Sentiment Classification of Chinese Contrast Sentences

Junjie Li[1], Yu Zhou[1], Chunyang Liu[2], and Lin Pang[2]

[1] National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
[2] National Computer Network Emergency Response Technical
Team/Coordination Center of China, Beijing, 100029
{junjie.li,yzhou}@nlpr.ia.ac.cn,
lcy@isc.org.cn, panglin_cncert@163.com

Abstract. We present the study of sentiment classification of Chinese contrast sentences in this paper, which are one of the commonly used language constructs in text. In a typical review, there are at least around 6% of such sentences. Due to the complex contrast phenomenon, it is hard to use the traditional bag-of-words to model such sentences. In this paper, we propose a Two-Layer Logistic Regression (TLLR) model to leverage such relationship in sentiment classification. According to different connectives, our model can treat different clauses differently in sentiment classification. Experimental results show that TLLR model can effectively improve the performance of sentiment classification of Chinese contrast sentences.

Keywords: Two-Layer Logistic Regression (TLLR) Model, sentiment classification.

1 Introduction

Sentiment analysis is an active research area and its main job is to identify the attitude of a text, a sentence or a phrase[1,6,10]. Both lexicon-based [4,12] and corpus-based [8,9] approaches exist for this task. In the corpus-based methods, the most common one for sentiment classification is the statistical model based on bag-of-words (BOW) [8]. Although the method is simple and effective, there are still many shortcomings, e.g., BOW ignores the relationship among words and that among sentences. Due to these defects, BOW method cannot handle many useful linguistic phenomena for sentiment classification, such as the contrast.

On one hand, in order to incorporate the relationship among words, [3] used contextual clues to disambiguate polarity. [14] proposed a dual training and dual prediction method to deal with the negation phenomenon.

On the other hand, in order to incorporate the relation among sentences, many researchers [2,11,13] make use of discourse structure to assist sentiment classification. Their experimental results show that the contrast relation is very important for sentiment classification. According to the discourse theory, a contrast sentence usually contains two clauses: the nuclei clause and the satellite clause. Previous works

C. Zong et al. (Eds.): NLPCC 2014, CCIS 496, pp. 205–216, 2014.

concentrate on treating these two parts differently in sentiment classification. [2,11,13] think different clauses in the contrast sentence have different functions in expressing sentiments. Their experiment results show that the nuclei clause is more important than the satellite one in the task. [7] studies sentiment analysis of conditional sentence systematically. [5] considers the contrast phenomenon as one kind of polarity shifting. Through splitting the sentence into polarity-shifted parts and polarity-unshifted parts, they designed three strategies (remove, shift and joint) to deal with the different parts. The most effective strategy is the joint strategy,which uses different BOW models for the two parts and learns the different BOW weights together and their experiment results support their method.

However, there are still two shortcomings in the previous works: Firstly, it is not adequate to assign different weights to the satellite and nuclei clause in contrast relation. Let's observe the following two example sentences (1) and (2).

(1) 酒店位置好|The hotel location is good, 只是设施糟糕|**yet** its facilities are poor.
(2) 酒店位置好|The hotel location is good, 但是设施糟糕|**but** its facilities are poor.

Except connectives, the two sentences consist of same words but they have opposite polarity. The reason is that different connectives in the same relation have different tendency degrees. The word "只是|yet" is a slight contrast connective. When two clauses have different polarities, the word usually makes the sentiment tend to the polarity of the former clause. However, the conclusion is different from the connective "但是|but", which always tends to polarity of the latter clause. Therefore, connectives are crucial for sentiment analysis, which are totally ignored in previous works.

Secondly, all the existing methods work as a pipeline schema. They learn word-level weights and clause-level weights separately. The pipeline approach often causes error propagation. Wrong word-level weights in the earlier stage harm the subsequent sentence sentiment label. To our best knowledge, there is no work to train word-level weights and clause-level weights jointly.

To address the issues mentioned above, we propose a Two-Layer Logistic Regression (TLLR) model in this paper, which jointly learn weights in both clause-level and word-level and consider the element of different connectives. Experimental results show that TLLR model can effectively improve the performance of sentiment classification of Chinese contrast sentences.

Although this paper only studies contrast sentences, it is also easy to integrate this method to an overall sentiment analysis or opinion mining system. The reason is that it is easy to identify contrast sentences using connectives in Table 1. When we find the input sentence is a contrast sentence, we can use the method in this paper to get the polarity.

2 Approach Overview

In this paper, we only focus on contrast sentences, which have explicit connectives (in Table 1) to split the sentence into different parts. Given such a sentence, Figure 1 depicts the general procedure of our approach. We first use connectives (in Table 1) to split the sentence into the nuclei clause and satellite clause, then the TLLR model works on the structure and returns the sentiment polarity of the sentence.

Table 1. Explicit connectives in contrast relation

Relation	Connectives
转折	不过|however 但是|but
|contrast	但|but 只是|yet 可是|however

The nuclei clause in a contrast sentence is the clause after the connective, while the satellite clause is the clause before the connective. From the following example, we can get more insights.

Input: 酒店位置好|The hotel location is good，但是设施糟糕|but its facilities are poor。

Output: Connective: “但是|but”, Relation: “转折|Contrast”. Satellite Part: “酒店位置好| The hotel location is good”, Nuclei Part: “只是设施糟糕| but its facilities are poor”.

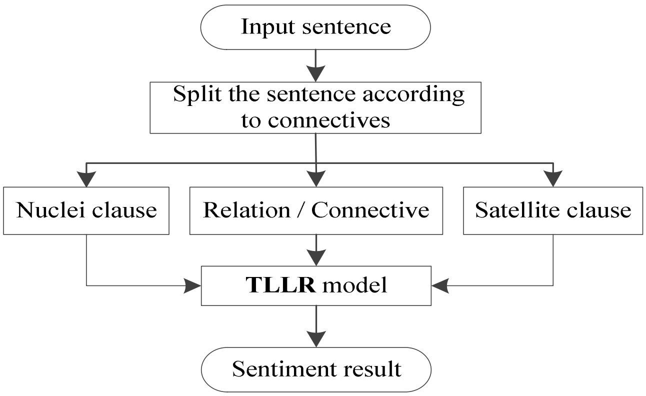

Fig. 1. Our approach overview

3 Two-Layer Logistic Regression Model

Logistic regression (LR) is a very famous model for two-class classification, which can deal with sentiment classification with BOW features. However, when we consider the contrast relation in sentiment classification, the traditional LR model based on BOW cannot handle the relation. Therefore, a Two-Layer Logistic Regression (TLLR) model is proposed to remedy the defect of LR model.

3.1 Logistic Regression Model

The structure of LR model is shown in Figure 2. The input for LR model is feature vector ($\vec{x}$) and the parameter is the weight vector ($\vec{\theta}$). Category (y) is the output corresponding to the feature vector. Sigmoid function is utilized to map the dot product of feature vector and weight vector into category. The mathematical formula about $\vec{x}, \vec{\theta}$, y can be seen in Equation (1) below.

$$P(y=1 \mid \vec{x}) = h(\vec{\theta} \cdot \vec{x}) = \frac{1}{1+e^{-\vec{\theta} \cdot \vec{x}}} \tag{1}$$

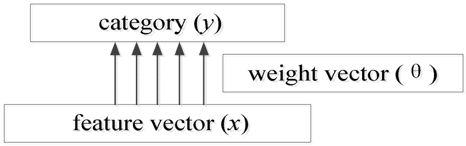

Fig. 2. The structure of Logistic Regression model

Loss Function: Suppose we have a dataset with N samples, the *k-th* sample has feature vector ($\overrightarrow{x_k}$), category (y_k). The loss function for the *k-th* sample is shown in Equation (2).

$$Cost(h(\vec{\theta} \cdot \overline{x_k}), y_k) = \begin{cases} -log(h(\vec{\theta} \cdot \overline{x_k})) & if \ \ y_k = 1 \\ -log(1-h(\vec{\theta} \cdot \overline{x_k})) & if \ \ y_k = 0 \end{cases} \tag{2}$$

Equation (3) gives the whole loss function for the dataset.

$$J(\vec{\theta}) = \sum_{k=1}^{N} Cost(h(\vec{\theta} \cdot \overline{x_k}), y_k) \tag{3}$$

We use L-BFGS algorithm to get the parameters.

3.2 Two-Layer Logistic Regression Model

TLLR model is proposed to make use of the contrast relation to improve sentiment classification. The structure of the model is presented in Figure 3.

For a contrast sentence, we use the connective in Table 1 to split the sentence into two clauses: the satellite clause and the nuclei clause. Then two feature vectors (satellite feature vector ($\overrightarrow{x_{sat,rel}}$) and nuclei feature vector ($\overrightarrow{x_{nu,rel}}$)) are employed to represent the two clauses in contrast sentence. We can get the nuclei score from the dot product of word weights vector in nuclei ($\overrightarrow{\theta_{nu}}$) and nuclei feature vector ($\overrightarrow{x_{nu,rel}}$). Similarly, we can also get the satellite score. Thus, the contrast sentence score can be computed by using the clause-level weights ($\alpha_{sat,rel}$and $\alpha_{nu,rel}$) to combine the nuclei score and the satellite score. The sigmoid function can be employed to map the contrast sentence score into category (y). The mathematical formula can be seen in Equation (4).

$$P(y=1 \mid \vec{x}) = h\left(\alpha_{sat,rel} \times (\overline{\theta_{sat}} \cdot \overline{x_{sat,rel}}) + \alpha_{nu,rel} \times (\overline{\theta_{nu}} \cdot \overline{x_{nu,rel}})\right) \tag{4}$$

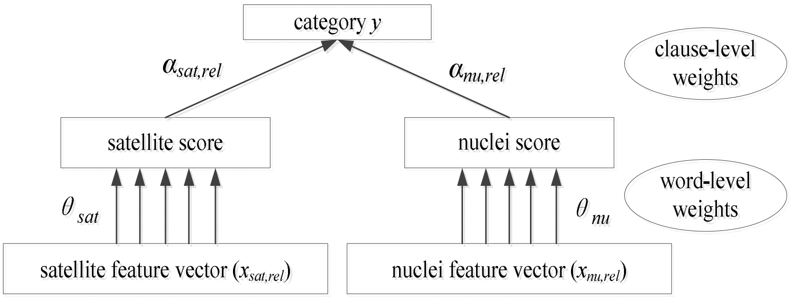

Fig. 3. The structure of Two-Layer Logistic Regression model

In Equation (4), $h(\bullet)$ is the sigmoid function. The variable *sat* and *nu* represents the satellite part and nuclei part of contrast sentences respectively. *rel* represents the relationship or connective in contrast sentences.

Here, we set a constraint for the nuclei and satellite part in the same *rel*.

$$\forall\, rel, \alpha_{sat,rel} + \alpha_{nu,rel} = 1 \tag{5}$$

Equation (4) is changed into Equation (6).

$$P(y=1 \mid \vec{x}) = h\left(\alpha_{sat,rel} \times (\overrightarrow{\theta_{sat}} \cdot \overrightarrow{x_{sat,rel}}) + (1-\alpha_{sat,rel}) \times (\overrightarrow{\theta_{nu}} \cdot \overrightarrow{x_{nu,rel}})\right) \tag{6}$$

Loss Function: For a dataset with N samples, the *k-th* sample has a feature vector ($\overrightarrow{x_k}$), category (y_k). The loss function for the *k-th* sample is shown in Equation (7).

$$Cost(P(y=1 \mid \overrightarrow{x_k}), y_k) = \begin{cases} -log(P(y=1 \mid \overrightarrow{x_k})) & if\;\; y_k = 1 \\ -log(1-P(y=1 \mid \overrightarrow{x_k})) & if\;\; y_k = 0 \end{cases} \tag{7}$$

The whole loss function for the dataset can be seen in Equation (8).

$$J(\vec{\theta}, \alpha) = \sum_{k=1}^{N} Cost(P(y=1 \mid \overrightarrow{x_k}), y_k) \tag{8}$$

Parameter Estimation: By minimizing the loss function, we will get all parameters. To simplify the formulation, we set:

$$g(\overrightarrow{x_k}, y_k) = P(y=1 \mid \overrightarrow{x_k}) - y$$

For the *k-th* sample, we get the gradient of the word weights in Equation (9) and (10).

$$\frac{\partial J(\vec{\theta}, \alpha)}{\partial \overrightarrow{\theta_{sat}}} = g(\overrightarrow{x_k}, y_k) \times \alpha_{sat,rel} \times \overrightarrow{x_{sat,rel}} \tag{9}$$

$$\frac{\partial J(\vec{\theta}, \alpha)}{\partial \overrightarrow{\theta_{nu}}} = g(\overrightarrow{x_k}, y_k) \times (1-\alpha_{sat,rel}) \times \overrightarrow{x_{nu,rel}} \tag{10}$$

Equation (11) gives the gradient of the clause-level weights. L-BFGS is used to get the parameters from the gradient.

$$\frac{\partial J(\vec{\theta},\alpha)}{\partial \alpha_{sat,rel}} = g(\overline{x_k}, y_k) \times (\overline{\theta_{sat}} \cdot \overline{x_{sat,rel}} - \overline{\theta_{nu}} \cdot \overline{x_{nu,rel}}) \tag{11}$$

4 Experiments

4.1 Distribution of Contrast Sentences and Connectives

We have crawled some cellphone reviews from 360buy[1]. After using sentence delimiter to split the reviews into sentences, we get 693,125 sentences. By using connectives in Table 1, we can get 41,629 contrast sentences, accounting for 6% in all sentences.

Table 2. Statistics of sentences with various contrast connectives

Contrast Connectives	Sentence Number	Sentence Proportion (%)
不过|however	14,869	35.72
但是|but	11,919	28.63
但|but	9,553	22.95
只是|yet	3,047	7.32
可是|however	2,241	5.38
All	41,629	100

A further analysis of the distribution of various connectives in contrast sentences is shown in Table 2. "不过|however", "但是|but", and "但|but" are the top three connectives in all contrast sentences. In the following experiment, we will find different connectives have different tendency degree for the sentiment classification.

4.2 Experiment Dataset

Reviews in 360buyare scored from 1 to 5 discretely. The higher score means more positive. We label the sentiment label of a comment review based on the score. If the score is 1 or 2, we label it negative. If the score is 4 or 5, we label it positive. We reject the review whose score is 3. After using contrast connectives to get contrast sentence review, we collect 2K contrast sentences as our experiment dataset. The dataset is balanced for two categories: positive and negative.

[1] http://www.360buy.com/

4.3 Experiment Setting

We have compared three groups of experiments to verify the effectiveness of our model. They are Baseline, Other Models, and Our Model.

- **Baseline:** After using BOW to represent the sentence, we treat Support Vector Machine (SVM) and Logistic regression (LR) models as baseline.
- **Other Models**
  - ✧ SNSS (Single Nucleus Single Satellite Method), following the idea of [2,11,13]. we get clause-level weights according to different clauses in contrast relation.
  - ✧ JS (joint strategy), the joint strategy used in [5]. Specifically, the joint strategy uses different BOW model to represent the different clauses of contrast sentences and learns the different BOW weights together.
- **Our Model:** we get the results of TLLR model. For the word-level weights, we first utilize the parameters of LR to initialize them, and then train weights according to Equation (9-10). The weights in clause level are set according to different parts in a contrast sentence with different connectives, and we use Equation (11) to get the parameters.

4.4 Experiment Results

The 5-fold cross-validation results are given in Table 3.

Baseline Group: We use the results of LR as the results of the comparative experiment in baseline system since our model is based on LR.

Other Models Group: Compared with the performance of LR baseline system (86.05%), the result of SNSS (86.00%) is decreased slightly. The reason is due to the drawbacks of the traditional pipeline method. As SNSS uses two-layer model which fixes word weights (from LR baseline) and just learns clause-level weights. The word weights in LR baseline are learned from contrast sentence dataset, which can be suitable for sentiment classification of contrast sentences. However, the word weights are probably not be suitable for clause sentiment classification. Therefore, the results of other models are slightly worse than the baseline. However, the result of JS is better than the baseline result, which shows different BOW model for different parts is useful for sentiment classification.

Table 3. Experiment results

Systems	5-Fold Cross Validation Results (%)					
	First	**Second**	**Third**	**Fourth**	**Fifth**	**Average**
SVM	86.00	86.00	83.00	85.00	85.50	85.10
LR	86.75	86.75	83.00	87.00	86.75	86.05
SNSS	86.25	88.00	83.50	86.00	86.25	86.00
JS	87.50	85.75	85.25	85.75	86.25	86.10
Our Model	86.58	87.70	84.95	86.73	87.73	**86.74**

Our Model Group: Compared with other models and the baseline system, our model gets better results by achieving an improvement by overall 0.7 point in the dataset. The reason our model is better than JS is that, JS is a simple version of TLLR. When we use one to fix the clause-level weights and just learn word-level weights, the TLLR model is equal to JS.

4.5 Parameter Analysis

In this section, the reasons why our model can work well will be presented. We set parameters in TLLR using the following methods. Since our model is a two-layer structure, we have to set the method to train word-level weights and clause-level weights separately.

For the word-level weights, we have two kinds of setting method. One is "Fix", which fixes word-level weights through the parameters of LR baseline. The other is "Diff", which first utilizes the results of LR to initialize them, and then trains the weights according to Equations (9) and (10). Compared with the two kinds of methods, we will confirm whether our joint model is better than the previous pipeline methods.

For the clause-level weights, we also have two kinds of setting methods. One is "Relation", which sets clause-level weights according to the contrast relation. The other is "Connective", which sets clause-level weights based on the different connectives. Compared with these two kinds of methods, we will find out whether the connectives are better than the contrast relation to capture which clause is more important for sentiment classification of contrast sentence.

- ✧ **FixRelation:** We fix our word-level weights through the parameters of LR baseline. We set the clause-level weights according to the different clauses in contrast sentence, and we use Equation (11) to get the parameters.
- ✧ **FixConnective:** We fix word-level weights through the parameters of LR baseline. The weights in clause-level are set according to the different clauses in contrast sentence for different connectives, and Equation (11) is used to get the parameters.
- ✧ **DiffRelation:** For the weights in word level, we first utilize the results of LR to initialize them, and then train the weights according to Equations (9) and (10). For the weights in clause level, the setting method is as the same as that in **FixRelation**.
- ✧ **DiffConnective:** For the weights in word level, the setting method is as the same as that in the **DiffRelation**. For the weights in clause level, the setting method is as the same as that in **FixConnective**.

Table 4 gives the results of the four methods in TLLR.

Table 4. Different parameter results in TLLR

Methods	5-Fold Cross Validation Results (%)					
	First	Second	Third	Fourth	Fifth	Average
FixRelation	86.25	88.00	83.50	86.00	86.25	86.00
FixConnective	86.25	87.75	84.25	86.75	87.25	86.45
DiffRelation	87.16	87.80	84.16	86.33	86.56	86.40
DiffConnective	86.58	87.70	84.95	86.73	87.73	86.74

Learning all Weights vs. Learning Clause-Level Weights: From Table 4, we can find the result of DiffRelation (86.40%) is better than the result of FixRelation (86.00%). The result of DiffConnective (86.74%) is also better than the result of Fix-Connective (86.45%). Based on the previous comparisons, we can find that the model, which learns word-level weights and clause-level weights jointly, is better than the model that only learns clause-level weights for the task of contrast sentence sentiment classification.

Connective vs. Relation: Which is the best parameter to model the clause sentiment label into sentence sentiment label, the connective or the relation? From Table 4, we can find the average results of FixConnective (86.45%) are better than the average results of FixRelation (86.00%), and the average results of DiffConnective (86.74%) are better than the average results of DiffRelation (86.40%). Therefore, we find the answer for that question is the connective, not the relation. Compared with the relation, connectives can be more correct and careful to reflect which part is more important for the contrast sentences. To analyze the problem more carefully, we get the average clause-level weights of 5-fold cross-validation in Table 5.

Table 5. Clause-level weights in TLLR Model

Type / Model		Satellite	Nuclei
Fix Relation	转折|Contrast	0.48	0.52
Fix Connective	不过|however	0.6	0.4
	但|but	0.46	0.54
	但是|but	0.41	0.59
	只是|yet	0.56	0.44
	可是|however	0.41	0.59
Diff Relation	转折|Contrast	0.45	0.55
Diff Connective	不过|however	0.58	0.42
	但|but	0.42	0.58
	但是|but	0.36	0.64
	只是|yet	0.55	0.45
	可是|however	0.34	0.66

For the model we fix word level weights (FixRelation and FixConnective), we can find that our model can capture that the nuclei clause (0.52) in sentence is more important than the satellite clause (0.48) for sentiment classification in the contrast relation. For the connectives, our model can capture more information. For connectives “但是|but” and “但|but”, the nuclei clause (0.67, 0.64) is more important than the satellite clause (0.33, 0.36), however, for connectives “只是|yet” and “不过|however”, the results are different. The satellite clause (0.53, 0.80) is more important than the nuclei clause (0.47, 0.20). The conclusion is also consistent with the situation when we use these connectives. We can also get the same conclusion from Table 5 for DiffRelation and DiffConnective.

5 Discussion and Analysis

Some case studies are discussed to illustrate the advantages of our model in this section, which are shown in Figure 4 and Figure 5.

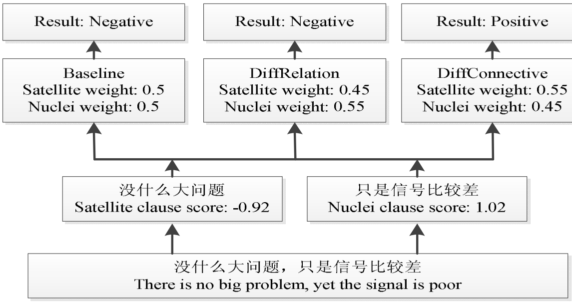

Fig. 4. An example about '只是|yet'

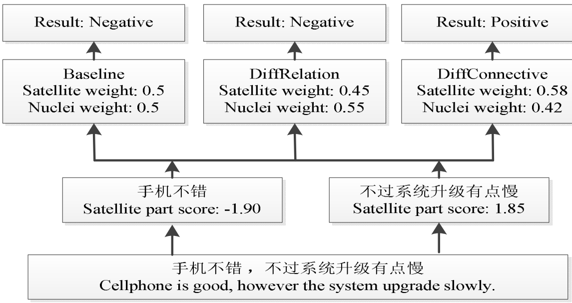

Fig. 5. An example about '不过|however'

The first example in Figure 4 is about the connective '只是|yet'. The standard sentiment label for the example is positive. Different scores result in different polarities. When the satellite clause score is bigger than 0, then the polarity of the clause is positive, and vice versa. Baseline parameter will set the satellite and nuclei weight to 0.5, then the sentence score will bigger than 0, therefore the result for Baseline is negative. DiffRelation parameter will also get the same result. However, when we set clause weight using DiffConnective parameter, the result will be reverse. For the connective '只是|yet', the satellite clause is important than the nuclei clause in DiffConnective parameter, which makes the method get the right answer.

The second example is about the connective '不过|however'. The standard sentiment label for the example is positive. We can get the similar result of analysis from Figure 5.

From the previous analysis, we can find the following conclusion. When the connective in a contrast sentence is '只是|yet' or '不过|however', our model can work better than related work. From the connective distribution in contrast relation in Table 3, the percentage of the two connectives are about 43%. Therefore, it is useful to consider different clause-level weights with different connectives.

6 Conclusion and Future Work

To address sentiment classification of contrast sentence, this paper proposes a TLLR model. Compared with the existing pipeline methods, TLLR model is a global model, which learns different level weights jointly. For the clause-level weights, TLLR model is more careful to capture the importance of the nuclei and satellite clause in the sentence for the contrast relation. For the word-level weights, TLLR model can learn more adequate weights for the clause-level sentiment label. From the experiment results, we find that our model is better than the baseline and all other existing models. Our contributions can be summarized as follows:

(1) TLLR model considers different clause-level weights with different connectives, not just relations.
(2) TLLR model can learn word weights and clause-level weights together, which can avoid the drawbacks in the pipeline method.
(3) Our experiment results confirm different connectives have different influence on overall sentiment classification.

The future work can be considered in two lines. The first line is to attempt to help sentiment classification by using other relations. The second line is to incorporate aspect into sentiment classification of contrast sentence. How to deduce the overall polarity from the polarity of each aspect is need to do for the future work.

Acknowledgements. We thank the three anonymous reviewers for their helpful comments and suggestions. We thank Haitong Yang for his help with model construction. We thank Chengqing Zong for his grateful comments on an earlier draft. The research work has been funded by the High New Technology Research and Development Program of Xinjiang Uyghur Autonomous Region under Grant No.201312103.

References

1. Balahur, A., Mihalcea, R., Montoyo, A.: Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and applications. Computer Speech & Language 28(1), 1–6 (2014)
2. Heerschop, B., Goossen, F., Hogenboom, A., Frasincar, F., Kaymak, U., de Jong, F.: Polarity analysis of texts using discourse structure. In: Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM), pp. 1061–1070 (2011)

3. Kennedy, A., Inkpen, D.: Sentiment classification of movie reviews using contextual valence shifters. Computational Intelligence 22(2), 110–125 (2006)
4. Kim, S.-M., Hovy, E.: Determining the Sentiment of Opinions. In: Proceeding of the Intenerational Conference of Computational Linguistics (COLING), pp. 1367–1373 (2004)
5. Li, S., Huang, C.-R.: Sentiment Classification Considering Negation and Contrast Transition. In: Proceedings of the Pacific Asia Conference on Language, Information, and Computation, PACLIC (2009)
6. Liu, B.: Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies 5(1), 1–167 (2012)
7. Narayanan, R., Liu, B., Choudhary, A.: Sentiment Analysis of Conditional Sentences. In: Proceedings of Empirical Methods in Natural Language Processing (EMNLP), pp. 79–86 (2009)
8. Pang, B., Lee, L., Vaithyanatha, S.: Thumbs up? sentiment classification using machine learning techniques. In: Proceedings of Empirical Methods in Natural Language Processing (EMNLP), pp. 79–86 (2002)
9. Pang, B., Lee, L.: A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In: Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), pp. 271–278 (2004)
10. Pang, B., Lee, L.: Opinion Mining and Sentiment Analysis. Foundations and Trends® in Information Retrieval 2(2), 1–135 (2008)
11. Taboada, M., Voll, K., Brooke, J.: Extracting sentiment as a function of discourse structure and topicality. Simon Fraser University, Tech. Rep, 20 (2008)
12. Turney, P.D.: Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), pp. 417–424 (2002)
13. Wang, F., Wu, Y., Qiu, L.: Exploiting Discourse Relations for Sentiment Analysis. In: Proceeding of the Inteneratioal Conference of Computational Linguistics (COLING), pp. 1311–1319 (2012)
14. Xia, R., Hu, X., Lu, J., Yang, J., Zong, C.: Instance Selection and Instance Weighting for Cross- Domain Sentiment Classification via PU Learning. In: Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), pp. 2176–2182 (2013)