10 Analyzing the discourse structure behind the text

- Hierarchical top-down RST parsing based on neural networks -

Abstract

Analyzing a discourse structure behind the document is crucial for context aware Natural Language Processing (NLP) tasks including machine translation and automatic summarization. We propose a neural discourse parsing method based on Rhetorical Structure Theory (RST) that regards a document as a constituent tree. Our parser builds RST trees at different levels of granularity in a document and then replace leaves of upper-level RST trees with lower-level RST trees that were already constructed. The parsing is performed in a top-down manner for each granularity level by recursively splitting a larger text span into two smaller ones while predicting nuclearity labels and rhetorical relations. Unlike previous discourse parsers, our parser can be fully parallelized at each granularity in a document and does not require any handcrafted features such as syntactic features obtained from full parse trees of sentences.

RST parsing by splitting text spans

Point Recursively splitting a large span into smaller ones

S_{11} : Snowfall accumulation reached 30cm.
S_{12} : The railway company has been considered to stop the operation.
S_{13} : The temperature dropped to -10 degrees.
S_{14} : Accidents occurred due to the road surface freezing.
S_{15} : The start of the entrance exam has been delayed for two hours.
S_{16} : The exam has been finished without any accident.

P_*: paragraph,
S_*: sentence

Red words: relation

Point

Document: a span consists of paragraphs
Paragraph: a span consists of sentences
Sentence: a span consists of elementary discourse units

Evaluation results on RST-DT corpus

	Unlabeled	Nuclearity	Relation
Proposed	72.0	58.6	46.7
w/o granularity	66.5	53.4	43.3
Previous	68.6	55.9	45.8
Human	78.3	66.8	57.1

Bracket F-score

This is a collaborative research project between Okumura Lab at Tokyo Institute of Technology and NTT CS labs.

References

[1] N. Kobayashi, T. Hirao, M. Okumura, M. Nagata, "Top-down RST Parsing Utilizing Granularity Levels in Documents," in *Proc. of 25th Annual Meeting of Natural Language Processing,* pp. 1002-1005, 2019.

Contact

Tsutomu Hirao   Email: cs-liaison-ml at hco.ntt.co.jp
Linguistic Intelligence Research Group, Innovative Communication Laboratory

Innovative R&D by NTT
Open House 2019