SESS: A Self-Supervised and Syntax-Based Method
for Sentiment Classification

Weishi Zhang[a,b], Kai Zhao[b], Likun Qiu[b,c], and Changjian Hu[b]

[a]School of Software, Tsinghua University, Beijing, China,
zhang-ws08@mails.tsinghua.edu.cn
[b]NEC Laboratories, China,
{zhaokai, qiulikun, huchangjian}@research.nec.com.cn
[c]Department of Chinese Language and Literature, Peking University, China.

Abstract. This paper presents a method for sentiment classification, called SESS (**SE**lf-**S**upervised and **S**yntax-Based method). SESS includes three phases. Firstly, some documents are initially classified based on a sentiment dictionary, and then the sentiments of phrases and documents are iteratively revised. This phase provides some accurately labeled data for the second phase. Secondly, a machine learning model is trained with the labeled data. Thirdly, the acquired model applies on the whole data set to get the final classification result. Moreover, to improve the quality of labeled data, the affect of compound and complex sentences on clause sentiment is examined. For three types of compound and complex sentences, i.e., coordination, concession or condition sentence, the clause sentiment is revised accordingly. Experiments show that, as an unsupervised method, SESS achieves comparative performance to state-of-the-art supervised methods on the same data.

Keywords: Self-supervised, syntax-based, opinion mining, sentiment classification.

1   Introduction

The task of sentiment classification is: given an opinionated piece of text, classify the opinion as falling under one of two opposing sentiment polarities (positive or negative) (Pang and Lee, 2008). The "piece of text" can refer to either a sentence or a document. In this paper, it refers to document, e.g., a movie review or a product review.

Generally, there are two types of approaches tackling the sentiment classification task: supervised (Dave *et al*., 2003; Yu and Hatzivassiloglou, 2003; Aue and Gamon, 2005; Read, 2005) and unsupervised (Pang, 2002; Turney, 2002; Gamon and Aue, 2005; Zagibalov and Carroll, 2008a). Supervised approaches usually employ machine learning methods to train a model based on some human-labeled data, and then apply the acquired model on the new data. On the contrary, unsupervised approaches usually employ a list of sentiment words, e.g., a sentiment dictionary or some seed words, to help decide the sentiment polarity of documents. Supervised approaches generally achieve better performance than unsupervised ones, because methods such as SVM or Naïve Bayes have been deeply studied in machine learning area, and the human-labeled data reveal a lot of clues about human classification. However, as a double-edged sword, human-labeled data also bring the disadvantage of domain-dependence. Although some researches have been done on domain adaptation (Aue and Gamon, 2005; Blitzer *et al*., 2007), the problem is far from resolved.

In this paper, a self-supervised method is proposed to share both the power of machine learning methods and the domain-independence property. The method is referred to as SESS

(**SE**lf-**S**upervised and **S**yntax-Based method). SESS takes three steps. Firstly, an unsupervised method is used on the data to label some documents (i.e., decide their sentiment polarities). Secondly, a machine learning method applies on these labeled documents to train a model. Thirdly, the model is used to label all the documents (notice that a document may change its label acquired in the first step to another one after the third step). SESS makes use of machine learning methods without need of any human-labeled data.

To ensure that the machine learning method can achieve good performance in the third step, the unsupervised method must provide accurately labeled documents in the first step. To satisfy that requirement, SESS makes two special designs in the first step. First, an iterative procedure is used to decide the polarities of documents and words. A general method may use a sentiment dictionary to decide the document polarity. However, the words in the dictionary may not be comprehensive and the sentiment of those words may not fit for current data set. The iterative method can find new sentiment words that are not in the dictionary and revises the polarity of words according to current data set. Second, the polarities of documents and words are revised by analyzing the relation of clauses of compound and complex sentences in documents. Particularly, seven types of compound and complex sentences are analyzed, while three of them, i.e., coordination (discourse markers such as *and* or *in addition*), concession (discourse markers such as *but* or *however*) and condition (discourse markers such as *if*) sentences, take effect on sentiment of clauses. The detailed effects of these sentences are examined.

The experiments show that SESS achieves an overall F_1-score of 81.7% on data sets of four domains, which is comparative to 83.3%, the best result of the supervised approach in previous studies (Li and Zong, 2008) on the same data set.

The rest of this paper is organized as follows. Section 2 surveys related work. The overview of our approach is presented in Section 3. Section 4, 5 and 6 describe the details of the SESS model. Experiments are shown in Section 7. The final section gives conclusions and proposes future work.

2   Related Work

Standard machine learning technologies such as SVM and Naïve Bayes are usually used by supervised approaches (Alpaydin, 2004). Different factors affecting the machine learning process were investigated. For instance, linguistic, statistical and n-gram features are used in (Dave *et al.*, 2003). Semantically oriented words are utilized to identify polarity at the sentence level (Yu and Hatzivassiloglou, 2003). Selected words and negation phrases are investigated in (Na *et al.*, 2004). Such approaches work well in situations where large labeled corpora are available for training.

But the performance of supervised approaches generally decreases when training data are insufficient or acquired from a different domain (Aue and Gamon, 2005; Read, 2005). To solve that problem, unsupervised or weakly supervised methods can be used to take advantage of a small number of annotated in-domain examples and/or unlabelled in-domain data. For instance, Aue and Gamon (2005) train a model on a small number of labeled examples and large quantities of unlabelled in-domain data. In (Blitzer, 2007), structural correspondence learning is applied to the task of domain adaptation for sentiment classification of product documents. Li and Zong (2008) integrate training data from multiple domains.

Unsupervised approaches usually assume that there are certain words people tend to use to express strong sentiment, so that it might suffice to simply produce a list of such words by introspection and rely on them alone to classify the documents. Pang (2002) checked this assumption by asking human to read movie reviews, selecting ten to twenty sentiment words (like *fantastic* or *terrible*), and using them to classify reviews. The results show that such a method performs worse than supervised models built on sufficiently large training sets in the movie review domain.

Later, such human-given-word-list method is extended. The words given by human are considered as seed words, and other sentiment words in the documents are picked out by some kind of "similarity" between the words and the seed words. Turney (2002) selected two seed words, *excellent* and *poor*. For every phrase in a document, the mutual information between the phrase and the two words are computed respectively, to reveal that the phrase is more positive (like *excellent*) or more negative (like *poor*). A document is classified as positive if the average sentiment of all its phrases is positive, and vice versa.

Fewer seed words imply less domain-dependency. Zagibalov and Carroll (2008a) select only one word *good* as seed positive word, and use negation words such as *not* to find initial negative expressions. In (Zagibalov and Carroll, 2008b), even the one word *good* is ignored. Instead, seed words are automatically generated based on a linguistic pattern which is called negated adverbial construction like *not very good*. In such way, the problem of domain dependency is completely avoided.

Discourse markers such as *but, and, or* have been explored to identify sentiment polarity of adjectives. Usually two adjectives have different sentiment polarity if they are connected by *but*, e.g., *elegant but over-priced*, while they usually have the same sentiment polarity if connected by *and*, e.g., *clever and beautiful*. Utilizing this property, Hatzivassiloglou and McKeown (1997) cluster the adjectives collected from a domain corpus and decide the sentiment polarity of these adjectives in that domain. Yao and Lou (2007) improved that method by combining it with the idea of Turney (2002). This paper is different from the rest in that, we concern with discourse markers between clauses, but not adjectives, and integrate the discourse marker analysis in the process of document polarity decision.

3 Overview of Our Approach

Figure 1 shows the flow chart of SESS. In phase 1, an unsupervised approach applies on the original data to automatically label some data. In phase 2, a supervised approach applies on the labeled data to acquire a model. In phase 3, the model applies on the original data to do classification.

In phase 1, the unsupervised approach adopts the method of (Zagibalov and Carroll, 2008b). That method initially selects some seed sentiment words automatically, and assigns initial sentiment score for those words. Then it employs an iterative procedure to update both the sentiment score of words and polarity of documents. First, the sentiment of sentences and documents is decided based on the sentiment words (and phrases) they have. A sentence is decided positive if the sum of the sentiment score of words (and phrases) in the sentence is greater than zero, and negative if the sum is less than zero. A document is judged as positive if it has more positive sentences than negative ones. Second, the sentiment score of words (and phrases) is updated according to the judged polarity of documents. Basically, if a word (or a phrase) occurs in more positive documents than negative documents, it is judged as positive, and the score is computed as the difference of the number of positive documents and negative documents the word occurs in.
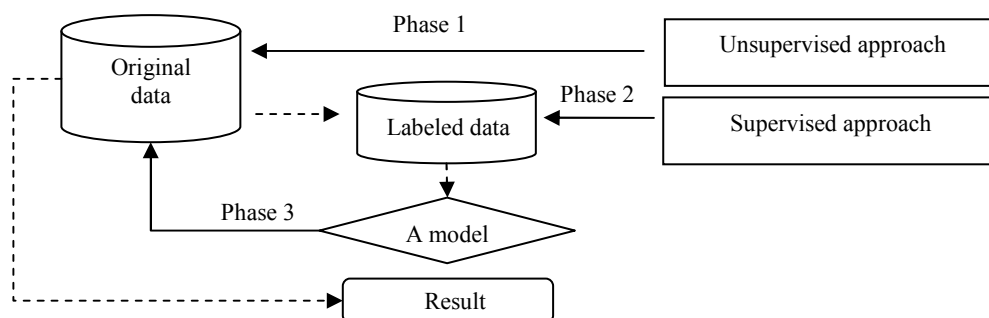
Figure 1: Flow chart of SESS (dash-lines refer to results of phases).

In this paper, we made three improvements over the method of (Zagibalov and Carroll, 2008b). First, positive/negative ratio control is introduced in the iterative procedure. Second, a sentiment dictionary is used to initialize seed words. Third, compound and complex sentences are examined to revise the sentiment of clauses and documents.

The supervised method in phase 2 requires the training data to be both adequate and precisely labeled. We can imagine that, if only a small percentage of data are labeled, or very low precision is acquired on the labeled data, the supervised method surely suffers bad performance no matter how powerful the method is. However, the quality of data labeling cannot be controlled as no human-labeled data is available. Therefore, what can be controlled here is only the amount of labeled data. Since the unsupervised approach takes an iterative style, it labels more and more documents when the iteration goes on. To make the control, a point is set on the percentage of the labeled data. In the experiment, we select the golden mean. That is, if 61.8% of documents have been labeled, the iteration procedure completes. And the labeled 61.8% of documents are provided as the training data of phase 2.

In phase 2, Naïve Bayes is selected as the realization of supervised approach. As a widely used method, Naïve Bayes achieves good performance in many areas. But in fact, the performance of phase 3 depends much more on the quality of labeled data provided by phase 1, while less on the particular machine learning method.

4   The Unsupervised Approach of SESS

The basic method of phase 1 adopts the method of (Zagibalov and Carroll, 2008b). The method keeps two lists, i.e., a sentiment vocabulary list and a sentiment document list. The sentiment vocabulary list is initialized by some seed words (see Initialization Step in the following). Then the sentiment vocabulary list is used to identify polarity of documents (Step 1), and the result is saved in the sentiment document list. Further, the sentiment document list is checked to reversely update the sentiment vocabulary list (Step 2). Such an iterative procedure completes when both lists remain unchanged. Our improvements on that method are introduced in Initialization Step and Step 1.

4.1   Initialization Step

Zagibalov and Carroll (2008b) identify seed sentiment words in the way: if a word such as *good*, occurs more frequently in the documents than its negated adverbial form such as *not very good*, then the word is judged as positive. That method has the advantage of domain independence, while the disadvantage is that only positive words can be found and no negative words are found. One consequence is that the method tends to classify negative documents as positive, because knowledge about negative expression is insufficient. To overcome that problem, we use a general sentiment dictionary to initialize the seed sentiment words. Since a sentiment dictionary contains a lot of positive words, as well as negative words, the bias on classification can be greatly lightened. In addition, since the general sentiment dictionary is applicable to many domains, the advantage of domain independence is still remained.

The sentiment vocabulary list, denoted by V_{sen}, maintains a list of items, each of which is a unigram or bi-gram, and assigned with a sentiment score. In the initialization step, +1 score is assigned to positive words while -1 for negative words. Some dictionaries such as Subjclueslen1-HLTEMNLP05 provide sentiment strength information, e.g., *great* is strong while *feasible* is weak. In such case, 1 is assigned to strong words while 0.5 to weak ones (for both polarities).

4.2   Step 1: Identify the Sentiment of Documents

To compute the polarity of a document D, for each item $w \in D$ and $w \in V_{sen}$, weight its score in D by the following formula:

$$S_w = \frac{L_w}{L_d} S_v N \tag{1}$$

where L_w denotes the length of w, L_d the length of D, S_v the sentiment score of w in V_{sen}, and N is -1 if an negation word precedes w, or 1 if none negations.

Divide D into clauses by comma and full stop. The sentiment score of a clause c, denoted by $CS(c)$, is defined as $CS(c)=\sum S_w$, for all $w \in c$. A clause c is positive if $CS(c)>0$ or negative if $CS(c)<0$. Zagibalov and Carroll (2008b) classify D to positive if it contains more positive clauses than negative ones. Then those documents compose of the sentiment document list.

We found a disadvantage of this method. Since there are usually different amount of classified positive and negative documents, when items are updated in step 2, their scores (S_v) may be biased. In detail, for formula (3), if there are more classified positive documents than negative ones, then F_p may be bigger than the value it should be. To overcome that bias, a ratio control is designed, which requires the number of positive and negative documents in the sentiment document list to be the same.

Denote the number of positive and negative documents in one round of iteration as $DN_{positive}$ and $DN_{negative}$ respectively. To realize the ratio control, first, rank all documents according to their sentiment score, denoted by $DS(D)$, where $DS(D)=\sum CS(c)$, for all $c \in D$. Second, take the smaller one of $DN_{positive}$ and $DN_{negative}$, i.e., $Min(DN_{positive}, DN_{negative})$, as a threshold, remain the positive and negative documents above the threshold in the sentiment document list, and remove others. To make the process stricter, a weight α, where $0<\alpha \leqslant 1$, can be added to the threshold of $Min(DN_{positive}, DN_{negative})$.

Figure 2 shows the whole process to classify the documents with ratio control. Those documents form the sentiment document list.

---

1. Let $DN_{min}=Min(DN_{positive}, DN_{negative})*\alpha$ $(0<\alpha \leqslant 1)$.
2. Rank all documents in descending order by their DS.
3. Document labeling:
    3.1 Label the top DN_{min} documents in the ranking list as positive.
    3.2 Label the tail DN_{min} documents in the ranking list as negative.
    3.3 Others are left unclassified.

Figure 2: Document sentiment classification with ratio control.

4.3 Step 2: Update the Sentiment Vocabulary List

For an item w, denote the number of positive documents containing w as F_p, and the number of negative documents containing w as F_n. Preceding by a negation makes the account reduce by one. E.g., if "not good" is found in a negative document, then $F_n= F_n-1$ for *good*. The idea of updating V_{sen} is: if F_p is much bigger than F_n, then w is very likely to be a positive item, and vice versa. The following formula is designed as a measure.

$$\text{DIF(w)} = \frac{\left| F_p - F_n \right|}{\left(F_p + F_n \right)} \tag{2}$$

If $DIF(w) \geqslant 1$, w is included in V_{sen} (current items in V_{sen} will be removed if they no longer satisfy this condition). The sentiment score of w is updated as

$$S_v(w) = F_p - F_n \tag{3}$$

4.4   Iteration Control

The unsupervised approach iterates between step 1 and 2. In (Zagibalov and Carroll, 2008b), the iteration completes when both V_{sen} and the sentiment document list do not change. Generally, when the iteration completes, almost all the documents are classified. For SESS, since the goal of the unsupervised approach is to provide accurately labeled data, it is not necessary to label that many documents. In addition, generally, more documents are classified, lower accuracy of the classification is acquired, for the errors generated in the former rounds of iteration will propagate to the following ones. Therefore, the iteration should complete at some early point of iteration. However, the iteration cannot complete too early, because the supervised approach still needs adequate data to train the model. A parameter β is set, where $0<\beta<1$. When $\beta*100$ percent of documents have been labeled, the iteration completes. In the experiments, β is set as 0.618 (i.e., golden mean).

5   Syntax-based Approach of SESS

In phase 1 of SESS, the sentiment score of a document is calculated as the sum of clause score of the document. But the relation of clauses was neglected. For instance, considering the following sentence,

The concept is a great one, but it's mostly a waste of time.

The former clause takes positive polarity while the latter one takes negative. If the former $CS(c)$ has bigger absolute value than the latter one, the whole effect of these two clauses on the document is positive. However, since the sentence emphasizes on the latter part, the effect should be negative. Considering that hint, the polarity of the former clause should be reversed (change positive to negative). This example reveals that, to correctly compute the sentiment of a document, the relation of clauses should be examined.

There are mainly seven types of compound and complex sentences, which are listed in Table 1. Among them, only three ones have effect on the sentiment of clauses, i.e., coordination, concession and condition sentences. Particularly, first, the polarity of two clauses of a coordination sentence should be consistent. If not, there may be an error in it. One principle to identify the error could be: the polarity of the clause having the smaller absolute value of $CS(c)$ is more likely to be an error. It should be adjusted to keep consistent with the polarity of the other clause. Second, a concession sentence usually emphasizes on the latter clause. Therefore, the polarity of the former clause should be reversed. Third, a condition sentence generally talks about an assumption. Thus, it should be ignored in the calculation of document sentiment. Figure 3 shows the realization.

The type of a compound or complex sentence is identified by discourse markers listed in Table 1. Figure 4 shows how to revise the frequency of documents an item occurring in step 2 of phase 1.

Table 1: Discourse markers of compound/complex sentences.

Type of compound/complex sentences	Discourse Markers
Coordination	*and, in addition, what's more, moreover*
Concession	*however, but, though, although*
Condition	*if, even if, in case, as long as, once*
Time/place	*When, while, after, since, where, wherever*
Purpose	*in order to, so that*
Result	*in order that, therefore, so, such that*
Reason	*because, since, as, for, now that*

For an adjacent clause-pair $<Cl_1, Cl_2>$,
1. If it is a coordination sentence with $CS(Cl_1) * CS(Cl_2) < 0$, denote i as the index of the clause whose absolute value of $CS(c)$ is bigger than the other one, and set $CS(Cl_1)= CS(Cl_2)=CS(Cl_i)$.
2. If it is a concession sentence, set $CS(Cl_1)= - CS(Cl_1)$.
3. If it is a condition sentence, set $CS(Cl_1)= CS(Cl_2)= 0$.

Figure 3: The revision of sentiment score of clauses (in step 1 of phase 1).

For an adjacent clause-pair $<Cl_1, Cl_2>$,
1 If it is a coordination sentence, calculate F_p or F_n as before (without revision) for any item $w \in Cl_1$ or Cl_2.
2 If it is a concession sentence, then
2.1   If current document is judged to be positive, set $F_p = F_p - 1$ for any item $w \in Cl_1$.
2.2   If current document is judged to be negative, set $F_n = F_n - 1$ for any item $w \in Cl_1$.
3 If it is a condition sentence, do not account current document in either F_p or F_n for any item $w \in Cl_1$ or Cl_2.

Figure 4: The revision of frequency of documents for an item (in step 2 of phase 1).

6   Supervised Approach of SESS

Naïve Bayes is chosen as the machine-learning method in phase 2. The items in sentiment vocabulary list V_{sen} in phase 1 are taken as features, which are weighted by TFIDF.

7   Experiments

7.1   Data and Tools

Experiments are carried out on the data set of reviews of four domains: books, dvds, electronics, and kitchen appliances[1]. All the documents are in English. Each domain contains 1,000 positive and 1,000 negative documents.

The sentiment dictionary required in the initialization step of phase 1 takes Subjclueslen1-HLTEMNLP05[2] Sentiment Dictionary, which contains 2294 positive words and 4146 negative words.

WEKA 3.4.11[3] is used as the implementation of Naïve Bayes classifier.

7.2   Results of the Supervised Methods

In (Li and Zong, 2008), the data in each domain are partitioned randomly into training data and testing data, with the portion of 70% and 30% respectively. The development data are used to train a meta-classifier. Four types of feature selection are used, i.e., 1Gram, 2Gram, 1+2Gram and 1Gram+2Gram. The 1Gram method takes unigram as feature candidates and optimally selects features by Bi-Normal Separation (BNS) method. The 2Gram and 1+2Gram are similar. For 1Gram+2Gram, it adopts the selected features in both 1Gram and 2Gram.

---

[1] This data set is collected by Blitzer *et al*. (2007): http://www.seas.upenn.edu/~mdredze/datasets/sentiment/

[2] OpinionFinder's Subjectivity Lexicon: http://www.cs.pitt.edu/mpqa/

[3] http://www.cs.waikato.ac.nz/˜ml/weka

7.3 Results of The SESS Model

In phase 1 of SESS, α is set as 0.618 (i.e., golden mean) for the first round of iteration, and 1.0 for the following rounds. β is set as 0.618. The following negation words are used:

> {*not, no, none, nothing, nor, neither, never, hardly, seldom, don't, doesn't, didn't, isn't, wasn't', aren't, weren't, won't, wouldn't, can't, cannot, couldn't*}

Table 2 shows the results of unsupervised, supervised and self-supervised methods. The result of unsupervised method is acquired in phase 1, while the iteration completes until sentiments of both V_{sen} and the sentiment document list do not change (phase 2 and 3 are ignored). The dictionary-based self-supervised method is different from SESS in that, in phase 1, the sentiment dictionary is directly used to label documents. That is, no iteration is taken. Then half top-ranked positive and negative documents are chosen to form the training data. Phase 2 and 3 are the same.

Table 2 shows that, 1) self-supervised and supervised methods are better than unsupervised one; 2) for self-supervised method, SESS is better than dictionary-based one; 3) SESS is better than three supervised methods while worse than one (1Gram+2Gram). Notice that the performance of supervised methods is achieved on 30% document, while SESS on the whole set of documents.

Different settings of β in SESS are examined. Table 3 shows that the best performance is achieved when β is set as 0.618 among the three settings: 0.5, 0.618 and 0.8. Generally the unsupervised method takes two or three rounds of iteration to label those data.

Table 2: Results of unsupervised, supervised and self-supervised approaches.

Methods		F_1				
		Books	Dvds	Electronic	Kitchen	*Average*
Unsupervised		72.4	71.5	71.2	74.4	72.4
Supervised	1Gram	75.0	84.0	80.0	82.5	80.4
	2Gram	75.0	73.0	81.5	78.5	77.0
	1+2Gram	76.5	81.0	82.5	80.0	80.0
	1Gram+2Gram	79.0	84.5	85.0	84.5	83.3
Self-supervised	Dictionary-based	77.6	77.4	80.3	82.5	79.5
	SESS	**79.4**	**80.1**	**82.6**	**84.8**	**81.7**

Table 3: SESS with different settings of β.

Domains	F_1		
	β= 0.5	β= 0.618	β= 0.8
Books	78.6	79.4	77.2
Dvds	78.2	80.1	76.7
Electronic	81.2	82.6	80.9
Kitchen	84.5	84.8	83.0
Average	**80.6**	**81.7**	**79.5**

7.4 The Improvement of Syntax-based Approach in SESS

The analysis of three compound and complex sentences takes effect on the performance of SESS simultaneously. To check their individual effect, five variant models were implemented. They are referred to as V_1, V_2, V_3, V_4 and V_5 respectively. In V_1, all the revision on three types of sentences is removed. In V_2, only the revision on coordinate sentence is remained. In V_3,

only the revision on condition sentence is remained. In V_4, only the revision on concession sentence is remained. In V_5, all the revisions are remained. Table 4 shows that all the three types of revision totally achieve 3.7% F_1-score improvement (from V_1: 78.0% to V_5: 81.7%). Among them, the revision on concession sentences plays the most important role (from V_1: 78.0% to V_4: 81.2%), condition as the second (from V_1: 78.0% to V_3: 78.7%), while coordination as the least (from V_1: 78.0% to V_2: 78.4%).

Compare the result of V_1 to Table 2, we can conclude that, for SESS, self-supervise way achieves 5.6% improvement over unsupervised way (from unsupervised: 72.4% to V_1: 78.0%), while the analysis on compound and complex sentences contributes 3.7% more (from V_1: 78.0% to V_5: 81.7%).

Table 4: Results of individual effect of three types of discourse markers (F_1=P=R).

Domains	F_1				
	V_1	V_2	V_3	V_4	V_5
Books	74.8	74.8	75.7	78.4	79.4
Dvds	75.9	76.5	76.1	79.6	80.1
Electronic	78.7	78.9	80.0	82.4	82.6
Kitchen	82.7	83.3	83.1	84.5	84.8
Average	**78.0**	**78.4**	**78.7**	**81.2**	**81.7**

8 Conclusion and Future Work

SESS is proposed in this paper to tackle the task of document sentiment classification. It uses an unsupervised method to automatically label some data, train a machine learning model with the labeled data, and classify the whole data by the acquired model. The contributions of this paper are: 1) propose a self-supervised method to do document sentiment classification; 2) use an iteration method to provide accurately labeled data for training; 3) improve the iteration method by revising clause sentiment of three types of compound and complex sentences.

Experiments show that SESS achieves performance better than the unsupervised method, the dictionary-based self-supervised method, and three supervised methods. It is just a little worse than a specially designed supervised method (1Gram + 2Gram).

In the future, there are still several avenues to be explored. Firstly, the use of linguistic knowledge in sentiment classification needs further study. For instance, the word "great" in the phrase "a great deal" is currently considered as sentiment word, but it contains no sentiment. Secondly, the none-opinioned part of a document should be separated from the opinioned part. For example, currently, the content of a book talking about a tragedy story mixes with the opinion of the reviews talking about happy feeling of the book, which introduces errors.

References

Alpaydin, E. 2004. *Introduction to Machine Learning*. The MIT Press, Cambridge.

Aue, A. and M. Gamon. 2005. Customizing sentiment classifiers to new domains: a case study. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, Borovets, BG.

Blitzer, J., M. Drezde and F. Pereira. 2007. Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pp. 440-447, Prague, CZ.

Cui, H., V. Mittal and M. Datar. 2006. Comparative experiments on sentiment classification for online product documents. In *Proceedings of the 21st International Conference on Artificial Intelligence*, pp.1265-1270, Boston, MA.

Dave, K., S. Lawrence and D. M. Pennock. 2003. Mining the Peanut gallery: opinion extraction and semantic classification of product documents. In *Proceedings of WWW2003*, pp. 519-528, Budapest, HU.

Gamon, M. and A. Aue. 2005. Automatic identification of sentiment vocabulary: exploiting low association with known sentiment terms. In *Proceedings of the ACL-05 Workshop on Feature Engineering for Machine Learning in Natural Language Processing*, Ann Arbor, US.

Hatzivassiloglou, V. and K. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the Joint ACL/EACL Conference*, pp. 174-181.

Li, S. S. and C. Q. Zong. 2008. Multi-domain Sentiment Classification. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*, pp. 257-260.

Na, J.C., H. Sui, C. Khoo, S. Chan and Y. Zhou. 2004. Effectiveness of simple linguistic processing in automatic sentiment classification of product documents. In I.C. McIlwaine, ed., *Knowledge Organization and the Global Information Society: Proceedings of the Eighth International ISKO Conference*, pp. 49-54, Wurzburg, Germany: Ergon Verlag.

Pang, B., L. Lee and S. Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of Conference on Empirical Methods in Natural Language Processing 2002,* pp.79-86.

Pang, B. and L. Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.

Read, J. 2005. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL-2005 Student Research Workshop*, Ann Arbor, MI.

Tan, S. B., G. W. Wu, H. F. Tang and X. Q. Cheng. 2007. A Novel Scheme for Domain-transfer Problem in the context of Sentiment Analysis. In *Proceedings of CIKM 2007*. pp.979-982.

Turney, P. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of documents. In *Proceedings of the 40th Annual Meeting of the Association of Computational Linguistics*, pp. 417-424.

Yao, T. F. and D. C. Lou. Research on Semantic Orientation Distinction for Chinese Sentiment Words. *ICCC2007*.

Yu, H. and V. Hatzivassiloglou. 2003. Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences. In *Proceedings of Conference on Empirical Methods in Natural Language Processing 2003*.

Zagibalov, T. and J. Carroll. 2008a. Unsupervised Classification of Sentiment and Objectivity in Chinese Text. In *Proceedings of the Third International Joint Conference on Natural Language Processing*, pp. 304-311.

Zagibalov, T. and J. Carroll. 2008b. Automatic Seed Word Selection for Unsupervised Sentiment Classification of Chinese Test. In *Proceedings of the 22nd International Conference on Computational Linguistics*, pp. 1073-1080.