Recognizing Macro Chinese Discourse Structure on Label Degeneracy Combination Model

Feng Jiang, Peifeng Li(✉), Xiaomin Chu, Qiaoming Zhu, and Guodong Zhou

School of Computer Science and Technology, Soochow University, Suzhou, China
{fjiang,xmchu}@stu.suda.edu.cn,
{pfli,qmzhu,gdzhou}@suda.edu.cn

Abstract. Discourse structure analysis is an important task in Natural Language Processing (NLP) and it is helpful to many NLP tasks, such as automatic summarization and information extraction. However, there are only a few researches on Chinese macro discourse structure analysis due to the lack of annotated corpora. In this paper, combining structure recognition with nuclearity recognition, we propose a Label Degeneracy Combination Model (LD-CM) to find the solution of structure recognition in the solution space of nuclearity recognition. Experimental results on the Macro Chinese Discourse TreeBank (MCDTB) show that our model improves the accuracy by 1.21%, compared with the baseline system.

Keywords: Label degeneracy · Combination model
Macro discourse structure

1 Introduction

Nowadays, the focus of most previous works on Natural Language Processing (NLP) has shifted from word to larger semantic levels, such as sentence and event. This trend makes discourse structure analysis more important because it is the foundation of many discourse-based NLP tasks. Discourse refers to a series of clauses, sentences or paragraphs as a whole [1]. It includes not only the text sequence but also the structural and logical relationships among them. Commonly, discourse structure analysis is divided into micro and macro structure analysis. The former studies the intra- or inter-sentence relationship, while the latter studies the discourse relationships among sentence groups, paragraphs and chapters [2], which pays attention to understanding the full text from higher-level semantics.

Macro discourse analysis uses paragraphs as elementary discourse units, and constructs a discourse structure tree between the paragraphs. It is a challenging task due to there is no connective between macro discourse units, and the length of the discourse unit is longer. In general, macro discourse analysis includes three tasks: structure recognition, nuclearity recognition, and relationship recognition. Structure recognition

M. Zhang et al. (Eds.): NLPCC 2018, LNAI 11109, pp. 92–104, 2018.
https://doi.org/10.1007/978-3-319-99501-4_8

is to identify whether there is a relationship between adjacent discourse units. Nuclearity recognition is to identify which is more important among discourse units. Moreover, relation recognition is to identify the logical relationship between discourse units. These three tasks constitute the analysis of macro discourse structure, and ultimately build a discourse structure tree of an article.

Table 1 is a sample text including a title and five paragraphs. Figure 1 is its macro discourse structure tree, in which the leaf nodes are paragraphs, and the parent nodes, the larger discourse units, connect the adjacent child nodes. A directed edge connected parent and child nodes, with the arrow's edge pointing to the important child node and the non-arrowed edge pointing to the secondary child node. In Fig. 1, P_2, P_3 and P_4 form a *Joint* relation, and they are equally important; P_1 and $DU_{2\text{-}4}$ form a *Commentary*

Table 1. Contents of chtb_0156.

Asia's Largest Fluorine-Free Coolant Manufacturing Factory Entered Production In Tianjin (亚洲最大无氟制冷剂生产厂在天津投产)

(P_1) Asia's largest fluorine-free coolant manufacturing base - GreenKel coolant (China) limited company completed … in Tianjin's development district. (亚洲最大的无氟制冷剂生产基地——格林柯尔制冷剂（中国）有限公司日前在天津开发区建成投产。)

(P_2) With headquarters established in Canada, GreenKel is a … company manufacturing and selling fluorine-free coolant. (总部设在加拿大的格林柯尔集团是生产和销售无氟新型制冷剂的大型跨国集团公司。) GreenKel coolant (China) limited company is … company, China's Tianjin … office and … Company, ltd. with a total investment value of … dollars, among which the foreign investment value is … dollars. (格林柯尔制冷剂（中国）有限公司是格林柯尔集团北美公司与中国天津开发区总公司和中国南方证券有限公司合建的合资企业，总投资额五千万美元，其中外方投资额四千一百八十五万美元。)

(P_3) The company occupies … meters, and introduced … advanced production line and testing equipment, to produce … fluorine-free coolants. (公司占地七万平方米，引进全套欧美先进的生产线和检测设备，生产格林柯尔无氟新型系列制冷剂。)

(P_4) The currently completed first phase project annual production volume that … , with annual production values reaching … dollars , is the fluorine-free coolant production base that … the most advanced equipment and technology in Asia .(目前竣工的一期工程年产量可达一万吨，全部出口外销，年产值达二亿美元，是目前亚洲生产规模最大，设备和技术最先进的无氟制冷生产基地。)

(P_5) The completion and entering into production of this production base, signifies that … the world's leading ranks. (该生产基地的建成投产，标志着中国无氟制冷生产跨进世界领先行列。)

relation, in which P_1 is more important; $DU_{1\text{-}4}$ and P_5 form an *Evaluation* relation, in which $DU_{1\text{-}4}$ is more important. A good article always has a good discourse structure tree. If we can construct a discourse structure tree, it will play an active role in downstream tasks such as automatic summarization and information extraction. In the task of automatic summarization, after constructing a macro structure tree, we can follow the arrow from the top down to the leaf node to get a more natural summary. For example, according to the Fig. 1, chtb_0156's abstract in Table 1 is the topic sentence of P_1.

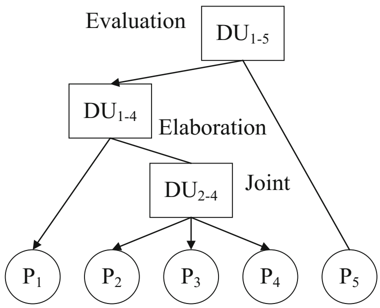

Fig. 1. The macro structure tree of chtb_0156.

Discourse structure recognition is the first step and the most important step in the tasks of discourse structure analysis. In this paper, we first use the features of structure information and semantic macro-information to form a combination model from different views. On this basis, it uses the label degeneracy to find out the solution of structure recognition in the solution space of the nuclearity recognition to obtain more detailed feature expression. Finally, we use the minimum probability post-editing to ensure that the model can automatically build a complete macro structure tree. The experimental results on MCDTB, a RST style macro Chinese discourse TreeBank, justify the effectiveness of our model.

2 Related Work

Discourse structure analysis is divided into micro and macro structure analysis. In micro discourse analysis, Hernault et al. [3] was the first to implement a complete discourse analyzer in English, and used SVM with rich textual features (including structure features, syntactic features and dominant sets) in structure recognition. Recent studies followed this research line and focused on effective features. Joty et al. [4] and Feng et al. [5] used sequence labeling instead of classification. The former used Dynamic Conditional Random Field (DCRF) model combining structure recognition with relationship recognition, and the latter used Conditional Random Field (CRF) model with post-editing avoiding meaningless sequence labeling. Li et al. [6] and Li et al. [7] used the recursive neural network and attention-based hierarchical neural network with the distribution representation of text for recognizing structure, respectively. Besides, Feng et al. [9] and Wang et al. [10] introduced N-gram and the tree features to recognize structure on the previous work [3, 8]. Due to the lack of unified theories and corpus, there are few researches on Chinese discourse analysis. Lv et al. [11] using Chinese FrameNet (CFN) and Li [1] using Connective-driven Dependency Tree (CDT) with maximum entropy model attempted to recognize the discourse structure on their self-built corpus, respectively.

There is less research at the macro level due to lack of corpus. In English, Sporleder et al. [12] used a maximum entropy model to recognize the macro discourse structure of RST-DT after correction and clipping. In Chinese, Jiang et al. [13] used the maximum entropy model to identify the macro discourse nuclearity with annotated discourse structures in MCDTB, the only one Chinese corpus labeled on the macro level.

3 Label Degeneracy Combination Model

Label Degeneracy Combination Model (LD-CM) is shown in Fig. 2. We first use structural information to train the Structure Model, and use semantic and macro information to train the Semantic and Macro Information Model to recognize the macro discourse nuclearity. Then, we use these models to form a combination model. The combination model can learn different features from different views, and decode the prediction results in a unified way to compensate for the bias caused by a single model, thereby improving system performance.

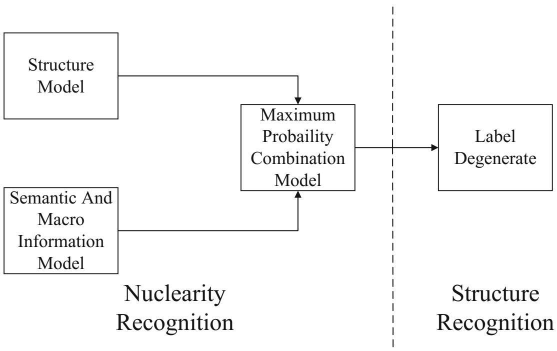

Fig. 2. The label degeneracy combination model.

Finally, we use the label degeneracy to map the nuclearity results to reveal the macro-structure. In previous work [1, 5], structure recognition and nuclearity recognition were mainly cascading tasks. That is to say, structure recognition is binary classification (whether the adjacent discourse units can be merged) and nuclearity is three-classification (*Nucleus Ahead, Nucleus Behind* and *Multi-Nucleus*) on the merging part that is obtained from structure recognition. We improve the nuclearity recognition and regard the *No-Relationship* as the fourth class. In this way, there is a mapping relationship between structure recognition and nuclearity recognition: the adjacent discourse units should be merged (*Nucleus Ahead, Nucleus Behind* and *Multi-Nucleus*) or not (*No-Relationship*). Therefore, it can make the model capture more detailed feature expression from three classes of nuclearity to be distinguished from *No-Relationship*, which helps to recognize structure.

3.1 Sequence Labeling on Chinese Discourse Structure Recognition

In Chinese discourse structure recognition, Li et al. [1] and Lv et al. [11] both regarded it as a classification problem, that is, to determine whether two given discourse units should be merged. However, their methods have certain disadvantages, such as the inability to consider contextual information and to recognize multi-relationship structure.

Therefore, we regard this problem as sequence labeling following Joty et al. [4] and Feng et al. [5], which has achieved good performance in English discourse structure recognition. For the task of predicting macro discourse structure in Chinese, the sequence labeling model has the following advantages: (1) it can consider the context information; (2) it can keep the original discourse structure; (3) it can achieve a balance between greedy algorithm and global optimization.

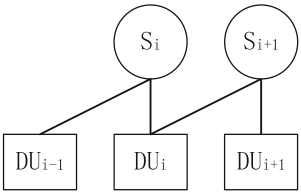

Fig. 3. Macro-structure sequence labeling model.

As shown in Fig. 3, DU_i is the i^{th} discourse unit and S_i is the structure label whether the i^{th} discourse unit is merged with the previous DU_{i-1}. On the one hand, when annotating S_i, we can take into account the information of DU_{i-1} and DU_{i+1}, so as to increase the accuracy of discourse structure recognition. On the other hand, there is no need to convert multi-relationship structure. For example, when DU_{i-1}, DU_i and DU_{i+1} form a multi-relationship structure, we only need to consider whether DU_i and DU_{i-1} should be merged and whether DU_{i+1} and DU_i should be merged. When both of them labeled as merged, three successive discourse units form a multi-relationship structure.

3.2 Feature Combination

Table 2 shows the features used in the previous studies on RST-DT, an English discourse corpus, where Sporleder et al. [12] focused on macro structure analysis and the rest focused on micro level. Table 2 illustrates that syntax information and dominant set are very useful in microstructure analysis. However, we cannot introduce them to macro structure analysis because the elementary of macro discourse unit is paragraph

Table 2. Statistics of the features used in recent studies.

Features	Sporleder [12]	Hernault [3]	Feng [9]	Joty [4]	Feng [5]	Wang [10]
Location information and distance information	√	√		√	√	√
Structure information						√
Syntax information, dominance set		√	√	√	√	√
N-gram			√	√	√	√
Number of paragraphs and sentences	√	√		√	√	√
Tree features				√	√	√
Status information						√
Word co-occurrence, semantic similarity	√		√			
Punctuation, tense	√					
Cue words, lexical chains.	√			√	√	
Contextual features				√	√	
Entity transfer					√	

that does not have syntax information and dominant set. Therefore, we select the other common features used in the previous studies as the structural features as follows:

- The position of the beginning and the end of a discourse unit;
- The number of sentences and the number of paragraphs contained in a discourse unit;
- The comparison of the number of sentences in a discourse unit to its previous unit;
- The comparison of the number of paragraphs in a discourse unit to its previous unit.

Inspired by the distributed representation of text [14], we improved the "word co-occurrence" feature adopted by Sporleder et al. [12] and used semantic similarity to measure the semantic connection between two discourse units. We used Word2Vec model to train word vectors on CTB8.0, and used the method proposed by Xu [15] to calculate semantic similarity.

In macro structure, it is not possible to display the connectives explicitly, but there may be a connective in the first sentence of a discourse unit. Therefore, we regard the first connective and its part of speech in the discourse unit as one of the features. In addition, because macro-structure analysis focuses more on macroscopic understanding, we add some macro information as features, such as whether a discourse unit is a leaf node and whether a discourse unit was merged at the previous round. Due to differences in language characteristics and discourse unit granularity, we do not use tense and N-gram features. Finally, we select the semantic and macro information features as follows:

- The semantic similarity between a discourse unit and its previous unit;
- The connective of the first sentence and it's part of speech in a discourse unit;

- Whether a discourse unit is a leaf node;
- Whether a discourse unit was merged in the previous round.

We first use the structural features to train Structure Model and use semantic and macro information features to train Semantic and Macro Information Model on the training set, respectively. Then we use the above two models to predict nuclearity of two given discourse units in the test set, respectively. For each nuclearity label in the prediction sequence, the one with the highest probability from the above two models is selected as the final result.

3.3 Label Degeneracy

Previous studies have considered nuclearity recognition as part of relationship recognition following the structure recognition. Different from them, we use label degeneracy to recognize the structure after nuclearity recognition. In nuclearity recognition, we added *No-Relationship* as the fourth class to form a model for recognizing structure and nuclearity simultaneously. In this way, there is a mapping relationship between structure recognition and nuclearity recognition: the adjacent discourse units should be merged *(Nucleus Ahead*, *Nucleus Behind* and *Multi-Nucleus*) or not (*No-Relationship*).

Our label degeneracy approach is as follows: when DU_i is merged with the previous discourse unit DU_{i-1}, the label S_i is 1, otherwise 0. In nuclearity recognition, when DU_i isn't merged with DU_{i-1}, the label N_i is 0. When DU_i is less important than DU_{i-1} (*Nucleus Ahead*), N_i is 1; when DU_i is more important than DU_{i-1} (*Nucleus Behind*), N_i is 2; when DU_i and DU_{i-1} are equally important (*Multi-Nucleus*), N_i is 3. The degenerate mapping relationship between the labels of the discourse structure recognition and the nuclearity recognition is shown in Eq. (1).

$$S_i = LD(N_i) = \begin{cases} 0, & N_i = 0 \\ 1, & N_i = 1, 2, 3 \end{cases} \tag{1}$$

Therefore, it can make the model capture more detailed feature expression from three classes of nuclearity to be distinguished from *No-Relationship*, which helps to recognize structure. Jiang's experiment [13] shows that the main mistakes of nuclearity recognition is that it always recognizes *Nuclearity Behind* as other two types by mistake because this class is scarce. In contrast, the other two relations *Nuclearity Ahead* and *Multi-Nucleus* can be well recognized. This result means that their models can capture differences between different nuclearity classes, making the characteristics of the merged part more detailed (not only labeled 1 in structure recognition, but also labeled 1, 2 and 3 in nuclearity recognition), thereby enhancing distinction with *No-Relationship* (labeled 0 in structure and nuclearity).

For example, if one relationship is labeled as 3 (*Multi-Nucleus*), usually it is a *Coordination* relation which may have more than two children. In this way, the label 3 (*Multi-Nucleus*) can strengthen the feature expression of this discourse structure (the adjacent discourse units should be merged). So that it is not only distinguished from *No-relationship* (label 0), but also distinguishable from the *Nuclearity Ahead* (label 1) and *Nuclearity Behind* (label 2), thus improving the accuracy of the structure recognition. In addition, even if the error is confused with three nuclearity classes (*Nuclearity Ahead, Nuclearity Behind* and *Multi-Nucleus*), label degeneracy can also make it not reduce the accuracy of the result.

4 Experiment

In this section, we first introduce the experimental setting, and then report the experimental results and analysis.

4.1 Experimental Setup

There are few genuine macro discourse corpora, especially in Chinese language. The Macro Chinese Discourse TreeBank (MCDTB) [13] is the only available corpus in Chinese. Different from RST-DT, MCDTB adopts the three categories of 15 classes of discourse relationships formed by the improved CDTB [1] in macro-level discourse relationship annotation. In addition, MCDTB also includes macro discourse information, such as paragraph topic sentences, summaries, abstracts and pragmatic functions. MCDTB annotated 720 news reports (0001-0325, 0400-0454, 0500-0554, 0600-0885 and 0900-0931) from CTB 8.0. As shown in Table 3, the corpus has 3,981 paragraphs, and the average number of paragraphs is 5.53 per document. Besides, a document contains at most 22 paragraphs and at least 2 paragraphs. The corpus contains 8,319 sentences, 398,829 words, with an average of 553.93 words per document.

Table 3. The statistical of MCDTB.

Items	Number
#documents	720
#paragraphs	3,981
#paragraphs of the longest document	22
#paragraphs of the shortest document	2
#sentences	8,319
Average length (paragraphs/document)	5.53
Average length (sentences/paragraph)	2.09

We use Conditional Random Field (CRF) to train the Structure Model and the Semantic and Macro Information Model, with the parameter C of 4, the feature window of 3 and other parameters are default. There are 8,863 samples in total, including 3,261 positive samples and 5,602 negative samples. We use five-fold cross validation to ensure the objectivity of the experiments. In particular, according to the article lengths, we divided the articles of different lengths into five sets, so that the size of each set is almost the same.

Feng et al. [5] pointed out that there were two constraints used in the serialization labeling method of discourse structure recognition: discourse units cannot be consecutively merged and there is no merger at all in the sequence result. For the above two constraints, we take the measures as following:

First, we maintain the multi-relationship structure. In RST-DT, the multi-relationship structure is relatively small. According to statistics, 95% of annotated relations are binary. In MCDTB, the multi-relationship structure accounts for 8.6%, reaching 246 and the multi-relationship has 16 child nodes at most. Using the right-branching tree to replace the multi-relationship tree, the number of newly generated relations will account for 19.66%. At the same time, it also increases the distortion of the structure tree. We use the original structure to ensure the objectivity of recognition.

Second, we use the minimum probability post-editing to ensure that the second case does not occur. When there is no merger at all in the sequence result, we traverse the probability values of each label in prediction, and replace the label 0, that has minimum probability, with label 1 to ensure that the prediction meets the constraint condition.

4.2 Experimental Results and Analysis

We use the features of Sporleder [12] that can be migrated to Chinese (distance information, location information, number of sentences and paragraphs, sentences and paragraphs' number comparison, semantic similarity and connectives) to form Sporleder Liked Model as a benchmark system. As shown in Table 4, the performance of the Structure Model (M1) and the Semantic and Macro Information Model (M2) in discourse structure recognition is slightly worse than the baseline. However, composed of M1 and M2, the Combination Model (M1 + M2) obtained 77.56% accuracy, which is 0.5% higher than Sporleder Liked Model. Our model LD-CM outperforms all other six models in accuracy and this justifies the effectiveness of the combination model and label degeneracy.

Table 4. The comparison of each model's accuracy. Significant differences between SLM and LD-CM* (with $p < 0.01$).

#	Model	Accuracy
M1	Structure Model (structure recognition)	76.09%
M2	Semantic and Macro Information Model (structure recognition)	77.01%
M3	Sporleder Liked Model (SLM)	77.06%
M4	Combine Model (M1 + M2)	77.56%
M5	Label Degeneracy Model (based on M2)	77.78%
M6	Label Degeneracy Combination Model (LD-CM)	**78.27%***
M7	LD-CM with Post-Editing (LD-CMWP)	78.24%

As shown in Table 5, there is 14.11% difference prediction between the Structure Model (M1) and the Semantic and Macro Information Model (M2). This proves that different features can learn different views, thereby enhancing the ability of discourse structure recognition.

Table 5. The difference between the predictions of M1 and M2.

Prediction	Wrong (M2)	Correct (M2)
Wrong (M1)	16.39%	**7.51%**
Correct (M1)	**6.60%**	69.49%

Label Degeneracy Model (M5) based on M2 achieved a correct rate of 77.78%, which was 0.72% higher than that of Sporleder Liked Model and was about the same as Combine Model (M4). Therefore, we also statistics the differences between the predictions of M4 and M5 as shown in Table 6. There is 11.94% difference prediction between these two models. This proves that different methods have improvement in the different direction, so it is effective to combine them into a Label Degeneracy Combination Model (LD-CM).

Table 6. The difference between the predictions of M4 and M5.

Prediction	Wrong (M5)	Correct (M5)
Wrong (M4)	16.36%	**6.08%**
Correct (M4)	**5.86%**	71.70%

Label Degeneracy Combination Model (LD-CM) achieves an optimal value of 78.27%, 1.21% higher than Sporleder Liked Model. We use LD-CM with Post-Editing (LD-CMWP) to ensure that it can build a complete tree automatically with only reduce the accuracy slightly. To find out how LD-CM can improve the performance of macro-structure recognition, we analyzed the experimental results and found two phenomena:

1. LD-CM can overcome overfitting caused by a large amount of samples from short articles. LD-CM is calmer than other models for whether the third and fourth discourse unit should be merged. In those short articles, the third and fourth discourse unit (usually paragraphs) are often merged. However, in those long articles, they are acted as higher-level discourse unit in high probability.
2. LD-CM is better at recognizing complex higher-level macro structure. Figure 4 shows the prediction results of each model for the high-level macro-structure of chtb_0112. LD-CM predicts the structure correctly (a). Structures (b), (c), and (d) are the results of Sporleder Liked Model, Combination Model and Label Degeneracy Model, respectively. In complex high-level macro structure, more attention should be paid to the information in the front part of the article, due to the fact that if the article is longer, its topic is more dispersed, and its important part is more likely to gather in the front part. Sporleder Liked Model did not show learning this. Combination Model shows such a predictive behavior: if a discourse unit is not merged with other discourse unit before, it is more likely to be merged with the next discourse unit. However, it does not take into account that the important part in the front of the article. Label Degeneracy Model also captures discourse units, which are more likely to be merged in the front part of the article, but does not consider the fact that the behind part has been merged.

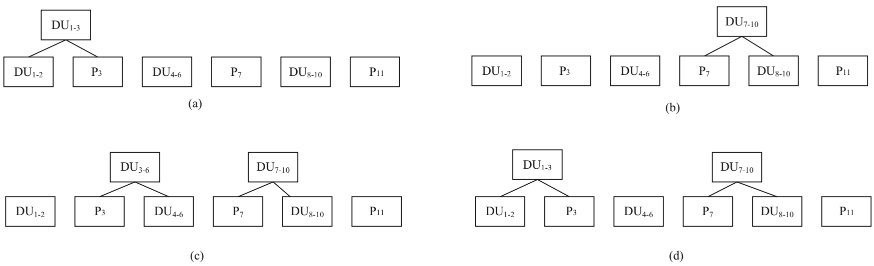

Fig. 4. High-level macro discourse structure of chtb_0112.

In addition, to study the performance of each model for articles on different lengths, we divide the corpus into two parts: the short articles whose paragraph lengths are less than or equal to 6 and the long articles whose paragraph lengths are more than 6. We model the corpora of 535 short articles and 185 long articles respectively and analyzed the results. Figure 5 shows the performance on different paragraph lengths. In those short articles, compared with the benchmark system M3, Combination Model (M4) and Label Degeneracy Model (M5) have improved 0.8% and 0.5% in accuracy, reaching 81.63% and 81.33%, respectively. The reason that M4 is better than M5 is that the structural information and semantic and macro information are equally important in the short article. It can capture different features from different aspects, to compensate for the bias caused by the single model.

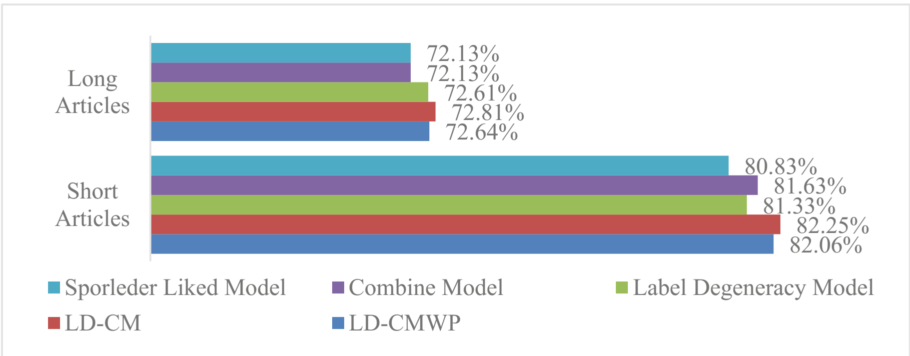

Fig. 5. The performance of models in different length articles.

In those long articles, M5 performed better than M4, reaching 72.61% accuracy. With the increasing of the article length, the structural information is more difficult to express for the structural characteristics of the discourse, and sometimes even excessively learning the low-level structure will cause the prediction error of the high-level structure. In those long articles, multi-relationships (usually *Joint* relation) appear more frequently, and this structure usually is labeled as 3 (*Multi-Nucleus*) in nuclearity, so it is easier to capture the special structure by the label degeneracy.

It is worth noting that LD-CM has achieved the best performance in both long and short article, reaching 72.81% and 82.25% respectively. Even LD-CMWP with slightly lower performance is better than other models, reaching 72.64% and 82.06% accuracy in long and short article respectively. This proves the effectiveness of our model.

5 Conclusions

In this paper, we propose a Label Degeneracy Combination Model (LD-CM) in recognizing macro discourse structure. We combine structural features with the semantic and macro-information features to form a combination model, and use the label degeneracy to find the solution of structure recognition in the solution space of nuclearity recognition. In this way, the model can capture a more detailed feature expression. The experimental results on MCDTB show that our LD-CM improves the accuracy by 1.21%, compared with the benchmark system. In particular, we use the post-editing to ensure generating a complete discourse structure tree automatically. In the future work, we will focus on recognizing discourse nuclearity and relationship, and eventually form an end-to-end macro discourse analyzer.

Acknowledgments. The authors would like to thank three anonymous reviewers for their comments on this paper. This research was supported by the National Natural Science Foundation of China under Grant Nos. 61772354, 61773276 and 61472265, and was also supported by the Strategic Pioneer Research Projects of Defense Science and Technology under Grant No. 17-ZLXDXX-02-06-02-04.

References

1. Li, Y.: Research on the structure of Chinese text structure and the construction of resources. Soochow University (2015)
2. Van Dijk, T.A.: Narrative macro-structures. PTL J. Descr. Poet. Theory Lit. **1**, 547–568 (1976)
3. Hernault, H., Prendinger, H., duVerle, D.A., Ishizuka, M.: HILDA: a discourse parser using support vector machine classification. Dialogue Discourse **1**(3), 1–33 (2010)
4. Joty, S., Carenini, G., Ng, R., Mehdad, Y.: Combining intra- and multi-sentential rhetorical parsing for document-level discourse analysis. In: ACL 2013, vol. 1, pp. 486–496. ACL, Sofia (2013)
5. Feng, V.W., Hirst, G.: A linear-time bottom-up discourse parser with constraints and post-editing. In: ACL 2014, pp. 511–521. ACL, Baltimore (2014)
6. Li, J., Li, R., Hovy, E.: Recursive deep models for discourse parsing. In: EMNLP 2014, pp. 2061–2069. ACL, Doha (2014)
7. Li, Q., Li, T., Chang, B.: Discourse parsing with attention-based hierarchical neural networks. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 362–371. ACL, Austin (2016)
8. Sagae, K.: Analysis of discourse structure with syntactic dependencies and data-driven shift-reduce parsing. In: International Workshop on Parsing Technologies, pp. 81–84. ACL, Paris (2009)

9. Feng, V.W., Hirst, G.: Text-level discourse parsing with rich linguistic features. In: ACL 2012, vol. 1, pp. 60–68. ACL, Jeju (2012)
10. Wang, Y., Li, S., Wang, H.: A two-stage parsing method for text-level discourse analysis. In: ACL 2017, vol. 2, pp. 184–188. ACL, Vancouver (2017)
11. Lv, G., Na, S., Li, R., Wang, Z., Chai, Q.: Frame-based discourse structure modeling and relation recognition for Chinese sentence. J. Chin. Inf. Process. **29**(6), 98–109 (2015)
12. Sporleder, C., Lascarides, A.: Combining hierarchical clustering and machine learning to predict high-level discourse structure. In: Coling 2004, pp. 43–49. ACM, Geneva (2004)
13. Jiang, F., Chu, X., Sheng, X., Li, P., Zhu, Q.: A macro discourse primary and secondary relation recognition method. J. Chin. Inf. Process. **32**(1), 43–50 (2018)
14. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. Comput. Sci. (2013)
15. Xu, S.: Research and Implementation of Paraphrase Recognition for Question Answering. Harbin Institute of Technology (2009)