Tree Representations in Transition System for RST Parsing

Jinfen Li
College of Arts and Sciences,
Syracuse University
jli284@syr.edu

Lu Xiao
School of Information Studies,
Syracuse University
lxiao04@syr.edu

Abstract

The transition-based systems in the past studies propose a series of actions to build a right-heavy binarized tree for RST parsing. However, the nodes of the binary-nuclear relations (e.g., Contrast) have the same nuclear type with those of the multi-nuclear relations (e.g., Joint) in the binary tree structure. In addition, the reduce action only construct binary trees instead of multi-branch trees, which is the original RST tree structure. In our paper, we design a new nuclear type for the multi-nuclear relations, and a new action to construct a multi-branch tree. We enrich the feature set by extracting additional refined dependency feature of texts from the Bi-Affine model (Dozat and Manning, 2016). We also compare the performance of two approaches for RST parsing in the transition-based system: a joint action of reduce-shift and nuclear type (i.e., Reduce-SN) vs a separate one that applies Reduce action first and then assigns nuclear type. We find that the new devised nuclear type and action are more capable of capturing the multi-nuclear relation and the joint action is more suitable than the separate one. Our multi-branch tree structure obtains the state-of-the-art performance for all the 18 coarse relations.

1 Introduction

It is expected that a document consists of text units that are logically connected within the context through discourse relations. Text-level discoursing parsing plays a significant role in many NLP downstream task, such as question-answering (Jansen et al., 2014), sentiment analysis (Mukherjee and Bhattacharyya, 2012) and abstractive summarization (Koto et al., 2019).

The Rhetorical Structure Theory (RST)(Mann and Thompson, 1988) is one of the widely explored text-level discourse parsing theory, which parses a document into a hierarchical discourse tree. The discourse tree is built over its smallest parsing unit, namely, elementary discourse unit (EDU), placed in the leaf nodes. The tree's non-terminal nodes bear the information of span, nuclearity and relation. The span represents the cover range of a sequence of EDUs; while the nuclearity status is the semantic role in a relation (i.e., nucleus or satellite). A text tagged by nucleus is more essential than the satellite. Conventionally, there are three nuclearity types including Nucleus-Satellite (NS), Satellite-Nucleus (SN) and Nucleus-Nucleus (NN). As for relation, there are mainly two types of relations: the mono-nuclear relation, such as "Attribution" or "Summary", with the the nuclearity type being either "NS" or "SN"; and the multi-nuclear relation, such as "Contrast" or "Same-Unit", with the nuclearity types of "NN".

Among all the parsing systems dedicated to RST parsing, the transition-based systems obtain comparative performance with the CKY-like system. The transition-based algorithms are first explored in (Marcu, 1999), which adopt a sequence of shift-reduce action to construct rhetorical structures of texts. The study of Wang et al. (2017) achieves the best performance among all the transition-based systems, by parsing the discourse tree in two stages, namely, parsing a naked discourse tree (identifying span and nuclearity) in the first stage and assigning relation in the second stage. The researchers use only one classifier for the joint action of shift-reduce and nuclear types (e.g., Reduce-NS) to parse the tree. In this

Proceedings of the 28th International Conference on Computational Linguistics, pages 6746–6751
Barcelona, Spain (Online), December 8-13, 2020

study, we explore whether two separate classifiers for shift-reduce action and nuclear type respectively (i.e., identifying span first, and then nuclearity comes later) is more effective in the RST parsing task.

Previous transition systems do not differentiate the nuclear type between multi-nuclear relations and binary-nuclear relations. Neither did they achieve satisfying performance on the multi-branch relation such as Topic-change. We find that there are 1.9% and 1.7% multi-branch nodes in the original RST tree in the training and test dataset respectively (the frequency details are shown in the Appendix). Because we utilize the Nucleus feature (Wang et al., 2017) in the relation classifier, a distinctive nuclearity type for the multi-branch relations would be helpful to differentiate them from the binary-branch relations.

In addition, previous studies represent RST by the right-heavy binary trees, which inevitably create the redundant nonexistent node in the original RST tree. To solve these problems, we come up with a new nuclear type "$\hat{\text{N}}$" for the multi-nuclear relation, and a new action "$\hat{\text{R}}$" to construct a multi-branch tree. Experimental results tell us that the newly designed nuclear type and action labels are capable of capturing multi-nuclear relations.

In sum, our paper mainly makes the following contributions:

- As Wang et al. (2017) proved that the dependency feature is helpful for span, nuclearity and relation identification in RST parsing. We refine the feature lists by adding the dependency feature generated by the deep Biaffine Attention model (Dozat and Manning, 2016) .
- We propose a new nuclear type ($\hat{\text{N}}$) to distinguish the multi-nuclear types from the binary-nuclear types.
- We come up with a new action: the Flat Reduce ($\hat{\text{R}}$) to construct a multi-branch tree.
- We examine that the joint action of shift-reduce and nuclear type is more suitable in the transition-based system comparing to the separate one.

2 Our Approach

We construct two tree structures: the binary tree and multi-branch tree structures, with a new nuclear type and a new action, respectively. To enrich the features for the small number of training examples of the new labels, we add the extra dependency features generated by the Bi-Affine model (Dozat and Manning, 2016), along with the features extracted by Wang et al. (2017). We also examine the performance of two approaches in our transition-based system: to identify span and nuclearity concurrently with one classifier vs to identify them separately using different classifiers.

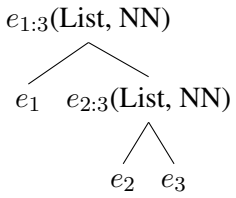
2.1 Binary Tree Structure

In the previous studies, researchers annotate the multi-nuclear type with a notion of "NN", and parse the RST tree into a right-heavy binary tree (as shown in Figure 1a). Such annotation cannot differentiate the multi-nuclear types from the binary-nuclear ones. Since the multi-nuclear relations have different tree structure in the original RST tree, we propose a new nuclear type of "$\hat{\text{N}}$" for the multi-nuclear nodes (as shown in Figure 1b,(Binary†)), corresponding to the nuclear type of "NN" for the binary-nuclear nodes. With the annotation of "$\hat{\text{N}}$", the nuclear types of the binary-nuclear relations (e.g., Explanation and Contrast) and the multi-nuclear relations (e.g., List and Same-Unit) are distinguishable. We present the example in Table 1.

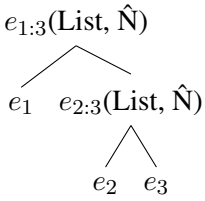
2.2 Multi-branches Tree Structure

In order to capture the distinct tree structure of the multi-nuclear relations, we keep the original RST tree structure (Multi†) by using a new action of Flat Reduce $\hat{\text{R}}$. As shown in Table 1, the action of $\hat{\text{R}}(\text{list}, \hat{\text{N}})_1$ forms a sub tree $\widehat{\mathbf{e_2 e_3}}$. However, instead of forming a binary sub tree, action $\hat{\text{R}}(\text{list}, \hat{\text{N}})_2$ flattens the tree structure so that EDU_1, EDU_2 and EDU_3 are in the same depth in a tree. A corresponding tree structure is shown in Figure 1c.

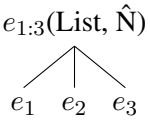
After proposing the new form and action labels, the frequency of the new joint-action R-$\hat{\text{N}}$ and $\hat{\text{R}}$-$\hat{\text{N}}$ is relatively low comparing to the others. In order to better fit for the sparse train examples, we enrich our feature by introducing the dependency feature explained in the following section.



(a) Binary tree structure of multi-nuclear relation in previous studies



(b) Binary tree structure of multi-nuclear relation



(c) Multi-branch tree structure of multi-nuclear relation

Figure 1: RST tree structures

Step	Stack	Queue	Binary		Multi	
			Action	Relation	Action	Relation
1	$\emptyset$	e_1, e_2, e_3, e_4	SH	$\emptyset$	SH	$\emptyset$
2	e_1	e_2, e_3, e_4	SH	$\emptyset$	SH	$\emptyset$
3	e_1, e_2	e_3, e_4	SH	$\emptyset$	SH	$\emptyset$
4	e_1, e_2, e_3	e_4	R(list, $\hat{\text{N}}$)$_1$	$\emptyset$	$\hat{\text{R}}$(list, $\hat{\text{N}}$)$_1$	$\emptyset$
5	$e_1, e_{2:3}$	e_4	R(list, $\hat{\text{N}}$)$_2$	$\widehat{\mathbf{e_2 e_3}}$	$\hat{\text{R}}$(list, $\hat{\text{N}}$)$_2$	$\widehat{\mathbf{e_2 e_3}}$
6	$e_{1:3}$	e_4	SH	$\mathbf{e_1}\widehat{\mathbf{e_{2:3}}}$	SH	$\mathbf{e_1}\widehat{\mathbf{e_2 e_3}}$
7	$e_{1:3}, e_4$	$\emptyset$	R(summary, NS)	-	R(summary, NS)	-
8	$e_{1:4}$	$\emptyset$	PR	$\widehat{\mathbf{e_{1:3}} e_4}$	PR	$\widehat{\mathbf{e_{1:3}} e_4}$

Table 1: A transition-based system with a new form of $\hat{\text{N}}$ and a new action $\hat{\text{R}}$

2.3 Refined Dependency Feature

(Wang et al., 2017) proposed the dependency feature, which is helpful for span, nuclearity and relation identification in RST parsing. We further refine dependency feature generated by the deep Biaffine Attention model (Dozat and Manning, 2016), and obtain the representations of head, dependent and arc-label for each EDU.

We present the architecture of Bi-Affine model in Figure 2. The Bi-Affine model is composed of several components including (1) a BLSTM layer extracting the representation of a sentence r_k, where k is the number of words; (2) two MLP layers obtaining hidden states of head $h_i^{(arc-head)}$, dependent $h_i^{(arc-dep)}$, respectively; (3) a Bi-Affine layer to extract the arc-label $s_{i,j}^{(arc)}$. After achieving the head, dependent and arc-label hidden states, we obtain the vector for an EDU D_e by concatenating them. The vector of the merged span $D_{e_{m:n}}$ is the average vector from D_{e_m} to D_{e_n}, where m and n is the start and end boundary of the span.

$$
\begin{aligned}
&r_k = BLSTM(x_k) \\
&h_i^{(arc-dep)} = MLP^{(arc-dep)}(r_i) \\
&h_j^{(arc-head)} = MLP^{(arc-head)}(r_j) \\
&s_{i,j}^{(arc)} = Bi-Affine(h_i^{(arc-dep)}, h_j^{(arc-head)}) \\
&D_e = [H^{(arc-dep)}; H^{(arc-head)}; S^{(arc)}] \\
&D_{e_{m:n}} = ave(\sum_{p=m}^{n} D_{e_p})
\end{aligned}
\tag{1}
$$

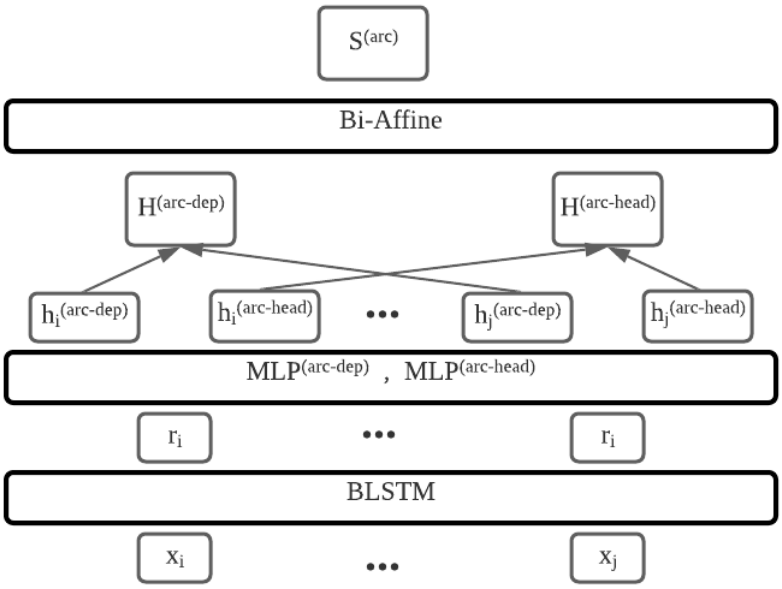


Figure 2: Overview of Bi-Affine Model

2.4 A Separate Classifier for Nuclear Type

Previous studies propose the information-rich actions (e.g., Reduce-NS-Contrast) to identify span, nuclearity and relation in the same stage, while (Wang et al., 2017) separate the relation from the actions and assign relation in a second stage, which obtains great success. In our study, we further explore whether separating the nuclearity from the action is more suitable. We assume an independent classifier for the binary action classification task, and a nuclearity classifier would achieve better performance on the separate subtasks.

We adopt all the features used by (Wang et al., 2017) including the *Status, Position, Structural, Dependency*(which is different from ours), *N-gram* and *Nucleus* for the action classifier, and an additional SVM nuclearity classifier (Separate†); while *Dependency, N-gram, Nucleus, Refined Structural* and *Tree* features are used for the relation classifier.

3 Experimental Setup

We use the default 385 documents (347 and 38 documents in training and test dataset respectively) from the Wall Street Journal (RST-DT) (Mann and Thompson, 1988). The 18 coarse-grained relations are evaluated in our experiment such as the binary-nuclear relation of "Contrast" and the multi-nuclear relation of "Joint" (Sagae and Lavie, 2005). We report the Micro-averaged F1 scores of RST-Parseval (Marcu, 2000) and labelled attachment decisions (LAS) with respect to span, nuclearity and relation.

4 Results

We only compare our results with the state-of-the-art transition-based system: two-stages system (Wang et al., 2017). As shown in Table 2, the additional refined dependency feature improves the performance in span, nuclearity and relation, measured in the two scores. The poor performance of Separate†implies that the joint action of shift-reduce and nuclear type is more suitable in the transition-based system. The reason might be the feature of Nucleus is essential for the span and relation identification. While the binary tree with the new nuclear type of $\hat{\text{N}}$ outperforms our baseline by around 2% in all three aspects. Our final model: multi-branch structure achieves the state-of-the-art performance in transition-based systems, for the reason that it reduces the redundant nodes in the RST parsing tree.

To further examine whether our approach improve the performance of the multi-nuclear relations, we present the scores of Joint, Same-Unit and Topic-Change in Table 3. Compared to the two-stage system, the F1-score of these relations using our methods improves. Specifically, the binary tree structure is more capable of capturing the relation of Joint and Same-Unit, and the multi-branch tree structure is more capable of capturing Topic-Change relation (the dramatic improvement of Topic-Change might due to its small frequency in the test dataset).

Model	RST-Parseval			LAS		
	Span	Nuclearity	Relation	Span	Nuclearity	Relation
Two Stages(Wang et al., 2017)	86	72.4	59.7	73	61.27	51.26
Two Stages†	87.63	74.52	60.33	75.26	64.34	49.62
Separate†	58.13	31.32	9.26	16.26	4.18	5.57
Binary†	87.9	74.86	62.6	75.45	63.57	53.18
Multi†	**88.39**	**75.78**	**63.05**	**76.47**	**65.32**	**53.93**
Human	88.7	77.7	65.8	78.7	66.8	57.1

Table 2: Micro-averaged F1 scores of RST-Parseval and LAS, †represents the incorporation of refined dependency features

Model	Joint	Same-Unit	Topic-Change
Two Stages(Wang et al., 2017)	38.9	72.45	4.76
Two Stages†	34.6	69.04	5.8
Binary†	**41.54**	**73.88**	8.7
Multi†	33.7	72.39	**28.57**

Table 3: F1-score of relations using LAS metric

5 Conclusions

In this paper, we come up with a new nuclear type "$\hat{\text{N}}$" for the multi-nuclear relation, and a new action "$\hat{\text{R}}$" to construct a multi-branch tree, with the additional refined dependency feature for the texts. We find that "$\hat{\text{N}}$" is effective in distinguishing from the binary-nuclear type, resulting a better performance on the multi-nuclear relations identifications. The newly designed action $\hat{\text{R}}$, remaining the original tree structure, helps to prevent from creating the redundant nodes in the right-heavy binary tree, thus achieving the state-of-the-art performance.

We find that the F1-score of Topic-Change is extremely low in comparison with to the other multi-nuclear relations, due to its sparsity; on the other hand, Topic-Change is a relation which benefits from global information. Therefore, we plan to design a top-down manner system to capture more global information in the future. We also find that the joint action of shift-reduce and nuclear type is more suitable in the transition-based system, and an analysis of which feature is essential in identifying the spans, nuclearity and relations is worth exploring in the future.

References

Timothy Dozat and Christopher D Manning. 2016. Deep biaffine attention for neural dependency parsing. *arXiv preprint arXiv:1611.01734*.

Peter Jansen, Mihai Surdeanu, and Peter Clark. 2014. 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)*, pages 977–986.

Fajri Koto, Jey Han Lau, and Timothy Baldwin. 2019. Improved document modelling with a neural discourse parser. *arXiv preprint arXiv:1911.06919*.

William C Mann and Sandra A Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.

Daniel Marcu. 1999. A decision-based approach to rhetorical parsing. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 365–372.

Daniel Marcu. 2000. *The theory and practice of discourse parsing and summarization*. MIT press.

Subhabrata Mukherjee and Pushpak Bhattacharyya. 2012. Sentiment analysis in twitter with lightweight discourse analysis. In *Proceedings of COLING 2012*, pages 1847–1864.

Kenji Sagae and Alon Lavie. 2005. A classifier-based parser with linear run-time complexity. In *Proceedings of the Ninth International Workshop on Parsing Technology*, pages 125–132.

Yizhong Wang, Sujian Li, and Houfeng Wang. 2017. A two-stage parsing method for text-level discourse analysis. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 184–188.

APPENDIX

SH	R-NN	R-$\hat{N}$ / $\hat{R}$-$\hat{N}$	R-NS	R-SN
19443	3276	1053	11702	3065

Table 4: Statistics of Five Joint-action. KEY: Shift(SH), Binary Reduce(R) and Flat Reduce($\hat{R}$)

Relation	Count
Attribution	357
Background	87
Cause	83
Comparison	37
Condition	27
Contrast	176
Elaboration	949
Enablement	62
Evaluation	66
Explanation	126
Joint	382
Manner-Means	24
Same-Unit	342
Summary	16
Temporal	94
Textual-Organization	40
Topic-Change	12
Topic-Comment	20

Table 5: Count of each relation in the test dataset