analysis

The method of MSFE_DP proposed in this paper was taken part in the three public evaluation tasks of NLPCC 2012 Chinese micro-blog sentiment analysis, namely the viewpoint sentence recognition, emotional tendencies judgment, and emotional elements extraction, seen in [21]. The data from Table 3, Table 4 and Table 5 are derived from NLPCC2012 public evaluation mission report. The experiment conducted the dependency analysis by using ctbparser open source toolkit [13], and adopted the standard Penn Chinese Treebank to achieve segmentation and POS tagging. The experiments is implemented using C + +language in Visual studio 2010. Finally, part of the results of NLPCC2013 [23] public evaluation were conducted comparative analysis. And the data from Table 6 are originated from the above two public evaluation reports.

A. Dataset and Thesaurus

a) Dataset

Data sets adopt the public evaluation data set which was issued on July 1, 2012, and the results of the evaluation was submitted before July 31. All the data of the evaluation are from Tencent Microblogging. The evaluation data includes 20 topics, 17 500 micro-Bo, and 32,000 sentences. The size of data set is 5.6M B, and the format of data is Unicode (utf-16) encoding xml, which has already pre-cut the sentences.

b) Thesaurus

Thesaurus includes emotional thesaurus, polarity thesaurus and negative thesaurus. In order to conduct experimental research and to participate in public evaluation, this paper constructed emotional thesaurus, polarity thesaurus and negative thesaurus respectively. The emotional thesaurus uses emotional vocabulary announced by "HowNet", "Common appraise Detailed dictionary meaning of the words", "Student Judgments righteousness Dictionary", "commendatory dictionary" and "derogatory dictionary" as a basis, deletes the emotional words of which the frequency of use is low, and increases network terms and oral emotional words, and constructs an emotional word library, which contains 5554 commendatory terms and 6321 derogatory terms.

We also marked a polarity thesaurus in the emotional thesaurus. There are some very strong

emotional words with quite strong polarity, especially some curse derogatory terms. In view of identifying the polarity of a view sentence, as long as those words appear, the polarity of the sentence will be determined as the polarity of the word (negative sentence in inverted). To distinguish the big emotional thesaurus, such thesaurus is called the polarity thesaurus, which contains 16 commendatory terms, and 262 derogatory terms. Emotion thesaurus is included with the polar thesaurus, so that emotional thesaurus contains all the words in the polarity thesaurus.

We also built a dedicated denial thesaurus, which contains the words in Chinese Pinyin: "Bu, wei, meiyou, qian" as in English "No, not yet, not, insufficient", scilicet ,and other 20 negative words.

B. View Sentence Recognition

Task 1, evaluation of identifying view sentences, is to recognize each sentence in every Microblogging is a perspective view of the sentence or not. There are 34 units who submitted 53 sets of results in total. Our results were denoted by NUE-M SFE_DP, seen in [21]. For the comparative study, we chose three groups of the comparative experimental evaluation results. The first one is the experimental results of Communication University of China who took the first place in the public evaluation, seen in [19], labeled CUCsas; the second comparison unit is the results of Beijing University of Aeronautics and Astronautics, seen in [22], labeled CUAA. They were using a method of SVM and conventional method respectively to carry out the experiments. The concrete results are shown in Table 3.

Analysis of the results: it can be concluded by the observation of the experimental data that our approach is with high accuracy but lower recall rate. One of the main reasons is that the emotion thesaurus is not comprehensive enough; the other one is the view sentences without any emotional words cannot be identified. The recall rate by using the SVM method is better, which can be promoted in future studies.

C. Judging of Emotional Tendency

Task 2, evaluation of determining the emotional tendency, is to judge the emotional tendencies of each view sentence in the Microblogging, and a total of 32 units with 48 sets of results were submitted. We still chose the results of CUC and CUAA to conduct the comparative study, see Table 4.

Analysis of the results: it can be learned by the observation of the experimental data that our approach is with medium accuracy and lower recall rate, mainly due to the method used to recall the rule set rate can only reach up to 0.45. If we want to improve recall rate, the method must be improved, and the SVM is an ideal one.

D. Microblogging Emotional Elements Extraction

The emotional elements extraction adopts two ways, namely strict evaluation and relaxed evaluation. In strict evaluation, the evaluation is correct only when the start and end positions of the emotional object in the whole Microblogging and answers are identical and have the same polarity. According to the evaluation outline, when the emotional object is being extracted, it is required to extract complete and clear evaluation object. For instance, in the sentence in Chinese Pinyin " ipad3 de pingmu hen bang " which can be translated into English "The screen of ipad3 is very good", scilicet, in accordance with strict evaluation, the emotional object we are supposed to extract is " ipad3 de pingmu" as in English " the screen of ipad3", scilicet, rather than "pingmu" in English "screen", scilicet. When we conduct the relaxed evaluation, the evaluation is always considered correct no matter what words or phrases extracted of the sentence " ipad3de pingmu" in English the screen of ipad3, scilicet. Therefore, the relaxed evaluation cannot be regarded as the evaluation basis of the task of extracting accurate emotional element.

Task 3, evaluation of emotional elements extraction, is asked to identify the evaluation object of each view sentence, that is the emotional object, and to determine the polarity of views for emotional objects. The submitting formats include: (Microblogging number, sentence number, emotional object, start position, end position, and views tendencies), a total of 15 units with 22 sets of results presented in. Table 5 gives the results of NUE, CUC and CUAA under the strict evaluation, see Table 5.

Analysis of the results: the precise evaluation of the best results evaluation in this Microblogging emotional element extraction, i.e. the F value is close to 30%. The F value of this method precise evaluation is 22.5%, ranking in the third place. Neither the classical method of SVM nor conventional method is an ideal evaluation in Task 3, indicating that the emotional element

extraction of Microblogging is a complex, difficult and comprehensive task. Meanwhile, it also reminds us that there is much room for improvement in terms of recall rate, F value, and so on.

E. Comparison of Experiments

In order to understand the evaluation work of emotional elements extraction of Microblogging better, the evaluation results of NLPCC2012 and 2013 NLPCC were compared, and we added the participation of Department of Computer Science and Technology in Huazhong Normal University (HNU) whose NLPCC2013 evaluation result can be seen in [23]. In addition, we also added the recent two years' NLPCC public evaluation data of Beijing University of Aeronautics and Astronautics, the specific results are shown in Table 6.

Two public evaluation results of Communication University of China (CUC) in NLPCC both took the first place. Analysis is concluded as follows: CUC work is carried out in the following characteristics, such as statistical methods which relies on manual Microblogging emotional word dictionary, language-based features to establish work rules, identification of the view sentences based on semantic rules, and extraction methods of evaluation objects, indicating that the method is still valid in the field of natural language processing. The results of the evaluation in 2013 are better than the year before, which indicated that the organization is continuing its efforts to improve the research methods. Therefore, when conducting language research and evaluation the methods based on rules and language, we need to pay much more attention in the future.

The methods Beijing University of Aeronautics and Astronautics (BUAA) used in the past two years in public evaluation are more traditional. The result is better when the methods are used respectively, while the effect of combination of the methods used in the evaluation of Microblogging emotional elements extraction is average.

TABLE 3 THE RESULTS COMPARISON OF THE EVALUATION TO IDENTIFY VIEW SENTENCES

| Unit | Method | Micro-averaged accuracy | | | Macro-averaged accuracy | | |
|------|----------|-------------------------|--------|-----------|-------------------------|--------|-----------|
| | | Precision | Recall | F-measure | Precision | Recall | F-measure |
| NUE | MSFE_DP | 0.828 | 0.537 | 0.651 | 0.826 | 0.52 | 0.629 |
| CUC | CUCsas | 0.756 | 0.812 | 0.783 | 0.757 | 0.797 | 0.773 |
| BUAA | SVM | 0.681 | 0.918 | 0.782 | 0.684 | 0.914 | 0.779 |
| BUAA | RuleSets | 0.707 | 0.577 | 0.635 | 0.715 | 0.577 | 0.632 |

TABLE 4 THE RESULTS COMPARISON OF THE EVALUATION TO DETERMINE THE EMOTIONAL TENDENCIES

| Unit | Method | Micro-averaged accuracy | | | Macro-averaged accuracy | | |
|------|----------|-------------------------|--------|-----------|-------------------------|--------|-----------|
| | | Precision | Recall | F-measure | Precision | Recall | F-measure |
| NUE | MSFE_DP | 0.772 | 0.451 | 0.54 | 0.776 | 0.404 | 0.523 |
| CUC | CUCsas | 0.844 | 0.685 | 0.756 | 0.842 | 0.672 | 0.745 |
| BUAA | SVM | 0.761 | 0.698 | 0.728 | 0.768 | 0.702 | 0.733 |
| BUAA | Rulesets | 0.764 | 0.440 | 0.559 | 0.758 | 0.445 | 0.559 |

TABLE 5 THE EVALUATION RESULTS COMPARISON OF MICROBLOGGING EMOTIONAL ELEMENTS EXTRACTION

| Unit | Method | Micro-averaged accuracy | | | Macro-averaged accuracy | | |
|------|-----------|-------------------------|--------|-----------|-------------------------|--------|-----------|
| | | Precision | Recall | F-measure | Precision | Recall | F-measure |
| NUE | MSFE_DP | 0.311 | 0.177 | 0.225 | 0.324 | 0.174 | 0.22 |
| CUC | CUCsas | 0.303 | 0.273 | 0.288 | 0.306 | 0.263 | 0.278 |
| BUAA | SVM | 0.066 | 0.147 | 0.091 | 0.07 | 0.144 | 0.093 |
| BUAA | Rule Sets | 0.065 | 0.102 | 0.08 | 0.068 | 0.101 | 0.08 |

TABLE 6 THE EVALUATION RESULTS COMPARISON OF OTHER EMOTIONAL ELEMENTS EXTRACTION

| Unit | Method | Micro-averaged accuracy | | | Macro-averaged accuracy | | |
|------|-----------|-------------------------|--------|-----------|-------------------------|--------|-----------|
| | | Precision | Recall | F-measure | Precision | Recall | F-measure |
| NUE | MSFE_DP | 0.311 | 0.177 | 0.225 | 0.324 | 0.174 | 0.22 |
| CUC | CUCsas | 0.303 | 0.273 | 0.288 | 0.306 | 0.263 | 0.278 |
| CUC | NLPCC2013 | 0.448 | 0.408 | 0.427 | 0.448 | 0.404 | 0.422 |
| HNU | NLPCC2013 | 0.077 | 0.081 | 0.079 | 0.078 | 0.08 | 0.078 |
| BUAA | SVM-2012 | 0.066 | 0.147 | 0.091 | 0.07 | 0.144 | 0.093 |
| BUAA | NLPCC2013 | 0.109 | 0.071 | 0.085 | 0.107 | 0.07 | 0.084 |

Conclusion

Emotional factors extraction is the main task of sentiment analysis. This paper presents a method of Microblogging emotional elements extraction based on dependency parsing, which first extends the emotional word with direct, indirect and dependent relationship, and identifies the evaluation unit automatically according to the emotional words and the interdependence mode of the evaluation object, and obtains a complete and clear evaluation object by merging algorithm of evaluation object. The proposed MSFE_DP method in NLP CC2012 Chinese Microblogging public sentiment analysis evaluation participated in Task 3 emotional element extraction, and ranked in the third place among all the participating groups. Compared with the results of NLPCC2013 public evaluation, the analysis shows that this method is still valid.

The follow-up work of this paper includes resolution of demonstrative pronouns within the context, construction of a comprehensive and detailed emotional word dictionary, and reinforce of the syntactic analysis researches, and further improvement of the accuracy and recall rate of Microblogging emotional elements extraction.

References

- [1] CNNIC 34th Statistical Report on Internet Development in China [EB/OL]. July 16(th) 2014. http://www.cnnic.net.cn/gywm/xwzx/rdxw/rdxx/201407/t20140721_47437.htm
- [2] Zhao Yanyan, Qin Bing and Liu Ting. Sentiment Analysis. Journal of Software, 21st ed., vol.8. 2010, pp. 1834-1848.
- [3] Yang Liang, Lin Yuan and Lin Hongfei. Micro-Blog Hot Detection Based on Emotion Distribution, 26th ed., vol.1. Journal of Chinese Information Processing, 2012, pp.84-90 +109.
- [4] Sun Yan, Zhou Xueguang and Fu Wei. Opinion Words Expansion based Dependency Parsing, 35th ed., vol.5. Journal of Beijing University of Posts and Telecommunications, 2012, pp. 90-93.
- [5] H. Kwak, C. Lee, H. Park, and S. B. Moon. What is twitter, a social network or a news media? //Proceedings of WWW, Raleigh North Carolina, USA, 2010, pp.591-600.
- [6] J. Weng, E. P. Lim, J. Jiang and Q. He. TwitterRank: Finding topic sensitive Influential Twitterers //Proceedings of WSDM. New York, USA, Feb. 2010.
- [7] Zhang Chenyi Yi, Sun jianling and Ding Yiqun. Topic mining for microblog based on MB-LDA Model, 48th ed., vol.10. Journal of Computer Research and Development, 2011, pp.1795-1802.
- [8] Li Si. Research on key problems in web opinion Mining. Beijing: Journal of Beijing University of Posts and Telecommunications, 2012.
- [9] Tesnière L, *Éléments de Syntaxe Structurale*, Paris. Klincksieck, 1959.
- [10] Liu Weiquan, Wang Minghui and Zhong Yixin. On study of hierarchy structure of dependency relations in Chinese , 10th ed., vol.2. Journal of Chinese Information Processing, 1996,

pp. 32-46.

[11] Huang Changning, Yuan Chunlaw and Pan Shimei. Corpus, knowledge acquisition and syntactic analysis, 6th ed., vol.3. Journal of Chinese Information Processing, 1992, pp. 3-8.

[12] Xia F. The part-of-speech tagging guidelines for the penn Chinese treebank(3.0) [EB/OL]. <http://www.cis.upenn.edu/~chinese/posguide.3rd.ch.pdf>.

[13] Qian Xian, Zhang Qi, Huang Xuanjing and Wu Lide. 2D Trie for fast parsing.//Proceedings of Coling 2010, Coling 2010 Organizing Committee. Beijing, 2010, pp. 904-912.

[14] Jian Ping and Zong Chengqing. Layer based dependency parsing by sequence labeling models, 24th ed., vol.6. Journal of Chinese Information Processing, 2010, pp. 14-22.

[15] Che Wanxiang, Zhang Meishan and Liu Ting. Active learning for Chinese dependency parsing, 26th ed., vol.2. Journal of Chinese Information Processing, 2012, pp. 18-22.

[16] Hatzivassiloglou V, Mckeown K. Predicting the semantic orientation of adjectives. //Proceedings of the 8th Conference of the European Chapter of the Association for Computational Linguistics (EACL1997). Stroudsburg: ACL, 1997, pp. 174-181.

[17] Kobayashi N, Inui K and Matsumoto Y. Extracting Aspect-evaluation and Aspect-of Relations in Opinion Mining. Proc. of the 2007 Joint Conf. on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Stroudsburg : ACL, 2007, pp.1065-1074.

[18] Qiu Guang, Liu Bing, Bu Jia jun and Bu Jiajun. Opinion word expansion and target extraction through double propagation, 37th ed., vol.1. Computational Linguistics, 2009, pp. 9-27.

[19] Hou M, Teng Y, Li X, Chen Y, Zheng S, Hou M, et al. Study on the Linguistic Features of the Topic-oriented Microblog and the Strategies for its Sentiment Analysis [DB / OL]. NLPCC2012 evaluation proceedings, <http://tcci.ccf.org.cn/conference/2012/dldoc/NLPCC2012papers/workshoppapers/sen/003.pdf>, Nov 1st 2012.

[20] Wang Qian, He Ting-ting, Wen Bin, Song Le, Zhang Mao-yuan. Research on Dependency Tree-Based Chinese Sentimental Elements extraction. Journal of Research Frontiers in China computational linguistics (2007-2009). Yantai, Shandong, 2009, pp. 624-629. TSINGHUA UNIVERSITY PRESS.

[21] 2012 CCF Natural Language Processing and Chinese computing conference Chinese Microblogging Sentiment Analysis Evaluation [DB / OL]. http://tcci.ccf.org.cn/conference/2012/pages/page_04_eva.html, Sep 12th 2012.

[22] Han Zhongming, Zhang Hui and Xie Youmeng. Effective sentiment classification method based on SVM for Microblogging. BAUU-SA[DB/OL]. [Http://tcci.ccf.org.cn/conference/2012/dldoc/NLPCC2012papers/workshoppapers/sen/001.Pdf](http://tcci.ccf.org.cn/conference/2012/dldoc/NLPCC2012papers/workshoppapers/sen/001.Pdf), Nov 1st 2012.

[23] 2013 CCF Natural Language Processing and Chinese computing conference Chinese Microblogging Sentiment Analysis Evaluation Result [DB / OL]. <https://www.softconf.com/e/nlpcc2013/>. Oct 10th 2013.