

博士学位論文審査等報告書

審査委員 主査 田伏正佳

副査 吉富康成

副査 細矢 憲

副査 岩崎雅史

- 1 氏名： 大内 智仁
- 2 学位の種類： 博士（理工学）
- 3 学位授与の要件： 学位規程第3条第3項該当
- 4 学位論文題目： 不要文削除及び重要文追加によるデータ拡張に関する研究
- 5 学位論文の要旨および審査結果の要旨

【学位論文の要旨】

別紙に記載

【論文目録】

別紙に記載

【審査結果の要旨】

本論文は、ディープラーニングを用いた生成型自動要約システムにおける学習データのデータ拡張手法について検討し、トピックモデルで評価した重要度により、最も重要でない一文を削除したデータを拡張データとして用いる手法が、比較検討した手法の中で最も効果的であることを明らかにした研究をとりまとめたものである。

第1章では、近年、ネット上の情報量は日々、増加していることを示し、できるだけ効率よく情報の内容を理解するために、自動要約の需要が高まっていることを述べた。そして、自動要約システムにおいて、自然な要約を作るためには、生成型自動要約システムが必須となるが、生成型自動要約システムを構築するには、大量の学習データが必要となる

こと、大量の学習データを作成するにはコストと時間がかかることを述べ、生成型自動要約システムにおいて、少ないデータ量でも、データ拡張手法によって生成型自動要約の精度を高めることが必要となると論じている。

第2章では、自動要約システムとして抽出型自動要約システムと生成型自動要約システムがあることを述べ、それぞれの特徴を示した。そして、本研究で用いた、Copy Mechanism と Coverage Mechanism を組み合わせた Pointer-Generator モデルを解説した。

第3章では、実験で用いたデータセット(CNN/DailyMail)の特徴と評価指標(ROUGE)について述べている。

第4章では、入力記事に対して、トピックモデルを用いて各文の重要度を算出し、最も重要度の低い一文を抜いたものを拡張データとし、Encoder-Decoder モデルに Attention Mechanism を加えたモデルを使って、提案したデータ拡張手法（不要文削除手法）の有効性が確認した。

第5章では、現在、生成型自動要約システムで最もよく使われている Pointer-Generator モデルを用いて、不要文削除手法の有効性を確認した。

第6章では、不要文削除手法と従来から使われているデータ拡張の手法(EDA、LexRank、Luhn)との性能比較をした。結果は、不要文削除手法が最も精度が高かった。

第7章では、不要文削除手法の他に、重要文を追加する手法を提案し、比較した。実験の結果、不要文削除手法が最も良い結果となった。

第8章では、不要文削除と重要文追加を組み合わせた手法を提案し、比較した。実験の結果、不要文削除手法が最も良い結果となった。

第9章では、記事数と拡張の効果の関係性について実験した。実験の結果、記事数が少なければ少ないほど拡張の効果が高いことが示された。

第10章では、第6章で生成した拡張なし手法と不要文削除手法の要約例を示した。

第11章は、本研究のまとめである。最も精度の良い結果が得られたのは、不要文削除手法であった。また、不要文を選択する手法も、Luhn や LexRank によって決めるよりもトピックモデルを用いて選択する手法の方が効果が高かった。

本論文は、生成型自動要約システムを構築する際の学習データの拡張法として新たな手法を提案した。従来の手法より簡易な手法で拡張データが得られ、かついずれの従来法より精度の良い生成型自動要約システムが得られた。以上より、本論文は博士論文の要件を十分に満たすものであると評価できる。

6 最終試験の結果の要旨

本論文の内容は、令和4年2月22日(火)午後1時から2時30分まで稲盛記念会館101講義室において公開の博士学位論文発表会で発表された。口頭発表後、質疑応答が行われた。質問の内容は、不要文削除手法とSR手法の比較、参照要約がない場合の活用法、生成

型自動要約で生成された文の文法的な正しさ、実験結果の評価、評価指標の妥当性、追加学習の可能性など、多岐にわたる内容であったが、いずれの質問に対しても、的確に回答した。最終試験の結果としては、審査委員全員一致で合格とした。

以上