

# 博士學位論文

内容の要旨及び審査結果の要旨

第47号

2020年3月

京都産業大学

は し が き

本号は、学位規則（昭和28年4月1日文部省令第9号）第8条の規定による公表を目的とし、令和2年3月21日に本学において博士の学位を授与した者の論文内容の要旨及び論文審査結果の要旨を収録したものである。

学位番号に付した甲は学位規則第4条第1項によるもの（いわゆる課程博士）であり、乙は同条第2項によるもの（いわゆる論文博士）である。

---

---

# 目 次

---

---

## 課程博士

1. 木村 <sup>キムラ</sup> 輔 <sup>タスク</sup> [博士 (先端情報学)] ..... 1
2. 天野 <sup>アマノ</sup> 瑠美 <sup>ルミ</sup> [博士 (生命科学)] ..... 5
3. CHATCHADAWALAI CHOKCHAITAWEESUK [博士 (生命科学)] ... 9

氏名（本籍）	木村 輔（京都府）
学位の種類	博士（先端情報学）
学位記番号	甲先 第1号
学位授与年月日	令和2年3月21日
学位授与の要件	学位規則第4条第1項該当
論文題目	クエリ指向テキスト要約のためのニューラル言語モデルに関する研究
論文審査委員	主 査 宮森 恒 教授
	副 査 平石 裕実 教授
	〃 中島 伸介 教授

## 論文内容の要旨

本論文では3件のクエリ指向型テキスト自動要約がまとめられている。本要旨ではまず研究に至るまでの背景を述べ、データセットの整備状況や既存研究の問題点を指摘する。次に各問題点を解決する提案手法として、文単位の入力文書と文単位の注意機構の導入および適応的注意機構を説明し、新規データセットにおける提案手法の精度検証した成果を報告する。

ビッグデータ時代が到来した昨今、企業が管理する構造化データと比べ、ビッグデータ中を占めるテキストや動画などの非構造化データの割合は80%にもなると言われている。特に1990年代を境に非構造化データが飛躍的に増加していることが報告され、今後も成長が続くと予想されている。テキスト自動要約は非構造化データであるテキストを対象としており、例えば新聞の1つの記事を入力とし、その記事の内容を「明確に表現するタイトル」や「概略をまとめたリード文」などを出力する研究分野である。今後のビッグデータの発展と共に、さらに重要な研究へ成長するといえる。

テキスト自動要約は、入力する文書の数や要約を出力する手法などによって分類でき、いずれの分野も盛んに研究されている。このうち本論文の研究は、単一の入力文書を対象とし、出力要約の内容に着目した生成型であるため、「生成型-クエリ指向型-単一-テキスト自動要約」に分類される。このクエリ指向型では、ユーザの情報要求（クエリ）に応じて、入力文

書中の重要な箇所を選定し、生成する要約を適切に変化させることが必要とされる。またクエリ指向型の入力文書は、文書中のトークン数が多い傾向にあると想定される。これはクエリが変化すれば要約内容も変化することを想定しているため、要約の元となる入力文書が様々なトピックを含む可能性が高いためである。

近年、生成型のテキスト自動要約において **Neural Network (=NN)** によって品質の高い要約を生成できることが報告されている。しかしクエリ指向型テキスト自動要約のデータセットはあまり整備されておらず、また **NN** の学習に用いることが可能な大規模データセットの数も少ないことが報告されている (問題点①)。また **NN** で多く用いられている系列データを長期記憶できる **LSTM** や、入力された系列データの各要素を注視できる注意機構を用いても、**60** トークンを超える文書の符号化の精度が低下する問題点が指摘されている (問題点②)。さらに生成型のテキスト自動要約は入力文書と要約間の内容の乖離が発生しやすい問題も指摘されている (問題点③)。

そこで筆者は、クエリ指向型テキスト自動要約の特性に従った新規データセットの構築 (問題点①の解決案)、トークン数の多い入力文書に頑健な手法 (問題点②の解決案) および入力文書の注視と俯瞰を制御する注意機構 (問題点③の解決案) の提案した。

まず筆者は、地方議会会議録と同会議のニュースレターの対応関係、および、**English Wikipedia** と **Simple English Wikipedia** の対応関係を用いて、それぞれ新規データセットを構築した。また **NN** を用いたクエリ指向型の既存研究では、単にトークン単位の入力文書を **LSTM** によって符号化しているため、依然②の問題点を抱えていた。そこでトークン単位の入力文書に加えて、文単位の入力文書を用いたクエリ指向型テキスト自動要約を提案した。さらに入力文書中の特定トークンへの注視に特化した注意機構は「木を見て森を見ず」の状態を誘発する恐れがあり、これが問題点③の一要因であると考えた。そこで入力文書の俯瞰と注視を制御できる新たな注意機構を提案した。構築した各新規データセットおよび既存研究のデータセットのそれぞれでも実験し、各提案手法の有効性を確認した。特に文単位の入力文書を用いる提案手法は、既存手法と比較して、入力文書のトークン数の増加に対し常に精度が改善することを示した。

## 論文審査結果の要旨

### 論文調査の手続きと調査結果に関して

学位申請者の木村輔（以下、学位申請者とする）から提出のあった博士後期課程学位論文の調査を以下の内容で実施した。

1. 主査 宮森 恒、副査 平石裕実、中島伸介の先端情報学研究科の教授3名により学位論文の査読を行い、研究目的の明確さ、研究方法の妥当性、研究結果の信憑性、考察の妥当性を審査した。
2. 令和2年2月10日に審査員3名の参加の下に、学位申請者自身により40分程度の学位論文の内容のプレゼンテーションを行い、その後1時間弱に渡り、口頭試問を3名の審査員が行った。
3. 学位申請者の退出後、主査の宮森を中心として、審査委員全員で学位論文の科学的妥当性、当該分野への貢献性などを審議した。宮森から学位申請者がこれまで外部に公表した研究業績の説明が行われた。最後に、審査員全員の審査結果を合わせて、以下の最終判断を行った。

審査員全員の合意として、本論文は以下に述べる理由により、博士後期課程学位に十分に値すると判断し、最終審査を合格と判定する。

### 学位論文の内容および本人によるプレゼンテーションに対する評価

学位論文における研究内容は、クエリ指向テキスト自動要約における種々の生成型モデルを提案し、その有用性を大規模なデータセットに基づいた実験と分析を通じて検証するものである。テキスト自動要約の研究は、特定の観点を考慮するかどうかにより、汎用型自動要約とクエリ指向自動要約に分類され、要約をどのように作成するかにより、抽出型と生成型に分類される。従来、汎用型あるいは抽出型についての研究は盛んに行われてきたものの、生成型については、生成した要約の品質を確保することが難しく、特にクエリ指向の生成型についての研究は、未だ新しい試みであり、挑戦的な研究である。

生成型のテキスト自動要約では、要約中の各単語をモデルが逐一生成する必要があるため、全体として意味の通った要約となるように制御することが容易でないという課題がある。また、近年様々な分野で利用と研究が進んでいる深層学習を使った言語モデルを使った場合も、トークン数が数十トークンより長い原文を適切に扱えていない問題点があることが指摘されている。

一方、テキスト自動要約のためのデータセットについては、これまで複数の評価型国際ワ

ークショップにおいてテキスト自動要約がタスクとして取り上げられてきたこともあり、多く整備されてきた。しかし、既存のデータセットは、汎用型テキスト自動要約を志向したものがほとんどであり、クエリ指向テキスト自動要約を想定したデータセットはほとんど存在していない。また、クエリ指向テキスト自動要約のための既存のデータセットは、原文のトークン数が数十トークン程度と比較的少ない。

本来、クエリ指向型では、同一の原文に対し、多様なクエリに応じた要約が求められるため、原文には複数のトピックやさまざまな観点からの内容が含まれることが期待され、必然的にトークン数が多い原文となるのが自然である。

学位申請者は、トークン数の多い原文にも頑健なクエリ指向テキスト自動要約の課題に対し、新たな機構の提案と系統的な実験・分析を大規模なデータセットを用いて行い、従来の研究では報告されていない知見を得た。学位論文において報告されている成果は大きく分けて以下の3点である。

### 1. トークン数の多い原文にも頑健なエンコーダの構築

クエリ指向テキスト自動要約における既存の生成型モデルの課題であった、トークン数の多い原文に対する生成要約の低い品質を向上させるため、文単位のエンコーダの導入と階層的な注意機構を用いる手法を提案し、生成要約の品質が改善することを確認した。

### 2. クエリ指向テキスト自動要約に適したデータセットの整備

これまで存在しなかったクエリ指向テキスト自動要約に適した、トークン数の多い原文を含むデータセットを新たに整備した。本データセットは広く一般に利用できるよう公開される予定である。

### 3. 内容の一貫性をより適切に維持するエンコーダの構築

原文と生成要約の内容の一貫性をより適切に保持することを念頭に、原文の注視と俯瞰の適応的な選択を目指した注意機構を提案し、生成要約の品質がさらに改善することを確認した。なお、提案手法は汎用型テキスト自動要約や機械翻訳にも導入可能な汎用的な機構となるよう配慮されている。

これらの成果は、クエリ指向テキスト自動要約の研究をより望ましい段階に発展させる新しい可能性を示唆するものとして注目される。学位論文においては、これらの知見を導く実験結果および分析内容が説明されていると判断する。

公聴会における学位論文内容のプレゼンテーションでは、研究の背景、目的に関して丁寧な説明が行われた。その後、提案手法と実験内容の説明およびその結果から考察される提案手法の特性の説明があり、最後にこの研究による従来の課題に対する貢献が提示された。審査委員全員の印象として、学位申請者は自らの研究内容を十分に理解しており、研究の目的、意義なども明確であった。また、口頭試問に対しても的確な応答がなされ、その内容も明確であった。学位論文の内容についても、適切であるとの評価であった。