

Sentiment Analysis Based on Expanded Aspect and Polarity-Ambiguous Word Lexicon

Yanfang Cao, Pu Zhang, Anping Xiong
Department of Computer Science and Technology
Chongqing University of Posts and Telecommunication
Chongqing, China

Abstract—This paper focuses on the task of disambiguating polarity-ambiguous words and the task is reduced to sentiment classification of aspects, which we refer to sentiment expectation instead of semantic orientation widely used in previous researches. Polarity-ambiguous words refer to words like “large, small, high, low”, which pose a challenging task on sentiment analysis. In order to disambiguate polarity-ambiguous words, this paper constructs the aspect and polarity-ambiguous lexicon using a mutual bootstrapping algorithm. So the sentiment of polarity-ambiguous words in context can be decided collaboratively by the sentiment expectation of the aspects and polarity-ambiguous words’ prior polarity. At sentence level, experiments show that our method is effective in sentiment analysis.

Keywords—polarity-ambiguous word; aspect; sentiment analysis

I. INTRODUCTION

In recent years, sentiment analysis became a hot research topic in the field of natural language processing due to its widely use. Previous work on sentiment analysis has covered a wide range of tasks, including polarity classification, opinion extraction[1], and opinion source assignment.

One fundamental task at word level for sentiment analysis is to determine the sentiment orientations of words. There are basically two types of approaches for word polarity recognition: corpus-based and lexicon-based approaches. Corpus-based approaches using constraints on the co-occurrence of words and statistical measures of word association in the large corpus to determine the word sentiments[2]. On the other hand, lexicon-based approaches use information about lexical relationships and glosses such as synonyms and antonyms in WordNet to determine word sentiments based on a set of seed polarity words.

Overall those two methods aim to generate a large static polarity word lexicon marked with prior polarities out of context. In fact, a word may indicate different polarities depending on what aspect it is modified, especially for the polarity-ambiguous words, such as “高|high”, which has a positive orientation in snippet “high quality” but a negative orientation in snippet “high price”. Though the quantity of polarity-ambiguous words is not large but polarity-ambiguous words cannot be avoided in a real-world application[1]. Unfortunately, polarity-ambiguous words are discarded by most research concerning sentiment analysis.

In this paper, the task of disambiguating polarity-ambiguous words is reduced to sentiment expectation of aspects. The sentiment expectation of aspects divide into two categories: positive expectation and negative expectation. A mutual bootstrapping algorithm is proposed in this paper to automatically construct the aspect and polarity-ambiguous words, utilizing relationships among aspects, polarity words and syntactic analysis. This algorithm is firstly initialized with a very small set of polarity-ambiguous words and syntactic patterns to retrieve a set of aspects. Then the sentiment expectation of an aspect is inferred utilizing the relations between aspects and polarity-ambiguous words in annotated reviews. Secondly, more polarity-ambiguous words is retrieved, utilizing the relations between aspects, syntactic patterns and annotated reviews. Finally, more syntactic patterns which are syntactic relations between aspects and polarity-ambiguous words, is retrieved. After several iterations the aspect and polarity-ambiguous word lexicon is constructed. Then the sentiment of polarity-ambiguous words in context can be decided collaboratively by the sentiment expectation of the aspects and the prior polarity of polarity-ambiguous words. At sentence level, experiments show that our method is effective.

II. RELATED WORK

Recently there has been extensive research in sentiment analysis and a large body of work on automatic SO prediction of words [2], but unfortunately they did not consider the SE of nouns in their research and regarded most of the nouns as “neutral”. Some studies try to disambiguate the polarity of the polarity-ambiguous word [3]. Some researchers exploited the features of the sentences containing polarity-ambiguous to help disambiguate the polarity of the polarity-ambiguous word. For example, intra-sentence conjunction rule in sentences from large domain corpora is taken into consideration. Many contextual information of the word within the sentence is taken into consideration, such as exclamation words, emoticons and so on [4]. In order to automatically determine the semantic orientation of polarity-ambiguous word within context, some researches reduce this task to sentiment classification of target nouns, by mine the Web using lexico-syntactic patterns to infer sentiment expectation of nouns, and then exploit character-sentiment model to reduce noises caused by the Web data [5]. A bootstrapping method to automatically discover CPs and predict sentiment expectation of nouns is proposed by Wu in order to improve both sentence and document level sentiment analysis results [6].

The disambiguation of polarity-ambiguous words can also be considered as a problem of phrase-level sentiment analysis. For example, analyze its surrounding sentences' polarities to disambiguate polarity-ambiguous word's polarity in the sentence[7]. Use a holistic lexicon-based approach to solving the problem by exploiting external evidences and linguistic conventions of natural language expressions[8]. An supervised three-component framework to expand some pseudo contexts from web is proposed by Zhao, which can obtain more useful context information to help disambiguate a collocation's polarity [9]. A set of subjective expressions to annotate the contextual polarity in the MPQA Corpus is used by Wilson[10].

III. THE PROPOSED APPROACH

A. Overview

The motivation of our approach is to disambiguate polarity-ambiguous words making full use of sentiment exception of aspects. First, a mutual bootstrapping algorithm is designed to automatically extract polarity words and aspects, utilizing relationships among aspects, polarity words and syntactic patterns. Each time, two sets of the three is fixed to constantly update the third set among aspect, polarity word and syntactic patterns. Secondly, infer the sentiment expectation of an aspect utilizing the relations between aspects and polarity-ambiguous words in annotated reviews. At the same time, more polarity-ambiguous words can be retrieved utilizing the relations between aspects and annotated reviews. Then construct two lexicons one is aspects with sentiment expectation and another is polarity-ambiguous words with prior polarity. Finally, the sentiment of polarity-ambiguous words in context can be decided collaboratively by sentiment expectation of the aspect modified by the polarity-ambiguous words and prior polarity of polarity-ambiguous words.

B. Mutual Bootstrapping Algorithm

Input: corpus S with sentence tag $SO(s)$; seeds polarity-ambiguous words (PAWs) set $W1$ and score $S2(wi)$ $wi \in W1$; syntactic patterns set $R1$ and score $S3(Ri)$, $Ri \in R1$; part-of-speech patterns set P ; iteration number M and candidate selection number $k1$, $k2$.

Output: Dic_aspect and Dic_word .

Initialize aspects set $Dic_aspect = \emptyset$; Initialize $Dic_PAWs = W1$; Initialize syntactic patterns $R_syntactic = R1$.

Tokenize each sentence $s \in C$ with lexical analysis using ICTCLAS and syntax analysis using ltp-cloud.

3.for $m = 1 \dots M$ do

4.Extract new aspects to Dic_aspect from corpus S as follows

1) For any word w in sentence $s \in S$, if $w \in Dic_PAWs$. Within the window of q words previous or behind to w :

If there is a noun phrase along with w meet patterns in P , put noun phrase into $Candi_aspect$;

If there is a noun phrase along with w meet the patterns in

$R_syntactic$, put noun phrase into $Candi_aspect$;

2) Use aspect pruning strategies to filter out error aspects in $Candi_aspect$.

3) Update aspect score $S1(Ai)$, where $Ai \in Candi_aspect$ based on (1), select the top $k1$ aspects to $A1$.

4) Infer the sentiment expectation of an aspect $a \in A1$ as follows:

For each $ap \in A1$, if $w \in Dic_PAWs$ and (ap, w) is a snippet in sentence $s \in S$.

If $SO(s)=1, SO(w)=1, SO(ap)=1, Freq(ap+)$ ++;

If $SO(s)=0, SO(w)=0, SO(ap)=1, Freq(ap+)$ ++;

If $SO(s)=1, SO(w)=0, SO(ap)=0, Freq(ap-)$ ++;

If $SO(s)=0, SO(w)=1, SO(ap)=0, Freq(ap-)$ ++;

5) If $Freq(ap+) < Freq(ap-), SO(ap)=0$, else $ap=1$. Add aspect ap with sentiment expectation into Dic_aspect and remove repeated aspects.

5.Extract new polarity-ambiguous words to Dic_word from corpus S as follows.

1) For any nouns phrase ap in sentence $s \in S$, if $ap \in Dic_aspect$.

Within the window of q words previous or behind to ap :

if there is a word along with ap is adjective or verb, put w into $Candi_word$;

if there is a word along with ap meet the patterns in

$R_syntactic$, put w into $Candi_word$;

2) Use polarity word pruning strategies to filter out error polarity words in $Candi_word$.

3) Update polarity score $S2(Wi)$, where $Wi \in Candi_word$ based on (2), select the top $k2$ words to $W2$.

4) Obtain polarity-ambiguous words from $W2$ as follows.

For each word $w \in W2$, if $ap \in Dic_aspect$ and (ap, w) is a snippet in sentence $s \in S$, $SO(w)$ is prior polarity of w in basic polarity lexicon.

If $SO(a)=0, SO(s)=1, SO(w)=0$ then w is polarity-ambiguous word;

If $SO(s)=0, SO(a)=0, SO(w)=1$ then w is polarity-ambiguous word;

5) Add polarity-ambiguous words to Dic_word and remove repeated words.

6.Extract new syntactic patterns to $R_syntactic$ as follows:

1) If $w \in Dic_word, ap \in Dic_aspect, (ap, w)$ is a snippet in sentence $s \in S$, extract syntactic pattern of w and ap to $R2$.

2) Update pattern score $S3(R_j)$, Where $R_j \in R2$, based on (3), and select top k3 patterns to the pattern set $R_syntactic$ and remove repeated syntactic patterns.

7.end for.

8.return the lexicon of Dic_aspect and Dic_word.

Using the above algorithm a number of PAWs and aspects in different domains can be abstracted. After the iterative process, incorrect PAWs and aspects may be involved in. So we'd better rectify the result manually.

Initiation

- The mutual bootstrapping begins with a seed polarity-ambiguous word set W1. W1 is grouped into two sets: positive-like adjectives (Pa) and negative-like adjectives (Na): Pa and Na are prior polarity of sentiment words in lexicon out of context, but the real positive or negative polarity in context will be evoked when they co-occur with target aspects.

Pa={ 高|high, 长|long, 重|heavy, 厚|thick, 深|deep, 多|many }

Na={ 低|low, 短|short, 轻|light, 薄|thin, 浅|shallow, 少|less }

- syntactic patterns set R1

These syntactic patterns in R1 are representative and manually selected from syntactic relations between aspects and polarity words using parsing machine. Score ranges from 1 to 10.

(1)NN<====>amod<====>JJ;(2)NN<====>nsubj<====>JJ;

(3)NN<====>doobj<====>VB;

(4)NN<====>conj_and<====>NN<====>amod<====>JJ;

(5)NN<====>doobj<====>VB <====>conj_and<====>JJ;

- part-of-speech patterns set P

These part-of-speech patterns are made up of two parts, one is the sequence of part-of-speech patterns set. We use these patterns to locate exactly target noun which is the component of an aspect modified by polarity word. Another part is the noun or verb phrase patterns set[8]. We use these patterns and the target nouns to find noun phrases or verb phrases which are candidate aspects.

As we all know an aspect consists of n characters $w=c1,c2, \dots \dots cn$, including nouns or verb. First a part-of-speech parser is applied to the reviews[14]. The noun is located as the target nouns if the tags of its surrounding consecutive words conform to any of the patterns in Fig.1 part A.

Then consecutive words including target nouns are extracted as candidate aspects from the review if their tags conform to any of the patterns in Fig.1 part B.

Part A :(1)NN+RB+JJ
(2)NN+JJ
(3)NN+VB+”的 of”+JJ
(4)JJ+NN

Fig. 1. part-of-speech patterns set P

^a. NN means nouns, RB means adverb, JJ means adjective.

- Formula in above Algorithm:

$$S1(A_i) = con(A_i) \times \log_2(Freq(A_i)) \times \sum_{R_k \in R} S3(R_k) \quad (1)$$

$$S2(W_j) = con(W_j) \times \log_2(Freq(W_j)) \times \sum_{R_k \in R} S3(R_k) \quad (2)$$

$$S3(R_k) = S1(A_i) \times S2(W_j) \quad (3)$$

$S1(A_i)$ is the score of each aspect, R is syntactic patterns set, $Con(A_i)$ (5) is the PMI of each aspect with aspect set A1, $Freq(A_i)$ is the frequency of aspect A_i in corpus, $S3(R_k)$ the score of syntactic patterns using which aspect is extracted. $Con(w_j)$ (7) is the PMI of each polarity with PAWs set W1. $Freq(w_j)$ is the frequency of w_j in corpus. $S3(R_k)$ is the score of syntactic patterns using which polarity word is extracted.

C. Sentiment Expectation of Aspects

1) Aspect pruning

Not all aspects extracted by syntactic patterns and part-of-speech patterns are useful or genuine aspects. There are also some uninteresting and redundant ones. Aspects pruning aims to remove these incorrect aspects. We use three types of pruning strategies [13].

a) word frequency filtrate: Filter out aspects with low frequency.

b) p-support (pure support): For each aspect t, assuming that the number of sentence including t is s and in these sentence the number of t alone as an aspect rather than a subset of another aspect phrase is k. So we define support=k/s, if the value of support is 0.5, then we recognize t is not a genuine aspect.

c) aspect filtrate based on PMI:

$$PMI(a, b) = N_{ab} / N_a \times N_b \quad (4)$$

$$Con(a_i) = \sum_{a_j \in Dic} PMI(a_i, a_j) \quad (5)$$

Here N_{ab} is the text number including aspects a and b. N_a is the text number only including aspect a. N_b is the text number only including aspect b. Dic is a set consists of 10 manually selected relevant aspects and product for each product domain as aspect set A.

2) Infer Sentiment Expectation of Aspects:

Sentiment expectation (SE) of aspects is divided into two categories: positive expectation and negative expectation. For a positive expectation aspect, people usually expect the thing referred to by the aspect to be bigger, higher or happen frequently. On the contrary, for a negative expectation noun, people usually expect the thing referred to by the aspect to be smaller, lower or don't happen. For example, “成本 cheng-ben|cost” is a negative expectation aspect. However, “质量 zhi-liang|quality” is a positive expectation aspect, as most people in most cases expect that their salaries become high.

The So of most snippets consists of aspects and polarity-ambiguous words can be determined by the sentiment expectation of aspects and prior polarity of the polarity-ambiguous words. If the polarity-ambiguous word has the same polarity as the SE of aspect, then the snippet has positive sentiment: if the polarity-ambiguous word has the opposite polarity to the SE of aspect, the snippet has negative sentiment. For example, snippet “成本高|high cost” has negative polarity, because the polarity-ambiguous word “高|high” has positive prior polarity opposite to the SE of aspect “成本|cost” which has negative polarity. While snippet “质量高|high quality” has positive polarity, because the polarity-ambiguous word “高|high” has positive prior polarity the same as the SE of aspect “成本|cost” which has positive polarity.

Relations among aspects, polarity-ambiguous words and snippets can be expressed by the Logic Truth Table below in Table 1. The value 1 on behalf of positive polarity and 0 on behalf of negative polarity.

TABLE I. LOGIC TRUTH TABLE

S(a)	S(w)	S(col)
1	1	1
0	0	1
0	1	0
1	0	0

Here S(a) is the SE of aspects, S(w) is the polarity of the PAWs, S(col) is the polarity of snippets. Combining the Logic Truth Table with polarity relations among aspects, PAWs and snippets, we can deduce formula as follows.

$$S^+(col)=S^+(a)\odot S^+(w)$$

$$S^+(col)=S^-(a)\odot S^-(w)$$

$$S^-(col)=S^-(a)\odot S^+(w)$$

$$S^-(col)=S^+(a)\odot S^-(w)$$

In order to derive the SE of aspects, we transform the above formulas as the following ones, which also meet the Logic Truth Table.

$$S^+(a)=S^+(col)\odot S^+(w)$$

$$S^-(a)=S^-(col)\odot S^-(w)$$

$$S^-(a)=S^+(col)\odot S^-(w)$$

$$S^-(a)=S^-(col)\odot S^+(w)$$

In the above formulas \odot means Not Exclusive Or, $S^+(w)$ means the positive category of PAWs, $S^-(w)$ means the negative category of PAWs; $S^+(a)$ means the positive sentiment expectation of aspects, $S^-(a)$ means the negative sentiment expectation of aspects; $S^+(col)$ means the positive category of snippets, $S^-(col)$ means the negative category of snippets. The polarity of snippets can be obtained by the annotated reviews. In this paper, we hold the assumption that all snippets in the same review have the same polarity as the review's. And the prior polarity of PAWs is fixed in PAWs lexicon.

Considering that one aspect may appear in different snippets and modified by different PAWs. So each aspect may have different SE in different snippets co-occurring with different PAWs. The way to accurately obtain the SE of aspects is based on statistical method. First, we extract snippets consisting of aspects and PAWs using the process in Algorithm from annotated reviews. Secondly, we compute the SE of aspect in each snippets using the formulas in Fig.2 and count the frequency of positive SE $Freq(i+)$ and negative SE $Freq(i-)$ of each aspect. Thirdly, the real SE of aspects can be calculated like this, if $Freq(i+)$ less than $Freq(i-)$, the SE of aspect is negative, otherwise the SE of aspect is positive.

D. Obtain Polarity-Ambiguous Words

The extraction of polarity word use the same syntactic patterns set $R_syntactic$ as aspects. while the part-of-speech patterns are just adjectives and verbs. We consider the adjectives surrounding aspects are candidate polarity word, but only the emotion verbs surrounding aspects are candidate polarity. The polarity of verbs can derived from the basic polarity lexicon.

1) Polarity word pruning based on PMI:

$$PMI(a, b) = N_{ab} / N_a \times N_b \tag{6}$$

$$Con(w_i) = \sum_{w_j \in Dic} PMI(w_i, w_j) \tag{7}$$

Here N_{ab} is the text number including polarity word a and polarity word b. N_a is the text number only including polarity word a. N_b is the text number only including polarity word b. Dic is set W1 used in Algorithm which contain 12 frequently used PAWs.

2) Infer Polarity-ambiguous words(PAWs)

Independent polarity word can accurately express the sentiment individually, such as “happy”, ”sad”. While the sentiment of PAWs in context should rely on the SE of aspects, such as “高|high”, “长|long”. So this is an indication of how to distinguish polarity words and PAWs. When inferring a polarity word we'd better take the snippet and aspect into consideration. If the polarity of snippet decided by the polarity individually, on the other hand the polarity of snippet agree with the polarity word, then we define the polarity word is an independent polarity word. If the polarity of snippet opposite to the polarity word, then we define the polarity word as PAWs. For example snippet “价格合理” means favorable price which has positive polarity, the same as “favorable”, so “favorable” is an independent polarity word. While snippet “价格高” means high

price, which has negative polarity, opposite to “高|high”, so “高|high” is a polarity-ambiguous word.

Special cases should be taken into consideration, when aspects are positive, the inference is not established. For example, snippet”质量好”means good quality which has positive polarity the same as “good”. As we all know “good” is an independent polarity word. While the snippet”质量高”means high quality, which also has positive polarity the same as “高|high”. But as we all know “高|high” is a polarity-ambiguous word. So the above assumption is valid when SE of the aspect is negative in snippets .This also prove the necessity of construct aspects lexicon with SE. Using this method we can find more PAWs and the polarity is its prior polarity.

IV. EXPERIMENTS AND RESULTS

A. Sentiment Analysis at Sentence Level

In order to test the performance of the PAWs lexicon and aspect lexicon constructed in this paper, we did some experiments.

1) Data and Preprocess

We collected data from popular forum sites it168, JingDong, DataTang. Reviews in different domains such as book, computer and so on are grabbed. In each domain we manually annotated 3000 positive reviews and 3000 negative reviews as train corpus, 500 positive reviews and 500 negative reviews as test corpus on sentiment analysis at sentence level. In order to concentrate on the disambiguation of PAWs, and reduce the noise introduced by the parser, we extracted sentences for test corpus containing at least one adjective and one aspect in a sentence.

The reviews were automatically word segmented and POS-tagged using the open software ICTCLAS [14].The reviews were also automatically syntactic analysed using software ltp-cloud[15].

2) Evaluation Metrics

Instead of using accuracy, we use precision (P), recall (R) and F1-value (F1) to measure the performance of sentiment analysis at sentence level. We establish the mixed matrix as shown in Table3. Mixed matrix is special to each category and it count the classification of each sentence.

TABLE II. TABLE2:MIXED MATRIX

	classified as positive	classified as negative
real positive	TP	FP
real negative	FN	TN

$$Re\ call = \frac{TP}{TP + FP} \quad (5)$$

$$Pr\ ecision = \frac{TP}{TP + FN} \quad (6)$$

$$F1 - value = \frac{2 \times Re\ call \times Pr\ ecision}{Re\ call + Pr\ ecision} \quad (7)$$

3) Methods

Our goal is not to propose a new method, but instead to test the performance of aspect and PAWs lexicon we constructed. We adopted the same algorithm with Wan (2008)[16], and we not only use Sentiment-HowNet but also NTUSD as basic polarity lexicon. But in our experiment, Intensifier_Dic didn't use.

Algorithm Compute_SO:

1) Tokenize each sentence $s \in S$ into word set W_s and PAWs;

2) For any word w in a sentence $s \in S$, compute its value $SO(w)$ as follows:

1) if $w \in PAWs$, compute $SO(w)$

a) In baseline1 method only use the PAWs lexicon

If $SO(w)=1$, $SO(w)=Dy_PosValue$;

If $SO(w)=0$, $SO(w)= Dy_NegValue$

b) In baseline2 method ,use the PAWs lexicon and aspect lexicon constructed by Zhou[18]

Within the window of q words previous or behind to w , if there is a term $a \in$ aspect lexicon.

If $SO(w)=1,SO(a)=1,SO(w)=Dy_PosValue$;

If $SO(w)=1,SO(a)=0,SO(w)=Dy_NegValue$;

If $SO(w)=0,SO(a)=1,SO(w)=Dy_NegValue$;

If $SO(w)=0,SO(a)=0,SO(w)=Dy_PosValue$;

c) In our method , use the PAWs lexicon and aspect lexicon constructed by this paper.

Within the window of q words previous or behind to w , if there is a term $a \in$ aspect lexicon A.

If $SO(w)=1,SO(a)=1,SO(w)=Dy_PosValue$;

If $SO(w)=1,SO(a)=0,SO(w)=Dy_NegValue$;

If $SO(w)=0,SO(a)=1,SO(w)=Dy_NegValue$;

If $SO(w)=0,SO(a)=0,SO(w)=Dy_PosValue$;

2)If $w \in Positive_Dict$, $SO(w)=PosValue$;

3)If $w \in Negative_Dict$, $SO(w)=NegValue$;

4)Otherwise, $SO(w)=0$;

5)Within the window of q words previous to w , if there is a term $w' \in Negation_Dict$.

$SO(w) = -SO(w)$;

3. $S(s) = \sum_{w \in W_s} SO(w)$

a) *Baseline1*: Not considering the context, assign all positive-like adjectives as positive, and all negative-like adjectives as negative.

b) *Baseline2*: Use aspect and PAWs lexicon constructed by Zhou[17]

c) *Our method*: Use aspect and PAWs lexicon lexicon constructed by this paper.

B. Result:

The performance of sentiment classification of product reviews in two domains which is book and computer was significantly improved. In each domain we use 500 positive reviews and 500 negative reviews, The result is shown in Table3:

TABLE III. THE EXPERIMENTAL RESULTS AT SENTIENCE LEVEL

		Baseline1		Baseline2		Our method	
		B	C	B	C	B	C
Pos	Pre	0.6740	0.7165	0.6899	0.7401	0.7123	0.7540
	Rec	0.7650	0.8310	0.7902	0.8645	0.8050	0.8990
	F1	0.7166	0.7696	0.7366	0.7974	0.7558	0.8201
Neg	Pre	0.7283	0.7797	0.7543	0.8284	0.7759	0.8687
	Rec	0.6301	0.6453	0.6450	0.6829	0.6750	0.6950
	F1	0.6756	0.7062	0.6954	0.7487	0.7219	0.7722
Aver		0.6916	0.7379	0.7160	0.7731	0.7539	0.7961

^b. B means book,C means computer, aver is the average of F1

Adding the disambiguation of PAWs, our method obviously outperforms the baseline1, especially in computer reviews. People usually use more PAWs in smart devices reviews. But in book domain the classification result is lower than in other domain, because in book reviews less PAWs and negative aspects are used. In some book reviews, there are words just describing the content of books which disturbs classify of reviews. And the classification result in negative reviews is lower than positive reviews. This is because in positive reviews people usually use more independent polarity words to express their emotion. While in negative reviews people tend to describe the property of products more frequently rather than express their emotion, so less independent polarity words are used. Our method also outperforms the baseline2 just a little bit, which prove that our method can recognize more PAWs and aspects with SE, though the quantity is not large.

V. CONCLUSION

This paper presents a mutual bootstrapping algorithm to construct aspect lexicon and polarity-ambiguous lexicon in order to disambiguate the polarity-ambiguous words in the context. When a polarity-ambiguous word appears in a

sentence, firstly extract the aspect around the PAWs, then find it's SE in aspect lexicon and find the polarity of the PAWs in Polarity-Ambiguous Word lexicon, finally compute the real polarity of the PAWs in sentence using the SE of aspect and prior polarity of the PAWs. For the sentiment analysis at sentence level, our method achieves promising result that is significantly better than baseline and automatically extract more polarity-ambiguous words rather than only 14 polarity words used in baseline. On the other hand compared to others manual extract methods, our method automatically extract aspects and polarity words which reduce the manually work and achieve obvious improvement in performance. This validates the effectiveness of our approach.

There leaves room for improvement. In this paper method of extracting the aspects and polarity words always generate some noises, so find out new methods to reduce noises is our future work. The mutual bootstrapping algorithm in this paper need annotated reviews which bring in manual operation. SO discover efficient unsupervised method without manual operation in inferring the SE of aspects and construct aspect and PAWs lexicon is the future work.

REFERENCES

- [1] Bo Pang and Lillian Lee, A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts, Proceedings of ACL 2004.
- [2] Peter Turney and MichaelL.Littman.2003. Measuring praise and criticism: Inference of semantic orientation from association. ACM TOIS, 21(4):315-346.
- [3] Ding, X., Liu, B. and Yu, P. 2008. A holistic lexicon-based approach to opinion mining. In Proceedings of WSDM'08.
- [4] Yasuhiro Suzuki, Hiroya Takamura, and Manabu Okumura. 2006. Application of semi-supervised learning to evaluative expression classification. In Computational Linguistics and Intelligent Text Processing, pages 502-513.
- [5] Wu,Y and Wen, M. 2010. Disambiguating Dynamic Sentiment Ambiguous Adjectives. In Proceedings of COLING2010.
- [6] Wen, M and Wu,Y.2011. Mining the Sentiment Expectation of Nouns Using Bootstrapping Method. Proceedings of the 5th IJCONL2011, 1423-1427.
- [7] Hu, Minqing and Bing Liu. 2004. Mining and summarizing customer reviews. In Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2004 (KDD-2004), pages 168-177, Seattle, Washington.
- [8] Ding, X, Liu, B. and Yu, P. 2008. A holistic lexicon-based approach to opinion mining. In Proceedings of WSDM'08.
- [9] Chunliang LI, Yanhui Zhu, Yeqiand Xu.Product reviews in Chinese word extraction method of research[J].Computer Engineering,2011,37(12):26-32
- [10] Yanyan Zhao, Bing Qin and Ting Liu. Collocation Polarity Disambiguation Using Web-based Pseudo Contexts. In Proceedings of the 2012 Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL 2012). 2012.07. Jeju, Republic of Korea. (long, oral)
- [11] Theresa Wilson, Janyce Wiebe, Paul Hoffmann. Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level. Computer Linguistics,25(3),399-433
- [12] Turney P. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews[C]Proceedings of ACL'02. Pennsylvania: ACL, 2002: 417-424.
- [13] Liu Hongyu,Zhao Yanyan,Qin B,Liu T. Aspect Extract and Analysis[J]. Journal of Chinese information.2010,24(1):84-89.
- [14] ICTParser: <http://nlp.ict.ac.cn/demo/ictparser/index.php>.
- [15] Harbin industrial university language technology platform.IR: <http://ir.hit.edu.cn/demo/ltp/>.

- [16] Wan, X. 2008. Using Bilingual Knowledge and Ensemble Techniques for Unsupervised Chinese Sentiment Analysis. Proceedings of EMNLP'08.
- [17] Zhou C, Sentiment analysis technology research for Chinese Web comments [D]. National University of Defense Technology. 2011.