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)	論文要旨
Type(English)	Summary

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

専攻： Department of	知能システム科学 専攻	申請学位（専攻分野）： Academic Degree Requested	博士（工学） Doctor of
学生氏名： Student's Name	石垣達也	指導教員（主）： Academic Supervisor(main)	高村大也
		指導教員（副）： Academic Supervisor (sub)	

要旨（和文 2000 字程度）
Thesis Summary（approx.2000 Japanese Characters ）

本研究では1つの文書から重要な情報を保持したままより短いテキストを出力する単一文書要約課題に取り組む．要約研究には「手法そのものの高度化」,「要約対象の多様化」という 2 つの研究の方向性がある．本研究ではこれらの方向性それぞれに対し，ニューラルネットワークに基づく手法に関する研究を行い貢献した．

1 章では本研究で取り組む課題である単一文書要約について，この技術の応用先，既存手法について
まとめる．

2 章では，ニューラルネットワークに基づく単一文書要約手法について，抽出型手法および生成型手法に分けて説明する．また，要約研究とは独立に研究されてきた談話構造解析や質問を対象とする言語処理課題など，本研究に関連する技術についても述べる．

3 章では，1 つ目の方向性である要約手法そのものの高度化に対する貢献する研究について述べる．この研究では，ニューラルネットワークに基づく抽出型要約モデルが入力文書の文間の意味的な関係を考慮しながら，文の重要度スコアを計算する手法について説明する．既存の談話構造解析機を用いて，共通のデータセットである DailyMail データセットに談話構造に関する情報を付与し実験を行った結果，既存手法よりも出力長制約 75 バイトの設定において既存手法よりも良い性能を示した．

4 章では，2 つ目の方向性である要約対象テキストの多様化という方向性に対し貢献する研究について述べる．この研究では，新たな要約対象として質問を対象とする要約課題を提案した．コミュニティ質問応答サイト Yahoo! Answers に投稿される質問とそのタイトルを，質問とその要約の対とみなし分析を行った．その結果，抽出型手法では要約することができず，生成型手法を必要とする事例の存在を確認した．規則に基づく抽出型手法，機械学習に基づく抽出型手法，ニューラルネットワークに基づく生成型手法をこのデータに対して適用し，ニューラルネットワークに基づく生成型手法が質問要約課題において良い性能を示すことを確認した．

5 章では，4 章までの内容をまとめ今後の方向性について議論する．今後の方向性として，ニューラルネットワークに基づく要約手法の今後の課題としては，学習に用いたドメイン以外のデータにおいて性能の劣化を抑える手法や，デコード時に談話構造を考慮しより文間の意味的なつながりが良い要約を出力するための改善などが挙げられる．

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

専攻： Department of	知能システム科学専攻　専攻	申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of　（工学）
学生氏名： Student's Name	石垣達也	指導教員（主）： Academic Supervisor(main)	高村大也
		指導教員（副）： Academic Supervisor(sub)	

要旨（英文 300 語程度）
Thesis Summary（approx.300 English Words ）

In this research, we tackle the problem of single-document summarization. In this problem, a system outputs a shorter text while keeping important information from a single-document. There are two research directions for this problem; 1. Improving methodology itself and 2. Diversifying target texts to be summarized. In this research, we contributed to each of these directions by conducting research on neural networks based approaches.

In Chapter 1, we introduce the detailed settings of single-document summarization task, the backgrounds and history of this research field.

Chapter 2 describes two different types of neural network-based models for single-document summarization. We first explain a basic extractive method to be extended in our research, and later describe an abstractive method. In addition, we also describe techniques related to our model extension, such as discourse structure analysis and some tasks targeting at questions, which have been studied independently of summarization research.

In Chapter 3, we will describe our contribution to advance the summarization method itself. The basic extractive method explained in Chapter 2 treats a source document as the sequence of sentences. However, sentences in the source documents semantically relate each other, and the relations are shown as important cues to calculate the important scores of sentences. Thus, we introduce a novel summarizer taking into account the relations between sentences; so called discourse-aware summarizer. Our summarizer achieved better scores in terms of ROUGE, a common evaluation metric for this task, in the setting with 75 bytes output summary length constraint.

In Chapter 4, we will describe my research that contributes to the second direction; diversifying targets texts to be summarized. In this research, we proposed a new summarization task targeting summarizing lengthy questions. There is no commonly used dataset for this task. Thus, we fist create the dataset for this task, analyze the data and compared various types of existing models. To create the dataset, we treated the pairs of the questions posted on a question answering site and their titles as pairs of questions and their summaries. In our analysis, we found that there are question-summary pairs that cannot be summarized by using extractive methods, but abstractive method is needed. We compared some rule-based approaches, machine learning-based approaches and the approaches based on neural networks. Our experiments showed that neural network-based approaches achieved higher performance in terms of ROUGE in question summarization task.

In Chapter 5, we conclude the thesis and give some future directions. We observed performance degradation when we apply neural network-based approaches for the text other than the domain used for training. Thus, a method to suppress performance degradation in out-of-domain data is one of the important problems to be solved.

備考 ： 論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。
Note : Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).