Chinese Discourse Parsing on Hierarchical Topic Graphs

Weihao Liu, Yaxin Fan, Xiaomin Chu, Peifeng Li and Qiaoming Zhu[✉]

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

Abstract. Discourse parsing aims to help understand the structure and semantics of discourse by mining the intrinsic structured information of the text. Most existing methods lack guidance from topic information in modeling discourse units, resulting in inconsistencies in semantic modeling at various levels. Therefore, we propose a Chinese discourse parsing method on hierarchical topic graphs, interacting with topic information and textual semantic information at different levels. In particular, we use GPT-4 to generate topic information at different levels. Then, we construct the topic information into a three-level hierarchical topic graph by referring to the original discourse unit division, allowing the core information at different levels to merge. The experiments on both Chinese UCDTB and English RST-DT demonstrate the effectiveness of our proposed method.

Keywords: Discourse Parsing, Topic Information, Hierarchical Topic Graph.

1 Introduction

Discourse parsing aims to uncover the structural information and semantic logic between discourse units, thereby facilitating the understanding of the structure and semantics of the whole discourse.

Rhetorical Structure Theory (RST) [1] is one of the most influential theories in discourse parsing, representing documents as a hierarchical tree structure called Discourse Trees (DTs). Most discourse parsing research concentrates on two levels: sentence-level and paragraph-level. The former investigates the relations between sentences or clauses; the latter examines the relations between paragraphs. Fig. 1 is an example of a complete discourse tree, where the leaf nodes at each level are referred to as Elementary Discourse Units (EDUs). The paragraph-level EDUs are the natural paragraphs of the text; the sentence-level EDUs are sentences or clauses. Adjacent leaf nodes are connected through nuclearity and relation labels, forming higher-level Discourse Units (DUs). Eventually, a document can form a complete DT. The relation labels represent the discourse relation between connected DUs, while the nuclearity labels describe the importance of the DUs. Finally, a document can form a complete DT.

Some studies [2,3] have divided discourse into multiple parsing levels and achieved certain success. However, the modeling of discourse units at each level lacks guidance from higher-level topic information. Jiang et al. [4] point out that there are semantic

logical relations between the title, paragraphs, and sentences of each article. The paragraphs throughout the text are closely linked to the theme of the discourse, and the sentences within each paragraph are also closely connected to the paragraph's theme. Thus, how to extract topic information from various levels and integrate it to enhance discourse parsing from a global perspective poses a challenge.
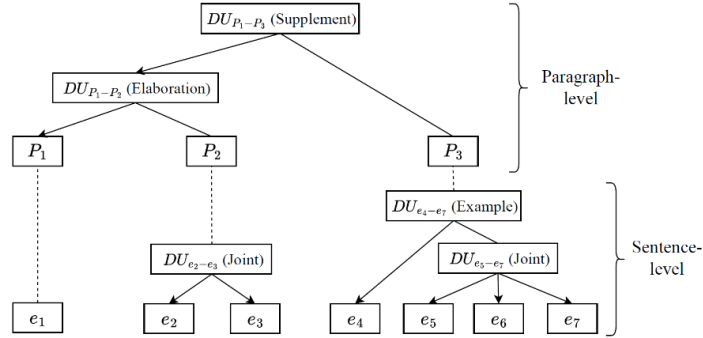

Fig. 1. Example of a discourse tree, where e and P represent the EDUs at the paragraph-level and sentence-level, respectively. In the brackets of a DU, the discourse relation between its child nodes is annotated (e.g., Elaboration and Supplement). The directed edge indicates that the node is a nucleus and the undirected edge indicates that the node is a satellite.

To address the above challenge, we propose a Chinese discourse Parser on Hierarchical Topic Graphs (HTGParser). For model structure, we take the MGIM of Liu et al. [3] as our basic model and incorporate topic information. For each article, we first utilize GPT-4 [5] to generate topic information at different levels, including discourse topic, paragraph topics, and sentence-level EDU keywords, to extract the core semantics from lengthy texts. Then, concerning the original discourse unit division, the topic information is constructed as a three-level hierarchical topic graph, integrating core information across different levels and exploring implicit relations between topics. Finally, Syn-LSTM [6] is utilized to enhance the whole discourse parsing ability by fusing the topic information with textual semantic information at each level. Experimental results on both Chinese UCDTB and English RST-DT demonstrate that our HTGParser outperforms the state-of-the-art baselines.

2   Related Work

Previous discourse parsing methods can be divided into three categories: sentence-level discourse parsing, paragraph-level discourse parsing, and multi-level discourse parsing.

Sentence-level discourse parsing Sentence-level discourse parsing uses clauses or sentences as EDUs. In English, RST-DT [7] is one of the most popular corpora. Hernault et al. [8] employed Support Vector Machines (SVM) as the identification model, constructing discourse trees bottom-up, and realized the first complete discourse parser HILDA. Yu et al. [9] proposed a second-stage EDU-level pre-training method to alleviate the mismatch between the model pre-training and the target task on the elementary

processing unit. In Chinese, CDTB [10] is a corpus annotated at the sentence level. Kong et al. [11] proposed an end-to-end Chinese discourse parser that outputs a complete sentence-level discourse tree. Zhang et al. [12] cast discourse parsing as a recursive split point ranking task and constructed discourse trees through a pointer network. On this basis, Zhang et al. [13] proposed an adversarial learning approach for the lack of global information in the split point ranking task and achieved SOTA performance.

Paragraph-level discourse parsing Paragraph-level discourse parsing uses paragraphs as EDUs. Sporleder et al. [14] studied paragraph-level discourse parsing on RST-DT, using a bottom-up approach to predict discourse structure between paragraphs. MCDTB [15] is the first corpus annotated with information on paragraph-level Chinese discourse structure. Jiang et al. [16] introduced a topic segmentation mechanism to divide articles into several parts, then independently constructed discourse subtrees within each part, and finally merged the subtrees to form a complete discourse tree. Fan et al. [17] enhanced the semantic representation of discourse units and semantic interactions between them by constructing internal topic graphs and interaction topic graphs, achieving SOTA performance.

Multi-level discourse parsing Several studies have divided the document into multiple levels for parsing. Based on HILDA, Feng et al. [18] divided discourse into two levels (intra-sentence and inter-sentence) and used rich linguistic features for discourse parsing. Joty et al. [19] also divided the discourse into two levels, employing the dynamic conditional random field for parsing. Kobayashi et al. [20] adopted a more detailed division, segmenting discourse structure parsing into three levels. The discourse trees were constructed top-down within each level. UCDTB [2] is the first unified corpus on Chinese discourse parsing. Liu et al. [2] proposed a unified document-level Chinese discourse parser that utilizes sentence-level information to assist with paragraph-level parsing. On this basis, Liu et al. [3] proposed a multi-granularity interaction method that utilized the interaction between different levels to facilitate discourse parsing, which achieved SOTA performance.

3   Basic Model: MGIM

Since our work is based on MGIM [3], we first introduce MGIM in this section. The basic model MGIM consists of five components: 1) a sentence-level parser, which constructs sentence-level DTs; 2) a paragraph-level parser, which constructs paragraph-level DTs, with a structure identical to the sentence-level parser; 3) Structure-Aware Graph Attention Network (SAGAT), which comprehensively represents structural and semantic information in DTs by combining distance and relation features between nodes; 4) Graph Contrastive Learning (GCL), which allows the model to distinguish the sentence-level DT of prediction errors and reduce error propagation; 5) Discourse Functional Pragmatics Recognition Auxiliary Task (DFPR), which guides the construction of sentence-level DTs from a paragraph-level perspective.

SAGAT takes sentence-level DT as its input and produces a dependency graph as its output. To maintain the tree structure's hierarchical information, SAGAT assigns two attributes to each edge: relation and distance. Meanwhile, GCL focuses on learning the

representation of the graph by maximizing the mutual information between the node representations and the overall representation of the graph. Finally, the sentence-level embedding G and the loss of graph contrastive learning L_g is obtained.

Discourse functional pragmatics aims to analyze the roles paragraphs undertake and fulfill within a document from a broader viewpoint. This task is to predict the discourse functional pragmatics at the paragraph level by leveraging the sentence-level graph embedding derived from SAGAT and obtain the loss L_f for DFPR.

4   Model

Our model HTGParser is based on MGIM. As shown in Fig. 2, the architecture of our HTGParser consists of five components: 1) Topic Information Construction Module, which uses GPT-4 to generate topic information; 2) Hierarchical Topic Graph (HTGraph), concerning the original discourse unit division, use a graph attention network to fuse the obtained topic information to construct a hierarchical topic graph; 3) Sentence-level Parser, which constructs sentence-level DTs. Based on the sentence-level parser in the MGIM model, we added Syn-LSTM for fusing topic information and textual semantic information at each level, and included the SAGAT and GCL modules; 4) Paragraph-level Parser, which adds Syn-LSTM to the paragraph-level parser in the MGIM model, is used to construct the discourse tree at the paragraph level; 5) Auxiliary Task, which uses the sentence-level embedding obtained from SAGAT to predict paragraph-level discourse functional pragmatics.
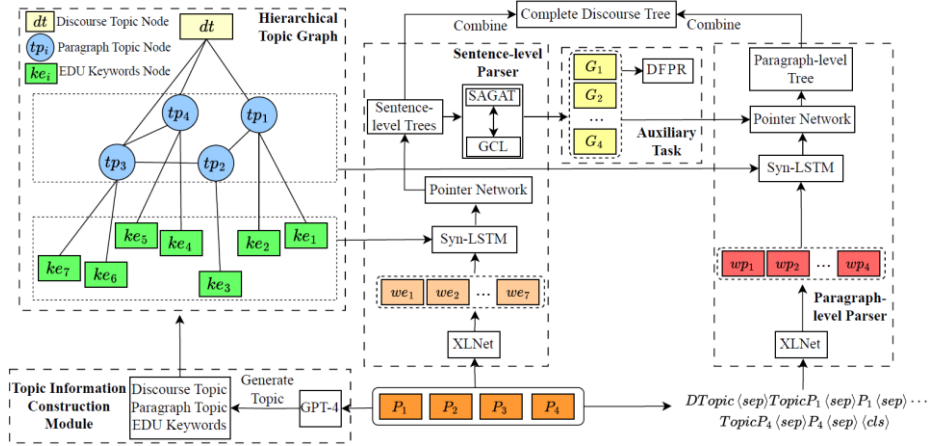

Fig. 2. The architecture of the HTGParser.

4.1   Topic Information Construction Module

We use GPT-4 to generate topic information to extract the core semantics of each level. The process of constructing topic information is shown in the left half of Fig. 3.

Formally, assume an article $Para = \{P_1, P_2, ..., P_n\}$ containing n paragraphs and m sentence-level EDUs, and for the discourse topic, the template we adopt for input to GPT-4 is "Describe the article with a short title", generate a discourse topic $DTopic$ for each article. For each paragraph, we use a template that "Describe the paragraph with a short title" to get the paragraph topic $TOPICP = \{TopicP_1, TopicP_2, ..., TopicP_n\}$. For each sentence-level EDU, we use the template "Describe the sentence with a keyword" to get the EDU keyword $WKE = \{wke_1, wke_2, ..., wke_m\}$. Eventually, one discourse topic, n paragraph topics, and m EDU keywords were generated for this article.

4.2   Hierarchical Topic Graph

The construction process of the hierarchical topic graph is shown in the right half of Fig.3. It references the original division of discourse units, uses GAT [21] to encode topic information, allowing information from different levels to influence each other.

$$S = DTopic\langle sep \rangle TopicP_1 \langle sep \rangle ... TopicP_n \langle sep \rangle P_n \langle sep \rangle \langle cls \rangle \tag{1}$$

Then, we feed the input S to XLNet for encoding and take out the vector at the $\langle sep \rangle$ position to get the semantic representation of paragraph text WP, the semantic representation of paragraph topics TP and the discourse topic representation dt. For the sentence-level EDU keyword sequence WKE, we use GloVe to encode it and obtain the semantic representation KE of EDU keywords.
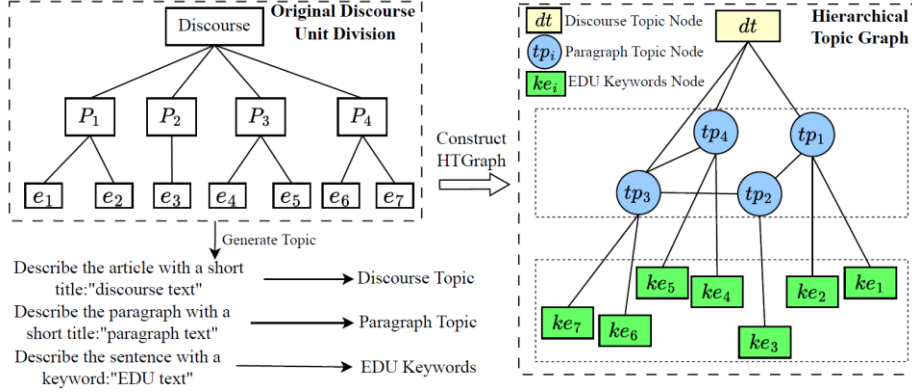

Fig. 3. Hierarchical topic graph construction process.

We refer to the original division of discourse units and construct a hierarchical topic graph. The upper layer is the discourse topic node, which introduces the overall topic information. The middle layer comprises paragraph topic nodes connected to the upper-level topic nodes. Additionally, each paragraph node is also connected to its neighboring nodes. The bottom layer consists of sentence-level EDU keyword nodes, each connected only to the paragraph node. The nodes in the hierarchical topic graph include discourse topic nodes dt, paragraph topic nodes TP, and EDU keywords nodes KE. Formally, let r_i represent the ith node in the graph, with its update as follows.

$$\alpha_{ij} = \frac{exp\big(\sigma(a^T[Wr_i\|Wr_j])\big)}{\sum_{k\in\aleph_i} exp\big(\sigma(a^T[Wr_i\|Wr_k])\big)} \tag{2}$$

$$h_i = \sigma\big(\sum_{j\in\aleph_i} \alpha_{ij}\, Wr_j\big) \tag{3}$$

where a and W are trainable parameters. σ denotes the activation function LeakyReLU, and $\aleph_i$ denotes the neighbor nodes of node i. h_i denotes the state of the ith node. We use HTP to represent the paragraph topic representations, and HKE to denote the sentence-level EDU keywords representations.

4.3   Sentence-level Parser

Given a sentence-level EDU sequence $E = \{e_1, e_2, ..., e_n\}$, where m denotes the number of sentence-lvel EDUs. The sentence-level parser first uses XLNet to encode the sequence of sentence-level EDUs, then extracts the vector at the $\langle sep \rangle$ position as the representation of the discourse unit, and finally obtains the sentence-level EDUs semantic representation $WE = \{we_1, we_2, ..., we_n\}$. To integrate the EDU keyword representation HKE with the EDU semantic representation WE, we introduce Syn-LSTM [6]. Syn-LSTM can model contextual information, computed as follows.

$$c_{hke_t} = tanh\big(W^{(k)}hke_t + U^{(k)}u_{t-1} + b_k\big) \tag{4}$$

$$c_{wke_t} = tanh\big(W^{(p)}we_t + U^{(p)}u_{t-1} + b_p\big) \tag{5}$$

$$c_t = f_t\odot c_{t-1} + i_{1_t}\odot c_{hke_t} + i_{2_t}\odot c_{wke_t} \tag{6}$$

$$u_t = o_t\odot tanh(c_t) \tag{7}$$

where $f_t, o_t, i_{1_t}, i_{2_t}$ are forget gate, output gate and two input gates. c_t denote the current cell states. u_{t-1} is the former hidden state. W, U, b are learnable parameters.

After obtaining the fused vector representation u_i, we input it into the pointer network to construct the sentence-level DT. The SAGAT and GCL modules are then used to obtain an overall representation G of each sentence-level DT.

4.4   Paragraph-level Parser

For the paragraph level, we first obtain the XLNet input as shown in Eq.1. Then, the input is fed into XLNet for obtaining the paragraph text semantic representation WP. Finally, WP and the paragraph topic node representation HTP are fed into the Syn-LSTM of the paragraph-level parser for fusion in the same way.

4.5   Model Training

We use Negative Log Likelihood loss (NLL loss) for DT parsing. The split point prediction loss $L_{\theta_{Sp}}$ for the paragraph-level parser and the split point prediction loss $L_{\theta_{Ss}}$ for the sentence-level parser can be obtained, as shown in Eq.8.

$$L_{\theta_{Ss}} = -\sum_{i=1}^{batch} \sum_{t=1}^{T} log P_{\theta_{Ss}}(y_t|y_{<t}, X) \qquad (8)$$

where $y_{<t}$ is the discourse units constructed by the decoder before the t-th step and X denotes the discourse units sequence.

Like Zhang et al. [13], we use one classifier to classify nuclearity and relation, obtaining the nuclearity and relation prediction loss, which we call N-R prediction loss. We use cross-entropy loss to obtain the paragraph-level N-R prediction loss $L_{\theta_{N-Rp}}$ and the sentence-level N-R prediction loss $L_{\theta_{N-Rs}}$. The DT parsing loss L_p can be obtained by summing the split point prediction loss and the N-R prediction loss as follows.

$$L_p = L_{\theta_{Ss}} + L_{\theta_{N-Rs}} + L_{\theta_{Sp}} + L_{\theta_{N-Rp}} \qquad (9)$$

Finally, the DT parsing loss L_p, graph contrastive learning loss L_g and discourse functional pragmatics recognition loss L_f are weighted and summed together to obtain the final model loss L. The final loss function is shown as follows.

$$L = \alpha_g L_g + \alpha_f L_f + \alpha_p L_p \qquad (10)$$

5   Experimentation

5.1   Datasets and Experimental Settings

Our model HTGParser is primarily evaluated on Unified Chinese Discourse TreeBank (UCDTB). Following previous work [2], we transform the non-binary tree of the original data into the right-binary tree. Finally, we report the micro-averaged F_1 score for span attachments in discourse tree construction (Span), span attachments with nuclearity (Nuclearity) and span attachments with rhetorical relation (Relation).

In both the sentence-level and paragraph-level parsers, we use the same XLNet-base for encoding, with a hidden layer dimension of 768, and employ AdamW to optimize the model. The dimension of GloVe encoding is set to 300. The number of layers in SAGAT is set to 2 and the dimensions of distance embedding and relation embedding in SAGAT are 128 and 768, respectively. The learning rate is set to 2e-5, batch size to 1, dropout rate to 0.5, and the number of training epochs is set to 50. α_f, α_g and α_p are set to 0.1, 0.1 and 0.8 respectively.

5.2   Baselines

To evaluate the performance of our HTGParser, we compared it with the following three types of baselines: Paragraph-level Chinese discourse parser, Sentence-level Chinese discourse parser and Multi-level Chinese discourse parser.

Paragraph-level Chinese discourse parser: 1) **MDParser-TS** [16]: It proposes a hierarchical method for constructing discourse structure trees based on topic segmentation; 2) **DGCNParser** [17]: It proposes a method that utilizes GCN to construct internal topic graphs and interactive topic graphs, which achieved the SOTA performance on MCDTB.

Sentence-level Chinese discourse parser: 3)**Top-DownParser** [12]: It proposes a method for casting discourse parsing as a recursive split point ranking task, using a pointer network to construct discourse structure trees in a top-down manner; 4) **AdverParser** [13]: It proposes an adversarial learning strategy to learn the relation between gold and fake tree diagrams based on Top-DownParser, which achieved the SOTA performance on CDTB.

Multi-level Chinese discourse parser: 5) **DCDParser** [2]: It proposes a method that utilizes sentence-level information to assist in paragraph-level parsing. We use **DCDParser(S)** to denote the method of embedding structure information and **DCDParser(S-N)** to denote the method of embedding structure and nuclearity information; 6) **MGIM** [3]: It proposes an approach that leverages the interaction between the paragraph and sentence levels , achieving SOTA performance on UCDTB.

5.3   Experimental Results

Table 1 shows the performance comparison between our HTGParser and all the baselines. It can be seen that our HTGParser outperforms all baselines. The model that performed best on Relation, DCDParser, utilized two XLNets, while our HTGParser used only one XLNet and achieved better performance with fewer parameters, demonstrating the method's effectiveness. Compared to the basic model MGIM, HTGParser enhances multi-level interaction by adding hierarchical topic information, allowing discourse units to focus on global information and receive guidance from upper-level thematic information. The experimental results show that HTGParser strengthens the core semantics of each discourse unit without increasing the number of model parameters while enabling them to focus on the global core information.

Table 1. The Performance Comparison on Chinese UCDTB.

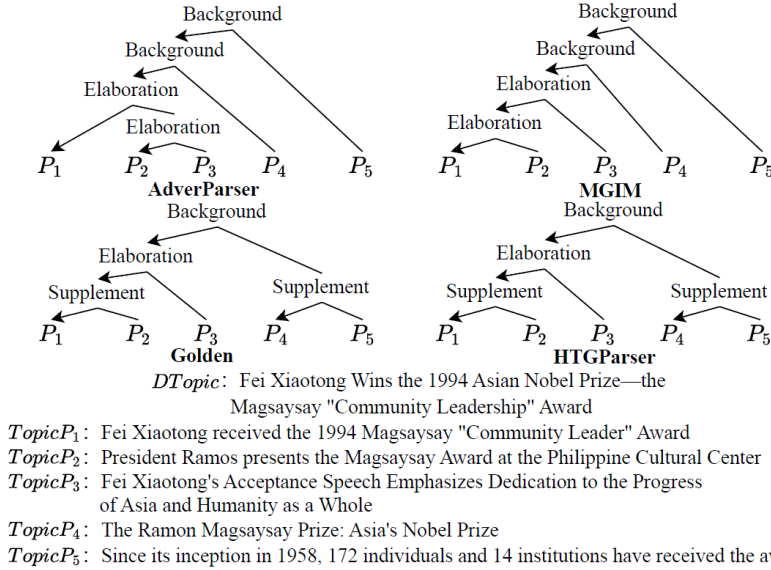| Model	Span	Nuclearity	Relation
DGCNParser	78.69	54.51	49.06
MDParser-TS	79.19	54.71	48.96
Top-DownParser	80.08	53.72	47.08
AdverParser	82.45	62.93	58.08
DCDParser(S)	85.03	64.32	61.15
DCDParser(S-N)	84.74	66.60	60.75
MGIM	85.13	66.70	59.86
HTGParser	**86.52**	**67.29**	**62.83**

6   Analysis

6.1   Ablation Study

To investigate the impact of each module on the discourse parsing performance in HTGParser, we reported the discourse parsing performance at different levels after removing various modules, as shown in Table 2.

Table 2. Performance of different levels on Chinese UCDTB.

Level	Model	Span	Nuclearity	Relation
Paragraph-level	HTGParser	**76.30**	**65.88**	**53.08**
	-Syn-LSTM	72.51	63.03	48.81
	-HTGraph	73.93	65.40	50.24
Sentence-level	HTGParser	**89.22**	**67.67**	**65.41**
	-Syn-LSTM	88.72	66.41	63.16
	-HTGraph	88.10	67.04	62.41
Overall	HTGParser	**86.52**	**67.29**	**62.83**
	-Syn-LSTM	85.33	65.71	60.16
	-HTGraph	85.13	66.70	59.86

"-Syn-LSTM" indicates that after removing Syn-LSTM, topic and textual semantic information are fused by summation. "-HTGraph" indicates the removal of the hierarchical thematic graph on top of removing Syn-LSTM, making the model structure the same as the basic model MGIM. Compared to MGIM, the overall performance of the model is not significantly improved after removing the Syn-LSTM, with only a slight improvement in span and relation recognition and a decrease in nuclearity recognition. In addition, there is some degradation in performance on the span, nuclearity, and relation recognition at the paragraph level. This suggests that Syn-LSTM can capture dependencies between two representations, which is more effective at the paragraph level with deeper semantic representations. Removing the hierarchical topic graph leads to a lack of global thematic information, which reduces the accuracy of identification at the sentence level, thereby resulting in a decrease in the overall discourse parsing performance.

DTopic: Fei Xiaotong Wins the 1994 Asian Nobel Prize—the Magsaysay "Community Leadership" Award
TopicP$_1$: Fei Xiaotong received the 1994 Magsaysay "Community Leader" Award
TopicP$_2$: President Ramos presents the Magsaysay Award at the Philippine Cultural Center
TopicP$_3$: Fei Xiaotong's Acceptance Speech Emphasizes Dedication to the Progress of Asia and Humanity as a Whole
TopicP$_4$: The Ramon Magsaysay Prize: Asia's Nobel Prize
TopicP$_5$: Since its inception in 1958, 172 individuals and 14 institutions have received the award

Fig. 4. Discourse trees parsed by various models.

6.2   Case Study

Fig. 4 shows an example of a paragraph-level discourse structure tree of an article parsed by our HTGParser, AdverParser, and MGIM, along with the discourse topic and paragraph topics generated by GPT-4. From the results, we can observe that Ad-verParser constructed an inaccurate discourse tree due to not considering multi-level interactions. MGIM makes an error in constructing paragraphs P_4 and P_5, incorrectly assuming that P_4 is the background of the preceding text. Compared to MGIM, HTGParser identifies a closer connection between P_5 and P_4 through their paragraph topics. Thus, these two paragraphs are merged into one discourse unit and form a back-ground relation with the preceding text. Furthermore, both AdverParser and MGIM incorrectly identified the relation between paragraphs P_1 and P_2, whereas HTGParser correctly recognized the relation through the topic information of the paragraphs.

6.3   Experimentation on English RST-DT

To verify the generalization of the proposed model, we also evaluate HTGParser on the English RST-DT. We processed cross-paragraph EDUs in the corpus using the same approach as Kobayashi et al. [20], the upper bound of the sentence-level performance is 95.15%. Following previous work [22], we transform all non-binary trees from the original data into right-binary trees and evaluate our model using the original Parseval. The results are shown in Table 3.

Table 3. The performance comparison on English RST-DT.

Level	Model	Span	Nuclearity	Relation
Paragraph-level	DCDParser(S)	52.21	44.16	31.17
	DCDParser(S-N)	55.06	45.71	29.61
	MGIM	54.81	44.68	29.61
	HTGParser	**57.14**	**46.75**	**31.42**
Sentence-level	DCDParser(S)	81.54	70.83	60.32
	DCDParser(S-N)	82.42	71.40	62.09
	MGIM	82.77	71.14	61.71
	HTGParser	**82.98**	**71.82**	**62.28**
Overall	DynParser	73.10	62.30	51.50
	AdverParser	76.30	65.50	55.60
	Parser-EDUPLM	76.40	66.10	54.50
	DCDParser(S)	76.65	66.38	55.46
	DCDParser(S-N)	77.86	67.11	56.67
	MGIM	78.07	66.70	56.32
	HTGParser	**78.64**	**67.61**	**57.11**

There are five baselines in Table 3 as follows: 1) **DynParser** [23]: It proposed a top-down parser with a dynamic oracle; 2) **AdverParser** [13]: It proposed an adversar-ial learning strategy based on the pointer network; 3) **Parser-EDUPLM** [9]: It pro-posed a second-stage EDU-level pre-training method; 4) **DCDParser** [2]: It proposes

a parser that utilizes sentence-level information to facilitate paragraph-level parsing; 5) **MGIM** [3]: It proposes a parser that utilizes multi-level interactions to achieve SOTA performance on RST-DT.

Table 3 shows that HTGParser outperforms all the baselines in recognition of span, nuclearity, and relation at different levels, proving that our proposed HTGParser is also effective in English discourse parsing. Significantly, HTGParser achieved a substantial improvement in paragraph-level discourse parsing. This is because in RST-DT, the discourse contains a larger number of paragraphs. Integrating upper-level discourse topic with lower-level sentence-level EDU keywords information is more beneficial for longer articles, assisting the model in capturing the core semantics within paragraphs while grasping the central theme of the entire text.

7   Conclusion

In this paper, we propose a Chinese discourse parsing method on hierarchical topic graphs, achieving an interaction between topic information and textual semantic information at different levels. Experimental results show that our proposed model achieves SOTA performance on both Chinese UCDTB and English RST-DT. In the future, we will explore how to utilize multi-modal information to enhance the performance of Chinese discourse parsing.

References

1. Mann, W.C., Thompson, S.A.: Rhetorical Structure Theory: A Theory of Text Organization. University of Southern California, Information Sciences Institute Los Angeles (1987)
2. Liu, W., Jiang, F., Fan, Y., Chu, X., Li, P., Zhu, Q.: A unified document-level Chinese discourse parser on different granularity levels. In: Proceedings of the 2023 International Conference on Document Analysis and Recognition. pp. 286–303 (2023)
3. Liu, W., Fan, Y., Chu, X., Li, P., Zhu, Q.: Discourse parsing on multi-granularity interaction. In: Proceedings of the 2023 International Joint Conference on Neural Networks. pp. 1–8 (2023)
4. Jiang, F., Liu, W., Chu, X., Li, P., Zhu, Q., Li, H.: Advancing topic segmentation and outline generation in chinese texts: The paragraph-level topic representation, corpus, and benchmark. arXiv preprint arXiv:2305.14790 (2023)
5. Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al.: Gpt-4 technical report. arXiv preprint arXiv:2303.08774 (2023)
6. Xu, L., Jie, Z., Lu, W., Bing, L.: Better feature integration for named entity recognition. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 3457–3469 (2021)

7. Carlson, L., Marcu, D., Okurowski, M.E.: Building a discourse-tagged corpus in the framework of rhetorical structure theory. Current and new directions in discourse and dialogue pp. 85–112 (2003)
8. Hernault, H., Prendinger, H., du Verle, D.A., Ishizuka, M.: Hilda: A discourse parser using support vector machine classification. Dialogue & Discourse 1(3), 1–33 (2010)
9. Yu, N., Zhang, M., Fu, G., Zhang, M.: Rst discourse parsing with second-stage edu-level pre-training. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. pp. 4269–4280 (2022)
10. Li, Y., Feng, W., Sun, J., Kong, F., Zhou, G.: Building chinese discourse corpus with connective-driven dependency tree structure. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. pp. 2105–2114 (2014)
11. Kong, F., Zhou, G.: A cdt-styled end-to-end chinese discourse parser. ACM Transactions on Asian and Low-Resource Language Information Processing 16(4), 1–17 (2017)
12. Zhang, L., Xing, Y., Kong, F., Li, P., Zhou, G.: A top-down neural architecture towards text-level parsing of discourse rhetorical structure. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 6386–6395 (2020)
13. Zhang, L., Kong, F., Zhou, G.: Adversarial learning for discourse rhetorical structure parsing. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. pp. 3946–3957 (2021)
14. 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. pp. 43–49 (2004)
15. 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)
16. Jiang, F., Fan, Y., Chu, X., Li, P., Zhu, Q., Kong, F.: Hierarchical macro discourse parsing based on topic segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 13152–13160 (2021)
17. Fan, Y., Jiang, F., Chu, X., Li, P., Zhu, Q.: Chinese macro discourse parsing on dependency graph convolutional network. In: Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing. pp. 15–26 (2021)
18. Feng, V.W., Hirst, G.: Text-level discourse parsing with rich linguistic features. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics. pp. 60–68 (2012)
19. 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. pp. 486–496 (2013)
20. Kobayashi, N., Hirao, T., Kamigaito, H., Okumura, M., Nagata, M.: Top-down rst parsing utilizing granularity levels in documents. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 8099–8106 (2020)
21. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. In: Proceedings of the 6th International Conference on Learning Representations. pp. 1–12 (2018)
22. Sagae, K., Lavie, A.: A classifier-based parser with linear run-time complexity. In: Proceedings of the Ninth International Workshop on Parsing Technology. pp. 125–132 (2005)
23. Koto, F., Lau, J.H., Baldwin, T.: Top-down discourse parsing via sequence labelling. arXiv preprint arXiv:2102.02080 (2021)