Front. Comput. Sci., 2022, 16(5): 165329
https://doi.org/10.1007/s11704-021-0500-z

REVIEW ARTICLE

A survey of discourse parsing

Jiaqi LI[1], Ming LIU[1,2], Bing QIN (✉)[1,2], Ting LIU[1,2]

1 School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China
2 Peng Cheng Laboratory, Shenzhen 518052, China

Abstract Discourse parsing is an important research area in natural language processing (NLP), which aims to parse the discourse structure of coherent sentences. In this survey, we introduce several different kinds of discourse parsing tasks, mainly including RST-style discourse parsing, PDTB-style discourse parsing, and discourse parsing for multiparty dialogue. For these tasks, we introduce the classical and recent existing methods, especially neural network approaches. After that, we describe the applications of discourse parsing for other NLP tasks, such as machine reading comprehension and sentiment analysis. Finally, we discuss the future trends of the task.

Keywords discourse parsing, discourse structure, RST, PDTB, STAC

1 Introduction

Discourse parsing is a task that can parse discourse structure in text automatically, including identifying discourse structure and labeling discourse relations. As a fundamental task in natural language processing (NLP), discourse parsing has been successfully applied in many other NLP tasks, such as question answering [1], machine reading comprehension [2], sentiment classification [3], language modeling [4], machine translation [5] and text categorization [6].

In this survey, we classify discourse parsing (DP) tasks into three main categories: RST-style DP, PDTB-style DP, and dialogue DP. The overview of discourse parsing is shown in Fig. 1. Among them, RST-style and PDTB-style discourse parsing are tasks for processing passages, but the inputs of dialogue discourse parsing are dialogue utterances. RST-style discourse parser aims to obtain the hierarchical rhetorical tree structure of an input document, but the PDTB discourse parser tries to get a flat discourse structure between sentences of clauses, not a tree. The discourse parser for multiparty dialogues parses the input dialog into a discourse dependency graph, and the discourse relations can exist between non-adjacent utterances which are different from RST-style DP.

Rhetorical Structure Theory (RST) is the theory of representing a document into the tree structure where elementary discourse units (EDUs) are represented as vertexes in the tree [7]. Discourse treebank plays an important role in discourse parsing. Inspired by RST, Rhetorical Structure Theory Discourse Treebank (RST-DT) [8] is released that is a typical English discourse treebank. Instead of the tree structure, [9] adopts graph structure to represent discourse and release Discourse Graphbank on LDC which contains 135 documents (105 documents from AP Newswire and 30 documents from the Wall Street Journal (WSJ)).

Different from RST building the full structure of discourse into an RST tree, PDTB-style discourse parsing mainly focuses on the discourse relations within the local structure between two arguments (Arg1 and Arg2). Penn Discourse Treebank (PDTB) [10] dataset is based on discourse lexical tree (D-LTAG) [11] and is released as a large-scale discourse treebank on LDC. The sense in PDTB is a three-layer hierarchical structure, including classes, types, and subtypes. The relations can be divided into two classes: explicit and non-explicit, considering the absence of connectives. RST-DT and PDTB have promoted most of the research on discourse parsing. One similarity between the two data sets is that they all derive from WSJ, a typical well-written text corpus. Figure 2 shows an example from both RST-DT and PDTB datasets. From Fig. 2, we can find that the RST discourse parser generates a hierarchical discourse tree, but the PDTB discourse parser only detects connectives, arguments, and the sense between arguments.

On the other hand, models that are trained in well-written passages dataset maybe not appropriate for spoken language or dialogue text. Furthermore, there are obvious differences in linguistic properties between passages and dialogue. A passage is a continuous text where there is a discourse relation between every two adjacent sentences. In contrast, there may be no discourse relation between adjacent utterances in a multiparty dialog. Utterances of a multiparty dialog are much less locally coherent than in prose passages. There are research papers that begin to focus on discourse parsing on multi-party dialogue, including handcrafted features based on shallow models [12,13], and deep sequential models [14].

There are two main challenges in discourse parsing.

- The first challenge is the difficulty of detecting discourse structure. Similar to other structure prediction

Received October 9, 2020; accepted March 9, 2021

E-mail: qinb@ir.hit.edu.cn

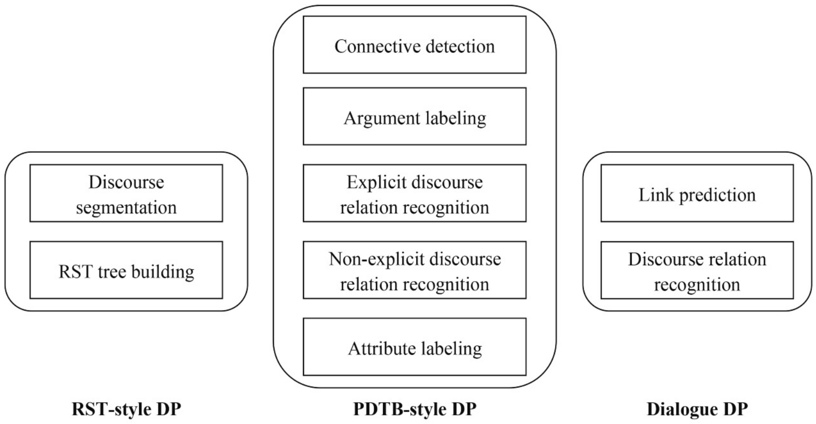

Fig. 1 An overview of discourse parsing (DP) task

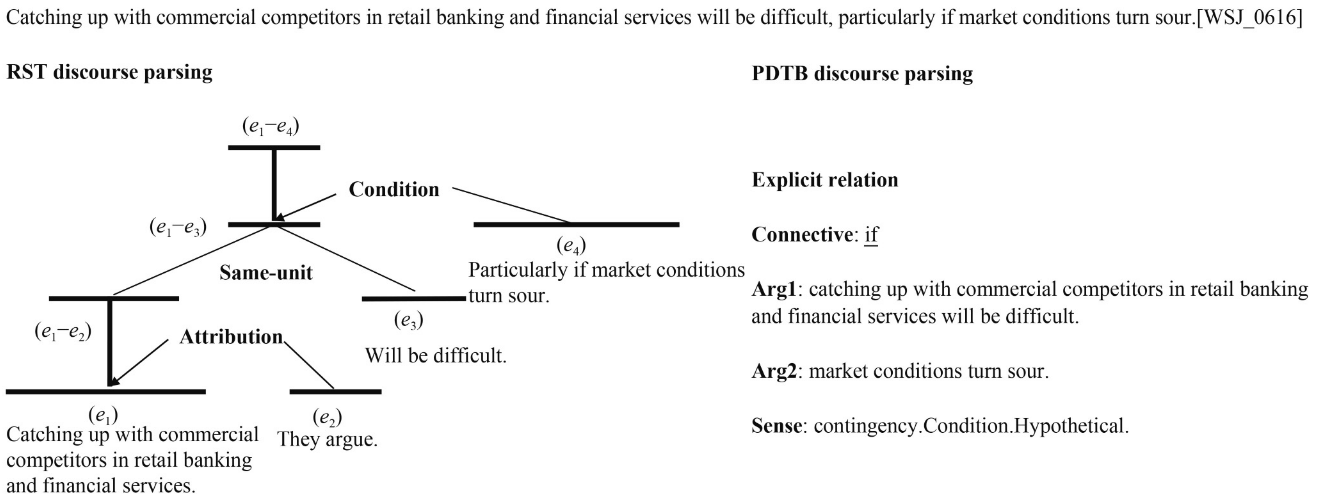

Fig. 2 An example between RST-style discourse parsing and PDTB discourse parsing

tasks, such as dependency parsing, detecting discourse structure is also a challenging task. In detail, detecting spans in RST-DT, labeling arguments in PDTB, and classifying links in STAC are still very difficult and cannot achieve satisfying results.

- The second challenge is classifying the specific discourse relations. Among RST-DT, PDTB, and STAC corpus, it still a challenging task to classify discourse relations given two texts, such as clauses or sentences. There are no indicative connectives in the text in most of the cases where the situations are called implicit discourse relation recognition in PDTB-style discourse parsing. The difficulty of discourse relation recognition is that models can hardly understand the full meanings of given short texts. Pre-trained models have greatly improved the ability to represent texts and successfully applied them to discourse parsing. However, there is a long way to get satisfying results for discourse parsing.

The rest of this paper is organized as follows. First, in Section 2 we introduce the discourse parsing task and available dataset. In Section 3, we introduce existing approaches for different discourse parsing tasks. Next, we introduce applications of discourse parsing in Section 4. Finally, we discuss the future trends of discourse parsing in Section 5.

2 Discourse structure

In this section, we will introduce different kinds of discourse structures and their related datasets, including the RST and RST-DT dataset, D-LTAG and PDTB dataset, and the STAC and Molweni dataset.

2.1 Rhetorical structure theory and RST-DT

Rhetorical Structure Theory (RST) is the framework to represent the structure of a document. RST-style discourse parsing aims to parse a document into a hierarchical tree structure. In the RST tree, leaf nodes represent the *elementary discourse unit* (EDU) which are usually clauses. The inner nodes are called a span that contains two or more adjacent EDUs. An instance of RST is shown in Fig. 3.

In RST, there are two categories of discourse relations: hypotactic (mononuclear) and paratactic (multi-nuclear). An inner node connects two EDU nodes in the mononuclear, more salient node is called *nucleus*, and another node is called *satellite*. In paratactic, all spans are equally salient.

The Rhetorical Structure Theory Discourse Treebank (RST-DT) is the discourse treebank based on RST. RST-DT contains 385 documents from WSJ, and 347 documents for training, and 38 for testing.

RST-style discourse parsing contains the following two tasks: discourse segmentation and tree building. Discourse

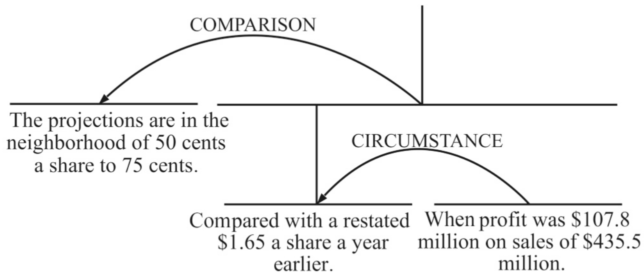

Fig. 3 An example of RST tree structure [15]

segmentation module can segment input text into EDUs, and the tree building module would build RST tree structure given EDUs. Because the performance of the EDU segmentation module has been close to 95%, most of the research of RST-style discourse parsing focus on the RST tree building task where gold EDUs are provided.

2.1.1 Discourse segmentation

EDU segmentation task aims to segment input text into EDUs. EDUs are the minimal discourse unit in RST-style discourse parsing, and EDUs usually are clauses. The annotators label EDUs with the following rules:

1. Clauses that are subjects or objects of the main verb are not treated as EDUs.
2. Clauses that are complements of the main verb are not treated as EDUs.
3. Complements of attribute verbs (speech acts and other cognitive acts) are treated as EDUs.
4. Relative clauses, nominal postmodifiers, or clauses that break up other legitimate EDUs, are treated as embedded discourse units.
5. Phrases that begin with a strong discourse marker, such as *because*, are treated as EDUs.

2.1.2 RST tree building

For the RST tree building task, golden EDUs are already provided and the task aims to construct an RST tree and label rhetorical relations on links. In detail, this task contains the following subtasks: span prediction, nuclearity indication, and relation classification.

Span prediction This subtask can be regarded as a binary classification task that aims to predict the tree structure of input text by classifying whether two EDUs or spans should be merged.

Nuclearity indication As mentioned above, there are two different kinds of nodes in the RST tree for hypotactic relations: *nucleus* and *satellite*. The nuclearity indication task aims to predict the nucleus or satellite given two EDUs or spans.

Relation classification This subtask aims to classify the specific rhetorical relations between given two EDUs or spans.

In RST-DT, there are 78 fine-grained rhetorical relations in total, including 53 mononuclear relations and 23 multi-nuclear relations. All 78 rhetorical relations can be divided into 16 relation classes as shown in Table A1. The definitions of a specific rhetorical relation are based on constraints on the nucleus, constraints on the satellite constraints on the combination of nucleus and satellite, and effect achieved on the text receiver. The sense hierarchy of RST-DT is shown in Table A1 of Appendix.

2.2 Discourse lexicalized tree adjoining grammar and PDTB

Different from rhetorical structure theory, discourse lexicalized tree adjoining grammar (D-LTAG) does not parse a discourse into a tree structure, but detect discourse relations within local text units, such as two clauses in a sentence or two adjacent sentences. Inspired by D-LTAG, PDTB dataset was released and attracted a widespread attention. The latest version of PDTB is the PDTB 3.0 [16].

The first PDTB-style discourse parser contains all modules of PDTB discourse parsing, including the connective classifier, argument labeler, explicit classifier, non-explicit classifier, and attribution span labeler [17]. In the sense hierarchy of PDTB 2.0, there are five categories of discourse relations: Explicit, Implicit, AltLex, Entel, and NoRel. In PDTB 3.0, AltLexC and Hypophora tokens are added. The sense hierarchy of explicit and implicit discourse relations in PDTB 3.0 is shown in Table A2.

Connectives Connectives are important clues for PDTB discourse parsing. If there are connective in the sentences, the discourse relation would be explicit. Otherwise, the relation would be non-explicit discourse relations. For non-explicit discourse parsing, people pay attention to the implicit discourse relation recognition.

Arguments In PDTB discourse parsing, argument labeling is also an important sub-task that aims to detect the boundary of two arguments anchored by the connectives.

Explicit discourse relations This sub-tasks aims to classify the specific discourse relations type. Due to the indication of connectives, explicit discourse relation recognition has been handled by simple machine learning methods [18].

Non-explicit discourse relations Due to the lack of connectives, recognizing non-explicit discourse relations today is still a challenging task. For all non-explicit discourse relation recognition, most people pay attention to classifying implicit discourse relations.

There is an example of implicit discourse relation:

- Arg1: But for the next few months, these boys of summers long past are going to be reveling in an Indian summer of the soul.
- Arg2: Now that the baseball season is officially over, you see, it's time for a new season to begin.
- Sense: Contingency.Cause.Reason.

Attributes PDTB labels attribute spans within discourse relations and annotate the sources, types, scopal polarities, and determinacy of Arg1, Arg2, and the relation.

Most researchers of PDTB-style discourse parsing pay attention to implicit discourse relation recognition task. It is a challenging task to recognize implicit discourse relations. Hong gives three reasons for the low performance of implicit discourse relation: the missing of connectives, the failure of finding effective features, and the data sparseness problem [19].

2.3 Discourse parsing for multiparty dialogue

Most existing research of discourse parsing is about news text. However, models that are trained in written datasets maybe not appropriate for spoken language. Therefore, the annotate scheme was proposed in spoken language, including telephone conversations and broadcast interviews, in the style of PDTB 3.0 and CCR [20]. They explored the differences between discourse relations in written language and spoken language. In written text, the distribution of explicit and implicit discourse relations are almost the same. But in spoken text, the number of explicit relations is almost twice as frequent as implicit relations. Another important finding is that more than 70% of discourse relations between PDTB 3.0 and CCR can be mapped into one other.

Different from previous work on the news or monologue dataset, there is little research focuses on discourse parsing on multi-party dialogue, including handcrafted features based on shallow models [12,13], and deep sequential models [14].

An example in STAC is shown in Fig. 4. The left side of the figure shows a multi-party dialogue, and the right side of the figure provides the ground truth of the dialogue. In the dialogue, there are three interlocutors, including A, B, and C. Three speakers sent six messages in total. Each utterance is an EDU and can be regarded as a vertex in the directed acyclic graph on the right side of the figure. The directed edge between two EDUs is a discourse dependency relation. For example, there is an Elaboration relation from u_1 to u_2. Different from discourse parsing on monologue or passages, there are non-projective relations between non-adjacent EDUs, such as the QAP relation from u_1 to u_3 and the QAP relation from u_4 to u_6.

We formulate the task of discourse parsing on multi-party chat dialogue as follows:

- **Input**: $D=\{u_1,u_2,...,u_n\}$, where D is a multiparty chat dialogue with n utterances. u_i is the ith utterance in the dialogue. Each utterance is regarded as an elementary discourse unit (EDU) in multiparty dialogue discourse parsing.

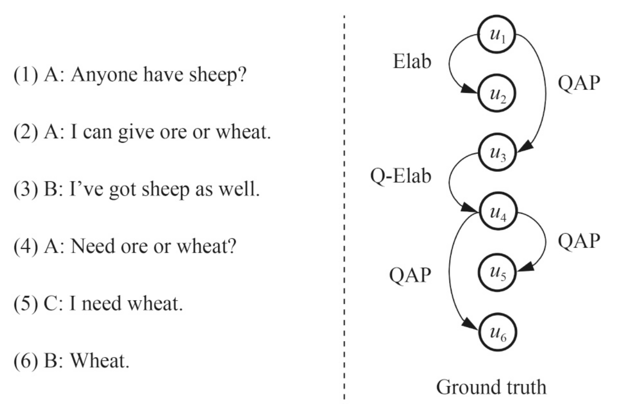

Fig. 4 A dialogue example from the STAC Corpus [14]. In the left part of the figure, there are six utterances from A, B, C three speakers. In the right part, nodes from u_1 to u_6 represent six utterances in the left. The links between nodes represent the discourse re lation, and the label on the link is the discourse relation type. Elab, QAP, Q-Elab are abbreviation for Elaboration, Question-answer Pair, Question-Elaboration

- **Output**: $G(V,E,R)$, where V represents vertex set consists of EDUs and $|V|=n$, and E represents edge set between EDUs, and R represents discourse relations.

The first dataset of discourse parsing for multiparty dialogue is the STAC corpus [21]. The corpus derives from an online game *The Settlers of Catan*. The game *Settlers of Catan* is a multi-party, win-lose game. More details for the STAC corpus are described in [21].

The overview of the STAC and Molweni corpus are shown in Table 1. From Table 1 we can know that there are more than 10K EDUs and relations and most of the EDUs are weakly connected.

The Molweni corpus is another dataset for multiparty dialogue discourse parsing [22]. The Molweni dataset is derived from the large-scale multiparty dialogue Ubuntu Chat Corpus [23]. The name *Molweni* is the plural form of “Hello” in the Xhosa language, representing multiparty dialogue in the same language as *Ubuntu*. The Molweni dataset contains 10, 000 dialogs with 88, 303 utterances and 32, 700 questions including answerable and unanswerable questions. All answerable questions are extractive questions whose answer is a span in the source dialogue. For unanswerable questions, we annotate their plausible answers from the dialogue. Most questions in Molweni are 5W1H questions – Why, What, Who, Where, When, and How. For each dialogue in the corpus, annotators propose three questions and find the answer span (if answerable) in the input dialogue.

2.4 Comparisons

Table 2 compares different discourse parsing task and existing datasets. RST-style datasets contain RST-DT [8], PDTB [10]. Furthermore, for multiparty dialogue discourse parsing, there are STAC [21] and Molweni [22] two datasets.

From Table 2, we can find the sources, theory, and sub-tasks of each dataset for discourse parsing. Furthermore, we list the statistical information for all datasets.

3 Existing methods

In this section, we will introduce existing methods for different styles of discourse parsing, including RST-style, PDTB-style, and multiparty dialogue discourse parsing.

3.1 RST-style discourse parsing

RST-style discourse parsing aims to parse the document into rhetorical tree structures, including discourse segmentation and RST tree building.

3.1.1 Discourse segmentation

There are two main kinds of methods for the discourse segmentation task, including binary classifier and sequence labeling.

The first method is to classify whether a token is the boundary of an EDU by a binary classifier. [24] introduced two probabilistic models to assign a probability for each word

Table 1 The statistic of the STAC and Molweni dataset

	Dialogues	Utterances	Relations
STAC	1, 091	10, 677	11, 348
Molweni	10, 000	88, 303	78, 245

Table 2 Comparisons among popular English discourse treebank

	RST-style	PDTB-style	Multiparty dialogue	DP
Datasets	RST-DT	PDTB	STAC	Molweni
Source	WSJ	WSJ	Settlers of Catan	Ubuntu Corpus
Theory	RST	D-LTAG	SDRT	SDRT
Scale	385 Docs 21, 789 EDUs	2, 159 Docs 40, 600 Relations	1, 091 Dialogue 10, 677 utterances 11, 348 relations	10, 000 Dialogue 88, 303 utterances 78, 245 relations
Annotation	(1) Discourse segmentation. (2) RST tree annotation.	1. Detect discourse connectives. 2. Arguments labeling. 3. Discourse relation recognition (Explicit, Non-Explicit). 4. Attribute labeling.	(1) Detect discourse dependency links. (2) Classify discourse relations.	(1) Detect discourse dependency links. (2) Classify discourse relations.

via incorporating syntactic and lexical features. As we know, [25] first proposed neural network for segmenting discourse into EDUs that trained a multi-layer perceptron binary classifier using lexical and context features. [26] trained a classifier using finite-state and context-free derived features. [27] applied the Dynamic Conditional Random Filed (DCRF) that is a probabilistic discriminative model addressing independence assumptions and sub-optimal limitation of the greedy algorithms.

The second method regards discourse segmentation as a sequence labeling task. The model will label each token to indicate whether the token is the boundary of an EDU. Usually, models assign B or C labels to a token. If the token is the beginning of an EDU, the token will be labeled as B. Otherwise, the token will be labeled as C [28–31]. [32] proposes a BiLSTM-CRF based model that achieves the state-of-the-art on F_1 measure. Compared to the two-pass model [31] and the SPADE model [24], the BiLSTM-CRF based model saves more time and obtain better results.

Models for the discourse segmentation task have achieved more than 95% of F_1 measure which is quite near to human performance (98%) in F_1 measure. Therefore, most RST-style discourse parsing researchers pay more attention to the RST tree building task.

3.1.2 RST tree building

Traditional methods for building RST tree adopt statistical machine learning methods using hand-craft features, such as context surface features and constituent features. [33,34] propose two heuristic rules to convert RST tree building task into discourse dependency parsing task, and [35] proves that the rule in [33] is more useful for text summarization task. [36] proposed two RST parsers that respectively adopt a constituent tree and dependency tree, and both two parsers achieved the state-of-the-art. [37] proposed a dependency perspective on RST discourse parsing and evaluation. They detect the similarities between dependency parsers and shift-reduce constituency parsers. Furthermore, the experiment results prove the effects of dependency parsing for RST discourse parsing. [38] proposed the CODRA model which adopted a binary classifier to detect the boundary of elementary discourse units and two Conditional Random Fields (CRF) to build both intra-sentential and multi-sentential discourse trees.

The first neural-based deep model was proposed for RST-style discourse parsing and outperformed statistical-based methods [15], and other recursive deep models were proposed following closely [39]. [40] proposed an attention-based hierarchical Bi-LSTM model with a tensor-based transformation module to learn more feature interactions. In recent years, there are more transition-based models for RST-style discourse parsing [41–43]. To address the limitation of the amounts of training data in RST-DT, [44] proposed a multi-view and multi-task framework to combine related tasks. Considering the similarities of RST-style discourse parsing among different languages, [45] proposed the first cross-lingual RST discourse parser and achieves the state-of-the-art for English RST parser.

The recent years' results of RST tree building are shown in Table 3. In the table, S, N, R, and F respectively represent Span, Nuclearity, Relation, and Full parser. From Table 3, we can find that most neural network-based models do not perform significant advantages over traditional hand-craft feature-based models. One reason for this phenomenon could be that the scale of RST-DT limits the training of complex neural models. How to effectively train a neural model on a limited RST-DT dataset would be still challenging for RST-style discourse parser researchers.

3.2 PDTB-style discourse parsing

As mentioned, PDTB-style discourse parsing contains several tasks, including connectives detection, argument labeling, and discourse relation recognition, attribute labeling. Most of the research on the PDTB dataset can be divided into classes: explicit discourse parsing and implicit discourse parsing. For explicit discourse parsing, the task aims to detect connectives, label arguments, and recognize explicit discourse relations. For implicit discourse parsing, two arguments are given, the

Table 3 The micro-F_1 score of RST tree building task [42]

Model	S	N	R	F
Feature-based models				
Hayashi et al., 2016	82.6	66.6	54.6	54.3
Surdeanu et al., 2015	82.6	67.1	55.4	54.9
Joty et al., 2015	82.6	68.3	55.8	55.4
Feng and Hirst, 2014	84.3	69.4	56.9	56.2
Neural network-based models				
Braud et al., 2016	79.7	63.6	47.7	47.5
Li et al, 2016	82.2	66.5	51.4	50.6
Braud et al., 2017	81.3	68.1	56.3	56.0
Ji&Eisenstein, 2014	82.0	68.2	57.8	57.6
Yu et al., 2018	85.5	73.1	60.2	59.9
Human	88.3	77.3	65.4	64.7

task is to classify implicit discourse relations.

3.2.1 Explicit discourse parsing

In explicit discourse relation recognition, since the connective can indicate discourse relations, recent methods got good performance. Pitler used an unsupervised method and got a good result only using the connective [46]. Besides, there are some supervised methods to recognize explicit discourse relations. For instance, Pitler used an approach based on some syntactic features related to the connective and got an improvement in explicit discourse relation recognition [46]. To reduce error propagation, a joint learning approach via structured perceptron for explicit discourse parsing was proposed, they got comparable results on relation classification and got an improvement on argument labeling [47].

3.2.2 Implicit discourse parsing

There are mainly three kinds of methods for recognizing implicit discourse relations. The first kind of method is separately modeling two arguments. Early research was mainly based on surface features and statistical machine learning methods [17,18,48–50]. With the success of the neural network, [51,52] respectively propose the recursive and recurrent model to learn the representations of arguments. [53] compares several different representations for implicit discourse classification. When they add features, they get a better result than previously. Ji proposed a novel method for implicit discourse relation classification based on latent variable recurrent neural network [4]. To prove the effect of BERT, [54] tries to prove the next sentence predict task for implicit discourse relation and achieves great improvements on the implicit discourse relation recognition task. Furthermore, [55] adopt the BERT model to represent arguments and focus on the connectives. The BERT-based model achieves the state-of-the-art and gets obvious improvements on *Temporal* and *Comparison* (very few instances on the datasets) two types on PDTB.

The second kind of method for recognizing implicit discourse relations is not only modeling each argument but also model the interactions between two arguments. Many papers have proved the effect of word pairs between two arguments to classify implicit discourse relations [56–58]. To solve the data sparsity and use the word-pair feature, Chen proposed new deep architecture with a gated relevance network (GRN) [59]. [60] proposed a new generative-discriminative framework that utilizes a new method to represent semantically and get a good result. [61] considers the linguistic characteristics including semantic interaction and the cohesion device (topic continuity and attribution) for three important discourse relations: *Comparison*, *Contingency* and *Expansion*. [62] proposed a neural tensor network with a sparse constraint to obtain deeper and more indicative pair patterns. [63] proposed a multi-level argument representation model that learns the representations of character, sub-word, word, sentence, and sentence pair. [64] adopts two factored tensor networks (FTN) to model interactions between two arguments and incorporate topic representations.

The third kinds of method adopt joint learning or multi-task architecture. In 2013, [50] firstly propose a method for implicit discourse relation classification based on multi-task. Inspired by her work, Liu trained a multi-task neural network that only uses PDTB as experimental data but also uses RST-DT and other data in auxiliary tasks [65]. [66] improved the inference of implicit discourse relations via classifying explicit discourse connectives. Different from common methods ignoring implicit discourse connectives, [67] proposed a novel model contains a discourse relation classifier and a sequence-to-sequence model to predict the implicit discourse connectives. [68] incorporated event knowledge and coreference relations into neural discourse parsing. [69] introduces knowledge information from WordNet to help classify discourse relations and proves the effect of knowledge. To deeper integrates the annotation information, [70] proposed a TransS-based method that learns a transition from Arg1 + relation to Arg2.

The latest results on implicit discourse relation recognition for the top four-class classification are shown in Table 4. From Table 4, with introducing the pretrained models for modeling arguments, the performance for detecting *Temporal* and *Comparison* relations has been significantly improved. The classifier for all four relations achieved more than 70% in the F_1 measure.

3.3 Dialogue discourse parsing

The first paper discourse parsing model for multi-party dialogue was proposed in 2015 [12]. As mentioned above, the task aims to parse discourse dependency structure in multi-party chat dialogue. In the paper, they adopted maximum entropy (MaxEnt) using hand-craft features to learn the local distribution. Instead of directly using probabilities from MaxEnt for classifier binary attachment and discourse relations, they used Maximum Spanning Trees (MST) for decoding.

In the paper, the authors adopted these three categories of features, including positional features, lexical features, and parsing features.

- Positional feature: speaker initiated the dialogue, the first utterance of the speaker, in the dialogue, position in dialogue, *distance between EDUs*, and *EDUs have the same speaker*.
- Lexical feature: ends with an exclamation mark, ends with interrogation mark, contains possessive pronouns, contains modal modifiers, contains words in lexicons, contains question words, contains a player's name, contains emoticons, and first and last words.
- Parsing feature: subject lemmas given by syntactic,

Table 4 The performance of implicit discourse relation recognition on PDTB

Model	One-Versus-All			
	Comp.	Cont.	Expa.	Temp.
Rutherford&Xue, 2015	41.0	53.8	69.4	33.3
Lei et al., 2018	43.24	57.82	72.88	29.1
Bai&Zhao., 2018	47.85	54.47	70.6	36.87
Shi&Demberg, 2019	41.83	62.07	69.58	35.72
Dai&Huang, 2019	45.34	51.8	68.5	45.93
Kishimoto et al., 2020	77.28	73.85	73.4	79.41

dependency parsing, and dialogue act according to predict model.

Because MST-based method only can predict tree structure discourse dependency structures, there are 9% in the dataset structure cannot be predicted. To predict non-tree structures in the DAGs, an integer linear programming (ILP) based method was proposed [13]. Besides local distribution between EDUs, the ILP-based method also compute global representation for decoding. They implemented the constraints in ILP as following equations:

$$\sum_{i=1}^{n} h_i = 1,$$

$$\forall j, 1 \leqslant nh_j + \sum_{i=1}^{n} a_{ij} \leqslant n.$$

Different from previous work on discourse parsing on multiparty dialogue, Shi and Huang first adopts the deep sequential model for discourse parsing on multi-party chat dialogue [14]. They also adopt an iterative algorithm to learn the structured representation and highlight the speaker's information in the dialogue. The ablation experiments prove the efforts of structured representation and speak the highlight mechanism. The architecture of their model is shown in Fig. 2. The model learns the dependency structure and discourse relations jointly and alternately. The structured representations are computed as follows:

$$g^S_{i,a} = \begin{cases} 0, & i = 0, \\ GRU_{hl}(g^S_{j,a}, h_i \oplus r_j i), & a_i = a, i > 0, \\ GRU_{gen}(g^S_{j,a}, h_i \oplus r_j i), & a_i \neq a, i > 0, \end{cases} \tag{1}$$

where *hl* and *gen* are *highlighted* and *general* respectively.

The results of the existing models of discourse parsing in the multiparty dialogue on the STAC dataset are shown in Table 5. Similar to the dependency parsing task, we adopt UAS and LAS to represent the performance of models that are short for unlabeled attachment score and labeled attachment score. For multiparty dialogue discourse parsing task, UAS and LAS respectively show the performance of models in identifying discourse structure and both identifying structure and labeling relations.

4 Applications

As a fundamental task in natural language processing (NLP), discourse parsing has been successfully applied in other NLP tasks, such as question answering (QA) [71], text summarization [72–74], sentiment classification [3], language modeling [4], machine translation [5,75,76], and text categorization [6]. In this section, we will briefly introduce the application of discourse parsing for QA. MRC and sentiment analysis task.

Table 5　The performance of discourse parsing on multi-party dialogues

Model	UAS	LAS
MST [12]	68.8	50.4
ILP [13]	68.6	52.1
Deep Sequential [14]	73.2	55.7

4.1 Question answering

Discourse information has been explored for question answering (QA) systems. Considering questions are often related in real QA systems, discourse information is used to model the relation between context questions. [77] first proposes a discourse-aware model for context question answering. To explore the use and role of discourse in context QA, they propose that the discourse status relates to the discourse role of entities and discourse transitions. [78] examined three models by Centering Theory to model question sequence as a discourse for question answering and achieved obvious improvements.

The above two papers all are about model question sequence as a discourse and adopt the discourse structure for answering questions. Another application of discourse is ranking answers for the non-factoid QA system [1]. They combine lexical semantics with discourse information and adopt two different methods to represent discourse information: a shallow discourse marker model and an RST discourse parser model. Experiments demonstrate the effect of two representations and prove modeling discourse structure is helpful for non-factoid questions.

4.2 Machine reading comprehension (MRC)

Different from question answering task, the machine reading comprehension (MRC) task aims to let the machine answer the questions given input passages or dialogues.

Discourse structure has been used for modeling input passages and detecting relations between passages and questions. The effectiveness of the discourse structure for MRC has been proved. [2] incorporates discourse relations for machine reading tasks. In the paper, they adopt a hidden variable to represent discourse relations, including *Causal*, *Temporal*, *Explanation* and *Other*. [79] proposed a novel method using an answer-entailing structure that models discourse structure within the text by RST and word alignments between text and hypothesis.

4.3 Sentiment analysis

Sentiment analysis is a classical text classification task that detects the sentiments or emotions of input text, such as positive and negative, or happiness and sadness.

Discourse structure has been successfully applied to sentiment analysis. [3] adopt discourse structure for classifying sentiment. In RST-style discourse parsing, nucleus nodes play more important roles than satellite nodes in hypotactic RST relations (RST subtree with two nodes). Considering the difference of sentiment between the satellite node and nucleus node, the final sentiment of inputs would be more affected by nucleus nodes instead of satellite nodes. [80] proposed another sentiment analysis neural model Discourse-LSTM based on RST.

4.4 Text summarization

Text summarization is the task that summarizes the input document into a summary. For neural network-based models of text summarization, it is an essential task for modeling the input document. Discourse structure has been proved for its improvements in text summarization.

As we know, [72] first proposed discourse-based framework for document summarization. [81] proved the benefits of discourse structure for content selection in text summarization task, including RST-base structural features and PDTB-based semantic features. [33] adopted RST discourse parser obtains discourse dependency relations of a document and trimmed the discourse dependency tree as a tree knapsack problem. [73] extracted the discourse structure of product reviews by off-the-shelf RST-style discourse parser to build the aspect rhetorical tree, and select important aspects for generating summary via a template-based framework. Different from [33], [82] can directly generate discourse dependency tree for text summarization without transforming the rhetorical discourse trees into dependency-based trees. [83] adopted both anaphora constraints and grammatical constraints including RST and syntactic trees. [84] examined the role of the EDUs from the RST discourse parser and proved the benefits of EDU segmentation for content selection in text summarization. [85] adopted RST discourse parser to segment discourse units and select content using different models, including RNN, transformer, and BERT. [74] proposes a discourse-aware neural model that captures the discourse structure using the RST tree and encodes discourse units with a graph neural network.

4.5 Machine translation

Machine translation is a traditional natural language processing task that aims to translate the source language into the target language. Discourse structure is used to model the semantic relations between discourse units and has been applied on machine translation [86,87]. Considering the importance and ambiguity of discourse connectives, Meyer introduces connectives to help machine translation [5,75,76].

After decades of development, the research on machine translation has achieved great progress from word-level and phrase-level translation tasks to sentence-level translation tasks. Document-level machine translation would be an important future tend where the discourse structure of the input document would play a more essential role in modeling text.

5 Future trends

5.1 Building a large-scale corpus

As mentioned above, the dataset of discourse parsing dialogue has been the bottleneck of this task. To further promote the development of the task, it is necessary to build a large-scale high-quality corpus. Two points need more attention.

- **Scale** Deep learning models are data-driven and enough training data is necessary. For example, for multiparty dialogue discourse parsing, the number of dialogue, EDUs, and discourse relations should big enough to train a powerful model. Besides, there should be enough instances in each discourse relations to avoid few-shot situations.
- **Consistency** Due to the difficulty of annotating discourse structure and relations, it would be a challenging task to ensure the consistency of annotation. The annotators should be well trained and fewer annotators would be better.

5.2 Deep graph-based method

There are two types of methods in the task of semantic dependency parsing, including the transition-based approach and graph-based approach. Because the task of discourse parsing for multiparty dialogue is non-projective, transition-based methods will not work for this task. There has been literature that adopted a transition-based method for semantic dependency graph parsing, but they did not achieve good results [88]. Furthermore, with the success of graph neural network (GNN) on NLP, it will be worth investigating exploring the GNN-based approach for this RST-style or multiparty dialogue discourse parsing.

5.3 Meta-learning based method

Due to the limitation of the discourse treebank, there are not enough instances for many discourse relations. From example, there are little instances of *Background* and *Alternation* in the STAC dataset. The meta-learning method can be a good solution that can be naturally applicable to a few-shot or one-shot phenomenon. Meta-learning can be regarded as learning to learn, which aims to fast adapt to new training data. We can learn meta-knowledge on other big datasets, and then apply the meta-knowledge in discourse treebank.

Meta-learning methods have been successfully applied to many tasks including regression, classification, and reinforcement learning. However, there are not many studies that apply meta-learning to natural language processing, especially structural prediction and text categorization.

5.4 Exploring pre-trained representations

Two pre-training approaches Elmo and BERT have attracted widespread attention, because these models can significantly improve the performance of many NLP tasks. Different from other approaches, Elmo is a general method to learn context-dependent representations from BiLSTM [89]. Followed by Elmo, BERT is proposed to learn word representations based on bidirectional transformer [90].

Because the RST, PDTB, and STAC are all small scale datasets, using pre-trained word representations trained from a large corpus may significantly improve the performance of discourse parsing for multi-party dialogue. It is difficult to build a large-scale dataset in some NLP tasks with complex structures. Therefore, using pre-trained representations on small data sets would become a trend in NLP. Exploring pre-training representations for multi-party discourse parsing should be worth studying.

The effect of pretrained models has been proved on the PDTB dataset and achieved great improvement [55]. Because of the scale of the existing corpus for discourse parsing, the use of pretrained models should be furthermore explored.

5.5 Multitask architecture with RST and PDTB

There are so many kinds of literature about discourse parsing on RST-DT and PDTB, and there has been work combining RST and PDTB [65]. It is necessary to combine previous methods and expand the dataset at the same time for the task of discourse parsing for multi-party dialogue.

There are two subtasks in discourse parsing for multi-party dialogue, including predicting edges between EDUs and labeling discourse relations on each edge. For predicting edges between EDUs, structural prediction on RST-DT can be an auxiliary task for our main task. RST aims to build a document into a tree structure, while discourse parsing on the STAC needs to construct a graph structure. Therefore, common approaches for RST parsing are not available for this task. But we can expand our dataset using multitask architecture considering the limited instances in datasets.

For labeling discourse relations, RST-DT and PDTB all can be used in a multitask architecture. The relations in STAC are quite similar to RST and PDTB, so discourse parsing on RST-DT and PDTB would improve the accuracy of labeling relations in the STAC. In particular, labeling relations in the dialogue is related to implicit discourse relations recognition in PDTB. Therefore, different relevant tasks could be beneficial to one another.

6 Conclusion

In this survey, we introduce the task of discourse parsing and related datasets, mainly including RST-DT, PDTB, and STAC. Furthermore, we introduce existing methods for discourse parsing. We describe the applications of discourse parsing and show our opinion on this task. At last, we introduce trends and related work.

Acknowledgements The research in this article is supported by the Science and Technology Innovation 2030 -"New Generation Artificial Intelligence" Major Project (2018AA0101901), the National Key Research and Development Project (2018YFB1005103), the National Natural Science Foundation of China (Grant Nos. 61772156 and 61976073), Shenzhen Foundational Research Funding (JCYJ20200109113441941), and the Foundation of Heilongjiang Province (F2018013).

Appendix: Sense hierarchy

The sense hierarchy of RST-DT, PDTB 3.0, and STAC are respectively shown in Tables A1–A3.

Table A1 The sense hierarchy of RST-DT corpus

Relation	Class
Attribution	attribution, attribution-negative
Background	background, circumstance
Cause	cause, result, consequence
Comparison	comparison, preference, analogy, proportion
Condition	condition, hypothetical, contingency, otherwise
Contrast	contrast, concession, antithesis
Elaboration	elaboration-additional, elaboration-general-specific, elaboration-part-whole, elaboration-process-step, elaboration-object-attribution, elaboration-set-member, example, definition
Enablement	purpose, enablement
Evaluation	evaluation, interpretation, conclusion, comment
Explanation	evidence, explanation-argumentative, reason
Joint	list, disjunction
Manner-Means	manner, means
Topic-Comment	problem-solution, question-answer, statement-response, topic-comment, comment-topic, rhetorical-question
Summary	summary, restatement
Temporal	temporal-before, temporal-after, temporal-same-time, sequence, inverted-sequence
Topic-Change	topic-shift, topic-drift

Table A2 The sense hierarchy of PDTB v3.0

Level-1	Level-2	Level-3
Temporal	Synchronous	–
	Asynchronous	Precedence
		Succession
Contingency	Cause	Reason
		Result
		NegResult
	Cause+Belief	Reason+Belief
		Result+Belief
	Cause+SpeechAct	Reason+SpeechAct
		Result+SpeechAct
	Condition	Arg1-as-cond
		Arg2-as-cond
	Condition+SpeechAct	–
	Negative-condition	Arg1-as-negCond
		Arg2-as-negCond
	Negative-condition+SpeechAct	–
	Purpose	Arg1-as-goal
		Arg2-as-goal
Comparison	Concession	Arg1-as-denier
		Arg2-as-denier
	Concession+SpeechAct	Arg2-as-denier+SpeechAct
	Contrast	–
	Similarity	–
Expansion	Conjunction	–
	Disjunction	–
	Equivalence	–
	Exception	Arg1-as-excpt
		Arg2-as-excpt
	Instantiation	Arg1-as-instance
		Arg2-as-instance
	Level-of-detail	Arg1-as-detail
		Arg2-as-detail
	Manner	Arg1-as-manner
		Arg2-as-manner
	Substitution	Arg1-as-subst
		Arg2-as-subst

Table A3 The sense hierarchy and distribution of the STAC corpus

Relation	Train	Dev	Test
Comment	1851	1684	167
Clarification_question	260	240	20
Elaboration	869	771	98
Acknowledgment	1010	893	117
Continuation	987	873	114
Explanation	437	407	30
Conditional	124	105	19
Question-answer_pair	2541	2236	305
Alternation	146	128	18
Q-Elab	599	525	74
Result	578	551	27
Background	61	58	3
Narration	130	116	14
Correction	212	189	23
Parallel	215	196	19
Contrast	493	449	44
Total	10513	9421	1092

References

1. Jansen P, Surdeanu M, Clark P. Discourse complements lexical semantics for non-factoid answer reranking. In: Proceedings of the 52nd

Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2014, 977–986
2. Narasimhan K, Barzilay R. Machine comprehension with discourse relations. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2015, 1253−1262
3. Bhatia P, Ji Y, Eisenstein J. Better document-level sentiment analysis from rst discourse parsing. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015, 2212−2218
4. Ji Y, Haffari G, Eisenstein J. A latent variable recurrent neural network for discourse-driven language models. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. June 2016, 332–342
5. Meyer T, Popescu-Belis A. Using sense-labeled discourse connectives for statistical machine translation. In: Proceedings of the Joint Workshop on Exploiting Synergies between Information Retrieval and Machine Translation (ESIRMT) and Hybrid Approaches to Machine Translation (HyTra). 2012, 129–138
6. Ji Y, Smith N A. Neural discourse structure for text categorization. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2017, 996–1005
7. Mann W C, Thompson S A. Rhetorical structure theory: Toward a functional theory of text organization. Text-Interdisciplinary Journal for the Study of Discourse, 1988, 8(3): 243–281
8. Carlson L, Marcu D, Okurowski M E. Building a discourse-tagged corpus in the framework of rhetorical structure theory. Springer, 2003
9. Wolf F, Gibson E, Fisher A, Knight M. Discourse graphbank. Linguistic Data Consortium. Philadelphia, 2004
10. Prasad R, Dinesh N, Lee A, Miltsakaki E, Robaldo L, Joshi A K, Webber B L. The penn discourse treebank 2.0. In: LREC. 2008
11. Webber B. D-ltag: extending lexicalized tag to discourse. Cognitive Science, 2004, 28(5): 751–779
12. Afantenos S, Kow E, Asher N, Perret J. Discourse parsing for multiparty chat dialogues. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015, 928–937
13. Perret J, Afantenos S, Asher N, Morey M. Integer linear programming for discourse parsing. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016, 99–109
14. Shi Z, Huang M. A deep sequential model for discourse parsing on multi-party dialogues. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 7007−7014
15. Ji Y, Eisenstein J. Representation learning for text-level discourse parsing. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2014, 13–24
16. Webber B, Prasad R, Lee A, Joshi A. The penn discourse treebank 3.0 annotation manual. Philadelphia, University of Pennsylvania, 2019
17. Lin Z, Ng H T, Kan M Y. A pdtb-styled end-to-end discourse parser. Natural Language Engineering, 2014, 20(2): 151–184
18. Pitler E, Louis A, Nenkova A. Automatic sense prediction for implicit discourse relations in text. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2. 2009, 683–691
19. Hong Y, Zhou X, Che T, Yao J, Zhu Q, Zhou G. Cross-argument inference for implicit discourse relation recognition. In: Proceedings of the 21st ACM international conference on Information and knowledge management. 2012, 295–304
20. Rehbein I, Scholman M, Demberg V. Annotating discourse relations in spoken language: A comparison of the PDTB and CCR frameworks. LREC, 2016
21. Asher N, Hunter J, Morey M, Benamara F, Afantenos S. Discourse structure and dialogue acts in multiparty dialogue: The STAC corpus. In: Proceedings of the 10th International Conference on Language Resources and Evaluation, 2016, 2721−2727
22. Li J, Liu M, Kan M Y, Zheng Z, Wang Z, Lei W, Liu T, Qin B. Molweni: A challenge multiparty dialogues-based machine reading comprehension dataset with discourse structure. In: Proceedings of the 28th International Conference on Computational Linguistics. 2020, 2642−2652
23. Lowe R, Pow N, Serban I, Pineau J. The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems. In: Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue. 2015, 285–294
24. Soricut R, Marcu D. Sentence level discourse parsing using syntactic and lexical information. In: Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics. 2003, 228–235
25. Subba R, Di Eugenio B. Automatic discourse segmentation using neural networks. In: Proceedings of the 11th Workshop on the Semantics and Pragmatics of Dialogue. 2007, 189–190
26. Fisher S, Roark B. The utility of parse-derived features for automatic discourse segmentation. In: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics. 2007, 488–495
27. 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. 2012, 904–915
28. Sagae K. Analysis of discourse structure with syntactic dependencies and data-driven shift-reduce parsing. In: Proceedings of the 11th International Conference on Parsing Technologies (IWPTÂąÂŕ09). 2009, 81–84
29. Hernault H, Prendinger H, Ishizuka M, others. Hilda: A discourse parser using support vector machine classification. Dialogue & Discourse, 2010, 1(3):
30. Bach N X, Le Nguyen M, Shimazu A. A reranking model for discourse segmentation using subtree features. In: Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue. 2012, 160–168
31. Feng V W, Hirst G. Two-pass discourse segmentation with pairing and global features. 2014, arXiv preprint arXiv: 1407.8215
32. Wang Y, Li S, Yang J. Toward fast and accurate neural discourse segmentation. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018, 962–967
33. Hirao T, Yoshida Y, Nishino M, Yasuda N, Nagata M. Single-document summarization as a tree knapsack problem. In: Proceedings of the 2013 conference on empirical methods in natural language processing. 2013, 1515−1520
34. Li S, Wang L, Cao Z, Li W. Text-level discourse dependency parsing. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2014, 25–35
35. Hayashi K, Hirao T, Nagata M. Empirical comparison of dependency conversions for rst discourse trees. In: Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue. 2016, 128–136
36. Surdeanu M, Hicks T, Valenzuela-Escárcega M A. Two practical rhetorical structure theory parsers. In: Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: Demonstrations. 2015, 1–5
37. Morey M, Muller P, Asher N. A dependency perspective on rst discourse parsing and evaluation. Computational Linguistics, 2018, 44(2): 197–235

38. Joty S, Carenini G, Ng R T. Codra: A novel discriminative framework for rhetorical analysis. Computational Linguistics, 2015, 41(3): 385–435
39. Li J, Li R, Hovy E. Recursive deep models for discourse parsing. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014, 2061−2069
40. 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. 2016, 362–371
41. 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). 2018, 438–443
42. Yu N, Zhang M, Fu G. Transition-based neural rst parsing with implicit syntax features. In: Proceedings of the 27th International Conference on Computational Linguistics. 2018, 559–570
43. Jia Y, Feng Y, Ye Y, Lv C, Shi C, Zhao D. Improved discourse parsing with two-step neural transition-based model. ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), 2018, 17(2): 11
44. Braud C, Plank B, Søgaard A. Multi-view and multi-task training of rst discourse parsers. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. 2016, 1903−1913
45. Braud C, Coavoux M, Søgaard A. Cross-lingual rst discourse parsing. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers. 2017, 292–304
46. Pitler E, Nenkova A. Using syntax to disambiguate explicit discourse connectives in text. In: Proceedings of the ACL-IJCNLP 2009 Conference Short Papers. 2009, 13–16
47. Li S, Kong F, Zhou G. A joint learning approach to explicit discourse parsing via structured perceptron. Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data, Springer, Cham, 2014, 70−82
48. Marcu D, Echihabi A. An unsupervised approach to recognizing discourse relations. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. 2002, 368–375
49. Wang X, Li S, Li J, Li W. Implicit discourse relation recognition by selecting typical training examples. In: COLING. 2012, 2757−2772
50. Lan M, Xu Y, Niu Z Y. Leveraging synthetic discourse data via multi-task learning for implicit discourse relation recognition. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2013, 476–485
51. Ji Y, Eisenstein J. One vector is not enough: Entity-augmented distributed semantics for discourse relations. Transactions of the Association for Computational Linguistics, 2015, 3: 329–344
52. Rutherford A T, Demberg V, Xue N. Neural network models for implicit discourse relation classification in english and chinese without surface features. 2016, arXiv preprint arXiv: 1606.01990
53. Braud C, Denis P. Comparing word representations for implicit discourse relation classification. In: Proceedings of Empirical Methods in Natural Language Processing (EMNLP 2015). 2015
54. Shi W, Demberg V. Next sentence prediction helps implicit discourse relation classification within and across domains. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019, 5794−5800
55. Kishimoto Y, Murawaki Y, Kurohashi S. Adapting bert to implicit discourse relation classification with a focus on discourse connectives. In: Proceedings of The 12th Language Resources and Evaluation Conference. 2020, 1152−1158
56. Rutherford A, Xue N. Discovering implicit discourse relations through brown cluster pair representation and coreference patterns. In: Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics. 2014, 645−654
57. McKeown K, Biran O. Aggregated word pair features for implicit discourse relation disambiguation. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. 2013, 69–73
58. Lei W, Wang X, Liu M, Ilievski I, He X, Kan M Y. Swim: A simple word interaction model for implicit discourse relation recognition. In: Proceedings of the 26th International Joint Conference on Artificial Intelligence. 2017, 4026−4032
59. Chen J, Zhang Q, Liu P, Qiu X, Huang X. Implicit discourse relation detection via a deep architecture with gated relevance network. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016, 1726−1735
60. Chen J, Zhang Q, Liu P, Huang X. Discourse relations detection via a mixed generative-discriminative framework. In: Proceedings of Thirtieth AAAI Conference on Artificial Intelligence. 2016, 30(1)
61. Lei W, Xiang Y, Wang Y, Zhong Q, Liu M, Kan M Y. Linguistic properties matter for implicit discourse relation recognition: Combining semantic interaction, topic continuity and attribution. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, 2018, 32(1)
62. Guo F, He R, Jin D, Dang J, Wang L, Li X. Implicit discourse relation recognition using neural tensor network with interactive attention and sparse learning. In: Proceedings of the 27th International Conference on Computational Linguistics. 2018, 547–558
63. Bai H, Zhao H. Deep enhanced representation for implicit discourse relation recognition. In: Proceedings of the 27th International Conference on Computational Linguistics. 2018, 571–583
64. Xu S, Li P, Kong F, Zhu Q, Zhou G. Topic tensor network for implicit discourse relation recognition in chinese. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019, 608–618
65. Liu Y, Li S, Zhang X, Sui Z. Implicit discourse relation classification via multi-task neural networks. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. 2016, 2750−2756
66. Rutherford A, Xue N. Improving the inference of implicit discourse relations via classifying explicit discourse connectives. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2015, 799–808
67. Shi W, Demberg V. Learning to explicitate connectives with seq2seq network for implicit discourse relation classification. In: Proceedings of the 13th International Conference on Computational SemanticsLong Papers. 2019, 188–199
68. Dai Z, Huang R. A regularization approach for incorporating event knowledge and coreference relations into neural discourse parsing. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019, 2967−2978
69. Guo F, He R, Dang J, Wang J. Working memory-driven neural networks with a novel knowledge enhancement paradigm for implicit discourse relation recognition. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 7822−7829
70. He R, Wang J, Guo F, Han Y. TransS-driven joint learning architecture for implicit discourse relation recognition. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020, 139–148
71. Verberne S, Boves L, Oostdijk N, Coppen P A. Evaluating discoursebased answer extraction for why-question answering. In: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. 2007, 735–736

72. Marcu D. The theory and practice of discourse parsing and summarization. MIT press, 2000
73. Gerani S, Mehdad Y, Carenini G, Ng R, Nejat B. Abstractive summarization of product reviews using discourse structure. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014, 1602−1613
74. Xu J, Gan Z, Cheng Y, Liu J. Discourse-aware neural extractive text summarization. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. July 2020, 5021−5031
75. Meyer T. Disambiguating temporal-contrastive connectives for machine translation. In: Proceedings of the ACL 2011 Student Session. June 2011, 46–51
76. Meyer T, Popescu-Belis A, Zufferey S, Cartoni B. Multilingual annotation and disambiguation of discourse connectives for machine translation. In: Proceedings of Association for Computational Linguistics-Proceedings of 12th SIGdial Meeting on Discourse and Dialogue, number CONF. 2011
77. Chai J, Jin R. Discourse structure for context question answering. In: Proceedings of the Workshop on Pragmatics of Question Answering at HLT-NAACL 2004. 2004, 23–30
78. Sun M, Chai J Y. Discourse processing for context question answering based on linguistic knowledge. Knowledge-Based Systems, 2007, 20(6): 511–526
79. Sachan M, Dubey K, Xing E, Richardson M. Learning answerentailing structures for machine comprehension. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2015, 239–249
80. Kraus M, Feuerriegel S. Sentiment analysis based on rhetorical structure theory: Learning deep neural networks from discourse trees. Expert Systems with Applications, 2019, 118: 65–79
81. Louis A, Joshi A, Nenkova A. Discourse indicators for content selection in summarization. In: Proceedings of the SIGDIAL 2010 Conference. 2010, 147–156
82. Yoshida Y, Suzuki J, Hirao T, Nagata M. Dependency-based discourse parser for single-document summarization. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014, 1834−1839
83. Durrett G, Berg-Kirkpatrick T, Klein D. Learning-based singledocument summarization with compression and anaphoricity constraints. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016, 1998−2008
84. Li J J, Thadani K, Stent A. The role of discourse units in near-extractive summarization. In: Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue. 2016, 137–147
85. Liu Z, Chen N. Exploiting discourse-level segmentation for extractive summarization. In: Proceedings of the 2nd Workshop on New Frontiers in Summarization. 2019, 116–121
86. Haenelt K. Towards a quality improvement in machine translation: Modelling discourse structure and including discourse development in the determination of translation equivalents. In: Proceedings of the 4th International Conference on Theoretical and Methodological Issues in Machine Translation. Mor-ristown: Association for Computational Linguiscs. 1992, 205–212
87. Mitkov R. How could rhetorical relations be used in machine translation? In: Proceedings of Intentionality and structure in discourse relations. 1993
88. Wang Y, Che W, Guo J, Liu T. A neural transition-based approach for semantic dependency graph parsing. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2018, 32(1)
89. Peters M, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L. Deep contextualized word representations. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). 2018, 2227−2237
90. Devlin J, Chang M W, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). 2019, 4171−4186

Jiaqi Li received the BS degree from the School of Computer Science and Technology, Heilongjiang University, China in 2015. He is currently working toward the PhD degree in the Harbin Institute of Technology, China. His research interests include discourse parsing for multiparty dialogues and its applications.

Ming Liu received the PhD degree from the School of Computer Science and Technology, Harbin Institute of Technology, China in 2010. He is a full professor/PhD supervisor of the Department of Computer Science, and the faculty member of Social Computing and Information Retrieval (HIT-SCIR), Harbin Institute of Technology, China. His research interests include knowledge graph, machine reading comprehension.

Bing Qin received the PhD degree from the School of Computer Science and Technology, Harbin Institute of Technology, China in 2005. She is a full professor of the Department of Computer Science, and the director of the Research Center for Social Computing and Information Retrieval (HIT-SCIR), Harbin Institute of Technology, China. Her research interests include natural language processing, information extraction, document-level discourse analysis, and sentiment analysis.

Ting Liu received the PhD degree from the Department of Computer Science, Harbin Institute of Technology, China in 1998. He is a full professor of the School of Computer Science and Technology, and the director of Faculty of Computing, Harbin Institute of Technology, China. His research interests include information retrieval, natural language processing, and social media analysis.