

Joint Modeling of Structure Identification and Nuclearity Recognition in Macro Chinese Discourse Treebank

Xiaomin Chu¹, Feng Jiang¹, Yi Zhou¹, Guodong Zhou¹, Qiaoming Zhu^{1,2,†}

¹School of Computer Science and Technology, Soochow University, China

²Institute of Artificial Intelligence, Soochow University, China

{xmchu, fjiang, yzhou}@stu.suda.edu.cn; {gdzhou, qmzhu}@suda.edu.cn

Abstract

Discourse parsing is a challenging task and plays a critical role in discourse analysis. This paper focus on the macro level discourse structure analysis, which has been less studied in the previous researches. We explore a macro discourse structure presentation schema to present the macro level discourse structure, and propose a corresponding corpus, named Macro Chinese Discourse Treebank. On these bases, we concentrate on two tasks of macro discourse structure analysis, including structure identification and nuclearity recognition. In order to reduce the error transmission between the associated tasks, we adopt a joint model of the two tasks, and an Integer Linear Programming approach is proposed to achieve global optimization with various kinds of constraints.

1 Introduction

A typical document is usually organized in a coherent way that each discourse unit is relevant to its context and plays a role in the entire semantics. Discourse structure analysis not only helps to understand the discourse structure and semantics, but also can benefit variety of downstream applications including question answering (Sadek and Meziane, 2016), machine translation (Guzmán et al., 2014), text summarization (Ferreira et al., 2014; Cohan and Goharian, 2017), and so forth.

There exist two hierarchical levels of discourse structures: micro level and macro level. The micro level structure refers to the structure and relation among the discourse units in a sentence, or consecutive sentences, or sentences groups. The macro level structure refers to the structure and relation among paragraphs, or chapters, or discourses. Corresponding to related research based on Rhetorical Structure Theory Discourse Treebank (RST-DT), the micro level is similar to the sentence-level discourse parsing, and the macro level is similar to the document-level discourse parsing.

To make a clearer explanation of the macro discourse structure, take the chtb_0019 as an example, which is a typical news article from Chinese Treebank 8.0 (Xue et al., 2013). The macro discourse structure of this article is shown as Figure 1. There are five paragraphs (P1, P2, P3, P4 and P5) in the news “Significant achievements in the construction of Ningbo Bonded Area”. The paragraphs are connected by discourse relations (*Elaboration*, *Background*, *Joint*). In this article, paragraph P1 points out the theme of the overall article. Depending on the direction of arrows from the root node to the leaf node in the discourse structure tree, the most important part can be quickly located.

Limited to the length of this paper, the full discourse text of this example is not included, please refer to the corpus. The main contents of the five paragraphs respectively are: P1) Ningbo Bonded Area achieved fruitful results after three years of construction; P2) the basic situation of the Ningbo Bonded Area; P3) the situation of import and export trade, warehouses, storage area, etc. P4) the situation of industrial processing projects and enterprises; P5) the situation of administrative services and information construction.

This work is licensed under a Creative Commons Attribution 4.0 International License. License details: <http://creativecommons.org/licenses/by/4.0/>

[†] The corresponding author is Qiaoming Zhu.

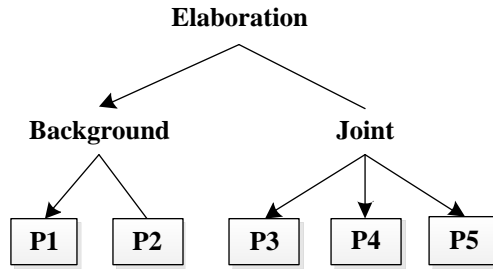


Figure 1: Macro structures tree of chtb_0019.

From the discourse structure tree of this example, we can see that the analysis of discourse structure is beneficial for the understanding of the content and the theme of the discourse. Based on the macro discourse structure analysis, we can further enhance the performance of natural language processing applications. For example, we can use the information of discourse structure to summarize the content and purpose of an article, use the information of discourse relation to assist the construction of question answering system, and improve the performance of automatic summarization by using the nuclearity information.

The existing research on discourse structure is mainly focused on the micro level, and the performance has not yet achieve the level for application. However, the macro level research still stays in the theoretical research, and there is no available corpus resource, nor a corresponding computational model. For the two reasons mentioned above, in this paper, we take the macro discourse structure as the main research object, that is different from the previous research. We explore a macro discourse structure presentation schema to present the macro level discourse structure, and propose a Macro Chinese Discourse Treebank (MCDTB) on the top of existing Chinese Discourse Treebank (CDTB) (Li et al., 2014). On the basis of the presentation schema and annotated corpus, we divide the macro level discourse structure analysis into four tasks, including structure identification, nuclearity recognition, relation classification and discourse tree building.

There are certain differences from the analysis of the micro level discourse structure. For example, macro discourse structure analysis takes paragraphs as the elementary discourse units, and the relations between the units are fairly loose, so it brings difficulties to the task of structure identification. Furthermore, there are virtually no connectives between the discourse units, and the texts of each discourse unit are relatively long, making discourse relation classification lack of effective cue phrase and lexical information.

The task of discourse structure identification is the first and the most crucial step in the macro discourse analysis and also the basis step of further tasks. Nuclearity recognition is only part of the discourse relation classification task in the existing research, and has not been given sufficient attention. However, in our study, we find that the performance improvement of nuclearity recognition contributes to the improvement of the overall performance. Therefore, in this paper, we concentrate on these two tasks of macro discourse structure, structure identification and nuclearity recognition, and take structure identification as the main task.

Our contribution is three-fold. First, we explore a macro discourse structure presentation schema to present the macro level discourse structure, and propose a macro discourse structure corpus, named Macro Chinese Discourse Treebank (MCDTB). The presentation schema and corpus resource can lay the foundation for macro discourse structure analysis. Second, we propose discourse structure identification and nuclearity recognition models on macro level discourse structure analysis. By using CRF models to label sequences of discourse units, we can incorporate contextual information in a more natural way, and achieve a satisfactory performance. Third, we propose a joint model of structure identification and nuclearity recognition to reduce the error transmission between the associated tasks, and achieve global optimization via Inter Linear Programming.

The rest of this paper is organized as follows. Section 2 overviews related work on discourse parsing.

Section 3 introduces our macro discourse structure presentation schema and corpus resources. Section 4 introduce the framework of our joint model of structure identification and nuclearity recognition. Section 5 describes the local models, and the joint approach we used with ILP is introduced in Section 6. Section 7 presents the experimental results. Section 8 gives the conclusion and future work.

2 Related Work

Discourse parsing is the task of discovering the presence and type of the discourse relations between discourse units. The existing discourse parsing researches are mainly based on Rhetorical Structure Theory Discourse Treebank (RST-DT). The RST-DT (Carlson et al., 2003) is built in the framework of Rhetorical Structure Theory, consisting of 385 Wall Street Journal articles from the Penn Treebank (Marcus et al., 1993) and representing over 176,000 words of text.

While recent advances in sentence-level discourse parsing have attained accuracies close to human performance (Joty et al., 2012), discourse parsing at the document-level still poses challenges.

The HILDA discourse parser (Hernault et al., 2010) is the first attempt at document-level discourse parsing on RST-DT. It adopts a pipeline framework, and greedily builds the discourse tree from the bottom-up. In particular, at each step of the tree-building, a binary Support Vector Machine (SVM) classifier is applied to determine which pair of adjacent discourse constituents should be merged to form a larger span, and then another multi-class SVM classifier is applied to assign the type of discourse relation between the chosen pair of constituents.

Joty et al. (2013) approach the document-level discourse parsing using a model trained by Conditional Random Fields (CRF). They decomposed the problem of document-level discourse parsing into two stage: intra-sentential and multi-sentential parsing. Specifically, they employed two separate models for intra- and multi-sentential parsing. They jointly modeled the structure and the relation for a given pair of discourse units, such that information from each aspect can interact with the other.

Feng and Hirst (2014) develop a much faster model whose time complexity is linear in the number of sentences. Their model adopts a greedy bottom-up approach, with two linear-chain CRFs applied as local classifiers. An approach of post-editing is performed, which modified a fully-built tree by considering information from upper-levels, to improve the accuracy.

There is no relevant research on document-level discourse parsing in Chinese so far. For micro level discourse structure analysis, Li (2015) proposes a Connective-driven Dependency Tree (CDT) schema to represent the discourse rhetorical structure in Chinese language, with elementary discourse units as leaf nodes and connectives as non-leaf nodes, largely motivated by the Penn Discourse Treebank and the Rhetorical Structure Theory. On this basis, a Chinese Discourse Treebank (CDTB) consisting of 500 discourses is annotated, and a Chinese discourse structure analysis platform is realized.

3 Macro Chinese Discourse Treebank

3.1 Macro Discourse Structure Representation Schema

Rhetorical Structure Theory (RST) (Mann and Thompson, 1987), one of the most influential theories of discourse, represents a discourse by a hierarchical structure, called discourse tree. The leaves of a discourse tree correspond to Elementary Discourse Units (EDUs). Adjacent EDUs are connected by rhetorical relations, forming larger discourse units. RST defines two different types of discourse units, in which the nucleus is considered as the central part, and the satellite is considered as the peripheral part.

A concept of “macrostructures” is put forward by Van Dijk (1980) in Macrostructure Theory. The point of “macrostructures” is that texts not only have local or micro structural relations between subsequent sentences, but also have overall structures that define their global coherence and organization. But even today, the crucial global structures (including macrostructures, superstructures) that define the overall meaning and form of texts are almost ignored.

Inspired by Van Dijk’s Macrostructure Theory and Rhetorical Structure Theory, we explore a macro discourse structure representation schema. In this representation schema, each discourse is represented as a hierarchical discourse tree (as shown in Figure 1). In the macro discourse structure tree, leaf nodes

represent paragraphs, and non-leaf nodes represent discourse relations. The edges connect the discourse units, with the arrows pointing to the “Nucleus” units.

Detailed definitions of macro discourse structure are described as follows.

Leaf nodes: Unlike the definition on the micro level (the elementary units are treated as leaf nodes), we directly treat the paragraphs which are naturally segmented in the discourses as leaf nodes on the macro level.

Non-leaf nodes: Discourse relations connect discourse units, which are treated as non-leaf nodes in our macro discourse structure. We classify the discourse relations into three categories and fifteen subcategories, including **Coordination** (Joint, Sequence, Progression, Contrast, Supplement), **Causality** (Cause-Result, Result-Cause, Background, Behavior-Purpose, Purpose-Behavior), and **Elaboration** (Elaboration, Summary, Evaluation, Statement-Illustration, Illustration-Statement).

Arrow pointing: A discourse unit linked by a discourse relation can be either a “Nucleus” or a “Satellite” depending on how central the message is. In the macro discourse structure, we use the arrows pointing to represent the “Nucleus-Satellite” relations. Specifically, the edges with arrows point to the “Nucleus” units, and the edges without arrows point to the “Satellite” units.

3.2 Corpus Annotating

Guided by the macro discourse structure framework defined in Section 3.1, we have carried out annotating work of macro Chinese discourse structure, which we call Macro Chinese Discourse Treebank (MCDTB)¹. In the process of annotating, the structure definition and annotating criteria are modified iteratively. After nearly a year of annotation, 720 news wire articles are annotated, the source of which is Chinese Treebank 8.0 (CTB 8.0). (Xue et al., 2002; Xue et al., 2013)

Because the discourse units are not isolated from the overall discourse, it’s difficult to judge whether the discourse units are important or not and what relations are between the discourse units simply from the units themselves. It is necessary to have a comprehensive understanding of the overall article before the annotating.

We divide the annotating work into three main stages. **The first stage** lasted four months, with three annotators participating. We selected the first 50 news articles from CTB 8.0, and annotated them together. After a lot of discussions, a preliminary annotating specification was formed. **The second stage** lasted for three months. Three annotators annotated articles independently and discussed the annotating result in groups. At the same time, the imperfect parts of the representation schema and annotating specification were discussed and amended. **The third stage**, which lasted four months, we added three new annotators to improve the efficiency of the annotating. Six annotators were divided into three groups, and each group was composed of a new staff and an old staff. The annotators of each group annotated independently and discussed in groups.

To ensure the quality of our corpus, we adopt the annotating consistency using agreement and kappa. Table 1 illustrates the annotating consistency in detail. We measure the agreement and kappa of discourse structures, nuclearity and discourse relations in the second and the third stage respectively. The method of consistency calculation used in this paper refers to the work of the corpus of RST (Marcu et al., 1999), and the appropriate adjustment is made according to the contents of our annotation.

Annotating Consistency of the Stage 2			Annotating Consistency of Stage 3		
Categories	Agreement	Kappa	Categories	Agreement	Kappa
Structure	88.54%	0.771	Structure	86.07%	0.671
Nuclearity	80.67%	0.694	Nuclearity	83.35%	0.647
Relation	83.05%	0.556	Relation	80.20%	0.597

Table 1: Annotating consistency

After the annotating work finished, the corpus consists of 720 newswire articles with a total of 398,829

¹The Macro Chinese Discourse TreeBank is available at <https://figshare.com/s/250474dba44e4161b040>.

Chinese characters. 3,981 paragraphs with 8,391 sentences are annotated. There are 5.53 paragraphs and 554 Chinese characters in each article on average. 2,870 discourse relations are annotated.

4 Overview of Framework

Figure 2 demonstrates the framework of our joint model of structure identification and nuclearity recognition. The test dataset is processed first, and it becomes groups of discourse units’ sequences to be labeled. Then we perform two CRF models to identify the structure and recognize the nuclearity using different feature sets respectively. In order to reduce the error transmission between the associated tasks, we adopt a joint model of the two tasks, and an Integer Linear Programming (ILP) approach is proposed to achieve global optimization with various kinds of constraints. In this way, a joint model of two layers is formed.

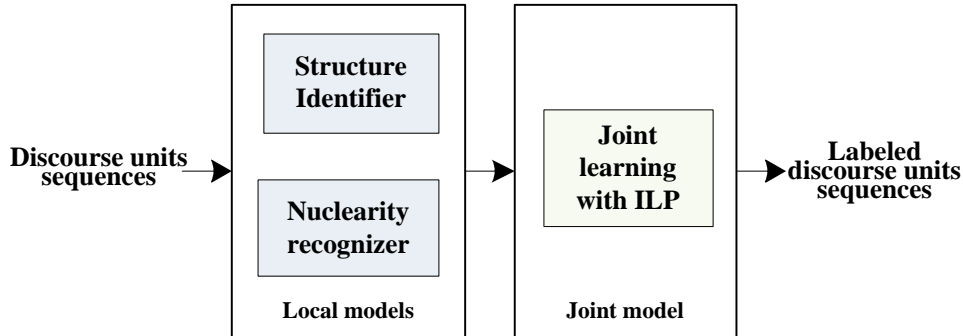


Figure 2: Joint model framework.

5 Local Models

We convert the macro discourse structure and the nuclearity prediction tasks into the sequence labeling problems. It has the following advantages of these conversions. 1) Context information can be fused conveniently. We use a window to capture the features of the previous and next discourse units. 2) Make the prediction process more naturally. In previous studies, in order to classify the structure and relations, methods of left join and right join were used to convert multiple relations to binary relations. There are two shortcomings of these approach, on the one hand, the number of non-original samples is added, and on the other hand, it is difficult to automatically build a complete real structure tree.

We build the local models of structure identification and nuclearity recognition respectively. Our local models are implemented using CRFs. In this way, we are able to take into account the sequential information from contextual discourse units, which cannot be naturally represented with Support Vector Machine (SVM) or Maximum Entropy (ME) as local classifiers.

As shown by Feng and Hirst (2012; 2014), for a pair of discourse units of interest, the sequential information from contextual units is crucial for determining structures. Therefore, it is well motivated to use CRF, which is a discriminative probabilistic graphical model, to make predictions for a sequence of units surrounding the pair of interest.

Figure 3 shows our structure identification model M_{struct} implemented with conditional random field algorithm. The first layer of the chain is composed of discourse units U_j 's, and the second layer is composed of nodes of S_j 's to indicate the probability of merging adjacent discourse units. When there is a relation between the discourse unit U_j and the previous one U_{j-1} , the structure of S_j is labeled as "1", and on the other hand, when there is no relation between the two consecutive discourse units, the structure of S_j is labeled as "0". To improve the accuracy of the structure identification, we enforce additional commonsense constraints in its Viterbi decoding. In particular, we disallow the existence of all-zero sequences (at least one pair must be merged).

The nuclearity recognition model $M_{nuclear}$ works in a similar way to M_{struct} , in which the first layer of the chain is composed of discourse units U_j 's, and the second layer is composed of nodes of N_j 's to

indicate the probability of nuclearity between the adjacent discourse units. When the current discourse unit U_j is more important than the previous discourse unit U_{j-1} , N_j is labeled as “1”. When the current discourse U_j is less important than U_{j-1} , N_j is labeled as “2”. When the two discourse units are equally important, the N_j is labeled as “3”.

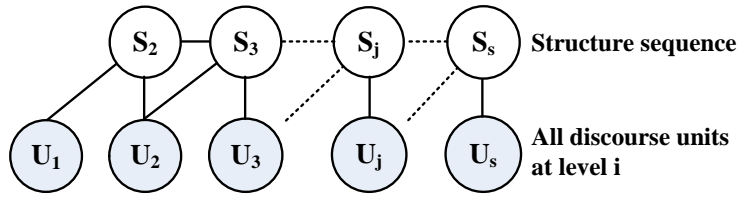


Figure 3: Local structure identification model.

In our local models, to encode two adjacent units, U_j and U_{j+1} , within a CRF chain, we use the following features listed in Table 2, some of which are modified from (Joty et al., 2013)’s and (Feng and Hirst, 2014)’s models.

Some of the helpful features of the RST discourse parsing cannot be used in macro discourse analysis or in Chinese discourse analysis. For example: 1) For macro level discourse analysis, the particles of the N-gram model are too small to represent the information of a paragraph, so the lexical features are not used in our tasks. 2) Syntactic information and dominance set features are very useful for micro level discourse analysis. However, the elementary units of the macro level are paragraphs, and these features are not applicable. 3) Since there is no tense in Chinese, we cannot use temporal features in macro level structure analysis.

Due to the appearance of word vector representation (Mikolov et al., 2013), the methods of co-occurrence (Sporleder and Lascarides, 2004) and word pairs (Feng and Hirst, 2012) are not necessary when the semantic similarity is calculated. We use the word2vec model to train word vector representation on the CTB 8.0, and use the method proposed by Jiang et al. (2018) to calculate the semantic similarity (including the semantic similarity between adjacent discourse units and the similarity between the discourse unit and the topic). In particular, in order to prevent the sparsity of the features, we discretize the semantic similarity into 10 levels.

Features	Used in SI	Used in NR
Organization features		
The beginning and end location of U_j .	Y	Y
Distances of U_j to the beginning and to the end.	Y	Y
Number of sentences (or paragraphs) in U_j .	Y	N
Whether U_j contains more sentences (or paragraphs) than U_{j+1} .	N	Y
Tree structure features		
Whether U_j is a bottom-level constituent.	Y	Y
Whether U_j is a combined unit in the previous step.	Y	Y
Similarity features		
The similarity between U_j and U_{j+1} .	Y	Y
The similarity between U_j and the topic of the discourse.	Y	Y
The similarity between U_{j+1} and the topic of the discourse.	Y	Y
Whether the similarity between U_j and the topic is greater than the similarity between U_{j+1} and the topic.	N	Y

Table 2: Features used in local models.

6 Joint Learning with Integer Linear Programing

While a pipeline model may suffer from the errors propagated from upstream tasks, a joint model can benefit from the close interaction between two or more tasks. Recently, joint modeling has been widely attempted in various NLP tasks, such as joint syntactic parsing and semantic role labeling (Li et al., 2010), joint argument identification and role determination (Li et al., 2013), joint structure identification and relation recognition (Joty et al., 2012), etc.

In our joint model, an ILP (Integer Logic Programming) -based inference framework is introduced to integrate two CRF-based local models, the structure identifier and the nuclearity recognizer. In this section, we propose a joint model of structure identification and nuclearity recognition with some intra-instance and contextual constraints.

We assume $p_{SI}(s_{\langle i,j \rangle} | seq_i)$ the probability of M_{struct} identifying $s_{\langle i,j \rangle}$ as a structure of an sequence seq_i , where $s_{\langle i,j \rangle}$ is the j th structure to be identified in the i th sequence seq_i . We define following assignment costs with $-\log$:

$$c_{\langle i,j \rangle}^{SI} = -\log(p_{SI}(s_{\langle i,j \rangle} | seq_i)) \quad (1)$$

$$c_{\langle i,j \rangle}^{-SI} = -\log(1 - p_{SI}(s_{\langle i,j \rangle} | seq_i)) \quad (2)$$

where $c_{\langle i,j \rangle}^{SI}$ and $c_{\langle i,j \rangle}^{-SI}$ are the cost of $s_{\langle i,j \rangle}$ whether or not a structure in sequence seq_i respectively.

There are three types of nuclearity labels between discourse units, including “NS”, “SN”, and “NN”. In addition, when there is no structure between the two successive units, we add a “NO-STR” label to distinguish it. The label of nuclearity could be represented as a 4-dimension vector and the value of each element in the vector is either 1 or 0 denoting whether the corresponding label is assigned or not. For instance, $n_{\langle i,j \rangle} = [1,0,0,0]$ denotes the assigned label is “NO-STR” and $n_{\langle i,j \rangle} = [0,1,0,0]$ denotes the assigned label is “NS”, where $n_{\langle i,j \rangle}[k]$ denotes the k th label in the vector and $p_{NR}(n_k | s_{\langle i,j \rangle})$ denotes the probability belonging the k th label. The cost of nuclearity recognition can be defined as follow:

$$c_{\langle i,j \rangle}[k]^{NR} = -\log(p_{NR}(n_k | s_{\langle i,j \rangle})) \quad (3)$$

where $c_{\langle i,j \rangle}[k]^{NR}$ is the cost of assigning or not assigning nuclearity n_k to $s_{\langle i,j \rangle}$.

Besides, we use indication variable $x_{\langle i,j \rangle}$ which is set to 1 if $s_{\langle i,j \rangle}$ is a structure of seq_i , and 0 otherwise. Similar to $x_{\langle i,j \rangle}$, we use another indicator variable $y_{\langle i,j,k \rangle}$ which is set to 1 if the $s_{\langle i,j \rangle}$ has the k th nuclearity label, and 0 otherwise. The objective function for the overall sample set can be represented as follows, where D denotes the overall sample set.

$$\min \sum_{seq_i \in D} \left(\sum_{s_{\langle i,j \rangle} \in seq_i} (c_{\langle i,j \rangle}^{SI} x_{\langle i,j \rangle} + c_{\langle i,j \rangle}^{-SI} (1 - x_{\langle i,j \rangle})) + \sum_{k=0}^3 c_{\langle i,j \rangle}[k]^{NR} y_{\langle i,j,k \rangle} \right) \quad (4)$$

Subject to

$$x_{\langle i,j \rangle} \in \{0, 1\} \quad (5)$$

$$y_{\langle i,j,k \rangle} \in \{0, 1\} \quad (6)$$

Constraints (5) and (6) are used to make sure that $x_{\langle i,j \rangle}$ and $y_{\langle i,j,k \rangle}$ are binary values.

Furthermore, we enforce following constraints (C1 and C2) on the consistency between SI and NR.

(C1) Nuclearity type constraints: the task of nuclearity recognition is a single-label classification problem. That is, the label of an instance could be only one option.

$$\sum_{k=0}^3 y_{\langle i,j,k \rangle} = 1 \quad (7)$$

(C2) Correlation constraints: if $s_{\langle i,j \rangle}$ is a structure, it must have a nuclearity label, otherwise, if $s_{\langle i,j \rangle}$ is not a structure, it must not have a nuclearity label.

$$\sum_{k=1}^3 y_{\langle i,j,k \rangle} = x_{\langle i,j \rangle} \quad (8)$$

Similar to local models, we disallow the existence of all-zero sequences, so we add the constraint C3 to the joint model. This constraint also lays the foundation for subsequent task of discourse tree building.

(C3) **Contextual constraints:** a sequence must have at least one structure. In order to avoid the ILP model optimizing the sequences into all-zero sequences, which has been already constrained by the Viterbi algorithm in the local models, this constraint is added.

$$\sum_{s_{\langle i,j \rangle} \in seq_i} x_{\langle i,j \rangle} \geq 1 \quad (9)$$

7 Experiments

In this section, we first describe the experimental setting and then evaluate our joint model of structure identification and nuclearity recognition on MCDTB.

7.1 Experimental Setting

Data: We use the the corpus annotated by ourselves for experiment, and the detailed corpus data is described in the Table 3.

Statistics Items	Value	Statistics Items	Value
Count of documents	720	Amount of sentences	8,391
Count of paragraphs	3,981	Average paragraphs (paragraphs/document)	5.53
Maximal of paragraphs	22	Average sentences (sentences/paragraph)	2.1
Minimal of paragraphs	2	Average characters (characters /paragraph)	554

Table 3: Corpus statistic data

There are 8,863 instances in MCDTB, and we use fivefold cross-validation to ensure the objectivity of the experiment. In particular, we divide the articles into 5 datasets and assign articles of the different lengths (length means the number of paragraphs of a discourse) into the 5 datasets relatively equally, so that the size of each data set is nearly the same. The number of discourses of different lengths is shown in Table 4.

Length	2	3	4	5	6	7	8	9	10	11	12	>13
Number	29	112	159	144	91	58	37	33	15	13	14	15

Table 4: Discourses of different lengths

Classification Algorithm: The CRF tool (*CRF++*)² is employed to train individual component classifiers and *lp_solver*³ is used to construct the joint model.

Evaluation Measurement: The performance is evaluated using the standard accuracy measurement.

7.2 Experimental Results

Table 5 compares the performance of local models when different feature sets are leveraged. The features used in local models are mentioned in Section 5. This table indicates the structure features make the greatest contribution to the local models. Although the performances of tree structure and similarity features is not good when used alone, combining with the organization features improves the performance of the model. Especially, when both these two kinds of features are used, the performance reaches the best

²<http://crfpp.googlecode.com/>

³<http://lpsolve.sourceforge.net/5.5/>

Features	Structure		Nuclearity	
	Accuracy	Macro-F1	Accuracy	Macro-F1
Organization	76.13	74.46	74.46	49.17
Tree Structure	74.48	73.38	70.21	36.86
Similarity	74.57	73.32	66.68	39.70
Organization + Tree Structure	76.31	74.64	74.70	47.25
Organization + Tree Structure + Similarity	77.52	75.98	75.50	49.83

Table 5: Comparison of experimental results of different feature sets

accuracy of 77.52% and 75.50% in SI and NR tasks respectively. The Macro-F1 values reach 75.98% and 49.83% in SI and NR tasks respectively. In the following experiments of joint model, we use the best performance for comparison.

Constraint	Accuracy	Macro-F1
Local model	77.52	75.98
ILP(C1)	77.95	77.68
ILP(C1+C2)	78.51	77.81
ILP(C1+C2+C3)	78.54	77.68

Table 6: Performance of structure identification with joint model

Table 6 shows the performance of the ILP approach when different constraints are used. From this table, we can see that the constraints C1, C2 and C3 are capable of improving the performance of structure identification. When all these constraints are utilized, the inference performance reaches the best of 78.54% in accuracy and 77.68% in Macro-F1, 1.02% in accuracy and 1.70% in Macro-F1 better than the best performance of local model. This indicates the beneficiary of label correlations intra- and multi-instance to the task of structure identification.

It is worthwhile to note that our ILP approach could also benefit the task of nuclearity recognition when the constraints are employed. Table 7 shows our ILP approach improve the performance of nuclearity recognition by 0.51% in accuracy and 1.86% in Macro-F1.

	Accuracy	Macro-F1
Local model	75.50	49.83
ILP approach	76.01	51.69

Table 7: Performance of nuclearity recognition with joint model

There are some discoveries in our experiment: 1) The lexical features of connective, such as “therefore”, “as a result” etc. are very useful on the micro level discourse analysis. But when used on macro level, the connective may confuse the model. That is because connectives usually occur between successive clauses or consecutive sentences, and are seldom used to express the relationship between paragraphs, especially in Chinese. 2) We have already tried some linguistic features, including lexical and syntactic, but the features are not outstanding in the experiment. There are several sentences in a paragraph, so syntactic information is not easy to use. We will explore other linguistic features and effective expression on macro level discourse structure analysis in the future.

8 Conclusion

In this paper, we present an efficient joint model of structure identification and nuclearity recognition on macro level discourse structure analysis. In particular, various kinds of feature sets are introduced to improve the performance of local models, and various constraints are introduced to improve the performance of joint model.

In future work, we will explore better joint modeling and effective linguistic features in discourse structure analysis. Furthermore, we wish to explore the other two tasks of macro discourse structure analysis, and build an End-to-End analysis platform.

Acknowledgements

We are grateful for the help of Jingjing Wang for his initial discussion. We thank our anonymous reviewers for their constructive comments, which helped to improve the paper. This work is supported by the National Natural Science Foundation of China (61773276, 61751206, 61673290) and Jiangsu Provincial Science and Technology Plan (No. BK20151222).

References

- Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. 2003. Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Current and new directions in discourse and dialogue*, pages 85–112. Springer.
- Arman Cohan and Nazli Goharian. 2017. Scientific document summarization via citation contextualization and scientific discourse. *International Journal on Digital Libraries*, pages 1–17.
- Vanessa Wei Feng and Graeme Hirst. 2012. Text-level discourse parsing with rich linguistic features. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, pages 60–68. Association for Computational Linguistics.
- Vanessa Wei Feng and Graeme Hirst. 2014. A linear-time bottom-up discourse parser with constraints and post-editing. In *ACL (1)*, pages 511–521.
- Rafael Ferreira, Luciano de Souza Cabral, Frederico Freitas, Rafael Dueire Lins, Gabriel de França Silva, Steven J Simske, and Luciano Favaro. 2014. A multi-document summarization system based on statistics and linguistic treatment. *Expert Systems with Applications*, 41(13):5780–5787.
- Francisco Guzmán, Shafiq Joty, Lluís Màrquez, and Preslav Nakov. 2014. Using discourse structure improves machine translation evaluation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 687–698.
- Hugo Hernault, Helmut Prendinger, David A. duVerle, and Mitsuru Ishizuka. 2010. Hilda: A discourse parser using support vector machine classification. *D&D*, 1(3):1–33.
- Feng Jiang, Xiaomin Chu, Sheng Xu, Peifeng Li, and Qiaoming Zhu. 2018. A macro discourse primary and secondary relation recognition method. *Journal of Chinese Information Processing*, 1(32):72–79.
- Shafiq Joty, Giuseppe Carenini, and Raymond T Ng. 2012. A novel discriminative framework for sentence-level discourse analysis. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 904–915. Association for Computational Linguistics.
- Shafiq R Joty, Giuseppe Carenini, Raymond T Ng, and Yashar Mehdad. 2013. Combining intra-and multi-sentential rhetorical parsing for document-level discourse analysis. In *ACL (1)*, pages 486–496.
- Junhui Li, Guodong Zhou, and Hwee Tou Ng. 2010. Joint syntactic and semantic parsing of chinese. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1108–1117. Association for Computational Linguistics.
- Peifeng Li, Qiaoming Zhu, and Guodong Zhou. 2013. Joint modeling of argument identification and role determination in chinese event extraction with discourse-level information. In *IJCAI*, pages 2120–2126.
- Yancui Li, Wenhe Feng, Jing Sun, Fang Kong, and Guodong Zhou. 2014. Building chinese discourse corpus with connective-driven dependency tree structure. In *EMNLP*, pages 2105–2114. Citeseer.
- Yancui Li. 2015. *Research of Chinese discourse structure representation and resource construction*. Ph.D. thesis, Suzhou: Soochow University.
- William C Mann and Sandra A Thompson. 1987. Rhetorical structure theory: A theory of text organization (no. isi/rs-87-190). marina del rey. CA: *Information Sciences Institute*.

- Daniel Marcu, Estibaliz Amorrortu, and Magdalena Romera. 1999. Experiments in constructing a corpus of discourse trees. *Towards Standards and Tools for Discourse Tagging*.
- Mitchell P Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. 1993. Building a large annotated corpus of english: The penn treebank. *Computational linguistics*, 19(2):313–330.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Jawad Sadek and Farid Meziane. 2016. A discourse-based approach for arabic question answering. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 16(2):11.
- Caroline Sporleder and Alex Lascarides. 2004. Combining hierarchical clustering and machine learning to predict high-level discourse structure. In *Proceedings of the 20th international conference on Computational Linguistics*, page 43. Association for Computational Linguistics.
- Teun Adrianus Van Dijk. 1980. *Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition*. Lawrence Erlbaum Associates.
- Nianwen Xue, Fu-Dong Chiou, and Martha Palmer. 2002. Building a large-scale annotated chinese corpus. In *Proceedings of the 19th international conference on Computational linguistics-Volume 1*, pages 1–8. Association for Computational Linguistics.
- Nianwen Xue, Xiuhong Zhang, Zixin Jiang, Martha Palmer, Fei Xia, Fu-Dong Chiou, and Meiyu Chang. 2013. Chinese treebank 8.0 ldc2013t21. *Linguistic Data Consortium, Philadelphia*.