Sentiment Classification with Polarity Shifting Detection

Shoushan Li[†‡] Zhongqing Wang[†] Sophia Yat Mei Lee[‡] Chu-Ren Huang[‡]

[†]Natural Language Processing Lab
Soochow University
Suzhou, China
{shoushan.li, wangzq870305}@gmail.com

[‡]Department of Chinese and Bilingual Studies
The Hong Kong Polytechnic University
Hong Kong
{churenhuang, sophiaym}@gmail.com

***Abstract*— Sentiment classification is now a hot research issue in the community of natural language processing and the bag-of-words based machine learning approach is the state-of-the-art for this task. However, one important phenomenon, called polarity shifting, remains unsolved in the bag-of-words model, which sometimes makes the machine learning approach fails. In this study, we aim to perform sentiment classification with full consideration of the polarity shifting phenomenon. First, we extract some detection rules for detecting polarity shifting of sentimental words from a corpus which consists of polarity-shifted sentences. Then, we use the detection rules to detect the polarity-shifted words in the testing data. Third, a novel term counting-based classifier is designed by fully considering those polarity-shifted words. Evaluation shows that the novel term counting-based classifier significantly improves the performance of sentiment analysis across five domains. Furthermore, when this classifier is combined with a machine-learning based classifier, the combined classifier yields better performance than either of them.**

***Keywords*: sentiment classification; emotion; semi-supervised learning**

I. INTRODUCTION

Sentiment classification aims to classify a text according to the sentimental polarity of opinions it contains (Pang et al., 2002). This task becomes a hot research issue in the community of natural language processing and has been widely used in many real applications.

One challenging isue in sentiment classification is the phenomenon of so-called polarity shifting (Li et al., 2010) which happens when the sentimental orientation of the whole text is different from its containing words or sentences. For example, in the sentence '*I do not like this book*', the polarity of the word '*like*' is different from the polarity of the whole sentence due to the polarity shifting caused by the trigger word '*not*'. It is one main reason why some bag-of-words based machine learning approaches fail under some circumstances.

Although polarity shifting has received more and more attention (Pang et al., 2002; Na et al., 2004; Kennedy and Inkpen, 2006; Ikeda et al., 2008; Li et al., 2010), the concerned structures causing polarity shifting are rather limited, mainly focusing on the negation structure. However, in linguistic studies, various linguistic structures or contextual clues could cause polarity shifting (Polanyi and Zaenen, 2004; Zhang and Li, 2011). Consequently, even with a rather sophisticated classification approach with compositional inference (Choi and Cardie, 2008), the improvement is still very limited (about 1.5 percent). Intuitively, systems considering polarity shifting should have performed much better. One possible reason for the low improvement is that the proportion of negation to all the shifting structures is not as high as people have imagined.

In this paper, we consider five structures possibly causing polarity shifting. They are *negation*, *contrastive transition*, *modality*, *implication*, and *irrelevance* which are summarized by Li et al. (2013). We collect all the trigger words, such as *not*, *without*, and *however*, from the corpus of polarity shifting (The corpus is described in detail in Li et al. (2013)) and leverage them to design several detection rules for detecting the happing of polarity shifting in the testing data.

To quickly check the effectiveness of the polarity shifting detection, we consider the polarity shifting into the classification process. Specifically, we employ a simple classification algorithm called term-counting which derives a sentiment measure by calculating the total number of negative and positive words. Apart from term-counting, our classification algorithm also takes the detection of polarity shifting into account. When calculating the number of positive (or negative) words, we regard the positive (or negative) word as negative (or positive) one if it is detected as polarity-shifted by our rule-based system.

The remainder of this paper is organized as follows. Section 2 introduces the related work on the application of polarity shifting on sentiment classification and related corpus construction. Section 3 introduces the corpus that containing the annotation of polarity shifting. Section 4 proposes a heuristic rule-based approach to detect polarity shifting. Section 5 explores the empirical study on sentiment classification when polarity shifting is considered. Finally, Section 6 concludes our paper.

II. RELATED WORK

A. *Corpus Construction on Polarity Shifting*

Most previous work concerns the annotation of different functional components (i.e., opinion holder, opinion expression, and opinion target), such as Multi-Perspective Question Answering MPQA (Wiebe et al., 2005) and customer reviews (Hu and Liu, 2004).

Some recently-proposed corpus aiming to annotate the discourse-level opinion and certain shifting structures are simultaneously annotated or implied. For instance, Somasundaran et al. (2008) propose opinion frames as a representation of discourse-level associations. Their main objective is to establish relations between targets and only few discourse-level structures of polarity shifting are identified in their annotated corpus. Toprak et al. (2010) present a corpus which considers the opinion expression at sentence-level from different aspects, such as *polarity*, *strength*, *modifier*, *holder*, and *target*. The annotation of polarity shifting is included yet very limited.

However, few annotated corpus has been directly proposed on polarity shifting for sentiment classification. Only one exception is the work by Li et al. (2013) where a corpus with five categories of polarity shifting structures is annotated. In this study, we mainly use this corpus to extract trigger words and to design the rules for detecting polarity shifting.

B. Sentiment Classification with Polarity Shifting

Polanyi and Zaenen (2004) present an extensive analysis on polarity shifting structures in movie reviews. The trigger words of polarity shifting are categorized into two types: sentence-based shifters, e.g., *not* and *never*, and discourse-based shifters, e.g., *but*, and *however*. The analysis is thorough but the proposed taxonomy is not specifically designed for computational applications, such as automatic detection of polarity shifting structures. Unfortunately, the shifting information is not annotated in the corpus.

Negation shifting, a specific structure of polarity shifting caused by negation, has been widely studied, such as Pang et al. (2002), Na et al. (2004), and Kennedy and Inkpen (2006). They usually detect negation structures based on certain rules with trigger words. Their results show that considering negation could only slightly improve the performance of machine learning approaches on sentiment classification. When other kinds of structures, such as content-based negation and contrastive transition, are also considered along with general negation structure, the improved performance becomes more significant (Choi and Cardie, 2008; Ding et al., 2008; Wilson et al., 2009).

Ikeda et al. (2008) and Li et al. (2010) introduce other shifting structures at sentence-level and document-level sentiment classification respectively. They automatically generate a pseudo corpus containing polarity-shifted sentences and propose machine-learning approach to detect polarity-shifted sentences. However, both studies train the detection model with the automatically-generated training data, which makes the accuracy suffer.

Unlike all above studies, our work considers much more structures on polarity shifting into the classification and these structures are given in a manual annotated corpus, which are more accurate and comprehensive.

III. CORPUS INTRODUCTION

Li et al. (2013) report a corpus which contains the product reviews from two domains: DVD and Kitchen (Blitzer et al., 2007). It was annotated by five categories of polarity-shifting structures. Figure 1 shows some examples of the annotated sentences in the corpus. Specifically, three main factors are annotated. The first one is the category of the polarity shifting structure. There are five main categories: *negation*, *contrastive transition*, *modality*, *implication*, and *irrelevance* and some sub-categories in each main category. The second one is the sentimental word in the sentence. The third one is the trigger word that possibly causes the happing of the polarity shifting. For example, in the first example in Figure 1, the main category of the polarity shifting is negation and the subcategory is functional. The sentimental word in this sentence is '*recommend*' while the trigger word is '*not*'. To better understand the structures of polarity shifting, we give a brief introduction on each of these categories and its subcategories.

A. Negation

Negation is a very popular structure that causes polarity shifting. Generally, there are two types of negators: function-word negators and content-word negators. Different from function-word negators (e.g., '*not*' and '*never*'), content-word negators themselves contain content meaning and can possibly be sentimental words (e.g., '*fail*' and '*eliminated*'). Accordingly, the negation can be further categorized into two subgroups: functional negation and contextual negation.

The corresponding trigger words can be *not*, *without*, *never*, *lack*, and so on.

B. Contrastive Transition

Contrastive transition is one special type of transition for expressing contradiction or contrast when connecting one paragraph, sentence, clause or word with another. It is distinguished from other types of transitions by different connectives. The subcategories contains: *intra-sentence*, *extra-sentence*, *and extra-paragraph* .

C. Modality

Modality is related to the attitude of the speaker towards her/his statements in terms of the degree of certainty, reliability, subjectivity, sources of information, and perspective (de Haan, 1995). This category is also a common structure that causes polarity shifting and yet underdeveloped in sentiment classification studies. The subcategories contain *Time*, *Hypothesis*, and *Uncertainty*.

The corresponding trigger words can be *possible*, *possibly*, *perhaps*, *probably*, and so on.

D. Implication

This category is specifically proposed to detect the polarity of the opinion expressed by a holder on a specific target. Sometimes, the holder or the target in the sentence might not be the one concerned. Although the opinion in this case is from other holders or about other targets, it sometimes implies opinions on the target concerned. The subcategories contain *Holder*, *Target*, and *Opinion*.

The corresponding trigger words can be: *other*, *instead*, *something*, and so on.

E. Irrelevance

The last category is called irrelevance. The sentences of irrelevance are not related to the topic concerned at all. Different from all of the above categories, the polarity-shifted sentences in this category contain no explicit trigger word.

(1) So i would not recommend # this film for families with young children , because of the nudity .
<NEGATION_functional, recommend, not>

(2) It might be better # for things that do not require as much water. <MODALITY, better, might be>

Figure 1: Some examples in the polarity-shifting corpus

Domain	Term Counting	Term Counting with Polarity Shifting				
		Negation Triggers	Contrastive Transition Triggers	Modality Triggers	Implication Triggers	All Triggers
Book	0.630	0.633	0.658	0.633	0.633	0.688
DVD	0.680	0.690	0.690	0.698	0.690	0.733
Electronic	0.675	0.703	0.698	0.675	0.683	0.743
Kitchen	0.713	0.728	0.728	0.713	0.720	0.763
Movie	0.633	0.638	0.633	0.645	0.633	0.670
Average	0.666	0.678	0.681	0.673	0.672	0.719

Table 1: Comparison of term counting and term counting with polarity shifting

IV. POLARITY SHIFTING DETECTION

Our basic idea to detect polarity shifting is to leverage the trigger words to design some heuristic rules. Our rules are categories into four types: intra-clause, intra-sentence, extra-sentence and extra-paragraph.

(1) Intra-clause Rule: when a trigger word from the categories *negation*, *modality*, and *implication* is found in a clause, the sentimental words in the clause is considered as polarity-shifted;

(2) Intra-sentence Rule: when a trigger word from the category of *intra-sentence Contrastive Transition* is found in a sentence, the sentimental words in the clause before (or after, decided by the trigger word) the trigger word is considered as polarity-shifted;

(3) Extra-sentence Rule: when a trigger words from the category of *extra-sentence Contrastive Transition* is found in a sentence, the sentimental words in the sentence before (or after, decided by the trigger word) the trigger word is considered as polarity-shifted;

(4) Extra- paragraph Rule: when a trigger words from the category of *extra-paragraph Contrastive Transition* is found in a sentence, the sentimental words in the sentence before the trigger word is considered as polarity-shifted;

V. EXPERIMENTAL STUDIES

A. *Experiment Setup*

Our experiments are conducted on the data sets from two sources. One is from the product reviews of four domains: book, DVD, electronics, and kitchen appliances, taken from the multi-domain sentiment classification corpus collected by (Blitzer et al., 2007)[1]. The other one is from the Cornell movie-review dataset[2] (Pang and Lee, 2004). In each of the four domains, 1,000 positive and 1,000 negative reviews are available. We randomly select 200 positive and 200 negative reviews as the testing data.

The baseline classification algorithm is based on the term-counting approach which derives a sentiment measure by calculating the total number of negative and positive words. To take the detection of polarity shifting into account, the basic term-counting is adjusted as follows: When calculating the number of positive (or negative) words, we regard the positive (or negative) word as negative (or positive) one if it is detected as polarity-shifted by our rule-based system. We call this novel approach term-counting with polarity shifting.

Besides the term-counting approach, we also implement the combination approach which combines the term-counting approach with machine learning approach (Kennedy and Inkpen, 2006). The machine learning algorithm is Maximum Entropy (ME) implemented with the help of the Mallet[3] tool. All parameters are set to their default values. As far as the features are concerned, Boolean word unigram features are employed, representing the presence or absence of a word in a document.

B. *Experimental Results*

The performances of the baseline term-counting approach and term-counting with polarity shifting are shown in Table 1.

We can see in Table 1 that the performance is generally improved when the trigger words from each category are used. The improvement by only using the negation triggers is small. This result is consistent with previous studies, such as Pang et al. (2002) and Kennedy and Inkpen (2006). When all the trigger words are used, the improvement becomes rather impressive (5.3 higher on average). Moreover, the improvements over the baseline are all significant (at the 0.01 level). Interestingly, we find that the improvement by using all triggers is much larger than the sum of the improvements by using triggers in each category separately. For example, in the movie domain, the

[1] http://www.seas.upenn.edu/~mdredze/datasets/sentiment/
[2] http://www.cs.cornell.edu/People/pabo/movie-review-data/
[3] http://mallet.cs.umass.edu/

improvement by using all triggers is 0.034 (from 0.633 to 0.670) while the sum of the separate improvements is 0.017 (0.005+0+0.012+0). This suggests that it would be better to consider more kinds of shifting structures at the same time when incorporating the polarity shifting knowledge into a sentiment classification system.

The performances of combining term-counting or term-counting with polarity shifting (denoted as TC or TC_shifting) and machine learning-based approach (denoted as ML) are shown in Table 2 and Table 3 where 200 and 1000 labeled samples are used to train the ME classifier respectively. From these two tables, we can see that term-counting, especially term-counting with polarity shifting could help improve the machine learning-based approach, even when the machine learning-based approach itself has achieved a good performance (when 1000 labeled samples are used). Note that because the term-counting approach doesn't need any annotated data, it is always costless to use it in real applications.

	ML	ML+TC	M+TC_Shifting
Book	0.648	0.703	0.728
DVD	0.660	0.710	0.718
Electronic	0.735	0.745	0.750
Kitchen	0.758	0.763	0.803
Movie	0.795	0.815	0.843
Aver.	*0.719*	*0.747*	*0.768*

Table 2: The performance of the combination approach when 200 labeled samples are used to train the ME classifier

	ML	ML+TC	ML+TC_Shifting
Book	0.788	0.808	0.813
DVD	0.753	0.775	0.785
Electronic	0.813	0.838	0.855
Kitchen	0.840	0.823	0.838
Movie	0.858	0.875	0.870
Aver.	*0.810*	*0.824*	*0.832*

Table 3: The performance of the combination approach when 1000 labeled samples are used to train the ME classifier

VI. CONCLUSION

In this paper, we leverage polarity shifting detection to improve the performance of sentiment classification. Specifically, we extract many triggers extracted from a corpus on polarity shifting and leverage them to design several heuristic rules to detect polarity shifting. Then, the detection of polarity shifting is used in the term-counting approach to sentiment classification. Empirical studies demonstrate that term-counting with polarity shifting yield much better performances than the basic term-counting approach. Furthermore, when it is combined with a machine-learning based classifier, the combined classifier performs better than either of them.

ACKNOWLEDGMENTS

This research work has been partially supported by two NSFC grants, No.61003155, and No.61273320, one National High-tech Research and Development Program of China No.2012AA011102, one General Research Fund (GRF) sponsored by the Research Grants Council of Hong Kong, No.543810.

REFERENCES

[1] Blitzer J., M. Dredze, and F. Pereira. 2007. Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification. In *Proceedings of ACL-07*, pp.440-447.

[2] Choi Y., and C. Cardie. 2008. Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis. In *Proceedings of EMNLP*, pp.793-801.

[3] de Haan F.1997. The Interaction of Modality and Negation: A Typological Study. Garland Publishing, Inc., New York: Garland.

[4] Ding X., B. Liu, and P. Yu. 2008. A Holistic Lexicon-based Approach to Opinion Mining. In *Proceedings of the International Conference on Web Search and Web Data Mining, WSDM-08*, pp.64-71.

[5] Hu M. and B. Liu. 2004. Mining and Summarizing Customer Reviews. In KDD'04: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.168-177, Seattle, Washington.

[6] Ikeda D., H. Takamura, L. Ratinov, and M. Okumura. 2008. Learning to Shift the Polarity of Words for Sentiment Classification. In *Proceedings of IJCNLP-08*, pp.296-303.

[7] Kennedy, A. and D. Inkpen. 2006. Sentiment Classification of Movie Reviews using Contextual Valence Shifters. *Computational Intelligence*, vol.22(2), pp.110-125.

[8] Li S., R. Xia, C. Zong, and C. Huang. 2009. A Framework of Feature Selection Methods for Text Categorization. *In Proceedings of ACL-IJCNLP-09*. pp.692-700.

[9] Li S., S. Lee, Y. Chen, C. Huang and G. Zhou. 2010. Sentiment Classification and Polarity Shifting. In proceedings of COLING-10, pp 635-643.

[10] Li S., S. Lee, and C. Huang. 2013. Corpus Construction on Polarity Shifting in Sentiment Analysis. In *proceedings of CLSW-13*.

[11] Na J., H. Sui, C. Khoo, S. Chan, and Y. Zhou. 2004. Effectiveness of Simple Linguistic Processing in Automatic Sentiment Classification of Product Reviews. In *Conference of the International Society for Knowledge Organization (ISKO-04)*.

[12] Narayanan R. and B. Liu, 2009. Sentiment Analysis of Conditional Sentences. In *proceeding of ACL-IJCNLP-09*, pp.180-189.

[13] Pang B., L. Lee, and S. Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. In *Proceedings of EMNLP-02.* pp.79-86.

[14] Pang B. and L. Lee. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In *Proceedings of ACL-2004*, pp. 271-278.

[15] Polanyi L. and A. Zaenen. 2006. Contextual Valence Shifters. Computing attitude and affect in text: Theory and application. Springer Verlag.

[16] Somasundaran S., J.Wiebe, and J. Ruppenhofer. 2008. Discourse level opinion interpretation. In *Proceedings of COLING-08*, pp.801-808.

[17] Toprak C., N. Jakob and I. Gurevych. 2010. Sentence and Expression Level Annotation of Opinions in User-Generated. In *proceedings of ACL-10*, pp.575-584.

[18] Turney P. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proceedings of ACL-02*, pp. 79-86.

[19] Wilson T., J. Wiebe, and P. Hoffmann. 2009. Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis. *Computational Linguistics*, vol.35(3), pp.399-433, 2009.