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.

References


Contact

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

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