

論文 / 著書情報
Article / Book Information

題目(和文)	修辞構造解析器の高度化に関する研究
Title(English)	
著者(和文)	小林尚輝
Author(English)	Naoki Kobayashi
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第12389号, 授与年月日:2023年3月26日, 学位の種別:課程博士, 審査員:奥村 学,熊澤 逸夫,中山 実,篠崎 隆宏,船越 孝太郎
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12389号, Conferred date:2023/3/26, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

(博士課程)

論文審査の要旨及び審査員

報告番号	甲第	号	学位申請者氏名	小林 尚輝		
論文審査 審査員		氏 名	職 名		氏 名	職 名
	主査	奥村学	教授	審査員	船越孝太郎	准教授
	審査員	熊澤 逸夫	教授			
		中山実	教授			
		篠崎隆宏	准教授			

論文審査の要旨（2000 字程度）

本論文は「修辞構造解析器の高度化に関する研究」と題し、全5章より構成されている。

第1章「序論」では、文書の構造を解析する言語処理課題である修辞構造解析について説明し、解析器を学習するにあたって必要となる要素を解析アルゴリズム、特徴量、データセットの3つに分けて整理している。また、解析器の高度化に向けた2つの提案手法についての概要をそれぞれ説明している。1つ目の提案手法では、文書中の係り受け構造を正確に解析するために重要となる木の上部における解析性能を改善するために、新しく下向きの解析アルゴリズムを修辞構造解析へと導入することに関して、2つ目の提案手法では、修辞構造解析における学習データの不足を補うために疑似正解データセットの構築に取り組み、部分木を活用した新しいデータ拡張手法を提案することに関して述べている。

第2章「関連研究」では、これまでの解析器の発展および変遷について、“解析アルゴリズムおよび分類器の変遷”と“データ拡張に関する研究”の2つに整理して説明している。また、上述した2つの提案手法に関して、これまでの既存研究の中から強く関連するものを詳しく述べ、提案手法と比較している。

第3章「スパン分割に基づく下向き修辞構造解析」では、従来のShift-Reduce法による上向きの解析手法の問題点として、解析誤りの伝播が引き起こす木の上部の解析性能の低下と、それによる係り受け構造の劣化を指摘し、文内の句構造解析で用いられるスパン分割に基づく下向きの解析手法を修辞構造解析へと導入している。また、解析誤りの伝播を抑制するために、文書に含まれる段落構造を活用した文と段落による解析空間の分割を提案している。提案手法の評価実験では、修辞構造解析におけるベンチマークとして用いられているRST-DTデータセットにおいてShift-Reduce法に基づく従来手法とスパン分割に基づく提案手法を比較し、木の構造および核性の推定において従来手法を上回る性能を発揮することを確認している。また、文と段落を用いた階層化においても階層化を行わない場合と比較して解析性能が改善されたことから段落を用いる階層化の有効性を示している。

第4章「部分木による疑似正解データセットを活用した修辞構造解析」では、修辞構造解析が文書の構造を扱う課題であることから、教師データの作成コストが大きくサイズが小さいという課題を解決するために人手のアノテーションを必要としない疑似正解データセットの構築に取り組んでいる。self-trainingの一種で複数の教師解析器を用いるtri-trainingを修辞構造解析のデータ拡張に適用するにあたって、木全体ではなく部分木を対象として複数の教師解析器間の一致を判定することにより、大規模かつ質の高い疑似正解データセットを構築する手法を提案している。また、一致した部分木の抽出は木の捜査に基づいて高速に列挙できることも示している。提案手法の評価実験では、部分木を利用した疑似正解データセットは従来手法に比べ短い学習時間で高い学習効果を発揮することを確認している。また、修辞関係ラベルごとの評価や疑似正解データセットの量を変化させた比較実験を通して疑似正解データセットが修辞関係ラベルの推定性能を改善することに有効であることを示している。

第5章「結論と今後の課題」では、本研究の結論と、第3章と第4章で提案した手法に対する課題と今後の展望について述べている。

以上を要するに、本論文は、文書の構造を解析する修辞構造解析器の高度化のために解析手法およびデータ拡張手法の二つの方向性で新しい手法を提案しその有効性を検証しており、工学上貢献するところが大きい。よって博士（工学）の学位を授与するに十分な価値を持つものと認められる。

注意：「論文審査の要旨及び審査員」は、東工大リサーチリポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容