

論文 / 著書情報
Article / Book Information

題目(和文)	ニューラル文書要約の高度化に関する研究
Title(English)	
著者(和文)	牧野拓哉
Author(English)	Takuya Makino
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11676号, 授与年月日:2020年12月31日, 学位の種別:課程博士, 審査員:奥村 学,船越 孝太郎,熊澤 逸夫,高村 大也,篠崎 隆宏
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第11676号, Conferred date:2020/12/31, 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	牧野 拓哉	
論文審査 審査員		氏名	職名	氏名	職名
	主査	奥村 学	教授	篠崎 隆宏	准教授
	審査員	熊澤 逸夫	教授		
		高村大也	教授		
	船越孝太郎	准教授			

論文審査の要旨 (2000 字程度)

文書要約技術は与えられた文書を簡潔にした短い文書を作成する技術であり、たとえば、テレビ局や新聞社における Web ページや電光掲示板といった各媒体向けの新聞記事の要約作成での応用が期待されている。これらの要約作成は媒体ごとに定義された要約長制約などに従った人手による作業であり、文書要約技術などによる負荷軽減が求められている。

本論文は「ニューラル文書要約の高度化に関する研究」と題し、全 6 章より構成されている。

第 1 章「序論」では、文書要約手法が抽出型手法、生成型手法に分類されることおよび、それぞれの特徴について説明している。また生成型手法の中でも特に、原文書からの単語のコピーおよび生成用語彙からの単語の生成によって要約を生成可能な生成型ニューラル要約モデルの発展により高い要約性能が得られるようになってきたことを述べている。一方で、既存の生成型ニューラル要約モデルは、実際の要約作成では要約長制約があるにもかかわらず要約長制約を超える要約の生成が少なくない点および、生成用語彙の構築方法によって要約特有な単語の生成割合が低下し、要約性能に影響しうると指摘し、本研究で解くべき課題を述べている。

第 2 章「関連研究」では、抽出型手法と生成型手法の先行研究、評価方法について説明している。抽出型手法では文分類に基づく手法、組み合わせ最適化手法、系列ラベリングに基づく手法についてまとめている。生成型手法では統計的要約手法、生成型ニューラル要約モデルについてまとめている。生成型ニューラル要約モデルに関する先行研究については、本論文で提案する手法との関係について述べている。

第 3 章「生成型ニューラル要約モデルの概要」では、本論文で用いる生成型ニューラル要約モデルの概要について述べている。生成型ニューラル要約モデルは原文書をベクトルにエンコードしたのち、逐次的に単語の生成確率を計算し、確率が最大となる単語を生成用語彙から出力することで要約を作成していることを説明している。

第 4 章「要約長制約下における生成型ニューラル要約モデルの学習」では、既存手法が実際の要約作成で設定される要約長制約を超える要約の生成が少なくないという課題に対して、文書要約の自動評価尺度である ROUGE 値が改善するように要約モデルを学習する minimum risk training に、要約長制約を超える要約の生成に罰則を追加する手法を提案している。提案手法は、従来手法と異なり、要約長制約内で ROUGE 値が高い要約の生成確率を向上させ、ROUGE 値が低い要約および、ROUGE 値が高くて要約長制約を超える要約の生成確率を低下させることを述べている。また、提案手法を英語・日本語のデータで評価し、従来手法よりも要約長制約内での要約生成の割合が増加するとともに、ROUGE 値が改善することを示している。さらには、自動生成要約の人手による後編集評価実験から、要約長制約内で要約を生成することが人手による後編集時間の削減に寄与することを示している。

第 5 章「原文書と要約対の差分に基づく生成用語彙構築」では、既存の生成用語彙構築手法は、要約側に出現する言い換えのような、要約特有な表現が生成用語彙から生成されにくく、要約性能に影響しうるという課題に対して、要約側にだけ出現する単語に着目した生成用語彙構築手法を提案している。提案手法は要約に高頻度で出現する単語を対象とした既存の生成用語彙構築手法と異なり、原文書からのコピーによって要約に含めることのできる単語を除きつつ要約特有の単語を生成用語彙に含めるために、原文書と要約の差分をとることで、要約にのみ出現する高頻度語を用いて生成用語彙を構築する方法であることを説明している。提案手法を 2 つの英語のベンチマークデータで評価し、従来の生成用語彙構築手法と比較し、高い ROUGE 値を得つつ、要約特有の表現を多く生成できることを示している。

第 6 章「結論と今後の課題」では、本研究の結論と今後の課題を説明している。結論では本論文で提案した 2 つの手法によって、生成型ニューラル要約モデルの課題を改善していることを述べている。今後の課題では、本論文では取り組んでいないものの、残された重要な課題を述べている。

以上を要するに、本論文は、要約長制約下での生成型ニューラル要約モデルの学習方法および多様な単語を生成するための生成用語彙構築方法を提案したうえ、提案手法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。