Expanding Corpora for Chinese Polarity Classification via Opinion Paraphrase Generation

Da Pan, Jiaying Song, and Guohong Fu(✉)

School of Computer Science and Technology, Heilongjiang University, Harbin 150080, China
pandacs@live.cn, jy_song@outlook.com, ghfu@hotmail.com

Abstract. Although much progress has been made to date on sentiment classification, lacking annotated corpora remains a problem. In this paper we propose to expand corpora for Chinese polarity classification via opinion paraphrase generation. To this end, we first exploit three strategies for opinion paraphrase generation, namely sentences re-ordering, opinion element substitution and explicit attribution implying. To improve the quality of the generated opinion paraphrases, we define four criteria for opinion paraphrase evaluation and thus present a filtering algorithm to discard improper opinion paraphrase candidates. To assess the proposed method, we further apply the expanded corpus to a SVM classifier for polarity classification. The experimental results show that the generated opinion paraphrases are beneficial to polarity classification.

Keywords: Sentiment analysis · Polarity classification · Paraphrase generation · Supported vector machines

1 Introduction

Recently, opinion mining has been attracting lots of attention in the community of natural language processing. As a pivotal sub-problem of opinion mining, sentiment polarity classification aims to predict opinionated documents or sentences as showing positive, negative or neutral opinions. To data, much progress has been made in sentiment polarity classification. However, lacking large scale annotated corpora is still a major issue. On the one hand, statistically-based methods become the majority in sentiment polarity classification. In general, a statistically-based polarity classifier needs labeled corpora for training. On the other hand, large-scale labeled corpora are still not available for polarity classification. Furthermore, opinion mining is always domain specific. Obviously, it is time and cost consuming to manually construct a large scale labeled corpus for each domain.

Over the past years, paraphrasing has proven to be an effective tool for improve the coverage of natural language processing systems such as machine translation, information retrieval and question answering [1–4]. More recently, paraphrase generation has been applied to enhance sentiment polarity classification [5]. Unlike opinion corpus

W. Che et al. (Eds.): ICYCSEE 2016, Part I, CCIS 623, pp. 362–373, 2016.
DOI: 10.1007/978-981-10-2053-7_32

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 polarity-labeled corpora and at the same time, to alleviate the data sparse problem in statistically-base polarity classifiers.

Following the line of [5], in this paper we explore opinion paraphrase generation to expand existing opinion-labeled corpora for Chinese sentence-level polarity classification. To approach this, we first construct a large number of paraphrase candidates via three strategies, namely sentences re-ordering, opinion element substitution and explicit attribution implying. In order to improve the quality of the generated opinion paraphrases, we define four criteria for opinion paraphrase evaluation and thus present a filtering algorithm to discard the improper opinion paraphrase candidates. To assess the proposed method, we further apply the expanded polarity-labeled corpus to a supported vector machine (SVM) classifier for polarity classification. The experimental results show that the generated opinion paraphrases are beneficial to polarity classification.

The rest of the paper proceeds as follows. Section 2 provides a brief review of the literature on sentiment classification and paraphrase generation. Section 3 defines the goals of paraphrase generation. Section 4 details the proposed method to Chinese sentence polarity classification via paraphrase generation. Section 5 reports our experimental results. Finally, Sect. 6 concludes our work.

2 Related Work

Sentiment classification is a fundamental problem in opinion mining, which is usually formulated as a binary classification problem [6, 7]. Most previous studies use supervised machine learning methods, including naïve Bayes model, support vector machines (SVMs), maximum entropy models (MEMS), conditional random fields (CRFs), fuzzy set, neural networks and so forth, to perform sentiment classification on different linguistic levels such as words, phrases, sentences and documents [7–11].

Lacking large scale labeled corpora is one major problem in supervised machine learning methods. To address this problem, some recent studies exploit bootstrapping or unsupervised techniques [6, 12–16]. Unfortunately, unsupervised sentiment classifiers usually have worse performance than the supervised methods. To leverage resources in the corpora to improve the sentiment classification performance, paraphrase generation approaches have been investigated.

For paraphrase generation based sentiment classification, the pervious works are just use the opinion element substitution method to enrich the data set [5, 17]. Besides the strategy of opinion element substitution for opinion paraphrase generation in [5, 17], in the present study we also explore two other strategies, namely sentences re-ordering and explicit attribution implying so as to generate more potential candidates for opinion paraphrases. Furthermore, in order to acquire paraphrases of high quality, we propose four opinion paraphrase evaluation criteria to discard the improper opinion paraphrase candidates.

3 Goals of Paraphrase Generation

Paraphrase generation is aiming at generating a set of target sentences in the same semantic with a given source sentence. In the present study, a proper generated paraphrase should satisfy four kinds of requirements as follows.

Semantic Equivalence: We hope that the generated paraphrases would help improve performance of sentiment polarity classification. Thus, the expanded sentences must be semantically equivalent and thus have the same polarity with their source ones.

Grammaticality: The paraphrase generation task can be treated as a processes of sentence-making with a certain semantic constraints, such that the constructed sentence is required to meet the basic syntax.

Frequency of Use: From the perspective of language understanding, for ease of both humans and machines to understand the meaning of sentence, we suppose that extremely common words should be used in paraphrase generation for applicability of domains.

Diversity of Language: Since the purpose of paraphrase generation is to avoid the problem of data sparseness, we believe that the changes in style and sentence patterns can enrich original presentation of semantic and improve the quality of data.

4 The Proposed Method

Figure 1 presents the general framework for Chinese polarity classification via opinion paraphrase generation.

As shown in Fig. 1, we performance paraphrase generation over both training test datasets. In addition to opinion element substitution [5], we also introduce sentences re-ordering and explicit attribution implying to produce more potential paraphrase candidates and then employ the re-ranking model to filtering the improper ones.

4.1 Paraphrase Candidate Generation

From the viewpoint of completeness and accuracy, three strategies are exploited for opinion paraphrase candidate generation. Firstly, we re-order source sentence by semantic chunks division. Secondly, replace the evaluation phrases by paraphrase knowledge base, and then make the explicit attribution implying.

Sentence Re-ordering. Typically, product reviews are composed of several chunks of attributes and the relation between the chunks is usually parallel because of its characteristics of stronger generality and concise style. Thus we treat opinion chunks as basic unit of word order adjustment, the specific process of division as follows: (1) Cut source sentence into several clauses by punctuation marks; (2) Mark the product attributes according to the product knowledge base [5]; (3) Divide clauses without attributes into corresponding chunks; (4) Combine the chunks starting with conjunctions; (5) Full permutation of parallel chunks as paraphrase candidate.

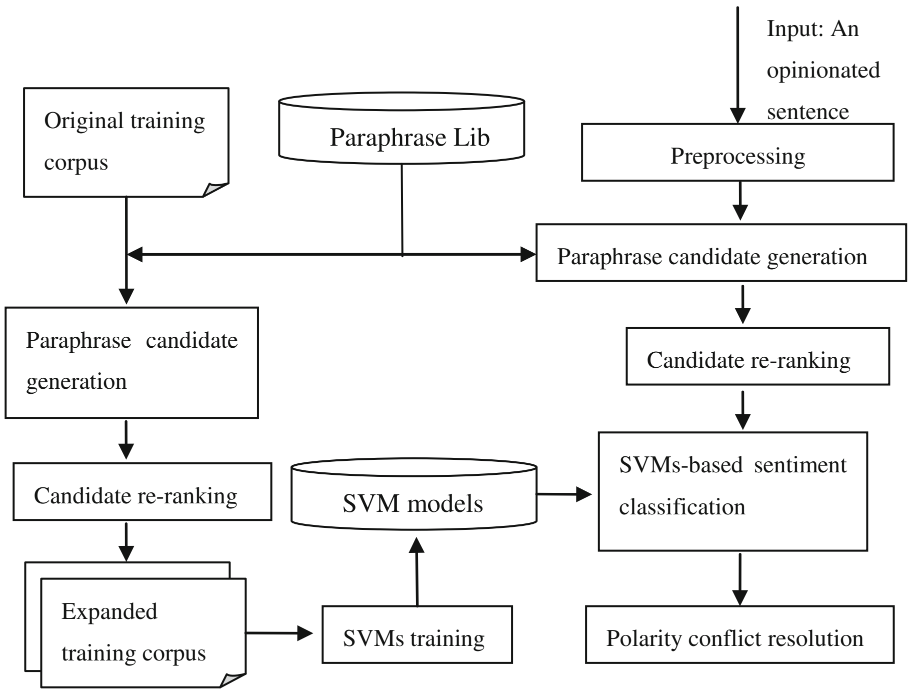

Fig. 1. The architecture of sentiment polarity classification based on paraphrase generation

Opinion Element Substitution. Since the study of domain-specific paraphrase generation is few, we construct a paraphrase knowledge base by the paraphrase recognition method by Fu *et al.* [5]. The generation approach which replacing evaluation phrase of similar attributes is simple but effective. Although the replacement method can avoid some grammatical errors and irrelevant information, but there are also some problems cannot be ignored. Due to the coarse granularity of paraphrase identification, there exists some noise in phrase knowledge base. Such as "屏幕大用着爽" (The screen is big and gives a cool feeling) and "屏幕分辨率高，看 "着好清楚啊" (The resolution of screen is high and watched clearly) are both positive evaluation for the screen, but they expressed different semantics. Other noise is introduced because of the source sentence's polarity is unknown that the candidates we make may have opposite polarity.

Explicit Attribution Implying. It is observed that attributes are usually implied in real product reviews. For example, "机子挺好看的" (the machine looks nice) has an implied attribute of appearance. Usually, mining implied attributes is very difficult in information extraction in that it depends on the guess of relevance contexts. But in the task of paraphrase generation, it is very easy to construct sentences with implied attributes by just deleting original attributes during the process of evaluation phrases substitution.

4.2 Paraphrase Candidate Filtering

In the paraphrase candidates generating, a lot of paraphrase candidates are produced for a given opinionated sentence. But some of them may be improper. The goal of the re-ranking module is thus to discard improper paraphrase candidates.

4.2.1 Criteria for Paraphrase Evaluation

We propose four criteria for opinion paraphrase evaluation as follows.

(1) Semantic Similarity of Key Strings. According to the first goal of paraphrase generation above, the primary is determining semantic equivalence between paraphrase candidates and their source sentences. Because of the generate method is based on phrase replacement, we focus on the semantic of substitutive evaluation in context and call it key string. We employ the word embedding to calculate semantic similarity of key strings [18], presented in Eq. (1).

$$Keyphrase = w_{i-3}w_{i-2}w_{i-1}p_iw_{i+1}w_{i+2}w_{i+3} \tag{1}$$

We regard the substitutive evaluation as center and three words window context to be the key string, we sum up corresponding dimension of words to synthesis vector of key string and use cosine distance to judge semantic equivalence relations.

(2) N-grams. Considering the second goal of paraphrase generation, we chose n-gram model as the basic grammar checking means. The probability of sentence in n-gram model is presented in Eq. (2).

$$p(s) = \prod_{i=1}^{m} p(w_i|w_{i-n+1}\cdots w_{i-1}) \tag{2}$$

(3) Usage Frequency of Phrases. Follow the third goal we hope the substitute phrase is common in domain reviews and add common weight for all phrases in paraphrase knowledge base. TF-IDF is a popular representation of weight in data mining and information retrieval. At the meanwhile, the matching degree of evaluations and attributes pairs is an important factor which affect the paraphrase generation. For example, both "配置-太低" (configuration is too low) and "配置-太少" (configuration is too little) are opinion collocation in paraphrase knowledge base, but people prefer to use "太低" (too low) in reviews. We take use of TF-IDF and co-occurrence frequency of attribute-evaluation pairs to express common degree of evaluation phrases.

$$Score = TF\text{-}IDF \times count(att, val) = n_{i,j} \times \log\left(\frac{D}{d_i}\right) \times count(att, val) \tag{3}$$

Where, $n_{i,j}$ is the frequency of evaluation phrase *val* exist in corpus, D is the total sentence in corpus, d_i is count of sentences which have *val* in it and *count (att, val)* represent the co-occurrence frequency of attribute *att*, evaluation *val*.

(4) Diversity of Key String Text. The morphological of word is not a primary concern and is used as a secondary filtering condition. On the base of similar semantic and correct grammar, in order to ensure the diversity of text we prefer to keep candidates whose morphological changes greatly. Thus we use Jaccard coefficient to measure the difference degree of key string (introduced in first assessment).

$$Score(morpho\log y) = 1 - jaccard(p_1, p_2) = 1 - \frac{|Set(p_1) \cap Set(p_2)|}{|Set(p_1) \cup Set(p_2)|} \tag{4}$$

Where, $Set(p_i)$ is the words set of key string p_i.

4.2.2 Strategies for Paraphrase Candidate Filtering

As we propose four evaluation criteria for opinion paraphrase evaluation above, two kinds of filtering strategies were designed to discard the candidates, namely the hierarchical filtering and the equal-Intersection filtering.

- **Hierarchical filtering**

These four evaluation criteria were sorted in line with major and minor standards. We set the most major standards called rank1, a slight secondary standards called rank2 and later followed by rank3, rank4. In filtering, we first sort all rank1 scores from highest to lowest and kept topN1 candidates. Then sort all candidates depending on rank2 scores and cut the rear ones except topN2. TopN3 and Top n-best candidates were kept later according to rank3, rank4 standards. Considering the first three criteria score are both very import, we design six kinds of order lists as follow. Using NG-score represents N-gram grammar assessment score, COS-score represents semantic similarity score, COMMON-score represents common degree score and JACC-score represents difference degree score. We set N1 300, N2 150, N3 50 in experiments of this screening strategy.

(1) order-1: NG-score → COS-score → COMMON-score → JACC-score
(2) order-2: NG-score → COMMON-score → COS-score → JACC-score
(3) order-3: COS-score → NG-score → COMMON-score → JACC-score
(4) order-4: COS-score → COMMON-score → NG-score → JACC-score
(5) order-5: COMMON-score → COS-score → NG-score → JACC-score
(6) order-6: COMMON-score → NG-score → COS-score → JACC-score

- **Equal-Intersection screening**

We suppose that all evaluation criteria is equal and necessary in this strategy and does not distinguish major or minor standards. Thus four lists of candidates in different order received by these criteria. We cut out rear half of each list and generate the final paraphrases using the intersection of remainders in four lists.

4.3 Sentiment Polarity Classification

With the paraphrase generation, we expand annotated corpus and obtain a larger scale annotated data. Whether the expanded corpus is useful for sentiment polarity classification or introduce too much noise would be verified by the sentiment classification experiment. After paraphrase generation model, we obtain *k-best* paraphrases for a review and predict polarity of these paraphrases by the polarity classifier. We need to avoid polarity conflict when the polarities of all paraphrases are in different type. In conflict resolution, we resolve it the same as Fu *et al.* [5] by voting mechanism. We totally have $k + 1$ opinionated sentence 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 $i = j$, the final polarity is same as that the original polarity of the input sentence during polarity classification.

5 Experiment Results and Discussion

To assess our approach, we have exploited the proposed paraphrase generation to two corpora of product reviews in car and cellphone domains, and further developed a SVM-based sentiment polarity classifier. This section reports the relevant experimental results.

5.1 Experimental Dataset

Table 1 shows the statistic of the experiment data.

Table 1. Statistic of the experimental data

Data	Car			Cellphone		
	Total	POS	NEG	Total	POS	NEG
Training set	2667	1318	1349	2668	1318	1350
Development set	333	183	150	332	182	150
Test set	1000	600	400	1000	600	400

5.2 Effects of Different Strategies on the Number of Generated Paraphrases

We construct a large number of paraphrase candidates for train set, develop set, test set based on sentence re-ordering and phrases replacing. The statistic of expanded corpus is listed in Table 2. The statistic show that scale of paraphrase we generated is quite large and proved our paraphrase generation method can construct a more complete candidate set. Such paraphrase candidate also shows that it may contain a lot of noise, the filtering step become quite necessary.

The filtering experiment is carried out according to the two strategies above and the statistics of 10-best paraphrase corpora of two methods are show in Tables 3 and 4.

Table 2. Statistics of the corpora with paraphrase

	Datasets	Total sentence	Positive sentence	Negative sentence
Car	Training set	466015366209	318679697922	147335668287
	Development set	54022308180	46323198637	7699109543
	Test set	144248170657	104827585126	39420585531
Cellphone	Training set	447955971558	266669529917	181286441641
	Development set	138582898809	122152527307	16430371502
	Test set	99247743050	69570802365	29676940685

Table 3. Statistics of the corpora with 10-best Rank-Cut paraphrase

	Data	Total sentences	Positive sentences	Negative sentences
Car	Training set	25160	12972	12188
	Development set	3389	1851	1538
	Test set	9990	6118	3872
Cellphone	Training set	26286	13221	13065
	Development set	3352	1898	1454
	Test set	9617	5847	3770

Table 4. Statistics of the corpora with 10-best Equal-Intersection paraphrase

	Datasets	Total sentences	Positive sentences	Negative sentences
Car	Training set	20915	11083	9832
	Development set	2974	1663	1311
	Test set	7954	4931	3023
Cellphone	Training set	21131	10749	10382
	Development set	2866	1689	1177
	Test set	7504	4557	2947

5.3 Effects of Different Paraphrase Generation on Polarity Classification

Because we have six orders in hierarchical filtering and the scale of corpora is quite different when retaining different n-best strategies. We consider five kinds of n-best paraphrases are investigated, namely 20-best, 10-best, 5-best, 2-best and 1-best. The performance of these sentiment classification experiments on development data sets are shown in Tables 5, 6, 7, 8 and 9.

Table 5. Effects of 20-best paraphrase generation on sentiment classification

Paraphrase		Order-1	Order-2	Order-3	Order-4	Order-5	Order-6	Intersection
Car	F_{pos}	**0.8777**	**0.8777**	0.8647	0.8647	0.8483	0.8483	0.8679
	F_{neg}	**0.8414**	**0.8414**	0.8235	0.8235	0.7870	0.7870	0.8339
	Acc	**0.8619**	**0.8619**	0.8468	0.8468	0.8228	0.8228	0.8529
Cellphone	F_{pos}	0.8764	0.8764	**0.8883**	**0.8883**	0.8780	0.8780	0.8711
	F_{neg}	0.8571	0.8571	**0.8693**	**0.8693**	0.8475	0.8475	0.8562
	Acc	0.8675	0.8675	**0.8795**	**0.8795**	0.8645	0.8645	0.8640

Table 6. Effects of 10-best paraphrase generation on sentiment classification

Paraphrase		Order-1	Order-2	Order-3	Order-4	Order-5	Order-6	Intersection
Car	F_{pos}	0.8757	0.8757	0.8639	0.8639	0.8504	0.8504	**0.8794**
	F_{neg}	0.8446	0.8446	0.8169	0.8169	0.8000	0.8000	**0.8464**
	Acc	0.8619	0.8619	0.8438	0.8438	0.8288	0.8288	**0.8649**
Cellphone	F_{pos}	0.8858	0.8858	**0.8864**	**0.8864**	0.8798	0.8798	0.8711
	F_{neg}	**0.8656**	**0.8656**	0.8647	0.8647	0.8523	0.8523	0.8562
	Acc	**0.8765**	**0.8765**	**0.8765**	**0.8765**	0.8675	0.8675	0.8640

Table 7. Effects of 5-best paraphrase generation on sentiment classification

Paraphrase		Order-1	Order-2	Order-3	Order-4	Order-5	Order-6	Intersection
Car	F_{pos}	**0.8930**	**0.8930**	0.8602	0.8717	0.8421	0.8421	0.8825
	F_{neg}	**0.8630**	**0.8630**	0.8231	0.8356	0.7902	0.7902	0.8410
	Acc	**0.8799**	**0.8799**	0.8438	0.8559	0.8198	0.8198	0.8649
Cellphone	F_{pos}	**0.8895**	**0.8895**	0.8840	0.8840	0.8635	0.8635	0.8718
	F_{neg}	**0.8675**	**0.8675**	0.8609	0.8609	0.8393	0.8393	0.8553
	Acc	**0.8795**	**0.8795**	0.8735	0.8735	0.8524	0.8524	0.8640

Table 8. Effects of 2-best paraphrase generation on sentiment classification

Paraphrase		Order-1	Order-2	Order-3	Order-4	Order-5	Order-6	Intersection
Car	F_{pos}	0.8865	0.8865	0.8717	0.8717	0.8429	0.8429	**0.8930**
	F_{neg}	0.8581	0.8581	0.8356	0.8356	0.7887	0.7887	**0.8630**
	Acc	0.8739	0.8739	0.8559	0.8559	0.8198	0.8198	**0.8799**
Cellphone	F_{pos}	**0.8736**	**0.8736**	0.8674	0.8674	0.8571	0.8571	0.8669
	F_{neg}	0.8467	0.8467	0.8411	0.8411	0.8267	0.8267	**0.8479**
	Acc	**0.8614**	**0.8614**	0.8554	0.8554	0.8434	0.8434	0.8580

As can be seen from Tables 5, 6, 7, 8 and 9, the re-ranking model performs well while using hierarchical filtering with order-1, order-2, order-3 and order-4 strategies. This illustrates that different paraphrase evaluation criteria have different importance in filtering improper paraphrases. So we need to determine the filtering order according to

Table 9. Effects of 1-best paraphrase generation on sentiment classification

Paraphrase		Order-1	Order-2	Order-3	Order-4	Order-5	Order-6	Intersection
Car	F_{pos}	0.8721	0.8721	0.8602	0.8602	0.8610	0.8610	**0.8760**
	F_{neg}	0.8269	0.8269	0.8231	0.8231	0.8207	0.8207	**0.8362**
	Acc	0.8529	0.8529	0.8438	0.8438	0.8434	0.8434	**0.8589**
Cellphone	F_{pos}	0.8774	0.8774	0.8387	0.8387	0.8556	0.8556	**0.8785**
	F_{neg}	0.8485	0.8485	0.7945	0.7945	0.8267	0.8267	**0.8533**
	Acc	0.8645	0.8645	0.8193	0.8193	0.8424	0.8424	**0.8671**

theirs importance. Our experimental results show that the order of importance for paraphrase evaluation criteria should be grammaticality, semantic equivalence, frequency of use, and diversity of language.

The strategy of Equal-Intersection also performs very well in several experiments. Through these experiments we determined the best strategies of two domains. The following experiment intends to investigate the effects of different paraphrase generation methods in sentiment classification. The former generation of Fu [5] were used as baseline and take n-gram score as the only criterion to acquire n-best paraphrase. We applied all paraphrase generated in baseline method and found the optimal results in 1-best to 20-best of baseline.

5.4 Comparison Results of Different Methods for Polarity Classification

The results are summarized in Table 10. The result show that if generate a large number of paraphrase only without any filtering, the result is not as good as classification on original corpus. It may show that resource in paraphrase knowledge base is unbalance and result in the number of two polarity samples is too unbalanced to modeling. On the other hand, all these paraphrase generation are on the base of baseline

Table 10. Comparison of polarity classification with/without paraphrase generation

Paraphrase		No-para	Base-para	Base-n-best	Our-n-best
Car	P_{pos}	0.8582	0.8283	0.8635	0.8785
	R_{pos}	0.8067	0.8117	0.8433	0.8433
	F_{pos}	0.8316	0.8199	0.8533	**0.8605**
	P_{neg}	0.7339	0.7257	0.7729	0.7778
	R_{neg}	0.8000	0.7475	0.8000	0.8246
	F_{neg}	0.7656	0.7365	0.7862	**0.8005**
Cellphone	P_{pos}	0.9394	0.9005	0.9372	0.9579
	R_{pos}	0.8267	0.8450	0.8700	0.8717
	F_{pos}	0.8794	0.8719	0.9023	**0.9127**
	P_{neg}	0.7797	0.7872	0.8239	0.8304
	R_{neg}	0.9200	0.8600	0.9125	0.9425
	F_{neg}	0.8440	0.8219	0.8660	**0.8829**

method, but after the sentence re-ordering and adding multiple criteria, we enriched the scale of candidate to ensure completeness and the filtering help us obtained better results. The filtering of baseline only assessed the reasonableness of grammar but did not evaluate semantic similarity from the essence of paraphrase.

6 Conclusion

In this paper, we have presented a paraphrase generation based method to corpus expansion for Chinese polarity classification. In particular, we introduce three strategies, namely sentences re-ordering, opinion element substitution and explicit attribution implying to produce potential paraphrases for a given opinionated sentence, and thus exploited four criteria to opinion paraphrase evaluation and filtering. We have also evaluated the proposed method under the framework of SVMs over two corpora of product reviews. The experimental results show that using opinion paraphrase generation is of great value to polarity classification.

Acknowledgments. This study was supported by Natural Science Foundation of Heilongjiang Province under Grant No. F2016036, National Natural Science Foundation of China under Grant No. 61170148, and the Returned Scholar Foundation of Heilongjiang Province, respectively.

References

1. Bhagat, R., Hovy, E.: What is a paraphrase? Comput. Linguist. **39**(3), 463–472 (2013)
2. Heilman, M., Smith, N.A.: Tree edit models for recognizing textual entailments, paraphrases, and answers to questions. In: Proceedings of NAACL 2010, pp. 1011–1019 (2010)
3. Zhao, S., Lan, X., Liu, T., et al.: Application-driven statistical paraphrase generation. In: Proceedings of ACL-IJCNLP 2009, pp. 834–842 (2009)
4. Fader, A., Zettlemoyer, L., Etzioni, O.: Paraphrase-driven learning for open question answering. In: Proceedings of ACL 2013, pp. 1608–1618 (2013)
5. Fu, G., He, Y., Song, J., Wang, C.: Improving Chinese sentence polarity classification via opinion paraphrasing. In: Proceedings of CLP 2014, pp. 35–42 (2014)
6. Turney, P.D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of ACL 2002, pp. 417–424 (2002)
7. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Found. Trends Inf. Retrieval **2**(1–2), 1–135 (2008)
8. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: Sentiment classification using machine learning techniques. In: Proceedings of EMNLP 2002, pp. 79–86 (2002)
9. Fu, G., Wang, X.: Chinese sentence-level sentiment classification based on fuzzy sets. In: Proceedings of COLING 2010, pp. 312–319 (2010)
10. Mikolov, T., Yih, W.-T., Zweig, G.: Linguistic regularities in continuous space word representations. In: Proceedings of HLT-NAACL 2013, pp. 746–751 (2013)
11. Tang, D., Wei, F., Yang, N., et al.: Learning sentiment-specific word embedding for Twitter sentiment classification. In: Proceedings of ACL 2014, pp. 1555–1565 (2014)

12. Mihalcea, R., Banea, C., Wiebe, J.: Learning multilingual subjective language via cross-lingual projections. In: Proceedings of ACL 2007, pp. 976–983 (2007)
13. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. Comput. Linguist. **35**(3), 399–434 (2009)
14. Speriosu, M., Sudan, N., Upadhyay, S., et al.: Twitter polarity classification with label propagation over lexical links and the follower graph. In: Proceedings of the First workshop on Unsupervised Learning in NLP, pp. 53–63 (2011)
15. Mehrotra, R., Agrawal, R., Haider, S.A.: Dictionary based sparse representation for domain adaptation. In: Proceedings of CIKM 2012, pp. 2395–2398 (2012)
16. Volkova, S., Wilson, T., Yarowsky, D.: Exploring sentiment in social media: bootstrapping subjectivity clues from multilingual twitter streams. In: Proceedings of ACL 2013, pp. 505–510 (2013)
17. Song, J., He, Y., Fu, G.: Polarity classification of short product reviews via multiple cluster-based SVM classifiers. In: Proceedings of PACLIC 2015, pp. 267–274 (2015)
18. Mikolov, T., Sutskever, I., Chen, K., et al.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, pp. 3111–3119 (2013)