

Improving Chinese Sentence Polarity Classification via Opinion Paraphrasing

Guohong Fu, Yu He, Jiaying Song, Chaoyue Wang

Heilongjiang University, Harbin 150080, China

ghfu@hlju.edu.cn, heyucs@yahoo.com, jy_song@outlook.com,
chaoyue.wang@yahoo.cn

Abstract

While substantial studies have been achieved on sentiment polarity classification to date, lacking enough opinion-annotated corpora for reliable training is still a challenge. In this paper we propose to improve a supported vector machines based polarity classifier by enriching both training data and test data via opinion paraphrasing. In particular, we first extract an equivalent set of attribute-evaluation pairs from the training data and then exploit it to generate opinion paraphrases in order to expand the training corpus or enrich opinionated sentences for polarity classification. We tested our system over two sets of online product reviews in car and mobilephone domains. The experimental results show that using opinion paraphrases results in significant performance improvement in polarity classification.

1 Introduction

With the explosive growth of the user-generated opinionated texts on the web over the past years, opinion mining has been attracting an ever-increasing amount of attention from the natural language processing community. As a key sub-problem of opinion mining, sentiment polarity classification aims to classify opinionated documents or sentences as expressing positive, negative or neutral opinions, and plays a critical role in many opinion mining applications such as opinion summarization and opinion question answering. Since sentence is usually considered as the smallest semantic unit for expressing the complete opinion, the current study focused on the sentence sentiment classification.

Although recent years have seen a great progress in sentiment classification, lacking large-scale opinion-annotated corpora is still a fundamental issue. On the one hand, statistically-based

methods become the mainstream in sentiment analysis. In general, a statistically-based polarity classifier needs an annotated corpus for training. So its performance heavily relies on the training corpus used. On the other hand, to date there are no any large-scale annotated corpora available for achieving reliable training process. Furthermore, opinion mining is usually domain specific. Obviously, it is time and cost consuming to manually construct a large-scale opinion-annotated corpus for each domain.

To address the above problems, in this paper we propose to improve polarity classification by enriching both training data and test data via paraphrasing. We have two motivations for this. Firstly, paraphrasing has proven to be an effective tool for improve the coverage of systems and has been successfully used in many applications such as machine translation, information retrieval and question answering (Bhagat and Hovy, 2013; Heilman and Smith, 2010; Zhao *et al.*, 2013; Fader *et al.*, 2013). However, to date, there has been very limited study on sentiment or opinion paraphrasing. Secondly, unlike opinion corpus annotation, paraphrases are relatively more flexible to acquire using different resources like synonym lexica, bilingual and parallel corpora, and so forth. Therefore, we believe that paraphrasing would be a feasible way to expand the training corpus and at the same time, to alleviate the data sparse problem in statistically-based systems. As such, the purpose of this study is to ascertain the effect of using opinion paraphrases in polarity classification at sentence level. To approach this, we first extract an equivalent set of attribute-evaluation pairs from the training data and then exploit it to generate opinion paraphrases in order to expand the training corpus or enrich opinionated sentences for polarity classification. Based on the generated opinion paraphrases, we also develop a polarity classification system for Chinese under the framework of support vector

machines (SVMs). Experimental results over two sets of online reviews on car and mobilephone products show that using the paraphrases generated by the proposed method can significantly improve the performance of sentence polarity classification.

The rests of the paper proceed as follows. Section 2 provides a brief review of the literature on sentiment classification and paraphrase generation. Section 3 describes in details the proposed method to Chinese sentence polarity classification via paraphrasing. Section 4 reports our experimental results on two sets of product reviews. Finally, section 5 concludes our work and discusses some possible directions for future research.

2 Related Work

Polarity classification is usually formulated as a binary classification problem (Turney, 2002; Pang and Lee, 2008). Most previous studies employ supervised machine learning methods, including naïve Bayes model, support vector machines (SVMs), maximum entropy models (MEMs), conditional random fields (CRFs), fuzzy sets, and so forth (Pang et al., 2002; Pang and Lee, 2008; Fu and Wang, 2010), to perform polarity classification on different linguistic levels such words, phrases, sentences and documents.

Lacking a large scale manually-annotated corpus is one of the major bottlenecks that supervised machine learning methods faced. To break this bottleneck, some recent studies exploit bootstrapping or unsupervised techniques (Turney, 2002; Mihaicea et al., 2007; Wilson et al., 2009,

Speriosu et al. 2011, Mehrotra et al. 2012; Volkova et al., 2013). Unfortunately, sentiment classifiers based on unsupervised methods usually yield worse performance compared to the supervised ones.

Different from most existing studies, in this study we attempt to enhance Chinese sentence polarity classifier by exploring opinion paraphrasing. We believe that paraphrasing provides us with an option to expand training corpora and to enrich opinion sentences for polarity classification, which would alleviate the problem of data sparseness and lack of annotated corpora for training. At this point, our current study is also relevant to paraphrasing tasks, including paraphrase recognition, paraphrase extraction and paraphrase generation. Although a variety of methods, from dictionary-based methods to data-driven methods (Madhani and Dorr, 2010), have been proposed for paraphrasing. Since in the present study we aim to answer the question whether the use of paraphrasing can enhance polarity classification performance, we do not want to look insight into paraphrasing issues. Instead, we just exploit some simple but efficient paraphrasing techniques to achieve opinion paraphrases for expanding training data and enriching text data for polarity classification, including opinion paraphrase extraction incorporating the Jaccard coefficient based literal similarity with the word embedding based semantic similarity, and opinion paraphrase generation with opinion element substitution.

3 The Proposed Method

3.1 Overview

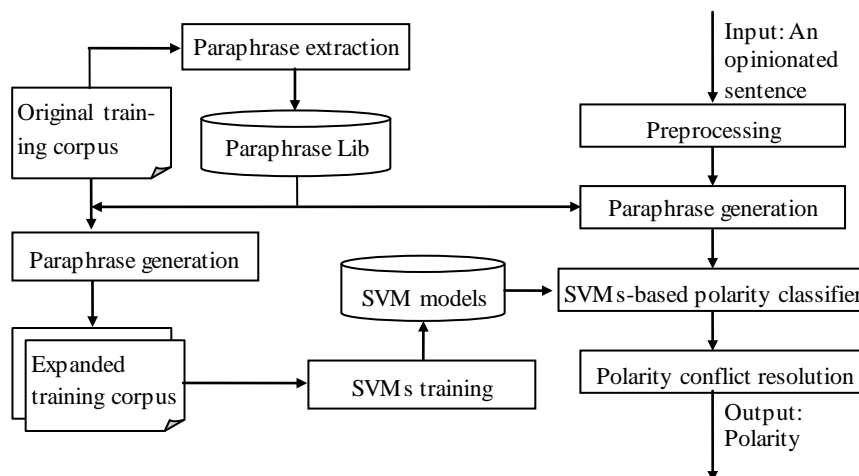


Figure 1. The overall framework of the proposed method to Chinese polarity classification.

Figure 1 presents the general framework for Chinese polarity classification via opinion paraphrasing, mainly including paraphrase extraction, training corpus expansion via paraphrase generation and the SVMs-based polarity classifier with paraphrasing.

Training corpus expansion. For each opinionated sentence from the original corpus for training, we first generate a set of suitable paraphrases and thus expand the training corpus by adding these generated paraphrases into it.

Paraphrase extraction. To achieve opinion element substitution based paraphrasing, we need to extract a set of equivalent attribute-evaluation pairs from the training corpus. In the present study, we incorporate literal similarity and word embedding-based semantic similarity between two coreferred product attributes with the polarity of the paired evaluation expressions to perform attribute-evaluation clustering.

Paraphrase generation. With regard to the focus of the current study, we generate sentential paraphrases by simply substituting opinion elements such as product attributes and their evaluations in the original sentence with their respective semantic equivalents.

SVMs-based polarity classifier. We perform sentence polarity classification using supported vector machines (SVMs) trained from the expanded training data via opinion paraphrasing.

Polarity conflict resolution. To avoid data sparseness, in the present study we perform paraphrasing on the input opinionated sentences in test before polarity classification. As a consequence, this may cause polarity conflicts between the original input sentences and their paraphrases after polarity classification. To address this problem, we employ a rule-based voting method.

In Sections 3.2 to 3.5, we provide the details of our implementation.

3.2 Paraphrases in Product Reviews

Before describing the techniques for paraphrase extraction and generation, it is necessary to clarify what a paraphrase is for product reviews. In linguistics literature, paraphrases are most often referred to as an approximate equivalence of meaning across sentences or phrases (Bhagat and Hovy, 2013). In the present study we characterize opinion paraphrases from the perspective of opinion elements. In general, opinion information consists of five main elements, namely opinion source (viz. opinion holder), opinion target, attribute, evaluation and polarity. Thus, the opinion element perspective defines paraphrases in terms of the kinds of opinion element changes that can take place in an opinionated sentence resulting in the generation of its paraphrases. Considering the characteristics of product reviews, here we focus on product attributes and their relevant evaluations within opinionated sentences in determining whether they are paraphrasing each other. Thus, two opinion sentences that contain the same or similar attribute-evaluation pairs are termed as opinion paraphrases.

With regard to semantic equivalence between attributions and evaluations within opinion expressions, we can thus classify paraphrases in product reviews into four main types, as shown in Table 1. Based on this, given two different opinionated sentences, if they involve identical or coreferred attributions, and at the same time, their corresponding evaluations are identical or approximately equivalent with respect to sentiment polarity, then the two opinionated sentences are considered to be paraphrastic.

Table 1. Categorization of opinion paraphrases in product reviews

Types	Attributes	Evaluations	Examples
1	exactly identical	exactly identical	操控性非常好(The controllability is very good.) 该车的操控性非常好。(The controllability of this car is very good.)
2	exactly identical	semantically equivalent	手感不错 (hand feeling is not bad) 手感好 (hand feeling good)
3	coreferent	exactly identical	性价比真高(The cost-performance ratio is really high) 性能价格比真高(The cost-performance ratio is really high)
4	coreferent	semantically equivalent	质地真不错 (The texture is really good) 材质挺好 (The material is very good)

3.3 Paraphrase Extraction

Since the definition of opinion paraphrase is based on the equivalence of attributes and their

corresponding evaluations within opinionated sentences, attribute-evaluation pairs are very important knowledge for substitution-based paraphrase generation. To obtain such knowledge for

opinion paraphrasing, we first extract all attribute-evaluation pairs from the training corpus and further cluster them in terms of attribute coreference relation and the polarity. Given two different attribute-evaluation pairs, if the attributes are coreferred each other and at the same time, the relevant polarity are identical, then the two attribute-evaluation pairs are paraphrastic and can be grouped to a cluster.

Due to the fact that polarity information has been manually annotated in the training corpora, attribute coreference resolution becomes the key to attribution-evaluation grouping. To address this problem, we combine two similarity measures, namely the literal similarity based on Jaccard coefficient and the semantic similarity based on word embeddings.

(1) Literal similarity. As shown in Equation (1), Jaccard coefficient measures (denoted by J) the literal similarity of two attribute expressions A_1 and A_2 by counting the number of identical characters contained in them.

$$SimJ(A_1, A_2) = \frac{|set(A_1) \cap set(A_2)|}{|set(A_1) \cup set(A_2)|} \quad (1)$$

Where, $set(A)$ denotes the set of characters that form the attribute A .

It should be noted that unlike the classical edit distance, Jaccard coefficient ignores the influence of character location in attributes. Considering two pairs of Chinese attributes (外表, 外形)

and (油耗, 耗油), their respective Jaccard coefficients are 0.33 and 1.

(2) Semantic similarity. Literal similarity measures rely on literal matching and work for product attributes with explicit literal connections. However, such information does not always exist in many co-referred feature expressions like 像素 (pixel) and 分辨率 (resolution). To address this problem, we introduce semantic similarity based on word embeddings. Actually, word embeddings map each word to an n -dimensional dense vector of real numbers and each dimension has certain latent semantic information (Mikolov, 2012; Mikolov *et al.*, 2013). Obviously, the data size has a strong relationship with the expression of semantic. Thus, we can obtain the similarity between two product attributes by calculating the cosine distance between their relevant vectors, as shown in Equation (2).

$$SimS(A_1, A_2) = \frac{\sum_{i=1}^n v_i(A_1) \times v_i(A_2)}{\sqrt{\sum_{i=1}^n v_i(A_1)^2} \times \sqrt{\sum_{i=1}^n v_i(A_2)^2}} \quad (2)$$

Where, $v_i(A_1)$ and $v_i(A_2)$ ($1 \leq i \leq n$) denote the respective word embeddings of product attributes A_1 and A_2 , and n denotes the number of dimensions in word embedding representation of product attributes.

Table 2 illustrates a sample of equivalent attribute-evaluation pairs extracted from the training corpora.

Table 2. A sample of equivalent attribute-evaluation pairs extracted from the training corpora

Product attributes	Positive evaluations	Negative evaluations
<i>Price</i> : 价 价格 价钱 价位 <i>Acceleration</i> : 加速 加速性 加速能力 ... <i>Touch screen</i> : 触摸屏 触屏 触控 触感 触控 ...	<i>Low</i> : 合适 适中 实惠 优惠 不高 公道 比较便宜 有优势 值 ... <i>Excellent</i> : 有推背感 一点不软 很好 很给力 令人满意 灵敏 很优秀 ... <i>Fast/Sensitive</i> : 不错 好 很好 灵敏 灵活 快 给力 挺流畅 反应快 好用 灵敏度高 ...	<i>High</i> : 高 太高 真高 偏高 有点高 贵 太贵 偏贵 有点贵 不合理 有点无语 ... <i>Weak</i> : 差 偏弱 有延迟 很突然 比较没劲 比较没力 ... <i>Slow/Insensitive</i> : 不太灵敏 不是很灵敏 比较慢 有点不灵活 反应太慢 不好用 迟钝 过于灵敏 ...

3.4 Paraphrase Generation

Given an opinionated sentence S , we generate paraphrases in two steps:

(1) Opinion element substitution. We first construct a set of equivalent utterances for each attribution or evaluation in S and store them with word lattice. For convenience, here we refer this word lattice as paraphrase word lattice.

The equivalent substitution of attributes or evaluations is essential to opinion paraphrase

generation. In the present study, we perform this task by substituting attributes and their evaluations using the extracted attribute-evaluation pairs shown in Table 2.

(2) n-best paraphrase decoding. Once the paraphrase word lattice is constructed, our problem is now to score all potential paraphrases within the lattice and select the most probable paraphrases as the equivalent expansion of the input sentence. For simplicity and efficiency of implementation, in this paper we employ bigram

language models to rank the paraphrase candidates and thus decode n -best paths from the paraphrase word lattice. Each path forms a probable paraphrase for the input sentence.

Table 3 shows some generated paraphrases and their bigram scores.

Table 3. Examples of generated paraphrases.

Original sentences	Generated paraphrases	scores
操控性特棒。(The controllability is excellent.)	操控性非常好(The controllability is very good)	1.12e-34
	操控性比较好(The controllability is OK)	6.81e-35
反应有点慢。(The reaction is a bit slow.)	反应比较慢(The reaction is relatively slow)	5.55e-05
	反应迟缓(The reaction is tardy.)	3.70e-05
价格最低！(Lowest price!)	价格合理！(Reasonable price!)	1.29e-11
	价格优惠！(Favorable price!)	8.40e-12

3.5 Polarity conflict resolution

Polarity conflict will arise when an input opinionated sentence and its paraphrases receive different polarity types during polarity classification. The reason may be due to inconsistent generation of paraphrases between the training data and the input opinionated sentences for polarity classification.

In order to avoid polarity conflicts, we employ a simple voting mechanism. Given an input opinionated sentence and its k -best paraphrases generated by the systems, then we have $k+1$ opinionated sentences for polarity classification. Let i ($0 \leq i \leq k$) be the number of sentences that are classified as positive by the system and j ($0 \leq j \leq k$, and $i + j = k$) be the number of sentences that are negative during polarity classification. Thus, we can take the following three rules to determine the final polarity of the original sentence.

- Rule 1. if $i > j$, then the final polarity is positive.
- Rule 2. if $i < j$, then the final polarity is negative.
- Rule 3. if $i = j$, then the final polarity is the same as that of the original polarity of the input sentence during polarity classification.

4 Experimental Results and Discussions

To assess our approach, we developed a SVM-based sentiment polarity classifier and conducted

experiments over car and cellphone product reviews. This section reports our experimental results.

4.1 Experimental Setup

The experimental data come from two domains of online product reviews, namely car reviews and mobilephone reviews. Both corpora are manually annotated with multiple linguistic and opinion information, such as word segmentation, part-of-speech tags, opinion elements and polarity classification, and are further divided into training datasets and test datasets, respectively. Table 4 presents the basic statistics of the experimental data.

Table 4. Basic statistics of the experimental data

Dataset	Car			Mobilephone		
	Total	Pos	Neg	Total	Pos	Neg
Training	1904	841	963	2042	1033	1009
Test	913	462	451	1021	516	505

Table 5. The equivalent attribute-evaluation pairs.

Training data	SimJ		SimJ + SimS	
	A-P	A-N	A-P	A-N
Car	137	177	109	161
Mobilephone	88	121	78	107

As shown in Table 5, we have constructed two knowledge bases, namely the equivalent pairs of attributes and their related positive evaluations (A-P pairs for short), and the equivalent pairs of attributes and their related negative evaluations (A-N pairs for short), for opinion paraphrase generation from the two training corpora, respectively. It should be noted that we consider two strategies for attribute clustering during paraphrase extraction, namely attribute clustering with Jaccard coefficient (SimJ for short) and attribute clustering incorporating Jaccard coefficient and the word embeddings based semantic similarity with linear interpolation (SimJ+SimS for short).

Furthermore, in this paper the performance of polarity classification is reported in terms of accuracy.

4.2 Effects of different paraphrasing

Our first experiment intends to investigate the effects of different paraphrasing strategies on polarity classification, including different n -best paraphrase generation and paraphrasing on different data. Note that in this experiment, we consider five cases (viz. $n = 1$ to 5) during n -best paraphrase generation, and compare the relevant polarity classification results. Furthermore, to

better understand the results for different n -best paraphrase generation, we also conducted an investigation on the relationship between the number of generated paraphrases for different data and the value of n in n -best paraphrases. It should be noted that in this experiment paraphrases are generated using equivalent attribute-evaluation pairs extracted with SimJ and SimS, as shown in Table 5. The results are summarized in Tables 6-9.

Table 6. Number of generated paraphrases for the training and test corpora in car domain

n-best	Dataset	Total	Pos	Neg
1	Training	3460	1708	1637
	Test	1702	805	792
2	Training	4914	2469	2296
	Test	2394	1141	1123
3	Training	6361	3229	2949
	Test	3083	1476	1452
4	Training	7796	3983	3596
	Test	3768	1809	1779
5	Training	9224	4735	4238
	Test	4450	2141	2104

Table 7. Number of generated paraphrases for the training and test corpora in mobilephone domain

n-best	Dataset	Total	Pos	Neg
1	Training	3768	1966	1802
	Test	1889	931	958
2	Training	5487	2897	2590
	Test	2751	1342	1409
3	Training	7187	3825	3362
	Test	3603	1749	1854
4	Training	8881	4751	4130
	Test	4447	2152	2295
5	Training	10568	5676	4892
	Test	5287	2555	2732

Table 8. Polarity classification over car reviews with different paraphrasing strategies

n -best	Para. on training data only	Para. on test data only	Para. on both training and test data
1	70.09	70.69	70.19
2	70.29	71.60	70.80
3	70.29	71.70	70.50
4	67.98	73.01	69.50
5	67.77	71.70	69.49

The results in Tables 7-8 reveal that the value of n in n -best paraphrase generation appears to be an important influence factor for polarity classification with paraphrases. As n increases, the number of generated paraphrases is going up, and at the same time, the polarity classification accuracy is also rising for the case of performing

paraphrasing on the training corpora. But in case of paraphrasing on the test data, the performance in polarity classification does not always rise with the number of generated paraphrases. The reason might be due to the fact larger number of generated paraphrases may introduce more polarity conflicts during polarity classification.

Table 9. Polarity classification over mobilephone reviews with different paraphrasing strategies

n-best	Para. on training data only	Para. on test data only	Para. on both training and test data
1	83.45	83.74	83.45
2	84.62	84.62	87.86
3	85.41	85.31	87.76
4	86.19	86.10	89.81
5	85.21	85.50	89.81

4.3 Comparison of polarity classification with/without paraphrasing

As we have mentioned above, paraphrasing provides us with an option for avoiding the problems of data sparseness in open applications. So our last experiment is designed to examine the effectiveness of using paraphrasing in polarity classification. The experiment is conducted by comparing the results produced by the SVMs-classifies with paraphrases to that of the systems trained with the original corpora in Table 5 only (viz. the baseline systems). Furthermore, we consider two strategies, namely SimJ and SimJ&SimS, for paraphrase extraction in this experiment. The results are presented in Table 10.

Table 10. Comparison of polarity classification with/without paraphrasing

Systems	Car	Mobilephone
Baseline	66.06	83.74
Para. on training data based on SimJ	69.99	86.39
Para. on test data based on SimJ	73.72	86.19
Para. both training and test data based on SimJ	70.90	89.62
Para. on training data based on SimJ&SimS	70.29	89.19
Para. on test data based on SimJ&SimS	73.01	86.10
Para. on both training and test data based on SimJ&SimS	70.80	89.81

As can be seen from Table 10, using paraphrases can significantly improve polarity classification performance. Take the system with paraphrasing on the training data only via

SimJ&SimS, the accuracy can be improved by more than 4 and 6 percents for car and mobilephone reviews, respectively, compared to the baseline without using any paraphrases, illustrating in as sense the effectiveness of the proposed method. Furthermore, it can be observed from Table 10 that the system yields better results for mobilephone reviews than for cars. Moreover, the results over mobilephone data shows the performance in polarity classification can be enhanced by incorporating word embeddings based semantic similarity with literal similarity for paraphrase extraction, while the experiments on car reviews do not illustrate similar results. The reason might be due to the fact that car products have more attributes than mobilephone products, which makes it more difficult to cluster product attributes. In addition, more attributes may results in more paraphrases and thus produce more polarity conflicts to polarity classification.

4.4 Polarity conflicts between paraphrases

As we have mentioned above, larger number of generated paraphrases may introduce more serious polarity conflicts to polarity classification. Our third experiment is thus to investigate the problem of polarity classification conflict between paraphrases. This experiment is conducted by counting the number of polarity class conflicts between each sentence in the test data and its paraphrases using different n-best paraphrase generation. In addition, here the system for polarity classification is trained using the expanded training data via 5-best paraphrase generation. The results are summarized in Table 11.

Table 11. Number of polarity conflicts in the test dataset yielded by systems using training datasets with/without paraphrasing

n-best	Car		Mobilephone	
	SimJ	SimJ&SimS	SimJ	SimJ&SimS
1	216	228	40	39
2	224	233	41	47
3	232	238	42	50
4	236	243	46	51
5	237	247	48	51

As can be seen from Table 11, the number of conflicts is also increasing with the rise of generated paraphrases. Also, we can observe from Table 11 that there are more polarity conflicts in the car data than in the mobilephone data. This illustrates again that the larger number of product attributes in car domain might be one potential reason for its relative lower performance in po-

larity classification, in comparison to the mobilephone domain.

Our in-depth analysis shows that there are three main possible causes for polarity conflicts, as shown in Table 12.

(1) Incorrect paraphrase generation. Wrongly-generated paraphrases possibly lead to polarity conflicts, as illustrated by the first example in Table 12.

(2) Dynamic polarity. In cases of opinionated and paraphrases with dynamic polar words, the classifier does not always works and thus cannot consistently yield correct polarity classes, as the second example in Table 12 shows.

(3) Explanatory opinionated sentence. The evaluation expressions in explanatory opinionated sentences usually have more complicated structures and most often have no explicit polarity words, as shown by the third example in Table 12. It is obviously very difficult for the system to produce correct paraphrases or perform consistent polarity classification for explanatory opinionated sentences (Kim, *et al.*, 2013).

Table 12. Examples of generated paraphrases with contradict polarity.

No.	Paraphrases with polarity conflicts
1	(a) 价格浮动频繁(The price fluctuation is frequent) (b) 价格很不给力 (The price is ungelivable) (c) 价格太高(The price is too high)
2	(a) 内置软件过多(There is too much built-in software) (b) 内置软件很多(There is very much built-in software)
3	(a) 电池一般三天左右(The duration of the battery is about three days) (b) 电池玩一段时间会发烫(The battery will be hot after a period of working) (c) 电池 1880 毫安(The battery capacity is 1880 mAh)

5 Conclusions and Future Work

In this paper, we have exploited opinion paraphrasing to enhance Chinese sentence polarity classification. We have demonstrated that paraphrasing on training corpora and test corpora can result in a significant improvement of performance in polarity classification.

The encouraging results of the present study suggest several possibilities for future research. With regard to the concentrate of our current work, we have only employed very simple techniques to perform paraphrase extraction and generation. To further enhance our system, in future

work we intend to exploit a more tailored method to achieve high-quality paraphrases for polarity classification. The present study focuses on Chinese polarity classification. In future, we also plan to extend our current system and apply it to other languages like English.

Acknowledgments

This study was supported by National Natural Science Foundation of China under Grant No.61170148 and No.60973081, the Returned Scholar Foundation of Heilongjiang Province, and Harbin Innovative Foundation for Returnees under Grant No.2009RFLXG007, respectively.

Reference

- A. Fader, L. Zettlemoyer, and O. Etzioni. 2013. Paraphrase-driven learning for open question answering. In *Proceedings of ACL'13*, pages 1608-1618.
- B. Pang, and L. Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2): 1-135.
- B. Pang, L. Lee, and S. Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of EMNLP-02*, pages 79-86.
- C.-C. Chang, and C.-J. Lin. 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(27): 1-27.
- G. Fu, and X. Wang. 2010. Chinese sentence-level sentiment classification based on fuzzy sets. In *Proceedings of COLING'10*, pages 312-319.
- M. Heilman, N.A. Smith. 2010. Tree edit models for recognizing textual entailments, paraphrases, and answers to questions. In *Proceedings of NAACL'10*, pages 1011-1019.
- M. Speriosu, S. Upadhyay, N. Sudan, and J. Baldridge. 2011. Twitter polarity classification with label propagation over lexical links and the follower graph. In *Proceedings of the First workshop on Unsupervised Learning in NLP*, pages 53-63.
- N. Madnani, and B. J. Dorr. 2010. Generating phrasal and sentential paraphrases: A survey of data-driven methods. *Computational Linguistics*, 36(3): 342-387.
- P.D. Turney. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of ACL'02*, pages 417-424.
- R. Bhagat, and E. Hovy. 2013. What is a paraphrase?. *Computational Linguistics*, 39(3): 463-472.
- R. Mehrotra, R. Agrawal, and S.A. Haider. 2012. Dictionary based sparse representation for domain adaptation. In *Proceedings of CIKM'12*, pages 2395-2398.
- R. Mihalcea, C. Banea, J. Wiebe. 2007. Learning multilingual subjective language via cross-lingual projections. In *Proceedings of ACL'07*, pages 976-983.
- S. Volkova, T. Wilson, D. Yarowsky. 2013. Exploring sentiment in social media: Bootstrapping subjectivity clues from multilingual twitter streams. In *Proceedings of ACL'13*, pages 505-510.
- S. Zhao, X. Lan, T. Liu, et al. 2009. Application-driven statistical paraphrase generation. In *Proceedings of the ACL-IJCNL'09*, pages 834-842.
- T. Mikolov. 2012. Statistical language models based on neural networks. Doctoral Thesis, Brno University of Technology.
- T. Mikolov, W. Yih, and G. Zweig. 2013. Linguistic regularities in continuous space word representations. In *Proceedings of NAACL-HLT'13*, pages 746-751.
- T. Nakagawa, K. Inui, and S. Kurohashi. 2010. Dependency tree-based sentiment classification using CRFs with hidden variables. In *Proceedings of HLT-NAACL'10*, pages 786-794.
- T. Wilson, J. Wiebe, and P. Hoffmann. 2009. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics*, 35(3):99-433
- H.D.Kim, M. Castellanos, M. Hsu, C.X. Zhai, U. Dayal, and R. Ghosh. 2013. Ranking explanatory sentences for opinion summarization. In *Proceedings of SIGIR'13*, pages 1069-1072