Joint Modeling of Recognizing Macro Chinese Discourse Nuclearity and Relation Based on Structure and Topic Gated Semantic Network

Feng Jiang, Peifeng Li and Qiaoming Zhu

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

Abstract. Nowadays, in the Natural Language Processing field, with the object of research gradually shifting from the word to sentence, paragraph and higher semantic units, discourse analysis is one crucial step toward a better understanding of how these articles are structured. Compared with micro-level, this has rarely been investigated in macro Chinese discourse analysis and faces tremendous challenges. First, it is harder to grasp the topic and recognize the relationship between macro discourse units due to their longer length and looser relation between them. Second, how to mine the relationship between nuclearity and relation recognition effectively is another challenge. To address these challenges, we propose a joint model of recognizing macro Chinese discourse nuclearity and relation based on Structure and Topic Gated Semantic Network (STGSN). It makes the semantic representation of a discourse unit can change with its position and the topic by Gated Linear Unit (GLU). Moreover, we analyze the results of our models in nuclearity and relation recognition and explore the potential relationship between them. Conducted experiments show the effectiveness of the proposed approach.

Keywords: Macro Chinese Discourse, Structure and Topic Gated Semantic Network, Gated Linear Unit, Nuclearity Recognition, Relation Recognition.

1 Introduction

In the field of Natural Language Processing (NLP), discourse analysis is becoming increasingly important as the object of research gradually shifts from the word to sentence, paragraph, and higher semantic units. Discourse analysis primarily examines the text coherence and cohesion, including the analysis on structure, nuclearity, and relation. There are two hierarchical levels of discourse analysis: micro level and macro level. The micro level takes a clause or a sentence as an Elementary Discourse Unit (EDU) and researches on intra-sentence or inter-sentence discourse relations. While the macro level takes a paragraph as a discourse unit, and researches on discourse relations between paragraphs and chapters [19] to revealing and insightful a higher level of text coherence above the sentence.

In this work, we focus on recognizing macro Chinese discourse nuclearity and relation, which helps to understand the central topic of the text better. To the best of our knowledge, Macro Chinese Discourse Treebank (MCDTB) [9] is only available macro Chinese discourse corpus, which annotated with macro discourse structure. Its annotation style is consistent with that of Rhetorical Structure Theory Discourse Treebank (RST-DT) [1], including the structure, nuclearity, and relation of macro discourse structure. In RST-DT, many existing studies associate structure recognition with nuclearity recognition, while fewer studies associate nuclearity recognition with relation recognition due to the explicit dependence of nuclearity and relation in RST-DT. However, in MCDTB, the nuclearity of a discourse unit is decided by whether it can better represent the theme of a document in a global view [3].

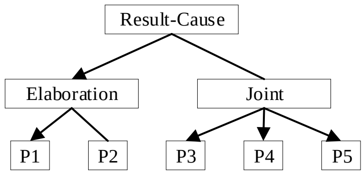

Fig. 1. The macro discourse tree of chtb0056 in MCDTB.

As shown in Fig. 1, to make a clearer explanation of the macro discourse structure, we take the chtb0056 in MCDTB as an example. There are five paragraphs (P1, P2, P3, P4, and P5) as DUs and three bigger discourse units with relations (Elaboration, Joint, and Result-Cause) in the article. The directed edge indicates that the child node is the primary discourse unit (Nucleus), and the undirected edge indicates that the child node is secondary (Satellite). Researching discourse nuclearity and relation recognition can benefit a variety of downstream applications including question answering, machine translation, text summarizing, and so forth. In the task of text summarizing, after constructing a macro discourse tree, we can follow the arrow from the top down to the leaf node to get a more natural summary. For example, according to Fig. 1, chtb0056's abstract is the topic sentence of P1.

On the other hand, macro Chinese discourse nuclearity and relation recognition faces tremendous challenges. First, different from micro-level, macro-level discourse unit has a larger granularity and longer length, and its topic should be grasped from a higher level. Therefore, it is crucial for macro discourse nuclearity and relation recognition that how to effectively combine semantic information and structure and topic information to represent discourse units. Second, previous works [8, 21] show there is a big gap between nuclearity and relation recognition. Therefore, how to join these two tasks suitably is another challenge.

In this study, we propose a joint model based on Structure and Topic Gated Semantic Network (STGSN) to recognize macro Chinese discourse nuclearity and relation. To obtain the macro semantic representation of a discourse unit, STGSN uses macro structure information of a discourse unit and the topic of the whole document to control the flow of information by Gated Linear Unit (GLU) [4]. Therefore, the semantic representation of a discourse unit can change with its position and the topic. In addition, we propose a joint model of nuclearity and relation recognition that reduces a single model's recognition errors by exploring the potential relation between the nuclearity of discourse units and relations among them.

Our key contributions are summarized as follows. First, to the best of our knowledge, we are the first to use neural network model on nuclearity recognition and relation recognition in macro Chinese discourse and propose a joint model to associate these two tasks. Second, we propose a Structure and Topic Gated Semantic Network (STGSN) for achieving the macro semantic representation of discourse unit changed with its position and the topic, which improves the performance by recognizing the type of fewer samples better. Third, we propose joint learning of nuclearity and relation recognition that reduces a single model's recognition errors and explore the potential relationship between the nuclearity of discourse units and relations among them.

The rest of the paper is organized as follows. Section 2 overviews the related work. Section 3 describes the proposed model in detail. Section 4 presents experiments and discussions. We conclude the paper in Section 5 and shed light on future directions.

2 Related Work

In English, previous studies of nuclearity and relation recognition mainly focus on full discourse parsing, with RST-DT [1] being one of the most popular discourse corpora. RST-DT is based on Rhetorical Structure Theory (RST) [14] and contains 385 documents from the Wall Street Journal. It is annotated with the discourse structure, nuclearity, and relation to representing the relationship between two or more discourse units. Since both micro and macro discourse structures with a document annotated as a tree, it does not explicitly distinguish between micro-level and macro-level discourse structure. In RST-DT, most existing approaches [5,6,20] either model discourse structure, nuclearity and relation recognition separately, while other studies regard nuclearity as subsidiary attributes of structure [7] or relation [10,11], ignoring the importance of nuclearity recognition and the implicit relationship among nuclearity and relation. However, a few studies focus on the macro level. Sporleder and Lascarides [17] used a maximum entropy model to identify the macro discourse structure, after pruning and revising the original discourse trees on RST-DT corpus, but they did not recognize nuclearity and relation on the macro level.

In Chinese, Li et al. [13] proposed Chinese discourse treebank (CDTB) on the micro level and there are some successful attempts [12,18] for discourse analysis tasks on this corpus. On the macro level, MCDTB [9] is the only available macro Chinese discourse corpus. Its annotation style is consistent with that of RST-DT, including the structure, nuclearity, and relation, but only annotated on the macro level. Currently, MCDTB contains 720 news documents annotated with 3 categories and 15 relations. Jiang et al. [8] proposed two topic similarity features as supplements to structural features and tried to use the maximum entropy model to identify the discourse nuclearity on MCDTB. Chu et al. [2] used Conditional Random Field (CRF) to build a local model, and then proposed a joint model of structure identification and nuclearity recognition by Integer Linear Programming (ILP) to reduce the error transmission between the associated tasks. Zhou et al. [21] proposed a distributed representation of macro discourse semantics on word vectors in a global view. Besides, he used some original features to improve the performance of relation recognition.

3 Overview of the Framework

In this section, we propose a joint model of nuclearity recognition and relation recognition based on Structure and Topic Gated Semantic Network (STGSN), and its high-level illustration is shown in Figure 2. It includes three modules: 1) Text Encoding, 2) Structure and Topic Gated Semantic Network, and 3) Joint Learning with Nonlinear Transformation Layer.

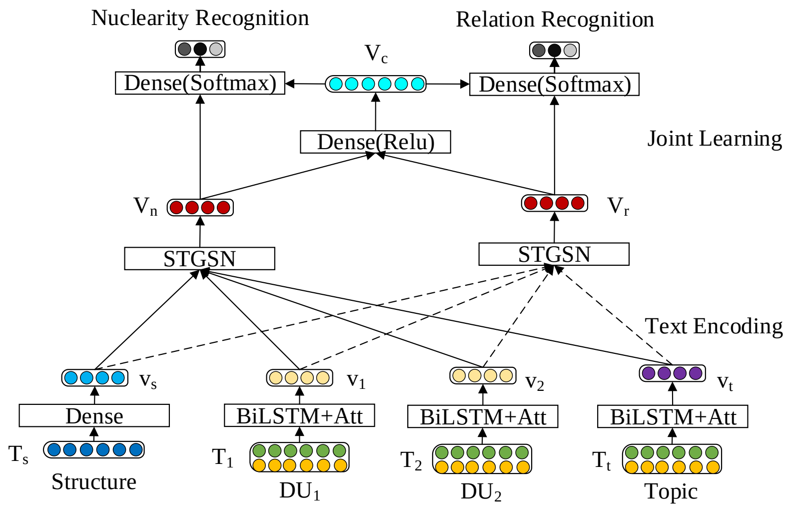

Fig. 2. The Joint model based on STGSN.

To recognize the nuclearity and recognition of two discourse units DU_1 and DU_2, we put their words with part-of-speech sequences as semantic information. What's more, we use DU_1 and DU_2's position features used in Jiang et al. [8] as structure information and the title of the document as topic information. All of them make up the input of our model. Then the Text Encoding module first encodes DU_1, DU_2, and topic into the semantic vectors V_1, V_2, and V_t by the BiLSTM and Attention layer. Besides, it encodes the position information (the number of the start and end of a discourse unit, the distance from the start and end of the document and so on) of DU_1 and DU_2 into the structure vector V_s. Then, these semantic and structure representations are separately fed into two STGSNs for nuclearity recognition and relation recognition. Finally, we feed V_n and V_r into a nonlinear transformation layer for capturing the relationship between nuclearity recognition and relation recognition and get a combined vector V_c which will be concentrated with V_n and V_r into each task in the joint learning module.

3.1 Text Encoding

In the Text Encoding module, DU_1, DU_2, and topic are represented as the sequence X= (w_1, w_2, ... , w_n), where n is the number of words in a discourse unit or the title of a document. We first use Word2Vec [15] to initialize the word embedding e_i of the word w_i and its part-of-speech embedding p_i. Then we merge all of the word embedding and part-of-speech embedding in a discourse unit to a sequence T= (t_1, t_2, ... , t_n) to represent this discourse unit and t_i is showed as Eq.1.

$$t_i = [e_i, p_i] \ \ 1 \leq i \leq n \tag{1}$$

We use BiLSTM to obtain the semantic representation of a discourse unit and use the attention mechanism (weighted summation of each time step) [16] to capture the more important parts of the discourse unit as Eq.2.

$$V_j = \text{Attention}\left(\text{BiLSTM}(T_j)\right) \tag{2}$$

Following Jiang et al. [8]'s structural features, we use randomly initialized hard-coded embedding s_k to represent the structural features (where k is the number of features), and use a concatenation layer to connect them and feed them into a dense layer for getting structure vector (V_s) as Eq.3.

$$V_s = W_s[s_1, s_2, \ldots, s_k] + b_s \tag{3}$$

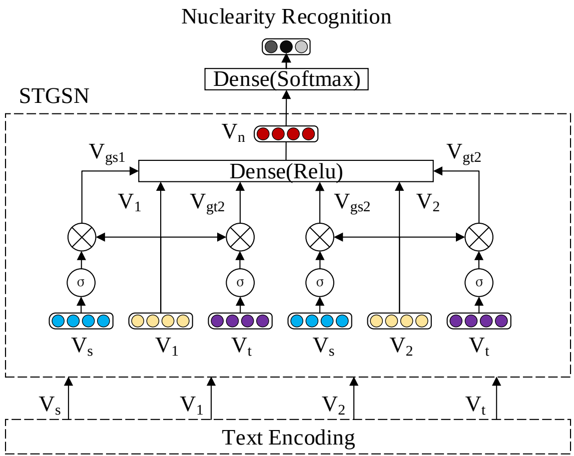

Fig. 3. The Structure and Topic Gated Semantic Network for nuclearity recognition.

3.2 Local Model

As shown in Fig.3, we illustrate STGSN with a local model for nuclearity recognition and there is the same model for relation recognition. After text encoding, we propose a Structure and Topic Gated Semantic Network. First, we use a discourse unit's structure information (V_s) and the topic of the document (V_t) to control the flow of semantic information (V_1 or V_2) by Gated Linear Unit (GLU) [4], to make the discourse unit's semantic representation can change with its position and the topic, which are formulated as Eq.4 and Eq.5.

$$V_{gsi} = V_i \otimes \sigma\left(W_{gs} V_s + b_{gs}\right) \quad i \in 1,2 \tag{4}$$

$$V_{gti} = V_i \otimes \sigma\left(W_{gt} V_t + b_{gt}\right) \quad i \in 1,2 \tag{5}$$

Then we concentrate them and feed them into a dense layer with Relu active function to get the final representation V_n as Eq.6. Finally, we feed the final vector V_n into a standard softmax layer for nuclearity recognition as Eq.7. During training, we use Adam optimizer to optimize the network parameters by maximizing the log-likelihood loss function between the predicted label $\hat{y}$ and the real label y.

$$V_n = W_n\left[V_1, V_{gs1}, V_{gt1}, V_2, V_{gs2}, V_{gt2}\right] + b_n \tag{6}$$

$$\hat{y} = softmax(W_{soft} V_n + b_{soft}) \tag{7}$$

3.3 Joint Learning with Nonlinear Transformation Layer

To make the information in the nuclearity and relation recognition interact and help each other, we use a dense layer with Relu active function to capture the implicit relationship between V_n and V_r. Then, we concentrate the output vector V_c with V_n and V_r in their own task and feed them to a standard softmax layer for recognition like in the local model. Besides, for paying more attention to relation recognition with lower performance, we joint learn for nuclearity recognition and relation recognition with weighting the sum of their losses, as Eq.8, where λ =0.8.

$$Loss = \lambda Loss_r + (1 - \lambda)\ Loss_n \tag{8}$$

4 Experimentation

To verify the performance of the proposed model, we conduct a set of experiments. We seek to answer the following research questions: (1). How does the STGSN perform on macro Chinese discourse nuclearity recognition and relation recognition? (2). How do nuclearity recognition and relation recognition interact with each other on macro Chinese discourse?

4.1 Experimental Setup

We evaluate our model on MCDTB [9], a Chinese macro discourse corpus. It contains 720 news from CTB 8.0 and annotated RST style discourse tree in each news. Following RST-DT and CDTB, MCDTB divides nuclearity into Nucleus-Satellite (NS), Satellite-Nucleus (SN), and Nucleus-Nucleus (NN), and it removes *Transition* and has three categories (*Elaboration*, *Causality*, *and Coordination*) and 15 types of relations.

To ensure the objectivity, we use five-fold cross-validation to experiment. In the processing of each article, we transform the non-binary trees of the original data into the left-binary trees and then extract their nuclearity and relation. To solve the problem of too few samples, we use topic sentences of discourse units as their semantic representation for re-sampling and get 6530 samples finally.

Table 1. The distribution of nuclearity and relation.

Nuclearity			Relation		
NS	SN	NN	Elaboration	Causality	Coordination
4060	160	2130	2406	828	3296

The distribution of nuclearity and relation is shown in Table 1. We use micro-averaged F1-score (Mic-F1) and include the macro-averaged F1-score (Mac-F1) to emphasize the performance of infrequent types. Considering the fewer samples, we use smaller hyper parameters to adjust parameters on the verification set. The key hyper parameters are as follows: lstmsize: 32, densesize: 64, batchsize: 64, epoch: 10, embeddingdim: 300, maxlength: 300, dropout: 0.2.

4.2 Experimental Results

To answer the research question (1), we compare the performance of our models with the following representative baselines. **BiLSTM [T]**: In this baseline, just like in local model based on STGSN, we apply an attention layer following a BiLSTM network on a sequence of word embeddings belonging to a discourse unit or a title. Then we concentrate three parts (two discourse units and a title) and feed them into a Multi-Layer Perceptron (MLP) for classification. **BiLSTM [S+T]**: This method is similar to **BiLSTM [T]** except that we add a discourse unit's structure information as a feature.

We additionally use two traditional machine learning models: Jiang et al. [8]'s Topic Similarity Model (**TSM**) for nuclearity recognition and Zhou et al. [21]'s Macro Semantics Representation Model (**MSRM**) for relation recognition as other baselines. In MSRM, we exclude some features (the depth of a discourse unit and the number of child node a discourse unit containing before binary processing) because these features cannot be extracted if we want to build a discourse tree from raw data.

Table 2. The performance of each model in nuclearity and relation recognition.

Models	Nuclearity		Relation	
	Mic-F1	Mac-F1	Mic-F1	Mac-F1
BiLSTM [T]	66.25	43.22	54.52	48.38
BiLSTM [S+T]	82.09	55.41	65.15	51.49
STGSN (Local)	**82.90**	**56.08**	66.45	56.81
Joint Model	81.95	55.23	**67.63**	**57.87**
TSM	82.41	55.73	-	-
MSRM	-	-	66.29	51.51

Table 2 shows the comparison results on the test set about nuclearity recognition and relation recognition separately. From Table 2, we make the following observations:

There is a great gap between the performance of BiLSTM [S+T] with structure information and BiLSTM [T] without it. It indicates that structural features are very important for nuclearity and relation recognition.

STGSN (Local) that we proposed outperforms all other single models including neural network and traditional machine learning model. This observation shows that STGSN is effectively rich for nuclearity and relation recognition with changing the representation of a discourse unit by its position and the topic. Compared with concentrating varieties of features simply, STGSN can grasp the meaning of each discourse unit more accurately, thus improving the accuracy of fewer samples type (Causality) recognition (See in Table 4).

The Joint Model recognizes relation better while has a slight degradation in nuclearity recognition. This is related to the joint distribution of the nuclearity and relation in the corpus. (Discussed in section 4.3).

Table 3. The performance of each model in various nuclearity.

Models	NS			NN		
	P	R	F1	P	R	F1
BiLSTM [T]	72.65	79.94	76.06	57.08	50.80	53.60
BiLSTM [S+T]	90.42	83.20	86.66	72.57	87.77	79.45
STGSN (Local)	91.98	82.95	**87.23**	73.14	90.80	**81.02**
Joint Model	87.59	**86.67**	87.06	**75.47**	82.20	78.64
TSM	**92.41**	81.79	86.78	71.90	**91.28**	80.43

Table 4. The performance of each model in various relations.

Models	Elaboration			Causality			Coordination		
	P	R	F1	P	R	F1	P	R	F1
BiLSTM [T]	56.82	49.68	52.79	49.09	3.05	5.23	59.92	79.27	68.17
BiLSTM[S+T]	64.63	**68.81**	66.53	14.04	1.86	3.25	73.45	87.29	79.70
STGSN(Local)	66.12	65.41	65.59	37.95	12.79	18.72	73.22	85.86	78.99
Joint Model	**67.65**	62.41	64.72	37.89	**25.55**	**29.55**	**74.79**	84.50	79.32
MSRM	66.45	68.16	**67.30**	**62.50**	3.62	6.85	73.29	**89.26**	**80.49**

4.3 Analysis and Discuss

In particular, STGSN (Local) significantly improves 5.32% Mac-F1 for relation recognition while only gets an improvement of 0.67% Mac-F1 for nuclearity recognition. To explore why it comes further, we make the statistic of each model's performance in different nuclearity and relation recognition as shown in Table 3 and Table 4. In relation recognition, the improvement of STGSN (Local) is mainly originated from recognizing the type that has fewer samples better. However, in nuclearity recognition, the STGSN (Local) does not significantly improve due to the number of SN (as shown in Table 1) is too small.

To figure out the research question (2), we have calculated the matrix of nuclearity and relation in the corpus, STGSN (Local)'s and Joint Model's predictions respectively, as shown in Table 5. We make two key observations based on Table 5. On the one hand, the NS-Elaboration (Ela.), NN-Coordination (Coo.) and NS-Coordination (Coo.) are the most types and our model recognized well as we expected. On the other hand, for STGSN (Local), there are many mistakes mainly suffer from two sides: recognizing more samples as NN-Elaboration (Ela.), and too many samples belong to NS-Causality (Cau.) are not recognized very well. While the Joint Model can handle them better, which reduces the errors from recognizing NN-Elaboration (Ela.) and recognizes more samples belong to NS-Causality (Cau.) correctly with making good use of the potential relationship between nuclearity recognition and relation recognition.

Table 5. The matrix of nuclearity and relation in corpus and our models' predictions.

	Corpus			STGSN (Local)			Joint Model		
	Ela.	Cau.	Coo.	Ela.	Cau.	Coo.	Ela.	Cau.	Coo.
NS	2316	730	1014	2133	321	1208	2167	523	1293
SN	72	58	30	0	0	0	0	0	0
NN	18	40	2252	242	10	2616	54	3	2490

5 Conclusions

In this work, we propose a joint model of macro Chinese discourse nuclearity and relation recognition based on the Structure and Topic Gated Semantic Network. On the one hand, we propose a Structure and Topic Gated Semantic Network (STGSN) instead of simply connecting semantic features with structure and topic features. On the other hand, we build a joint model of nuclearity recognition with relation recognition and explore the implicit relationship between them to improve recognition performance. Experimental results on the MCDTB corpus show that our model achieves the best performance. Our future work will focus on how to build an end-to-end macro discourse analysis system for helping other NLP tasks.

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. 61836007, 61773276 and 61472354.

References

1. Carlson, L., Marcu, D., Okurowski, M.E.: Building a discourse-tagged corpus in the framework of rhetorical structure theory. In: Current and new directions in discourse and dialogue, pp. 85-112. Springer (2003).
2. Chu, X., Jiang, F., Zhou, Y., Zhou, G., Zhu, Q.: Joint modeling of structure identification and nuclearity recognition in macro Chinese discourse treebank. In: Proceedings of the 27th International Conference on Computational Linguistics. pp. 536–546 (2018).
3. Chu, X., Zhu, Q., Zhou, G.: Discourse primary-secondary relationships in natural language processing. Jisuanji Xuebao/Chinese Journal of Computers 40, 842–860(04 2017). https://doi.org/10.11897/SP.J.1016.2017.00842.

4. Dauphin, Y.N., Fan, A., Auli, M., Grangier, D.: Language modeling with gated convolutional networks. In: Proceedings of the 34th International Conference on Machine Learning-Volume 70. pp. 933–941. JMLR. org (2017).
5. Feng, V.W., Hirst, G.: A linear-time bottom-up discourse parser with constraints and post-editing. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). vol. 1, pp. 511–521 (2014).
6. Hernault, H., Prendinger, H., Ishizuka, M., et al.: HILDA: A discourse parser using support vector machine classification. Dialogue & Discourse 1(3) (2010).
7. Jia, Y., Ye, Y., Feng, Y., Lai, Y., Yan, R., Zhao, D.: Modeling discourse cohesion for discourse parsing via memory network. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). pp. 438–443 (2018).
8. Jiang, F., Chu, X., Xu, S., Li, P., Zhu, Q.: A macro discourse primary and secondary relation recognition method based on topic similarity. Journal of Chinese Information Processing 32(1), 43–50 (2018).
9. Jiang, F., Xu, S., Chu, X., Li, P., Zhu, Q., Zhou, G.: Mcdtb: A macro-level Chinese discourse treebank. In: Proceedings of the 27th International Conference on Computational Linguistics. pp. 3493–3504 (2018).
10. Joty, S., Carenini, G., Ng, R.: 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. pp. 904–915 (2012).
11. Joty, S., Carenini, G., Ng, R., Mehdad, Y.: Combining intra-and multi-sentential rhetorical parsing for document-level discourse analysis. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). vol. 1, pp. 486–496 (2013).
12. Kong, F., Zhou, G.: A cdt-styled end-to-end Chinese discourse parser. ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP) 16(4), 26 (2017).
13. Li, Y., Kong, F., Zhou, G., et al.: Building Chinese discourse corpus with connective-driven dependency tree structure. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 2105–2114 (2014).
14. Mann, W.C., Thompson, S.A.: Rhetorical structure theory: A theory of text organization. University of Southern California, Information Sciences Institute (1987).
15. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems. pp. 3111–3119 (2013).
16. Rocktäschel, T., Grefenstette, E., Hermann, K.M., Kočiskỳ, T., Blunsom, P.: Reasoning about entailment with neural attention. arXiv preprint arXiv:1509.06664 (2015).
17. Sporleder, C., Lascarides, A.: Combining hierarchical clustering and machine learning to predict high-level discourse structure. In: Proceedings of the 20th international conference on Computational Linguistics. p. 43. Association for Computational Linguistics (2004).
18. Sun, C., Kong, F.: A transition-based framework for Chinese discourse structure parsing. Journal of Chinese Information Processing 32(12), 48 (2018).
19. Van Dijk, T.A.: Narrative macro-structures. PTL: A journal for descriptive poetics and theory of literature 1, 547–568 (1976).
20. Wang, Y., Li, S., Wang, H.: A two-stage parsing method for text-level discourse analysis. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). pp. 184–188 (2017).
21. Zhou, Y., Chu, X., Zhu, Q., Jiang, F., Li, P.: Macro discourse relation classification based on macro semantics representation. Journal of Chinese Information Processing 33(3), 1–7 (2019).