A Unified RvNN Framework for End-to-End Chinese Discourse Parsing

Chuan-An Lin[1], Hen-Hsen Huang[1], Zi-Yuan Chen[1], and Hsin-Hsi Chen[1,2]

[1]Department of Computer Science and Information Engineering
National Taiwan University, Taipei, Taiwan
[2]MOST Joint Research Center for AI Technology and All Vista Healthcare, Taipei, Taiwan
{calin, hhhuang}@nlg.csie.ntu.edu.tw
{b02902017, hhchen}@ntu.edu.tw

Abstract

This paper demonstrates an end-to-end Chinese discourse parser. We propose a unified framework based on recursive neural network (RvNN) to jointly model the subtasks including elementary discourse unit (EDU) segmentation, tree structure construction, center labeling, and sense labeling. Experimental results show our parser achieves the state-of-the-art performance in the Chinese Discourse Treebank (CDTB) dataset. We release the source code with a pre-trained model for the NLP community. To the best of our knowledge, this is the first open source toolkit for Chinese discourse parsing. The standalone toolkit can be integrated into subsequent applications without the need of external resources such as syntactic parser.

Title and Abstract in Chinese

適用於點對點中文語篇剖析的遞迴類神經網路統一架構

中文語篇剖析有四項子任務，包含初級語篇單元分割、剖析樹建立、主次關係識別、語篇關係辨識等。本文展示一個點對點中文語篇剖析器，並提出一套統一架構，可以對輸入之中文篇章直接產生完整的中文語篇剖析結果。我們的剖析器以遞迴類神經網路為基礎，同時對四項子任務進行學習，在中文語篇樹庫（CDTB）資料集上，達到最先進的效能。我們釋出了這個剖析器的原始碼與預先訓練完成的模型，立即可用。據我們所知，這是第一個開放原始碼的中文剖析工具集，而且這套獨立的工具集不須依賴外部資源（如句法剖析器），便於下游應用的整合。

1 Introduction

Discourse parsing is aimed at identifying how the discourse units are related with each other, forming the hierarchical structure of an article. As pointed out by Mann and Thompson (1988), no part in an article is completely isolated. The discourse structure provides critical information to understand an article. NLP tasks such as summarization (Louis et al., 2010), information retrieval (Lioma et al., 2012), and text categorization (Ji and Smith, 2017) have been shown benefited from the information extracted by discourse parsing.

Prior work of Chinese discourse parsing focuses on intra-sentential parsing (Huang and Chen, 2012). The CoNLL 2016 Shared Task deals with shallow parsing (Xue et al., 2016). So far, there is quite less work on complete hierarchical Chinese discourse parsing at paragraph or article level (Kang et al., 2016). The subtasks in Chinese discourse parsing depend on each other. In a pipelined system, there may be a severe issue of error propagation among elementary discourse unit (EDU) segmentation, connective recognition, parse tree construction, and relation labeling (Kang et al., 2016). The other problem is that prior Chinese discourse parser relies on linguistic features extracted by external third party packages. This is an important issue especially for a pipeline system. Extracting feature from free text is also an issue, while most systems rely on external syntactic parser for providing informations to do the above tasks. For a toolkit targeting real-world applications, a standalone system is more robust and easy to

deploy. Inspired by Li et al. (2014a), in this work we propose an end-to-end Chinese discourse parser that performs EDU segmentation, discourse tree construction, and discourse relation labeling in a unified framework based on recursive neural network (RvNN) proposed by Goller and Kuchler (1996). The RvNN model learns to construct the structured output through merging children nodes to parent nodes in the bottom-up fashion. Within the RvNN paradigm, recurrent neural network (RNN) is employed to model the representations from word segments, discourse units, to the whole paragraph. RNN like long short-term memory (LSTM) neural network (Hochreiter and Schmidhuber, 1997) is reportedly successful in learning the text representation. In the prediction stage, we use the CKY algorithm to deal with both local and global information during the construction of discourse parse tree, eliminating the gap between the bottom-up approach and top-down annotation schemes.

The contribution of this work is three-fold: (1) We release a ready-to-use toolkit for end-to-end Chinese discourse parsing. To the best of our knowledge, this is the first publicly available toolkit for Chinese discourse parsing.[1] (2) We propose a novel unified RvNN framework for end-to-end discourse parsing. Experimental results show our model achieves the state-of-the-art performance. (3) Without the need for external resources like syntactic parser, our standalone end-to-end parser can be easily integrated into subsequent applications. The open source package can be even adapted to other languages.

The rest of this paper is organized as follows. We present our unified RvNN framework in Section 2. In Section 3, the performance of our system is evaluated and compared with that of previous work. Section 4 concludes the remarks.

2 System Description

The architecture of our united framework for end-to-end Chinese discourse parsing is shown in Figure 1. For a given text, we first segment the text into m text segments $\mathbf{w}^1, \mathbf{w}^2, \mathbf{w}^3, ..., \mathbf{w}^m$ by using punctuation marks as delimiter, where $\mathbf{w}^i = (w_1^i, ..., w_{n_j}^i)$ forms the sequence of words in the ith text segment. The words are fed into an embedding layer, and $\mathbf{w}^i$ is then represented as $\mathbf{e}^i = (e_1^i, ..., e_{n_j}^i)$. Then, an LSTM encoder is trained to convert $\mathbf{e}^i$ into the segment representation $\mathbf{s}^i$, and $\mathbf{s}^1, \mathbf{s}^2, \mathbf{s}^3, ..., \mathbf{s}^m$ serve as the input for the RvNN. Through the RvNN, segments are hierarchically joined to discourse units (DUs) in the bottom-up fashion. Finally, a single discourse parse tree is constructed, and the sense and the centering relations of each join are labeled.
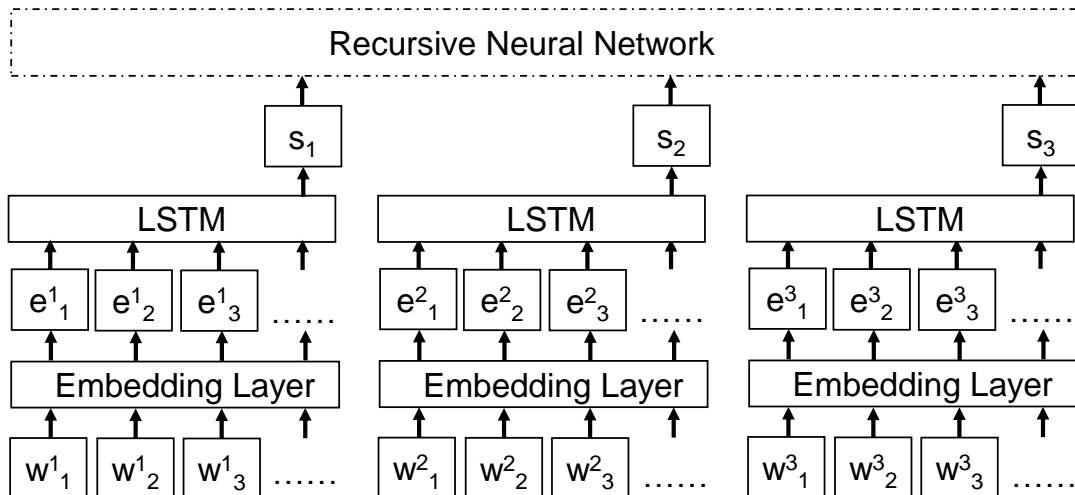

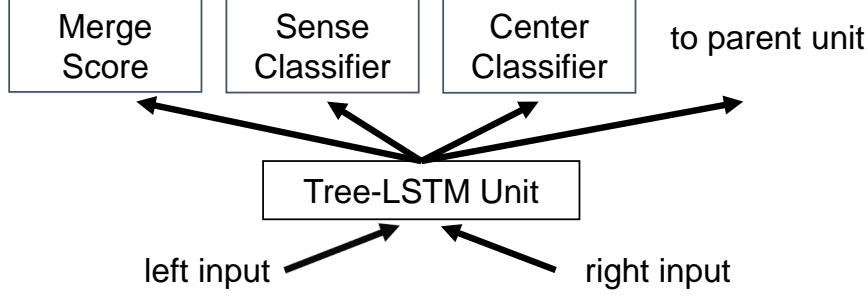Figure 1: Architecture of our unified RvNN discourse parser.

Figure 2: Tree-LSTM unit used in the recursive neural network.

2.1 Recursive Neural Network

Figure 2 illustrates the unit in our RvNN based on the Tree-LSTM unit (Tai et al., 2015). Given the left and the right inputs (i.e. two text segments or two DUs), the Tree-LSTM composition function produces a representation for the new tree node. The Tree-LSTM unit generalizes the LSTM unit to the tree-based input. Similar to LSTM, Tree-LSTM makes use of intermediate states as a pair of an active state representation $\vec{h}$ and a memory representation $\vec{c}$. We use the version similar to (Bowman et al., 2016) as the formula:

$$(1) \quad \begin{bmatrix} \vec{i} \\ \vec{f_l} \\ \vec{f_r} \\ \vec{o} \\ \vec{g} \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \sigma \\ \sigma \\ \tanh \end{bmatrix} \left(W_{\text{comp}} \begin{bmatrix} \vec{h_s^1} \\ \vec{h_s^2} \end{bmatrix} + \vec{b}_{\text{comp}} \right)$$

$$(2) \quad \vec{c} = \vec{f_l} \odot \vec{c}_s^2 + \vec{f_r} \odot \vec{c}_s^1 + \vec{i} \odot \vec{g}$$

$$(3) \quad \vec{h} = \vec{o} \odot \tanh(\vec{c})$$

where σ is the sigmoid activation function, $\odot$ is the element-wise product, and the pairs $\langle \vec{h}_s^1, \vec{c}_s^1 \rangle$ and $\langle \vec{h}_s^2, \vec{c}_s^2 \rangle$ are input from its two children tree nodes. The output of Tree-LSTM is the pair $\langle \vec{h}, \vec{c} \rangle$. Note that the Tree-LSTM unit is designed for binary tree. We handle the multinuclear with the same scheme as Kang et al. (2016).

The representations $\vec{h}$ and $\vec{c}$ produced by Tree-LSTM are taken for the following four usages: merge scoring, sense labeling, center labeling, and the input for the upper Tree-LSTM unit. In the prediction stage, the representation will be first sent into the merge scorer to measure the probability of the join of its two children tree nodes:

$$(4) \quad \vec{p_m} = \text{softmax}(W_m \begin{bmatrix} \vec{h} \\ \vec{c} \end{bmatrix} + \vec{b}_m)$$

The output $\vec{p_m}$ is a 2-dimensional vector, representing the probabilities of to-merge and not-to-merge.

Similarly, the sense classifier and the center classifier compute the probability distributions $\vec{p_s}$ and $\vec{p_c}$, respectively, as follows:

$$(5) \quad \vec{p_s} = \text{softmax}(W_s \begin{bmatrix} \vec{h} \\ \vec{c} \end{bmatrix} + \vec{b}_s)$$

$$(6) \quad \vec{p_c} = \text{softmax}(W_c \begin{bmatrix} \vec{h} \\ \vec{c} \end{bmatrix} + \vec{b}_c)$$

For sense labeling, $\vec{p_s}$ consists of 6 values constituting the probabilities of the following six senses: causality, coordination, transition, explanation, subEDU, and EDU. Our end-to-end parser constructs the

discourse parse tree from the text segments, EDUs, and to non-leaf DUs in a united framework. The first four of the six senses are used to label the discourse relation between DUs, while the last two senses are used for EDU segmentation. For center labeling, $\vec{p_c}$ consists of 3 values constituting the probabilities of the three center categories including front, latter, and equal. Center labeling is only performed on DUs.

2.2 Parser Training

To train the RvNN, the positive instances are the tree nodes extracted from the discourse parse trees in Chinese Discourse Treebank (CDTB) dataset developed by Li et al. (2014b). On the other hand, the negative instances are derived from the ill-joined discourse trees. We select arbitrary two neighboring subtrees and merge them into a new tree. The new tree is regarded as a negative instance if it is inconsistent with the ground-truth. The losses of the merging scorer, the sense classifier, and the center classifier, denoted as $\mathcal{L}_m$, $\mathcal{L}_s$, and $\mathcal{L}_c$, respectively, are measured with cross-entropy. The loss function is defined as:

$$(7) \quad \mathcal{L} = \begin{cases} \mathcal{L}_m, & \text{if the instance is negative} \\ \mathcal{L}_m + \mathcal{L}_s + \mathcal{L}_c, & \text{otherwise} \end{cases}$$

We use stochastic gradient decent (SGD) with the learning rate of 0.1 for parameter optimization.

2.3 Parse Tree Construction

In the prediction stage, we construct the discourse parse tree based on the predictions made by Tree-LSTM. The Cocke–Younger–Kasami (CKY) algorithm (Younger, 1967) is employed to maximize the probability of the whole parse tree. The dynamic programing algorithm simulates the recursive parsing procedure, considering local and global information jointly.

3 Experiments

We compare our model LSTM-RvNN with the baseline model proposed by Kang et al. (2016). To the best of our knowledge, it is the only existing Chinese discourse parser at the paragraph level. We also evaluate our model given the golden EDUs. The standard evaluation tool PARSEVAL (Carlson et al., 2001) is performed to measure the F-score of the tree structure prediction.

Table 1 shows the experimental results. The F-scores of EDU segmentation, parse tree construction (Structure), parse tree construction with sense labeling (+Sense), parse tree construction with center labeling (+Center), and parse tree construction with both sense and center labeling (Overall) are reported. In general, our model outperforms the baseline model in every aspect except EDU segmentation. Even so, the final discourse parse trees constructed and labeled by our model are more accurate.

Model	EDU	Structure	+Sense	+Center	Overall
Baseline with golden EDU		52.3%	33.8%	23.9%	23.2%
LSTM-RvNN with golden EDU		**64.6%**	**42.7%**	**38.5%**	**35.0%**
Baseline	**93.8%**	46.4%	28.8%	23.1%	20.0%
LSTM-RvNN	87.2%	**49.5%**	**32.6%**	**28.8%**	**26.8%**

Table 1: System performances in F-score.

4 Conclusion

This paper demonstrates an end-to-end Chinese discourse parser that performs the CDT-style parsing without the need of external resources such as syntactic parser. We propose a unified framework based on RvNN to model the subtasks jointly. Experimental results show our parser achieves the state-of-the-art performance in the CDTB dataset. We release the source code of our parser with a ready-to-use pre-trained model for the NLP community. To the best of our knowledge, this is the first toolkit for Chinese discourse parsing.

Acknowledgements

References

Samuel R. Bowman, Jon Gauthier, Abhinav Rastogi, Raghav Gupta, Christopher D. Manning, and Christopher Potts. 2016. A fast unified model for parsing and sentence understanding. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL'16)*, pages 1466–1477, August.

Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. 2001. Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue (SIGDIAL'01)*, pages 1–10.

C. Goller and A. Kuchler. 1996. Learning task-dependent distributed representations by backpropagation through structure. In *Proceedings of the 1996 IEEE International Conference on Neural Networks*, volume 1, pages 347–352 vol.1, Jun.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Comput.*, 9(9):1735–1780, November.

Hen-Hsen Huang and Hsin-Hsi Chen. 2012. Contingency and comparison relation labeling and structure prediction in chinese sentences. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL'12)*, pages 261–269.

Yangfeng Ji and Noah A. Smith. 2017. Neural discourse structure for text categorization. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL'17)*, pages 996–1005, July.

Xiaomian Kang, Haoran Li, Long Zhou, Jiajun Zhang, and Chengqing Zong. 2016. An end-to-end chinese discourse parser with adaptation to explicit and non-explicit relation recognition. In *Proceedings of the CoNLL-16 shared task*, pages 27–32, August.

Jiwei Li, Rumeng Li, and Eduard Hovy. 2014a. Recursive deep models for discourse parsing. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP'14)*, pages 2061–2069, October.

Yancui Li, Wenhe Feng, Jing Sun, Fang Kong, and Guodong Zhou. 2014b. Building chinese discourse corpus with connective-driven dependency tree structure. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP'14)*, pages 2105–2114, Doha, Qatar, October. Association for Computational Linguistics.

Christina Lioma, Birger Larsen, and Wei Lu. 2012. Rhetorical relations for information retrieval. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'12)*, pages 931–940.

Annie Louis, Aravind Joshi, and Ani Nenkova. 2010. Discourse indicators for content selection in summarization. In *Proceedings of the SIGDIAL 2010 Conference*, pages 147–156, Tokyo, Japan, September. Association for Computational Linguistics.

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

Kai Sheng Tai, Richard Socher, and Christopher D. Manning. 2015. Improved semantic representations from tree-structured long short-term memory networks. *Association for Computational Linguistics*.

Nianwen Xue, Hwee Tou Ng, Sameer Pradhan, Attapol Rutherford, Bonnie Webber, Chuan Wang, and Hongmin Wang. 2016. Conll 2016 shared task on multilingual shallow discourse parsing. In *Proceedings of the CoNLL-16 shared task*, pages 1–19, TOBEFILLED-Ann Arbor, Michigan, August. Association for Computational Linguistics.

Daniel H Younger. 1967. Recognition and parsing of context-free languages in time n3. *Information and control*, 10(2):189–208.