

<p>どんな研究</p>	<p>文書において、文書内のそれぞれの文どうしは意味を構成するために関わりがあります。こうした文と文間の意味関係を明らかにする修辞構造理論は、文書をラベル付き木としてあらわします。本展示ではそれを自動で解析する修辞構造解析技術についてご説明します。</p>
<p>どこが凄い</p>	<p>修辞構造解析技術は教師あり学習を用いているため、性能の向上には人手で作成した正解データが大量に必要でした。提案法では既存の自動解析結果を組み合わせ疑似的な学習データを作ります。これにより、人手コストをかけることなく解析性能を向上できます。</p>
<p>めざす未来</p>	<p>文脈を解釈することでより自然な要約や翻訳を実現することができます。修辞構造解析技術はこうした文脈を解釈する処理の基盤となります。本技術をさらに発展させ実用に耐えうる解析器を実現することで、要約や翻訳の性能向上をめざします。</p>

文章の修辞構造

修辞構造理論: 文章中の隣接する単文間の修辞関係をラベル付き2分木として表現

木の葉ノード: 単文
木の内部ノード: 単文のまとまり
木の内部ノードラベル: 単文のまとまりの役割
→2種 (中心/周辺)
※周辺→中心に修辞関係
木のエッジラベル: 単文のまとまりの間の**修辞関係**
→18種 (詳細化、根拠、背景など)

要約、翻訳、質問応答など文章を対象とした自然言語アプリケーションで利用される

修辞構造解析の難しさ

修辞関係ラベルの推定
18値分類
ラベル分布に大きな偏り
最大規模のデータですら385文書しかない

データが足りない

提案法: 疑似正解データの活用

正解データ (385文書) 人手作成

ラベルなしデータ (9万文書)

先生1 先生2.....先生K

自動解析結果

疑似正解データ 約17万個

正解データ (385文書)

1. 事前学習
生徒は修辞関係ラベル推定が苦手だけどそれ以外は優秀

2. 追加学習
大量の疑似正解データで事前学習、少量の正解データで追加学習することで解析器を改善

提案法の効果

人間が作成した木に対する正答率

手法 \ 評価指標	内訳			
	全ラベル	木の形	+中心/周辺	+修辞関係
解析器先生 [Wang+17]	58.8	86.0	72.4	59.7
解析器生徒 [Kobayashi+20]	59.6	87.1	74.6	60.0
提案法	62.6 UP!	87.1	75.0	63.4 UP!

修辞関係ラベリングの性能が向上し、全体の性能が向上

翻訳や要約の文章としての首尾一貫性が向上

本研究は東京工業大学、奥村研究室との共同研究の成果です。

関連文献

[1] N. Kobayashi, T. Hirao, H. Kamigaito, M. Okumura, M. Nagata, "Improving Neural RST Parsing Model with Silver Agreement Subtrees," in Proc. 2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2021.

連絡先

平尾 努 (Tsutomu Hirao) 協創情報研究部 言語知能研究グループ
Email: cs-openhouse-ml@hco.ntt.co.jp