

10

文や段落の修飾関係を賢く判断

～ニューラルネットによる階層的トップダウン修辭構造解析～

どんな研究

文書の背後にある談話構造を木として表現する研究を紹介します。文書を段落系列と文系列で構成された入れ子の階層構造としてとらえ、それぞれの階層での修辭構造をニューラルネットを用いて推定します。そして、それらをつなぎ合わせることで文書全体をあらわす修辭構造木を構築します。

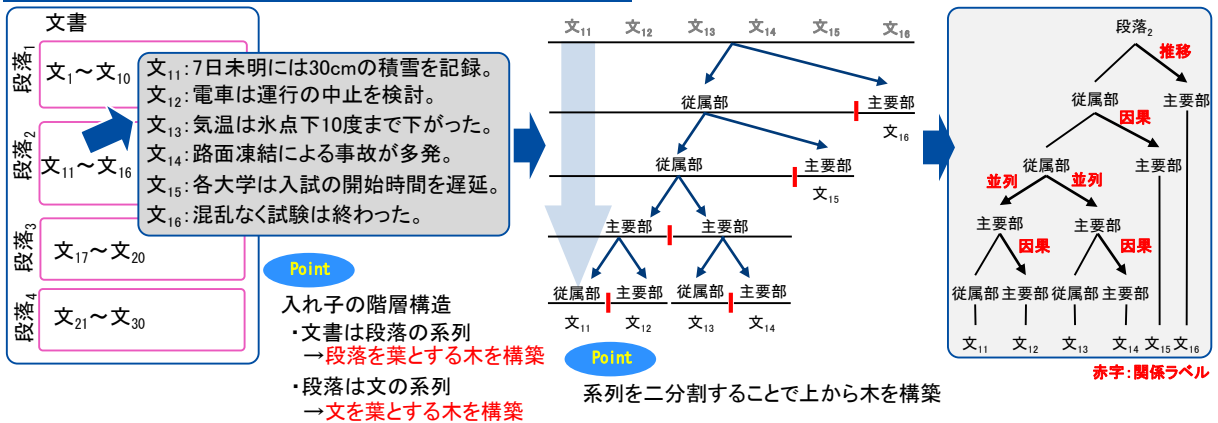
どこが凄い

段落、文系列を独立かつ再帰的に二分割することで木を構築するため、並列化が容易であり長い文書であっても高速に動作することが期待できます。ニューラルネットを用いてテキストデータを自動的に特徴ベクトルに変換するため、人手で複雑な特徴を考える必要がありません。

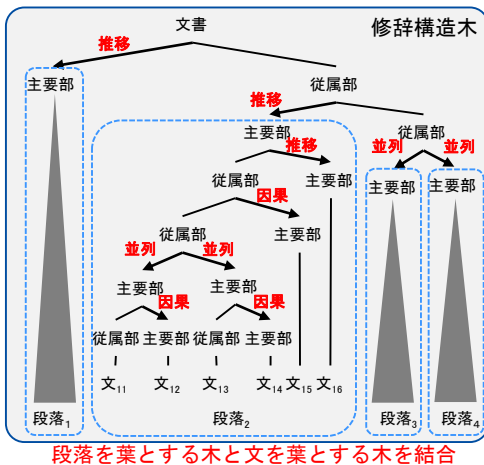
めざす未来

自然言語処理技術は「文」だけではなく「文章」も処理対象として扱う必要があります。例えば、文書を翻訳する際には文脈に応じて適切な訳語を選ぶ、要約する場合には原文書の話題の一貫性を損なわぬようにすることが重要です。本研究はこうした自然言語処理の基盤として必要不可欠です。

テキスト系列の分割によるトップダウンパーズング



修辭構造木の構築



修辭構造解析のベンチマークセットによる評価結果

	木の形	+主/従ラベル	+関係ラベル
提案法	72.0	58.6	46.7
提案法 階層化なし	66.5	53.4	43.3
従来法	68.6	55.9	45.8
Human	78.3	66.8	57.1

Bracket F値: 正解の修辭構造木との一致の割合

本展示は東京工業大学 奥村研究室との共同研究の成果です

関連文献

[1] 小林尚輝, 平尾努, 奥村学, 永田昌明, “階層構造を考慮したトップダウン談話構造解析,” 言語処理学会第25回年次大会, pp. 1002-1005, 2019.

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-liaison-ml at hco.ntt.co.jp



Innovative R&D by NTT

オープンハウス 2019

Copyright © 2019 NTT. All Rights Reserved.