

論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Citation Block Determination in Academic Texts
著者(和文)	Kaplan Dainan
Author(English)	Dainan Kaplan
出典(和文)	学位:博士(学術), 学位授与機関:東京工業大学, 報告番号:甲第10254号, 授与年月日:2016年3月26日, 学位の種別:課程博士, 審査員:徳永 健伸,徳田 雄洋,宮崎 純,村田 剛志,藤井 敦
Citation(English)	Degree:Doctor (Academic), Conferring organization: Tokyo Institute of Technology, Report number:甲第10254号, Conferred date:2016/3/26, 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	Dainan Kaplan	
論文審査 審査員		氏名	職名	氏名	職名
	主査	徳永 健伸	教授	藤井 敦	准教授
	審査員	徳田 雄洋	教授		
		宮崎 純	教授		
村田 剛志		准教授			

論文審査の要旨 (2000 字程度)

本論文は“Citation Block Determination in Academic Texts”と題し、英文 9 章から構成されており、学術論文において先行研究を引用している部分をテキストの一貫性の性質を用いて同定する手法とその評価について論じている。

第 1 章 “Introduction”では、本研究の背景と動機およびその目的について述べている。インターネット上の情報が爆発的に増加し、必要な情報へのアクセスがますます困難になっている「情報洪水」の状況は学術論文の世界にもあてはまり、学術論文の要約技術は学術の進歩のためには必要不可欠であると述べている。そのためのひとつのアプローチとして、要約の対象となる学術論文が他の論文中で引用されている部分を複数の論文から抽出し、これらを集積することにより被引用論文を要約することが可能であると述べている。このような要約は従来の対象論文から重要な情報を抽出するアプローチに比べ、研究分野への貢献や多視点による洞察といった有益な情報を含むことができると論じている。本研究はこのような引用に基づく要約を作成するための基盤として、通常、通番や著者名と出版年の組み合わせで表示される引用標識 (citation anchor) とその論文に言及している引用部分 (citation block) を学術論文の中から抽出することを目的としていると述べている。

第 2 章 “Problem Setting and Examples”では、本研究で用いる諸概念を定義し、研究の主な対象となる引用部分について学術論文コーパスを分析・分類した結果を例とともに示している。その上で、引用部分が引用標識を含む文にとどまらず、複数の文にまたがることもあるため、引用部分と非引用部分の境界を同定することが重要であると述べている。

第 3 章 “Related Work”では、学術論文の引用関係を利用した関連研究について、引用分析、引用内容の抽出・要約、引用関係を利用した内容抽出、引用関係を利用した研究領域のサーベイの自動生成などに分類し、調査をおこなっている。その結果、本論文で対象としている引用部分の同定を主目的とする研究はほとんどないことと、各研究の問題点を明らかにしている。また、本研究で利用するテキストの一貫性についても、その基礎的な概念について解説している。

第 4 章 “Experimental Setup”では、以下の章でおこなう種々の実験の基盤となる実験環境について述べている。特に実験に用いる学術論文のコーパスの諸性質、利用する機械学習の手法、実験結果の評価に用いる評価方法について述べている。また、研究の目的がもっとも類似している先行研究を紹介し、これを評価のための基準として採用すると述べている。

第 5 章 “CBD with Citation-intrinsic Features”では、評価基準となる研究で利用している引用標識の特徴をさらに詳細化・語彙化した素性を利用する手法を提案し、これによって引用部分の同定性能を基準となる手法と比較して F 尺度で 0.1 改善できたと報告している。

第 6 章 “CBD with Coreference Features”では、同じ引用部分に含まれる文の間には、照応関係を含む傾向があることに着目し、テキストの照応解析をおこなった後、抽出した照応関係を利用して引

引用部分の同定をおこなう手法を提案している。実験結果の考察では、照応関係を利用する手法は 3 文以上からなる引用部分の同定に有効であること、現在の照応解析器の性能が十分でないために引用部分の同定の性能が十分に引出せないことを議論している。

第 7 章 “CBD with Textual Coherence” では、前章の結果を受けて、照応関係を含むより広い概念であるテキストの一貫性に関する言語学の知見を機械学習のモデルに導入し、引用部分を同定する手法を提案している。本章では、テキストの一貫性を表現する具体的な素性として、前章の照応関係に加え、文内で言及されている各実体の文法役割の推移 (Entity-grid) に基づく文間の類似度、相互情報量に基づく文間の類似度、トピックモデルに基づく文間の類似度、文間の談話関係を導入している。評価実験の結果、相互情報量に基づく文間の類似度がもっとも効果が高く、評価基準より F 尺度で 0.058 の改善ができたと報告している。

第 8 章 “CBD with Combined Features” では、第 5 章で導入した引用標識の特徴と第 7 章で導入したテキストの一貫性に関する特徴を組み合わせる手法を提案している。評価の結果、両者を組み合わせることによって、5 章、7 章のそれぞれの結果と比較して F 尺度では顕著な改善がみられなかったが、再現率が改善できることがわかったと述べている。さらに誤り分析の結果、テキストの一貫性を利用する際に注意すべき点として、テキスト全体の大域的な一貫性と引用部分の局所的な一貫性を区別することが重要だと述べている。

第 9 章 “Conclusion and Future Work” では、本論文が学術論文の引用解析の研究分野に貢献した項目をまとめ、結論として要約している。特にテキストの一貫性の性質を用いて引用部分を同定する手法を提案した点が本論文の貢献であると述べている。また、本論文の研究成果をさらに発展させるための方向性についても示している。

以上要するに本論文は、テキストの一貫性の性質を利用して学術論文の引用標識とその引用部分を抽出する新しい手法を提案し、実データによる評価を通してその有効性を示している。本論文の成果は、学術論文の引用関係を利用した要約を作成するための基盤技術として位置付けることができ、学術上貢献することが大きい。よって本論文は、博士(学術)の学位論文として十分価値があるものと認める。

注意: 「論文審査の要旨及び審査員」は、東工大リサーチポジトリ(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で作成してください。