

論文 / 著書情報
Article / Book Information

題目(和文)	単一文書要約の高度化に関する研究
Title(English)	
著者(和文)	菊池悠太
Author(English)	Yuta Kikuchi
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10377号, 授与年月日:2016年12月31日, 学位の種別:課程博士, 審査員:高村 大也,新田 克己,寺野 隆雄,奥村 学,小野 功
Citation(English)	Degree:., Conferring organization: Tokyo Institute of Technology, Report number:甲第10377号, Conferred date:2016/12/31, 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

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	菊池悠太	
論文審査 審査員		氏名	職名	氏名	職名
	主査	高村 大也	准教授	小野 功	准教授
	審査員	新田 克己	教授		
		寺野 隆雄	教授		
		奥村 学	教授		

論文審査の要旨 (2000 字程度)

文書要約は、入力として単一の文書あるいは複数の関連した文書集合を受け取り、その内容を短くまとめた要約を出力する課題である。長い歴史をもつ文書要約研究のうち、本研究が具体的に取り組む手法は、単一文書を対象とした抽出に基づく要約手法（抽出型要約）である。抽出型要約の抽出単位としては、文からはじまり、近年ではより高い圧縮率を実現するために文圧縮などの技術で文の一部のみを要約に組み込む手法が盛んに研究されている。また、古典的な要約の生成手段は、抽出単位の重要度の高いものから選択していくものであったが、今日においては要約を整数計画問題（ILP）として定式化し大域最適解を求めることで情報の網羅性を高める枠組みが支配的な手段となっている。また、ILP の登場以前から蓄積されていた知見やアイデアを、ILP という新しい枠組みに適用する試みも近年行われるようになってきている。また抽出単位の重要度を決定する方法として、機械学習を利用する試みは古くからあるものの、大量の訓練データの不足という問題が常に存在してきた。

本論文は、単一文書を対象とする要約課題において、これまでの研究の歴史を踏まえた上で残された重要な課題のうち、大きく二つに焦点を当てている。本論文は、それぞれの問題に有効な貢献をすることで単一文書要約を高度化することを目的としており、「単一文書要約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、本研究の背景および目的を述べている。これまでの文書要約研究の歴史と残されている課題について、単一文書要約と複数文書要約の違いを中心に振り返り、それらを踏まえ、文間の依存関係を利用した重要箇所の同定、および大規模要約資源の活用という、本研究で取り組む二つの課題について説明している。

第2章「関連研究」では、本研究に関連する従来の研究について紹介している。従来提案されてきた ILP による定式化の例や文圧縮の方法、談話構造の利用などいくつかの技術のほか、文書要約における機械学習の利用事例を中心に説明している。

第3章「文間の依存関係を考慮した文抽出と文圧縮の同時最適化手法」では、本研究で取り組んだ一つ目の課題について述べている。従来高い精度が確認されている文抽出と文圧縮の同時最適化モデルに、新たな構造的制約として修辭構造に基づく文と文の間の依存関係を組み込む新たなモデルを提案している。提案手法を従来の同時最適化モデルや木制約付きナップサック問題による要約手法と比較評価し、文書要約の自動評価指標である ROUGE において最高精度が得られることを確認している。また、高い圧縮率が要求される要約設定へのさらなる柔軟性を獲得するため、文圧縮の制約を緩和するよう拡張を加え、その有効性を示している。

第4章「要約器の効果的な訓練のための大規模要約資源の活用手法」では、本研究で取り組んだ文書要約研究における二つ目の課題について述べている。ターゲットとなる少数の整備された訓練データに加え、大規模であるが整備されていない要約資源である New York Times Annotated Corpus (NYTAC) がある状況で、後者を有効に要約器の訓練に活用するための手法を提案している。ドメイン適応の分野から標準的な5つの手法を取り上げ、さらに訓練する要約器の特性に併せた事例のフィルタリングをオプションとして用意することで、それらの組み合わせが訓練に与える影響を確かめている。実験の結果、要約器を NYTAC で一度訓練したあと、そのパラメータを初

期値として所望のターゲットデータで追加的に訓練する手法が有効に働くという知見を得ている。加えて、使用する要約器の特性に合わせた事例選択を事前に行うことで、一部の評価データセットでは更に精度が大きく向上することを確認している。

第5章「結論と今後の課題」では、本研究において単一文書要約の高度化のために取り組んだ二つの課題について達成した内容をまとめ、今後の展望や残された課題について述べている。ここでは、今回取り組んだ課題にかぎらず、文書要約研究全体を通して今後取り組むべき問題についてまとめている。

以上を要するに、本論文は、文間の依存関係を利用した重要箇所同定の同定、および大規模要約資源の活用という、単一文書要約の高度化に関する二つの課題に焦点を当て、前者については文間依存木と単語間依存木の入れ子依存木上の最適化問題として要約を定式化することにより、後者についてはドメイン適応の手法を援用しかつ有用な訓練事例を選択する機構を考案することで、単一文書要約の性能を向上させた。文間の依存関係と単語間の依存関係を同時に有効利用した研究、および大規模要約資源を訓練データとして有効利用した研究はこれまでになく、これらの枠組みにより単一文書要約を高度化したことは、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も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で作成してください。