Polarity Shifting: Corpus Construction and Analysis

Xiaoqian Zhang, Shoushan Li*, Guodong Zhou
Natural Language Processing Lab
Soochow University
Suzhou, China
{xiaoqian.z047,shoushan.li} @gmail.com,
gdzhou@suda.edu.cn

Hongxia Zhao
State Key Laboratory of Intelligent Control and Management of Complex Systems
Institute of Automation, Chinese Academy of Sciences
Beijing, China
hongxia.zhao@ia.ac.cn

***Abstract*—Polarity shifting has been a challenge to automatic sentiment classification. In this paper, we create a corpus which consists of polarity-shifted sentences in various kinds of product reviews. In the corpus, both the sentimental words and shifting trigger words are annotated. Furthermore, we analyze all the polarity shifted sentences and categorize them into five categories: *opinion-itself*, *holder*, *target*, *time* and *hypothesis*. Experimental study shows the agreement of annotation and the distribution of the five categories of polarity shifting.**

***Keywords*: sentiment classification; polarity shifting; corpus annotation**

I. INTRODUCTION

Sentiment classification is a special task of text classification whose objective is to classify a text according to the sentimental polarity of opinions [8]. This task has received considerable interest and been widely studied in the community of computational linguistics [11,13,14].

One challenge problem in sentiment classification is the phenomenon of so-called polarity shifting [5] 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'. This is one main reason why bag-of-words based machine learning approaches fail under some circumstances.

Generally, there are two main steps for considering polarity shifting in a sentiment classification system: detecting polarity-shifted words or sentences and designing specific classification algorithms. Apparently, the first step is the key in the whole design of system. To handle this, some machine-learning based approaches have been proposed recently [3,5]. However, the corpus used to train the machine learning model is normally obtained automatically and thus unavoidably contains much noise. Such noise much harms the performance of sentiment classification low. Therefore, it is urgent to build a high-quality manually-annotated corpus on polarity shifting to better understand the polarity-shifting phenomenon and help build a more accurate detection system.

* Corresponding Author

In this paper, we manually annotate the polarity-shifted sentences in the document-level reviews. Furthermore, we categorize the polarity-shifted sentences into five categories (manually annotated): opinion-itself, holder, target, time and hypothesis. We believe that this corpus will be beneficial for both pure theoretical linguistic studies on polarity shifting and computational linguistic studies on sentiment classification.

The remainder of this paper is organized as follows. Section II introduces the related work. Section III describes the annotation scheme and our analysis. Section IV presents the empirical study on this corpus. Finally, Section V presents the conclusion and future work.

II. RELATED WORKS

Early work on polarity shifting mainly focuses on negation shifting, such as, Pang et al. [8], Na et al. [6], Ding et al. [2], and Kennedy and Inkpen [4]. They usually use some rules with some trigger words to tackle the negation shifting problem. Their results show that considering negation shifting could improves the performance of machine learning approaches on sentiment classification.

Both Ikeda et al. [3] and Li et al. [5] automatically generate a pseudo corpus containing polarity-shifted sentences and propose machine learning driven approaches to detect polarity-shifted sentences. Ikeda et al. [3] mainly focus on sentence-level sentiment classification with a manually-constructed dictionary containing positive and negative sentimental words while Li et al. [5] focus on document-level sentiment classification using some feature selection methods. However, both of them train the detection model with the automatically-generated training data. This much harms the detection performance due to the noise in the automatically-generated training data.

Note that Toprak et al. [9] presents a manually annotated corpus which considers the opinion expression in the sentence level from many sides, such as polarity, strength, modifier, holder, and target. The concept of polarity shifting is also mentioned. However, their annotation scale on polarity shifting is very small. In comparison, our annotation contains more kinds of situations under which polarity shifting happens.

III. CORPUS ANNOTATION AND ANALYSIS

In this study, polarity shifting means that the polarity of a word is different from the polarity ex-pressed by the whole document. The sentences containing such words are considered as polarity-shifted sentences. We annotate the polarity-shifted sentences from the product reviews of two do-mains: DVD and KITCHEN [1], where the polarity of each document has been given already.

In order to reduce the annotation cost, we first annotate a sentimental word list and then use it to automatically generate a rough initial corpus which contains many possible polarity-shifted sentences. The annotation work for polarity shifting is then conducted on those possible sentences rather than all the ones, motivated by Ikeda et al. [3] and Li et al. [5].

A. Collection and Annotation of the Sentimental words

We collect the sentimental words from two resources: one is the dictionary proposed by Wilson et al. [10] which are domain-independent, and the other is the word list selected by performing the information gain (IG) selection method on the documents from all domains [12] and pick the top 1200 words as sentimental words. Then we merge the two resources together. These initial sentimental words are rough and might not be entirely correct. Therefore, we manually re-label the polarities of all the words to make sure that their polarities are correct.

B. Initial Corpus Generation

After obtaining the sentimental words, we first segment the document into sentences with punctuations, e.g., '*?*', '*.*', etc, and some simple conjunctions, e.g., '*and*', '*but*', '*so*', '*although*', and so on. Then, we automatically generate the initial corpus based on the following strategy: if a sentence in the negative/positive document contains a positive/negative word, we assume it as polarity-shifted. Note that these sentences might not be really polarity-shifted. For example, '*like*' is a positive word in our dictionary and '*It looks like a dog*.' is a sentence in a negative document. Since the word in this sentence does not contain any sentimental meaning, we mistakenly label it as polarity-shifted. Nevertheless, this strategy can reduce the annotation cost and we only need to manually check the pseudo polarity-shifted sentences instead of checking all sentences.

C. Analysis on Polarity Shifting

Through analyzing the corpus, there are five popular cases that a sentimental word could appear in a document expressing a different sentiment orientation. That is to say, we categorize the impact of a sentimental word on polarity shifting into following five categories:

1) Opinion-itself: where a clear specific structure, e.g., negation or contrast transition, changes the polarity of a sentimental word, e.g. (the sentimental word is followed by the sign of #):

'It isn't a good # table.'

'I have never been happy # with it.'

Typical trigger words include not, never, lack, fail, little, and so on.

2) Holder: where the holder of sentimental expression is not the author, e.g.

'Other people think it is a good # product.'

Typical trigger words include other people, someone, another one and so on.

3) Target: where the opinion is not expressed on the target object but on something else, e.g.

'I recommend # buy something instead.'

Typical trigger words include other, instead, something and so on.

4) Time: where the opinion is not expressed at the present but in the past. For example:

'I thought it was a good # product.'

Typical trigger words include should, would, and thought and so on.

5) Hypothesis: where the phenomenon of an assuming sentence [7] is described or the sentimental expression of assuming is a hypothetical situation. For example:

'If they change the color, I will recommend # it.'

Typical trigger words include wish, if, hope and so on.

In our annotation, we mark both the category information and corresponding trigger words in the corpus. Sometimes, it happens that there is not a clear trigger word in the polarity-shifted sentence. In this case, we call this phenomenon '*implicit shifting*', and the sentence is labeled as '*implicit*'. For example, in the sentence of *'Susan thinks it is a good # product'*, *Susan* is the opinion holder and cannot be annotated as a definite trigger word. In this case, we annotate it as an implicit polarity-shifted sentence with the specific format introduced in the next subsection.

D. Specific Label Format

Each polarity-shifted sentence in the corpus is annotated using a vector in the following format:

<*Category Index*, *Sentimental word*, *shifting trigger word*>

The vector contains three elements. The first element is the category index expressing the category information as described in Section III.D. For example, *Category Index=2* means polarity shifting happens because the holder of the sentimental expression is not the author, but somebody else. The second element indicates the sentimental word. The third element indicates the trigger word causing polarity shifting. Following are four examples for better understanding.

1) *'It isn't a good # table.*
<1, good, isn't>'

2) *'I thought it was great #, but I'm wrong.'*
<4, great, thought>

3) *'I recommend # buy another one.'*
<3, recommend, another>

4) *'Others think it is very interesting #.'*
<2, interesting, others >

IV. EXPERIMENTAL STUDIES

Totally, 1200 product reviews (600 positive and 600 negative) are annotated. Each review has been annotated by two annotators.

A. Annotation Agreement

As discussed in Section III.A, we annotate the sentimental words from two resources: one is the dictionary [10] and the other is a word list generated from some feature selection methods. Table I gives the kappa value (κ value) for each resource, the agreement of our annotation on the polarities of the words. The disagreement often happens when the polarity of the word is hard to judge, e.g. '*struggle*', '*marvelous*', '*surprise*', and '*financial*'.

TABLE I. ANNOTATION AGREEMENT ON SENTIMENTAL WORDS

Source of the Sentimental Words	**Makeable**	**Agreement** (κ value)
Feature Selection	1200	0.65
Dictionary	8200	0.71

B. *Analysis on the Initial Corpora*

As reported in Section III.B, we use a strategy to generate the initial corpus to reduce the annotation cost. Manual inspection indicates that 67.72% of annotated polarity-shifted sentences in the generated initial corpus are really polarity-shifted with the errors caused by the uncertainty on the meaning of some sentimental words. For example, the sentence ' *It doesn't looks like # a cat*' is wrongly generated as a polarity-shifted sentence. Here, the word '*like*' does not express any opinion, although it is in our sentimental word list.

C. *The Proportion of Shifting Happening*

Table II shows the number of reviews that contains polarity-shifted sentences and the proportion of polarity-shifted sentences. We can see that polarity shifting is a popular phenomenon in the reviews and is prone to appear in the negative ones. Such high proportion of polarity-shifted sentences highly supports our research on this issue in sentiment analysis.

TABLE II. PROPORTION OF THE REVIEWS WHERE POLARITY SHIFTING HAPPENS

Domain	**Polarity**	**Number of shifting**	**Proportion**
DVD	*Negative*	233	0.77
	Positive	162	0.54
Kitchen	*Negative*	211	0.70
	Positive	151	0.50

D. *Category Statistics*

As discussed in Section III.C, we categorize polarity shifting of sentimental words into five categories. Table III shows the proportion of each category.

From this table, we can see that the percentage of type 1 reaches half of all. This is because the phenomenon of negation shifting is very common. The third category also accounts for about 30%, which demonstrate that polarity shifting often hap-pens when the opinion is not expressed on the real target object. The overall proportion of other three categories is only about 20%.

TABLE III. DISTRIBUTION ON POLARITY SHIFTING CATEGORIES

Categories	**Trigger Words/Phrases**	**Proportion (%)**
Opinion-itself	not, no, without	0.50
Holder	someone, other people	0.10
Target	another, instead	0.30
Time	thought, should	0.06
Hypothesis	if, wish	0.04

V. CONCLUSION AND FUTURE WORK

In this paper, we create a polarity-shifted product review corpus in the sentence level, annotated with sentimental word, shifting trigger word, and shifting category. Empirical studies demonstrate that the polarity shifting phenomenon indeed appears very commonly in review text.

Our future work will focus on the annotation of the long-distance shifting phenomenon, i.e., the polarity shifting phenomenon involves multiple sentences. For example, the paragraphs separated by the trigger word '*however*' are usually contains several polarity-shifted sentences simultaneously. Another future research issue is to train a high-performance polarity-shifting detector with our corpus and apply it to sentiment classification.

ACKNOWLEDGMENTS

The research work described in this paper has been partially supported by three NSFC grants, No. 61003155, No. 60873150 and No. 90920004 and Open Projects Program of National Laboratory of Pattern Recognition.

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.

[2] 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.

[3] 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.

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

[5] Li S., S. Lee, Y. Chen, C. Huang and G. Zhou. 2010. Sentiment Classification and Polarity Shifting. In proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), pp 635-643.

[6] 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).

[7] Narayanan R., B. Liu, 2009. Sentiment Analysis of Conditional Sentences. In proceeding of ACL and AFNLP 2009, pp 180-189.

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

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

[10] 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.

[11] Wu Q., S. Tan, and X. Cheng. 2009. Graph ranking for sentiment transfer. In proceedings ACL-IJCNLP, short paper, pp. 317–320.

[12] Yang Y. and J. O. Pedersen. 1997. A Comparative Study on Feature Selection in Text Categorization. In Proceedings of ICML, pp.412-420.

[13] Yessenalina A., Y. Yue, and C. Cardie. 2010. Multi-Level Structured Models for Document-Level Sentiment Classification. In Proceedings of EMNLP, pp. 1046–1056.

[14] Zhao J.， K. Liu, and G. Wang. 2008. Adding redundant features for CRFs-based sentence sentiment classification. In Proceedings of EMNLP, pages 117–126.