T2R2 東京工業大学リサーチリポジトリ
Tokyo Tech Research Repository

論文 / 著書情報
Article / Book Information

題目(和文)	ニューラルネットワークを用いた単一文書要約に関する研究
Title(English)	
著者(和文)	石垣達也
Author(English)	Tatsuya Ishigaki
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11264号, 授与年月日:2019年9月20日, 学位の種別:課程博士, 審査員:高村 大也,奥村 学,小野 功,中本 高道,長谷川 晶一
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第11264号, Conferred date:2019/9/20, 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

Powered by T2R2 (Tokyo Institute Research Repository)

（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		石垣達也	
論文審査 審 査 員		氏　名	職　名		氏　名	職　名
	主査	高村大也	教授	審査員	長谷川晶一	准教授
	審査員	奥村学	教授			
		小野功	准教授			
		中本高道	教授			

論文審査の要旨（2000 字程度）

文書要約は，入力された文書に対し，その重要な内容を含むより短いテキストを生成する課題である.文書要約は自然言語処理の分野において長らく研究されている課題であり,数多くの既存研究が存在する．本論文ではその中でもニューラルネットワークに基づく自動要約モデルに着目し，要約手法の高度化を目指す方向，適用対象テキストの多様化を目指す方向という，要約研究における二つの方向に対しそれぞれ貢献している.

本論文では要約手法の高度化という方向に対し，入力文書の談話構造を考慮するニューラルネットワークによる要約モデルを新たに提案し，良い性能を示すことを報告している．また，要約対象の多様化という方向に対し，質問を対象とする新たな要約課題を提案し，考えうる様々な要約モデルを適用し，ニューラルネットワークに基づく生成型要約モデルが良い性能を示すことを明らかにしている.

本論文は「ニューラルネットワークを用いた単一文書要約に関する研究」と題し，全 5 章により構成されている.

第 1 章「序論」では，最初に文書要約の必要性および既存の要約研究で用いられている問題設定や手法の概要を述べている．次に要約研究のこれまでの発展は「要約手法の高度化」および「適用対象テキストの多様化」という二つの方向に分類されることを説明している．その後，これら二つの方向に対する本論文の貢献をまとめ，最後に本論文全体の構成を紹介している.

第 2 章「関連研究」では，まず要約手法の分類として抽出型モデルと生成型モデルが存在することを説明している.その後,ニューラルネットワークを用いて抽出型モデルを構築する方法およびその部品であるエンコーダ，出力層について説明している．さらに，ニューラルネットワークを用いて生成型モデルを構築する方法と，その部品であるエンコーダおよびデコーダについて説明している．その上で，本研究に関連する談話構造解析に関する研究についてまとめている．最後に，質問を対象とする要約課題と関連する既存手法について述べている.

第 3 章「談話構造を考慮するニューラル抽出型要約モデル」では，本研究で取り組んだ要約手法の高度化という一つ目の方向に対する研究について説明している．この研究では，文書を文の系列とみなすのではなく，文間の談話構造に着目した木構造として表現し，談話構造木と文の重要度を同時に学習する新たな抽出型要約モデルを提案している．抽出型要約モデルの評価として用いられる DailyMail データセットおよび DUC2002 データセットを用いた評価実験において，自動評価指標である ROUGE-1 値，ROUGE-2 値および ROUGE-L 値が，既存手法の値よりも有意に向上したことを報告している．また，人間にシステム要約を提示し採点させる評価においても，要約に含まれる文間の一貫性に関する平均点数が，談話構造を考慮する機構を導入することにより有意に向上することを確認している.

第 4 章「質問を対象とする要約課題の提案とニューラル要約モデルの適用」では，適用対象テキストの多様化という方向に関し，本研究で取り組んだ質問を対象とした要約課題について述べている．要約手法の高度化の方向において，要約モデルは新聞記事をもとにした共通のデータセットで性能評価されることが多いとしている．そのため，質問など新聞記事とは異なる特徴を持つ

テキストに適用した場合の性能は明らかではないとしている．本研究では，Yahoo! Answers データセットに含まれる質問をフィルタリングし，質問を対象とした要約モデルの学習および評価に用いることのできるデータセットを提案している．このデータセットを用い，規則に基づく要約手法，機械学習に基づく要約手法およびニューラルネットワークに基づく要約手法を比較し，とくに原文書中の単語を要約にもより含めやすくするコピー機構を備え持つ生成型ニューラル要約モデルが，質問を対象とした要約課題に対し良い性能を示すことを確認したとしている．また，従来の要約課題において強いベースラインとして知られている先頭文を抽出する手法が，質問を対象とした要約課題においては質問文を抽出するといった単純な手法よりも性能が低いことを報告している．

第 5 章「結論と今後の課題」では，本研究における課題に対する提案手法の有効性について実験結果に基づく考察を実施し，その結論を述べている．さらに本研究において残された課題について，学習したドメインとは異なるテキストへの適用，生成型要約器の出力テキスト制御，マルチモーダルテキストへの対応，質問に特化した要約モデルに分け，議論している．

以上を要するに，本論文は，「要約モデルそのものの高度化」および「適用対象テキストの多様化」という二つの方向に焦点を当て，前者においては談話構造を考慮する抽出型ニューラル要約モデルを提案することで，後者においては質問を対象とする要約課題の提案およびニューラル要約モデルの適用を行うことで貢献している．また，これらの研究はニューラルネットワークに基づく要約モデルの要約課題における重要性を示すものであり，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充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。