DFDS: A Domain-Independent Framework for Document-Level Sentiment Analysis based on RST

Zhenyu Zhao[1,3], Guozheng Rao[1,3,*], Zhiyong Feng[2,3]

1 School of Computer Science and Technology, Tianjin University, Tianjin, China
2 School of Computer Software, Tianjin University, Tianjin, China
3 Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin, China
{zhaozhenyu_tx, rgz, zyfeng }@tju.edu.cn

Abstract. Document-level sentiment analysis is among the most popular research fields of nature language processing in recent years, in which one of major challenges is that discourse structural information can be hardly captured by existing approaches. In this paper, a domain-independent framework for document-level sentiment classification with weighting rules based on Rhetorical Structure Theory is proposed. First, original textual documents are parsed into rhetorical structure trees through a preprocessing pipeline. Next, the sentiment score of elementary discourse units is computed via sentence-level sentiment classification method. Finally, according to the rhetorical relation between neighbor discourse units, we define weighting schema and composing rules based on which scores of elementary discourse units are summed recursively to the whole document. Experiment results show that our approach has better performance on datasets in different domains, compared with state-of-art document-level sentiment analysis systems based on RST, and the best result is 15% higher than baseline.

Keywords: Sentiment Analysis, Rhetorical Structure Theory, Domain Independent.

1 Introduction

With internet integrated everywhere in daily life, more and more people share their opinions such as movie reviews, service feedback and so on, on social media or online communities. Researches has proved that authors' sentiments implicated by such data can be of great value to business activities as well as public security [1, 2, 3, 4].

Although sentiment analysis and opinion mining has been attractive in recent years and yielded a great number of excellent results in different respects [5, 6], there are still lots of challenges to be settled. This work mainly focuses on the classification of sentiment polarities of textual documents, especially long texts.

Compared with sentiment classification at sentence-level, that of document-level has its own special challenges. For example: there is a common phenomenon that several opposite sentiments may inhere in the same article, where the opposite opinion may

[*] Corresponding author

serve as a foil to the main opinion or the author just wants to make his review more comprehensive, which are illustrated as follows:

"[It could have been a great movie.]^A[it does have beautiful scenery,]^B[some of the best since Lord of the Rings.]^C[The acting is well done,]^D[and I really liked the son of the leader of the Samurai.]^E[He was a likeable chap,]^F[and I hate to see him die.]^G[But, other than all that, this movie is nothing more than hidden rip-offs.]^H" [7].

Although the review has a great length of praise, from sentence B to G, the author's main idea is to disparage the movie, which can be generated just by the sentences A and H. However it is easy for human to make such a judgement but hard for a computer.

The main reason is that the current approaches can hardly comprehend the structure and rhetorical manners of the article. To address this problem, an intuitive idea is to compose sentences' scores to the whole text based on Rhetorical Structure Theory (RST; Mann and Thompson [8]), where relations between neighbor sentences are defined and several connected discourse units are separated into Nucleus (the more important part) and Satellite(the less important part). Specifically, the example showed in last paragraph can be parsed as Fig.1 according to RST.

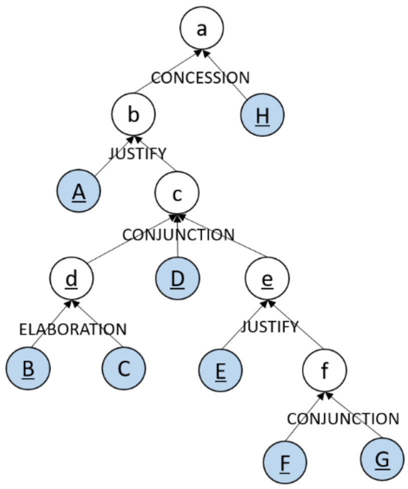

Fig. 1. Representation of the example after RST parsing

The leaf nodes are Elementary Discourse Units (EDUs) and the nodes with a line below the label are nucleus. From Fig. 1, although the discourse has a lot of positive words, we still know the discourse unit H is the main opinion of the author because H is the Nucleus towards the composition of discourse units from A to G, which have a great meaning for document-level sentiment analysis.

As automatic RST parser has improved considerably, for example, the state-of-the-art automatic RST parsing system DPLP (also employed in this paper) proposed by Y Ji [9], it is time to think about the possibility and practicability of combining discourse parsing with sentence-level sentiment analysis approaches to deal with document-level sentiment classification.

In this paper, we proposed a Domain-independent Framework for Document-level Sentiment analysis based on RST (DFDS), which has been proved to have more competitive results on datasets in different domains. The main contribution of this work is listed as follows:

- A novel domain-independent document-level sentiment analysis framework is presented, which combines Rhetorical Structure Theory on discourse parsing and recursive neural network on sentiment classification of elementary discourse units;
- Effective weighting schema and composing rules of discourse units are proposed;
- We explore the possibility and performance of composing different sentence-level sentiment analysis methods with RST parser. And we believe that with improvement of automatic discourse parsing, the contribution of this work will have referential significance.

2 Preliminaries

2.1 Rhetorical Structure Theory

Rhetorical Structure Theory is a compositional framework of discourse parsing that defines 24 relations of discourse structure, such as condition, concession and so on. Elementary discourse units (EDUs) are combined into larger discourse units according to RST, recursively covering the whole text in the end. In each relation, there may be a nucleus and a satellite which is called “Nucleus-Satellite” relation or several nucleus which is called “Multinuclear”. The nucleus discourse unit plays a more important role in the larger discourse unit, while the satellite is less important.

RST has been widely use in discourse structure parsing. In the early work on discourse parsing, hand-crafted rules and heuristics were applied to build discourse parsing trees according to Rhetorical Structure Theory [10]. Soricut and Marcu introduced probabilistic models to identify elementary discourse units and built discourse parsing trees, which results in the birth of automatic parser SPADE [11]. But SPADE can only parse discourse within sentence-level. Feng and Hirst developed an RST discourse parser, based on the HILDA discourse parser and rich linguistic features [12].By combining the machinery of large-margin transition based structure prediction, the automatic discourse parser named DPLP has reached a satisfactory accuracy about 62%, while the accuracy of human-annotated relation judgement is just 65.8% [9]. S Li et al. first proposed to use dependency structure to represent the relations between EDUs and got competitive performance [13].

2.2 Document-level sentiment analysis

There has been plenty of sentiment analysis approaches proposed for document-level sentiment polarity classification. Bo Pang et al. proposed a novel machine-learning method with text-categorization techniques to classify the subjective portions of the document [14]. A Sharma et al. used three popular sentiment lexicons to extract sentiment representing features and used BPANN to do classification [15]. Tang D et al. proposed Gated Recurrent Neural Networks to overcome the challenge of encoding the intrinsic relations between sentences in the semantic meaning of a document [16]. Xu J et al. presented Cached Long Short-Term Memory neural networks to capture the overall semantic information and get the state-of-the-art results on three publicly available document-level sentiment analysis dataset [17]. Approaches based on supervised learning seem to have more competitive accuracy than unsupervised methods, but they always have strong dependence on domain and scale of datasets, which makes most document-level sentiment analysis methods inefficient and domain-dependent.

Contrary to consistent improvements on discourse parsing, there are few efforts to incorporate RST into sentiment analysis. Voll et al. proved that RST could improve the results of lexicon-based sentiment analysis with manually-annotated RST parse trees [18]. However, it's very time-consuming and expensive to annotate relation between spans. Heerschop et al. proposed Pathos, a framework based on a document's discourse structure [19]. Wang et al. incorporated hierarchical discourse structure into an unsupervised sentiment analysis framework, but they used manually-annotated discourse parses and a small dataset (604 reviews) [20]. Zhou et al. focused on the automation of recognizing intra-sentence level discourage relations for polarity classification with unsupervised method [21].

The researches summarized above specialized in document-level classification, but just applied RST on intra-sentence level. Hogenboom et al. employed a weighting scheme as well as HILDA on just one dataset [22], which is similar to our approach, but our schema and rules do not only differentiate satellite and nucleus weights by types of relations in RST, but also the sentiment polarities of satellite and nucleus. Moreover we conduct experiments on four datasets in different domains and get better results. Bhatia et al. showed that RST can improve text-level sentiment analysis and proposed three methods, which employed the dependency-based discourse tree. But their work based on lexicon just considered dependency of satellite and nucleus instead of the rhetorical relation they belong to, and machine learning based methods they proposed had little improvement compared with their baseline [7].

3 Overall Framework

The overall framework is shown in Fig. 2. Firstly, original texts are preprocessed through a NLP pipeline, during which tokenization, part-of-speech tagging and sentence splitting have been done. We apply Stanford CoreNLP[2] in this step in view of its good performance. Secondly, texts are parsed into rhetorical structure trees and structural information in a text can be extracted. At the same time, the sentiment scores of sentences are computed via a sentence-level analysis method. Finally, according to rhetorical structure trees and relations, weighting rules are proposed to sum up the scores

[2] Available at http://stanfordnlp.github.io/CoreNLP/

of sentences recursively to the whole document and then the sentiment polarities will be classified.

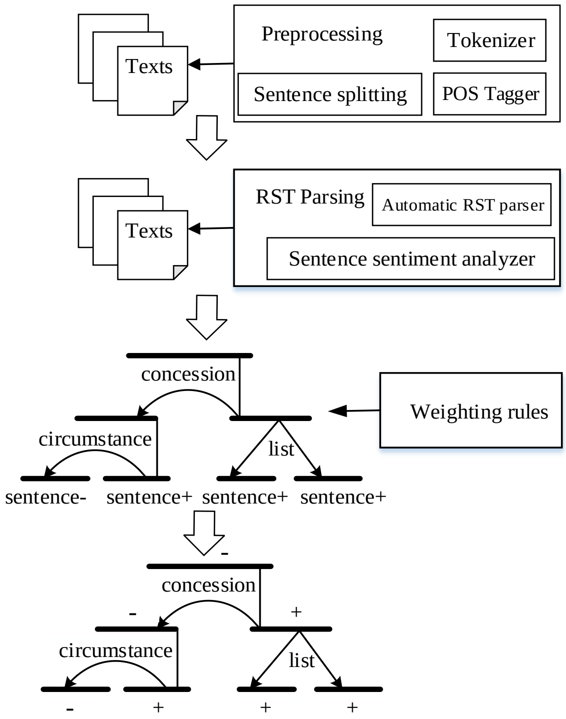

Fig. 2. The overall framework of our method

In a RST parsing tree, sentiment scores of nodes except leaves, which correspond to each elementary discourse unit, are computed via the following formula:

$$S_i = \sum_{j=1}^{k} \varphi(s_{ij}) \quad (1)$$

Where s_{ij} is the jth child node of the node n_i, which have k child nodes. The sentiment scores of child nodes functioned in weighting rules φ sum up to get the sentiment score S_i. If S_i>0, we believe the node n_i is positive, and if S_{root}>0, we believe the document is positive, otherwise n_i is negative except the case of S_i=0, which is called neutral. Weighting rules φ of each relation will be defined and introduced in detail in section 4.

For leaf nodes, sentence-level sentiment analysis approaches are introduced in to get the sentiment scores of elementary discourse units. To select a more effective approach, we compare two methods based on lexicon with Stanford-Sentiment Annotator[3]. For lexicon based methods, the score of a sentence is computed via the following formula:

[3] http://nlp.stanford.edu/sentiment

$$S = \sum_{i=1}^{n} w_i \tag{2}$$

Where w_i is the sentiment score of the ith word of the sentence according to the lexicon.

4 Weighting Rules

Weighting rules include composing rules of discourse units in different rhetorical relations and weighting schema w for each composing rule, both of which are defined and explained in this section.

4.1 weighting schema

There are 24 rhetorical structure relations defined in RST where 20 relations are "Nucleus-Satellite" and the others are multinuclear. In "Nucleus-Satellite" relations, we define five types weights of w: w_{vs}, w_s, w_l, w_{vl}, w_e.

$w_s \in [0.3, 0.4]$ and $w_l \in [0.6, 0.7]$ are used to weaken or strengthen the influence of discourse units to the overall document; w_{vs} is used to eliminate the influence of the discourse units that have not much to do with the author's attitude, but if $w_{vs} = 0$, there will be lots of neutral results, so we define $w_{vs} \in (0,0.2]$ and accordingly define $w_{vl} \in [0.8, 1)$ to retain the important units. These four types of weights are all below 1, thus with sentiment scores summed recursively from leaves to the root in a RST tree, sentences more far away from the root, which means they are less important, will have less influence on the final sentiment score of the overall document. Particularly, w_e is defined more than 1 to outstand the key discourse unit, such as for conclusion sentences.

4.2 Composing rules of rhetorical structure relations

In "Nucleus-Satellite" relations, given a node n_i and sentiment scores of its two child nodes, $s_i(nucleus)$ and $s_i(satellite)$, we separate 20 RST N-S relations into 7 categories functioned by different composing rules:

Categories 1.
Relations: *antithesis, concession*
Composing rules*:*

If $s_i(nucleus)$* $s_i(satellite)$<=0

$$S_i = w_l * s_i(nucleus) - w_s * s_i(satellite)$$

else

$$S_i = w_l * s_i(nucleus) + w_s * s_i(satellite)$$

Explanation: The nucleus span is thought to be the point of the authors, while the satellite often has the opposite polarity like" Even though the picture is fine, I don't like the film without the plot". In other cases, the score of nucleus part can be 0, but we can take use of the opposite number of the satellite part to get the sentiment polarity of larger discourse unit. If the satellite and nucleus have the same polarity, we just add them up but put the nucleus at a prominent position via a larger weight.

Categories 2.
Relation: *circumstance, background*
Composing rules:

$$S_i = w_{vl} * s_i(nucleus) + w_{vs} * s_i(satellite)$$

Explanation： As for the relations of *circumstance* and *background*, the satellite always indicate the time or the place in which the nucleus part takes place, which has not much to do with the opinion of authors, so we set w_{vs} as the weight of satellite part to eliminate its influence, and w_{vl} as the weight of nucleus part.

Categories 3.
Relation: *condition*
Composing rules:

$$S_i = -w_{vl} * s_{ia}(nucleus) + w_{vs} * s_i(satellite)$$

Explanation： It seems hard to tell whether the polarity of nucleus in relation of *condition* can represent author' true opinion, but in the domain of reviews, the satellite part always states the case that actually does not happen, so the nucleus part most likely stands in the opposite position with the author.

Categories 4.
Relation: *motivation*, *purpose*
Composing rules:

If $s_i(nucleus) * s_i(satellite) <= 0$

$$S_i = w_{vl} * s_i(nucleus) + w_{vs} * s_i(satellite)$$

else

$$S_i = w_l * s_i(nucleus) + w_s * s_i(satellite)$$

Explanation： In most cases, the satellite of relation *motivation* and *purpose* has the same polarity with that of the nucleus, and we just sum them up multiplying different weights. If their polarities are opposite, we only retain the scores of the nucleus in order to eliminate the uncertainty of the satellite.

Categories 5.
Relation: *evidence*, *justification*, *restatement*, *reason*, *result*, *enablement*
Composing rules:

If $s_i(nucleus) * s_i(satellite) <= 0$

$$S_i = w_{vs} * s_i(nucleus) + w_{vl} * s_i(satellite)$$

else

$$S_i = \max\{s_i(nucleus), s_i(satellite)\}$$

Explanation： As for these relations, in most cases, the satellite should have the same polarity with the nucleus, and we choose the maximum as the score of the parent span. Since the accuracy of sentence-level sentiment analysis method still doesn't reach 100%, if the polarities of the satellite and the nucleus are different, the satellite are the detailed description of the author's attitude.

Categories 6.
Relation: *evaluation*, *conclusion*
Composing rules:

$$S_i = w_e * s_i(nucleus) + w_{vs} * s_i(satellite)$$

Explanation： The nucleus part in relations of evaluation and conclusion is always the summary of previous article, thus it is reasonable to enlarge the proportion of the nucleus.

Categories 7.
Relation: *other N-S relations in RST*
***Composing rules**:*

$$S_i = w_l * s_i(nucleus) + w_s * s_i(satellite)$$

Explanation： With regard of the others N-S relations in RST, we just add the scores of the nucleus and the satellite up, although there may be several relations which may make some difference to the sentiment classification of whole texts.

What's more, for multinuclear relations, we follow the fomula:

$$S_i = \sum_{j=1}^{k} S_{ij} \tag{3}$$

Where S_{ij} is the sentiment score of *j*th child node among k child nodes of node n_i.

5 Experiments

We conduct experiments for document-level sentiment classification, based on sentence-level method and weighting rules. We describe the details and results of the experiments in this section.

5.1 Datasets and setup

Experiments are conducted on four datasets[4] in different domains, which were collected by Blitzer et al. The datasets contain book reviews denoted as "books", DVD reviews

[4] Available at http://www.cs.jhu.edu/~mdredze/datasets/sentiment/

denoted as “dvd”, electronics reviews denoted as “elec” and houseware reviews denoted as “houseware”. All four datasets have both 1000 positive documents and 1000 negative documents.

Before weighting rules are used, original texts have to be preprocessed into RST trees. Firstly, Stanford CoreNLP is applied to do tokenization, POS tagging and splitting the texts into sentences. Secondly, we use an automatic parser named DPLP[5] proposed by Ji Y et al. [13] to parse the texts to RST trees. In order to use the weighting rules we defined, we restructure the RST trees with java and for leaf nodes, different sentence-level sentiment classification methods are introduced to compute the sentiment scores of EDUs

Since there are few related work on document-level sentiment analysis with RST, we employed the state-of-the-art document-level sentiment classification system based on RST, Discourse Depth Reweighting (DDR) proposed by Bhatia et al. [7], as our baseline. It should be noted that there are three methods proposed in Bhatia’s work, but only DDR based on lexicon, which is the baseline we choose in this paper, has great improvement then their baselines while the other two methods seem to make little contribution. What’s more, most machine learning methods do not have comparability with our work since all weighting schema, composing rules and sentence-level sentiment in our work do not need to be trained on the datasets used in experiments, which makes our method domain-independent and efficient. However, most document-level sentiment analysis methods based on machine learning, such as neural network, need large scale of data to train a model and the test performance have strong dependence on the domain of train datasets.

5.2 Determination of weights w

We have defined 5 types of weights (w_{vs}, w_s, w_l, w_{vl}, w_e) in section 4 to adjust the influence of discourse units in different categories, but specific value will be determined in this section.

8 weighting schemas denoted as Ⅰ, Ⅱ, Ⅲ,…are presented and tested on the same dataset (book reviews). We apply two lexicons to get sentiment scores of sentences and evaluate the performance of each schema. The first lexicon[6], which is proposed by Wilson et al. and has also been applied in our baseline, contains 2721 positive words and 4914 negative words. The second lexicon denoted as SWN[7] (sentiWordNet) has 21714 positive words and 25173 negative words. Distribution of Part of speech of words contained in both lexicon is shown in Table 1

5 Available at https://github.com/jiyfeng/DPLP

6 Available at http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/

7 Available at http://sentiwordnet.isti.cnr.it/

Table 1. Distribution of Part-of-Speech of words.

Lexicon	Noun		Adjective		Verb		Adverb	
	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Wilson et al.	1030	2002	1541	2462	473	1061	0	0
SentiWord-Net	10082	13171	6330	7960	2928	3415	2674	627

The 8 schemas are listed in detail in the Table 2. It should be noted that we don't want to train a schema by methods of machine learning, which may perform better accuracy with a sentence-level sentiment analysis method on this dataset, because this schema will have strong dependence on that dataset and can hardly adapt datasets in different domains.

Table 2. Proposed weighting schemas

Number	w_{vs}	w_s	w_l	w_{vl}	w_e
Ⅰ	0.2	0.4	0.6	0.8	4
Ⅱ	0.1	0.4	0.6	0.9	4
Ⅲ	0.1	0.3	0.7	0.9	4
Ⅳ	0.2	0.35	0.65	0.8	4
Ⅴ	0.1	0.35	0.65	0.9	2
Ⅵ	0.2	0.3	0.7	0.8	2
Ⅶ	0.15	0.35	0.65	0.85	2
Ⅷ	0.15	0.4	0.6	0.85	2

Two lexicon based methods are employed to evaluate the 8 schemas and choose the best one among them as the final schema to be used in DFDS. We also compare the weighting schemas we proposed with the weighting strategy that our baseline used. The comparison of accuracy among 9 weighting methods are illuminated as Fig. 3.

Fig. 3 shows the performance of each schema with two lexicon based methods. We can see that schema Ⅲ has better performance, which motivated us to apply it as final weighting schema of our framework. In addition, with the same lexicon the schema Ⅲ has better results than our baseline DDR, which means the schema we present on RST is more reasonable.

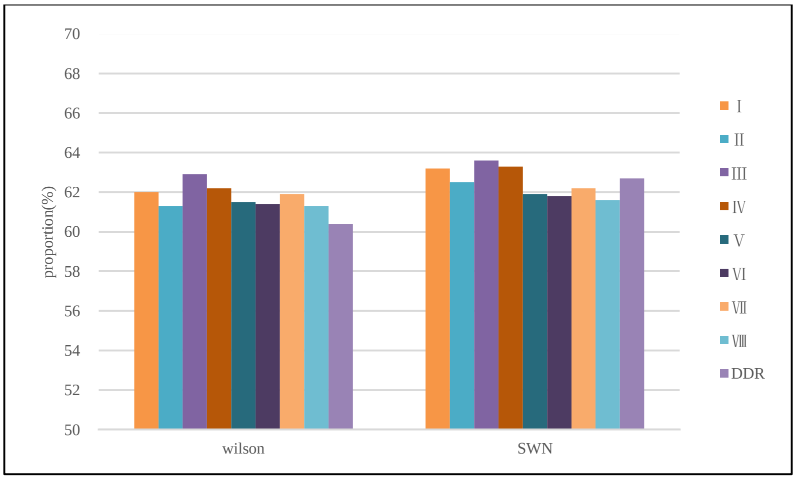

Fig. 3. Comparison of accuracy of weighting schemas with two lexicon based methods

What's more, we also find that all results of 9 weighting methods are improved after applying SWN lexicon. Investigation in results shows that lots of sentences with sentiments are judged to neutral. This is because some sentiment words are not collected in the lexicon collected by Wilson, which results in large amount of neutral classification results, especially in short documents. As shown in Table 1, SWN lexicon contains far more sentiment words than Wilson lexicon, thus gets better accuracy using the same weighting schema. This result also means better lexicon or sentence-level method may also result in better accuracy, so we apply Stanford-Sentiment Annotator as the sentence-level sentiment analysis method of our framework because of its good performance on sentence-level sentiment polarity classification. It must be noted that this annotator does not have the ability to classify the document-level textual corpus.

5.3 Experiments results and analysis

Experiments of document-level sentiment classification are conducted on four different domain datasets. We evaluate and compared our method with the baseline on F-score and accuracy. The results have been shown in Fig. 4 and Fig. 5.

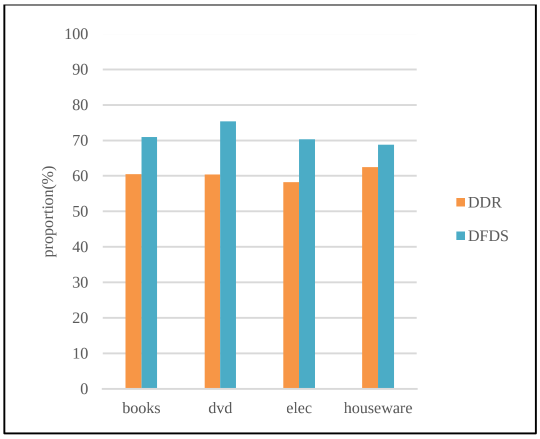

Fig. 4. Comparison of accuracy on all datasets

From Fig. 4 we can see that our method outperforms better accuracy on all four datasets than our baseline. The best result got on DVD reviews dataset has reached more than 75% and has about 15 percent better than DDR. Even on the dataset of houseware reviews, where DFDS gets the lowest accuracy, it still has about 7 percent better than DDR. Although four datasets are from different domains, the accuracy of our method maintains at a level about 70%, which indicates good adaptability for different domains.

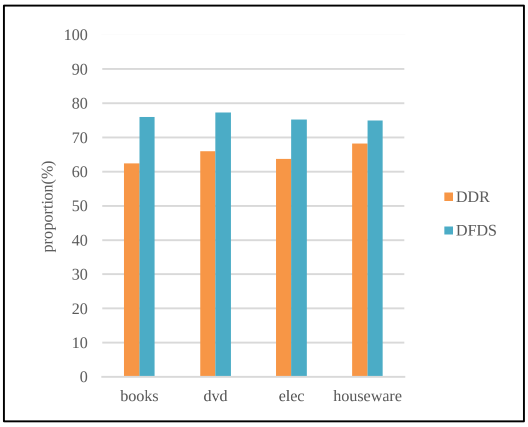

Fig. 5. Comparison of F-Score on all datasets

Fig. 5 shows the comparison of F-score on all datasets of both methods, and results show that DFDS outperforms better than DDR. On dvd review dataset, DFDS reach about 77 percent with about 15 percent better than the baseline. On both book review and electronic datasets, F-score of DFDS maintains about 70 percent, while DDR is just about 60 percent.

The results above show that our method perform better than our baseline DDR, but the method of sentence-level sentiment classification is based on Wilson lexicon, which has been shown to have much fewer sentiment words. On the contrary, we proposed novel weighting schemas and composing rules and employed better sentence-level analysis method in DFDS, so it is necessary to prove whether the weighting schemas and composing rules we used make sense. Therefore, we extended the work of DDR, and use Stanford-Sentiment Annotator to do sentence-level sentiment classification, denoted as DDRS. We evaluate the accuracy of DDRS on all four datasets with comparison to DDR and DFDS, which is shown in Table 3.

We also compare the proportion of neutral results that the three methods have judge out and the detail statistics are also listed in table 3. The results showed that our method have better results on all four datasets than DDRS, which have the same sentence-level sentiment analysis approach. That means the weighting schemas and composing rules proposed in this paper performs better on semantic relations capture and document-level sentiment classification. In addition, the proportion of neutral of DFDS is much less than our baseline and DDRS, which also definitely results in the improvement of our method. What's more, DDRS performs better than DDR on all datasets, which means the accuracy of overall document classification based on RST has positive correlation with performance of sentence-level sentiment analysis method.

Table 3. Comparison of accuracy for DDR, DDRS and DFDS (%).

Datasets	DDR		DDRS		DFDS	
	Accuracy	Neutral	Accuracy	Neutral	Accuracy	Neutral
books	60.5	10.4	68.9	9.2	71.0	3.0
dvd	60.4	12.2	73.7	6.9	75.4	1.3
elec	58.2	17.8	63.9	9.4	70.3	4.8
hourseware	62.5	17.4	66.0	10.2	68.8	4.8

6 Conclusion

To capture the semantic relation between sentences is still a great challenge in document-level sentiment classification. Improvement of discourse structure parsing in recent years motivated us to present a document-level sentiment analysis framework DFDS. We also proposed novel weighting schema and composing rules of discourse units and apply sentence-level sentiment classification method in our framework. Experiments has shown that our method gets competitive results on datasets in different domains.

7 Acknowledgement

This work is supported by the National Natural Science Foundation of China (NSFC) (61373165, 61373035 and 61672377).

Reference

1. O'Connor, B., Balasubramanyan, R., Routledge, B R., et al.: From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. ICWSM (11), 122-129 (2010).
2. Musto, C., Semeraro, G., Lops, P., et al.: CrowdPulse: A framework for real-time semantic analysis of social streams. Information Systems (54), 127-146 (2015).
3. Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. Journal of Computational Science 2(1), 1-8 (2011).
4. Smailović, J., Grčar, M., Lavrač, N., et al.: Stream-based active learning for sentiment analysis in the financial domain. Information Sciences 285(1), 181-203 (2014).
5. Liu, B., Zhang, L.: A survey of opinion mining and sentiment analysis. In: Mining text data, 415-463. Springer, US (2012).
6. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and trends in information retrieval 2(1-2), 1-135 (2008).
7. Bhatia, P., Ji, Y., Eisenstein, J.: Better document-level sentiment analysis from rst discourse parsing. arXiv preprint arXiv:1509.01599 (2015).
8. Mann, W C., Thompson, S A.: Rhetorical Structure Theory: Description and Construction of Text Structures. In: Natural language generation, 85-95. Springer, Netherlands (1987).
9. Ji Y, Eisenstein J.: Representation Learning for Text-level Discourse Parsing. Meeting of the Association for Computational Linguistics, 13-24. USA (2014).
10. Corston-Oliver, S H.: Beyond String Matching and Cue Phrases: Improving Efficiency and Coverage in Discourse Analysis. The AAAI Spring Symposium on Intelligent Text Summarization, 9-15 (1970).
11. Soricut, R., Marcu, D.: Sentence level discourse parsing using syntactic and lexical information. Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, vol. 1, 149-156. Association for Computational Linguistics (2004).
12. Feng, V W., Hirst, G.: Text-level discourse parsing with rich linguistic features. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers, vol. 1, 60-68. Association for Computational Linguistics (2012).
13. Li, S., Wang, L., Cao, Z., et al.: Text-level Discourse Dependency Parsing. Meeting of the Association for Computational Linguistics, vol. 1, 25-35 (2014).
14. Pang, B., Lee, L.: A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. Meeting on Association for Computational Linguistics, 271. Association for Computational Linguistics (2004).
15. Sharma, A., Dey, S.: A document-level sentiment analysis approach using artificial neural network and sentiment lexicons. ACM SIGAPP Applied Computing Review 12(4), 67-75 (2012).
16. Tang, D., Qin, B., Liu, T.: Document Modeling with Gated Recurrent Neural Network for Sentiment Classification. Conference on Empirical Methods in Natural Language Processing, 1422-1432. Portugal (2015).
17. Xu, J., Chen, D., Qiu, X., et al.: Cached Long Short-Term Memory Neural Networks for Document-Level Sentiment Classification. arXiv preprint arXiv:1610.04989 (2016).
18. Voll, K., Taboada, M., Not All Words Are Created Equal: Extracting Semantic Orientation as a Function of Adjective Relevance. Australasian Joint Conference on Artificial Intelligence. Springer, 337-346. Berlin Heidelberg (2007).
19. Heerschop, B., Goossen, F., Hogenboom, A., et al.: Polarity analysis of texts using discourse structure. ACM Conference on Information and Knowledge Management. DBLP, 1061-1070.Glasgow, United Kingdom (2011).

20. Wang, F., Wu, Y., Qiu, L.: Exploiting Discourse Relations for Sentiment Analysis. COLING: Posters, 1311-1320 (2012).
21. Li, J., Zhou, Y., Liu, C., et al.: Sentiment Classification of Chinese Contrast Sentences. Natural Language Processing and Chinese Computing 496, 205-216. Springer. Berlin Heidelberg (2014).
22. Hogenboom, A., Frasincar, F., De Jong, F., et al.: Using rhetorical structure in sentiment analysis. Communications of the Acm 58(7), 69-77 (2015).