Exploiting Discourse Relations for Sentiment Analysis

Fei Wang[1] Yunfang Wu [1] Likun Qiu[2]
(1) Institute of Computational Linguistics, Peking University
(2) School of Chinese Language and Literature, Ludong University
sxjzwangfei@163.com, wuyf@pku.edu.cn, qiulikun@gmail.com

ABSTRACT

The overall sentiment of a text is critically affected by its discourse structure. By splitting a text into text spans with different discourse relations, we automatically train the weights of different relations in accordance with their importance, and then make use of discourse structure knowledge to improve sentiment classification. In this paper, we utilize explicit connectives to predict discourse relations, and then propose several methods to incorporate discourse relation knowledge to the task of sentiment analysis. All our methods integrating discourse relations perform better than the baseline methods, validating the effectiveness of using discourse relations in Chinese sentiment analysis. We also automatically find out the most influential discourse relations and connectives in sentiment analysis.

TITLE AND ABSTRACT IN CHINESE

基于句际关系的情感分析方法

文档情感与其篇章结构息息相关。将一篇文档切分成具有不同句际关系的文本语段，可以自动训练并获得表征不同句际关系重要性的权重，进而利用这些篇章结构信息来提升情感分析的性能。本文利用显性关联标记来预测句际关系，继而提出了多种不同的方法利用句际关系来进行情感分析。实验结果表明，融合句际关系的方法均优于前人的情感分析方法，证明了句际关系对于汉语篇章情感分析的重要作用。本文还自动发现了情感计算中显要的句际关系类型和显要的关联标记。

KEYWORDS: sentiment analysis; discourse relation; connective.
KEYWORDS IN CHINESE：情感分析；句际关系；关联标记

Proceedings of COLING 2012: Posters, pages 1311–1320,
COLING 2012, Mumbai, December 2012.

1 Introduction

Sentiment analysis has attracted considerable attention in the field of natural language processing. Previous work on this problem falls into three groups: opinion mining of documents, sentiment classification of sentences and polarity prediction of words. Recently, the importance of discourse relations in sentiment analysis has been increasingly recognized.

In traditional lexicon-based methods, all words and sentences are treated equally, ignoring the structural aspects of a text. However, discourse structure knowledge is vital to some texts for polarity prediction. Take (1) as an example:

(1) 诺基亚5800屏幕很好| Nokia 5800's screen is very good, 操作也很方便| the operation is convenient, 通话质量也不错| the call quality is good, 但是外形偏女性化| but the shape is feminine, 而且电池不耐用| and the battery life is short, 总之我觉得不值| in general, I think it is not worth buying.

Three words "很好|very good", "方便|convenient" and "不错|good" are positive, and three words "女性化|feminine", "不耐用|short" and "不值|not worth" are negative. The overall sentiment of document (1) would be predicted as neutral using the lexicon-based method, however, it is negative.

By analysing a text's discourse structure, a text is split into spans with different semantic relations. With this discourse knowledge, we assign text spans with different weights in accordance with their contribution to the overall sentiment of a document. For example in document (1), the span introduced by connective "但是|but" has higher degree of importance, denoting a Contrast relation; the span introduced by connective "总之|in general" has the highest degree of importance, denoting a Generalization relation. This leads to the overall negative sentiment.

This paper exploits discourse relations by using explicit connectives for sentiment classification of texts, achieving better results than state of the art method. Our contributions are: (1) For the first time, we propose a relatively complete discourse relation hierarchy and list their corresponding connectives in Chinese, and validate their effectiveness in sentiment analysis; (2) We conduct weighting schemes at various granularities of discourse relations; (3) We find out the influential discourse relations and connectives that contribute most to the overall meaning of texts.

2 Related Work

In sentiment analysis, we can refer to Pang and Lee (2008) for an in-depth survey. For discourse parsing, we can refer to Joty et al. (2012), Hernault et al. (2010) and Wang et al. (2010) for recent progresses. Polanyi and Zaenen (2006) argue that polarity calculation is critically affected by discourse structure. In applying discourse relations to sentiment analysis, previous work can be divided into two groups: constraint-based approaches and weight-based schemes.

Somasundaran et al. (2008) and Somasundaran et al (2009) represent reinforce and non-reinforce relations in opinion frame. For example, text spans targeted at the same entity with reinforce relations are constrained to have same polarities, while text spans targeted at opposing entities with reinforce relations are constrained to have opposite polarities. Narayanan et al. (2009) apply conditional relations to improve sentiment analysis. Zhou et al. (2011) describe several constrains to eliminate the intra-sentence polarity ambiguities. For example, a sentence holding Contrast relation contains two text spans with opposite polarities.

Taboada et al. (2008) hypothesize that sentiment words expressed in nuclei are more important than words in satellites, and thus give different weights (1.5 vs. 0.5) to words in nuclei and satellites. Heerschop et al. (2011) hypothesize that not only nuclei and satellites should be weighted differently; satellites of different discourse relations should also be weighted differently.

In this paper, we adopt Rhetorical Structure Theory (RST) (Mann and Thompson, 1988) as the basis of discourse relations, and we follow the weighting scheme. Different from previous work, we hypothesize: (1) nuclei of different relations and satellites of different relations should all be weighted differently; (2) some relations are more important than other relations in sentiment classification.

3 Our Method

3.1 Overview

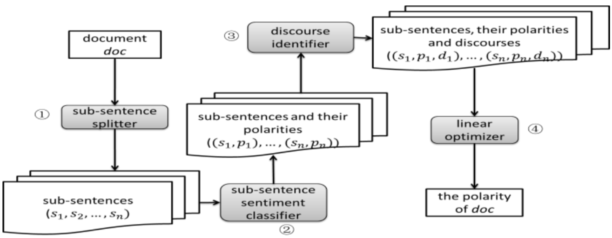

FIGURE 1 – Overview of our method

The proposed method consists of 4 main steps, as shown in Figure 1. First, a document doc is divided into sub-sentences $(s_1, s_2, \ldots s_n)$ by sub-sentence splitter. Second, a polarity p_i is assigned to each sub-sentence s_i by sub-sentence sentiment classifier. Third, discourse identifier identifies the discourse type d_i holding by sub-sentence s_i. Last, linear optimizer generates the polarity of the document by calculating the weighted sum of sub-sentences in accordance with their discourse types.

3.2 Sub-sentence Splitter

Sub-sentence splitter utilizes punctuation marks, including comma, period, semicolon, exclamation mark and question mark, to divide a sentence into sub-sentences. A document consists of one or more sentences, and a sentence consists of one or more sub-sentences. For example in document (1), it consists 6 sub-sentences. We treat intra-sentence and inter-sentence relations equally, because Chinese comma can signal both intra- and inter- sentence boundaries (Yang and Xue, 2012).

3.3 Sub-sentence Sentiment Classifier

The polarity of each sub-sentence p_i is generated by the Basic SELC model proposed by Qiu et al. (2009), a state of the art work. In Basic SELC model, some documents are initially classified based on a sentiment dictionary (HowNet[1]); and then more sentiment-bearing n-grams are learned and more documents are classified through an iterative process with negative/positive ratio control.

For each sub-sentence s_i, the polarity p_i is assigned with +1 if the sub-sentence is positive, and -1 if the sub-sentence is negative, and 0 if the sub-sentence is neutral.

Our method is a little different from the work of Taboada et al. (2008) and Heerschop et al. (2011), where discourse weights are multiplied with individual sentiment-bearing words (word-based method for short). However in our method, relation weights are multiplied directly with sub-sentences (sub-sentence-based method for short).

We also conduct a word-based method using HowNet, but it gives us a poor baseline with an F-score of 56.9% without using discourse relations. So we adopt the sub-sentence-based method that provides a relatively high baseline with an F-score of 83.55% (as shown in Table 3). What's more, the sub-sentence-based method is more consistent with people's intuition on discourse structure.

3.4 Discourse Identifier

Discourse identifier tags each sub-sentence s_i with a discourse type d_i. Discourse relation defines the relationship between two adjacent sub-sentences, while discourse type represents the relationship from the view of each component sub-sentence. For example, there is a Contrast relation between the two sub-sentences in sentence (2).

(2) 虽说是大牌| although it is a famous brand, 但是没感觉出大牌的味道| (but) I didn't feel anything extraordinary.

In sentence (2), the second sub-sentence is the Head (nucleus) of the sentence while the first sub-sentence is the Modifier (satellite). Thus, we will assign discourse type ContrastH to the second sub-sentence and ContrastM to the first one. For those relations with multi-nucleus, we assign all the component sub-sentences with the same relation type.

The research on Chinese discourse parsing has just begun, and there isn't a gold standard for Chinese discourse relation annotation in previous work. So we develop a specification of Chinese discourse relation hierarchy, as shown in Table 1. In this task, we remove a few connectives that may cause relation ambiguities, and we only list the discourse types that have explicit connectives. In the absence of Chinese discourse parser, we exploit explicit connectives to predict the discourse types. Sub-sentences introduced with specific connectives (Table 1) will be assigned with the corresponding discourse types, and sub-sentences without explicit connectives will be tagged with None.

For single-nucleus relations (except List), a head sub-sentence can appear by itself, while a modifier sub-sentence must co-occur with its corresponding head. So, if one modifier sub-sentence appears alone, we will guess the subsequent sub-sentence as its head. For example:

（3）如果拿它看书| if you want to read e-books on this mp4，眼睛会非常累|your eyes would be very tired.

The first sub-sentences is tagged as HypotheticalM because of connective "如果|if". Though the second sub-sentence contains no connective, it would still be guessed as the head of the first sub-sentence and thus labelled as HypotheticalH. As a result, for a specific relation, there are more head instances than modifier ones, as shown in Table 2.

[1] http://www.keenage.com/download/sentiment.rar

	Discourse relation	Discourse type	Connectives
联合关系 Multi-nucleus	等立Coordinate	Coordinate	同时|于此同时|另外|此外|再则|另一方面|一边|时而
	时序 Temporal	Temporal	尔后|接下来|起初|而后|随即|随后|继而
	选择 Alternative	Alternative	或|或者|或是|或者说|抑或|要么|或则|宁可|宁肯|宁愿|不如说|不如
	递进 Progression	Progression	不但|不光|不仅|不止|且不说|并且|何况|况且|而且|再说|在这|并|甚至
	重述 Equivalence	Equivalence	换言之|就是说|事实上|实际上
	顺承 Succession	Succession	N/A
主从关系 Single-nucleus	转折Contrast	ContrastH	不过|但|但是|而是|反之|可是|然而|转而|恰恰相反|反倒|反而|却|仍旧|仍
		ContrastM	虽说|固然|非但|虽然|尽管
	让步 Concession	ConcessionH	也
		ConcessionM	即便|即使|即或|即令|即若|纵然|纵使|就算
	因果 Cause	CauseH	之所以|因此|故而|那么|那末|所以|于是|进而|则|乃至于|因而|难怪|显而
		CauseM	因|因为|由于|既然|是因为|既|也许|或许|兴许
	结果 Result	ResultH	从而|以至|以致|以至于|致使
	目的 Purpose	PurposeH	以免|以便
	假设 Hypothetical	HypotheticalM	假如|假若|假使|倘若|如果|如若|要是|如果说|万一|一旦
	条件Condition	ConditionH	否则|要不|要不然|不然
		ConditionM	要不是|除非|不管|不论|无论|只要|只有|任|哪怕|多亏|幸而|幸好|幸亏
	解证 Explanation	ExplanationM	具体地说|具体来说|具体来讲|一方面
	分述 List	ListM	首先|其次|然后
	总括Generalization	GeneralizationH	总之|总的来说|总的看|综上所述|总的来看|总而言之

TABLE 1 – Discourse relation, discourse types and explicit connectives

3.5 Linear Optimizer

Linear optimizer generates the polarity of a document by calculating the weighted sum of its sub-sentences in accordance with their discourse types.

$$\text{score} = \left(\sum_{(s_i, p_i, d_i)} \text{weight}(d_i) \times p_i\right) + b \tag{1}$$

where $\text{weight}(d_i)$ is the weight of discourse type d_i, p_i is the polarity score of sub-sentence s_i, and b is an offset adjustment factor. The offset corrects a possible bias in sentiment scores caused by people's tendency to write negative reviews with positive words. Both Taboada et al. (2008) and Heerschop et al. (2011) validated that an offset can improve experiment results. We use a linear kernel SVM to train $\text{weight}(d_i)$ and b. The document would be classified as positive/negative/neutral if score is larger than/less than/equals zero.

4 Influential Discourse Relation Detecting

Intuitively, some discourse relations are more influential on the overall sentiment of a document. We apply a greedy search method to detect the most influential discourse relations, and the corresponding discourse types are considered as influential discourse types. When predicting the sentiment of a document, sub-sentences of influential discourse types are identified and weighted differently; the weight of remained sub-sentences are constrained to be equal. Figure 2 shows the procedure of detecting influential discourse relations.

Definitions: Define DR as the set of discourse relations, $DR = [dis_1, dis_2 .. dis_m]$

Define IDR as the set of influential discourse relations, initialized as $IDR = \varnothing$

Algorithm: while true:

$dis = \arg\max_{dis_i \in DR}(Performance(IDR \cup [dis_i]) - Performance(IDR))$ which meets

$Performance(IDR \cup [dis_i]) - Performance(IDR) > 0$

//find dis which could get the highest performance gain

if $\exists dis$: $IDR = IDR \cup [dis]$; $DR = DR - [dis]$

else: break

return IDR

FIGURE 2 – Greedy search for influential discourse relations

5 Experiment

5.1 Data

Our data is collected from 360buy (http://www.360buy.com/). Reviews on 360buy are structured, elaborating the strong points and shortcomings of the products. Reviews collected from the strong point column are automatically tagged as positive, and reviews collected from the shortcoming column are automatically tagged as negative. In our task, all the extracted reviews should meet two requirements: (1) contain at least two sub-sentences; (2) contain at least one connective. There are 53,040 reviews in our collected corpus, including 24,532 positive reviews and 28,508 negative reviews. Each review consists of 5.06 sub-sentences on average. Table 2 illustrates the occurrences of each discourse type in our collected data.

Discourse type	#times	discourse type	#times	discourse type	#times	discourse type	#times
None	46645	HypotheticalH	4060	ConditionM	2023	Alternative	365
ContrastH	22303	HypotheticalM	4060	Coordinate	1840	ConcessionM	231
Progression	8339	CauseM	3257	Equivalence	1460	ResultH	60
ConcessionH	5541	ContrastM	2727	ListM	789	PurposeH	47
CauseH	5062	ConditionH	2147	GeneralizationH	595	Temporal	35

TABLE 2- Distribution of Discourse Types

Discourse types whose occurrence is less than 100 are merged into "None" type. 3/4 of the collected data is used to train the linear optimizer, and the rest is used as test data. In the case of detecting influential discourse relations, we further divide 1/4 from the training data as the development data (development data is used to tune the most influential discourse relations), and the test data remains the same.

5.2 Experiment Set

We conduct 6 types of experiments, described as follows.

Baseline1. We implemented Qiu et al. (2009) as our baseline.

Baseline2. All the sub-sentences are equally weighted in Formula (1). No discourse knowledge is applied.

SNSS. (**Single Nucleus Single Satellite Method.**) Following the idea of Taboada et al. (2008), all discourse types denoting the Heads of relations are grouped as “nucleus”, and other discourse types are grouped as “satellite”. In this method, we have only two distinguishing categories: nucleus and satellite.

SNMS. (**Single Nucleus Multiple Satellites Method.**) Following the idea of Heerschop et al. (2011), all discourse types denoting the Heads of relations are grouped as “nucleus”, while all discourse types denoting the Modifiers of discourse relations are reserved. In this method, we hypothesize that nucleus types contribute equally while different satellite types contribute differently to the overall polarity of documents.

MNMS. (**Multiple Nuclei Multiple Satellites Method.**) All the discourse types specified in Table 1 are reserved and weighted differently in calculating a document’s sentiment. In this method, we hypothesize that both different nucleus types and different satellite types contribute differently to the overall polarity of the documents.

GDR. (**Greedy Discourse Relation Method.**) Following Figure 2, influential discourse relations are identified. The corresponding discourse types are reserved and weighted differently, and others are grouped as “None”.

GCW. (**Greedy Connective Word Method.**) Explicit connectives are objective language usage, while relation types are subjective induction. Following the same procedure as Figure 2, we hypothesize that the weight of each sub-sentence depends directly on its connective. That means, only influential connectives are identified and weighted differently in calculating a document’s sentiment, while others are grouped as “None”.

5.3 Experiment Results

Performance is evaluated in terms of Precision (Pre), Recall (Rec) and F score.

	Positive			Negative			Overall	Comments &
Method	Pre	Rec	F	Pre	Rec	F	F	Influential discourse relations or connectives
Baseline1	85.0	**84.8**	84.9	**81.9**	82.0	81.9	83.55	The performance gain of baseline2 than baseline1 indicates the effectiveness of offset b in Formula (1).
Baseline2	**91.5**	77.6	84.0	77.2	**91.4**	83.7	83.86	
SNSS	89.1	82.0	85.4	80.4	88.0	84.0	84.76	The performance gain of SNSS, SNMS, MNMS, GDR and GCW than baseline2 indicates the effectiveness of our weighting scheme which exploits discourse knowledge.
SNMS	89.4	82.7	**85.9**	80.9	88.2	84.4	85.20	
MNMS	89.2	82.2	85.5	80.5	88.1	84.1	84.86	
GDR	89.9	82.2	**85.9**	80.7	89.0	**84.6**	**85.28**	Contrast, Cause, Condition, Generalization
GCW	90.4	81.4	85.6	80.1	89.6	**84.6**	85.13	不过|however, 虽然|although,但|but, 同时|at the same time, 总的来说|in general,但是|but

TABLE 3- Experiment results

As shown in Table 3, all our methods integrating discourse knowledge perform better than both baselines in overall F score. To test for significance, we conduct t-test which meets $p<0.01$. GDR achieves the best result, 1.42% higher than baseline2. This validates the effectiveness of using discourse relations in Chinese sentiment analysis. In English data, Heerschop et al. (2011) yield an improvement of 4.7% in F score when using discourse structure, but their baseline is quite low with an F score of 68.7%.

The overall F value of SNSS is 0.9% higher than baseline2, validating the effectiveness of the simple distinction between nuclei and satellites. Both SNMS and MNMS perform better than SNSS, indicating that more discourse knowledge is helpful in calculating the overall polarity. Note that MNMS performs slightly worse than SNMS, perhaps this is because too many weights have to be trained in MNMS.

GDR, which differentiates Contrast, Cause, Condition and Generalization from other discourse relations, harvests the best result. It is consistent with our intuition that these relations have great impact on the meaning of the texts. The influential discourse relations that we find out are partly consistent with previous work: Narayanan et al. (2009) exploit Conditional sentences for sentiment analysis; Zhou et al. (2011) focus their attention on Contrast, Condition, Cause, Continuation and Purpose in polarity classification.

To our surprise, GCW, which utilizes only 6 explicit connectives, obtains a rather promising result, with a performance of 1.27% higher than baseline2. Among these 6 connectives, "不过|however", "虽然|although","但|but", "但是|but" denote a Contrast relation; "同时|at the same time" denotes a Coordinate relation; and "总的来说|in general" denotes an Generalization relation.

Conclusion and Future Work

In this paper, we utilize explicit connectives to predict discourse relations, and then conduct several methods to incorporate discourse structure knowledge to the task of sentiment analysis. We define discourse relations in different granularities: nucleus-satellite, nucleus-different satellites and different nuclei-different satellites. The experimental results validate the effectiveness of using discourse relations in Chinese sentiment analysis. Furthermore, we automatically detect the most influential discourse relations and connectives. Experimental results show that Contrast, Cause, Condition and Generalization are the most influential relations, and "不过|however", "虽然|although", "但|but", "同时|at the same time", "总的来说|in general", "但是|but" are the most influential connectives.

This is only a preliminary study on discourse relation and Chinese sentiment analysis. The future work includes the following aspects. (1) We would like to develop a Chinese discourse parser to automatically parse the discourse structure, to get both explicit and implicit relations and their argument spans. (2) We will apply more sophisticated methods to get more reliable polarity scores for sub-sentences. (3) We will incorporate discourse structure knowledge to other tasks such as summarization.

Acknowledgments

This paper was supported by National High Technology Research and Development Program of China (No.2012AA011101), 2009 Chiang Ching-kuo Foundation for International Scholarly Exchange (No.RG013-D-09) and National Natural Science Foundation of China (No. 61103089).

References

Heerschop, B., Goossen, F., Hogenboom, A., Frasincar, F., Kaymak, U. and de Jong, F. (2011). Polarity analysis of texts using discourse structure. In Proceedings of CIMK-2011, pages 1061-1070.

Hernault, H., Bollegala, D. and Ishizuka, M. (2010). A semi-supervised approach to improve classification of infrequent discourse relations using feature vector extension. In Proceedings of the EMNLP-2010, pages 399-409.

Joty, S., Carenini, G. and Ng, R.T. (2012). A novel discriminative framework for sentence-level discourse analysis. In Proceedings of EMNLP-2012, page 904-915

Mann, W.C. and Thompson, S.A. (1988). Rhetorial structure theory: towards a functional theory of text organization. Text, 8(3):243-281.

Narayanan, R., Liu, B. and Choudhary, A. (2009). Sentiment analysis of conditional sentences. In Proceedings of EMNLP-2009, pages 180-189.

Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. Information Retrieval, 2(1-2): 1-135.

Polanyi, L. and Zaenen, A. (2006). Contextual valence shifters, Computing attitude and affect in text: Theory and applications, page 1-10.

Qiu, L., Zhang, W., Hu, C. and Zhao, K. (2009). SELC: a self-supervised model for sentiment classification. In Proceedings of CIKM-2009, pages 929-936.

Somasundaran, S., Namata, G. , Wiebe, J. and Getoor, L. (2009). Supervised and unsupervised methods in employing discourse relations for improving opinion polarity classification, In Proceedings of the ENMLP-2009, pages 170-179.

Somasundaran, S., Wiebe, J. and Ruppenhofer, J. (2008). Discourse level opinion interpretation, In Proceedings of COLING-08, pages 801-808.

Taboada, M., Voll, K. and Brooke, J. (2008). Extracting sentiment as a function of discourse structure and topically. Technical Report No. 2008-20, Simon Fraser University.

Wang, W.T., Su, J. and Tan, C.L. (2010). Kernel based discourse relation recognition with temporal ordering information. In Proceedings of ACL-2010, pages 710-719.

Yang, Y. and Xue, N. (2012). Chinese comma disambiguation for discourse analysis. In Proceedings of ACL-2012, pages 786-794.

Zhou, L., Li, B., Gao, W., Wei, Z. and Wong, K.F. (2011). Unsupervised discovery of discourse relations for eliminating intra-sentence polarity ambiguities, In Proceedings of EMNLP-2011, pages 162-171.