

審査の結果の要旨

氏 名 趙 陽

流通するテキスト文書の量が急激に増加する中、その内容を簡潔に提示する要約技術の重要性が増している。伝統的なテキスト文書要約の研究では、比較的長いテキストから重要な文を選び出す文抽出手法がさかんに検討されてきたが、要約の性能をさらに向上させるためには、文自体を短くする「文圧縮 (sentence compression)」の技術が必要となる。このような文圧縮は、限られた表示空間上に情報を提示するための手段として、今後、多くの場面での活用が期待される。

本論文における文圧縮とは、入力されるテキストから重要な情報を拾い上げ、自然言語として適格な1つの文として再構成する技術である。ここでの中心的な課題は、要約の長さが与えられているという想定のもとで、自然言語としての流暢さを表す指標である「可読性 (readability)」と、要約としての情報量の多さを表す指標である「有用性 (informativeness)」の2つの指標について、出力される圧縮文の質を高めることである。従来の文圧縮の研究は、人手で経験的に作成した変換ルールを最適な方法で組み合わせることで入力文を短くする方法と、事前に準備した参照圧縮文を訓練データとして文圧縮モデルを学習する方法の2つに大別されるが、前者では変換ルールの文脈依存性を考慮できない点が、後者では可読性を考慮した最適化が難しい点が、それぞれ問題となっていた。本論文では、従来研究の抱える上記課題の解決を目的として、ニューラル文圧縮モデルに言語的知識を取り込むための手法を体系的に論じ、実験を通して有用性を検証している。

文圧縮の問題に対する本論文の貢献は以下のようにまとめられる。まず、ニューラル文圧縮において単語レベルの言語的知識を活用する基本的な枠組みを示した。また、言語モデルを用いて文レベルの可読性を数量的に捉えることを提唱し、言語的な知識を活用することで、文圧縮の性能が改善できることを示した。さらに、複数文要約における可読性の問題を解決するため、得られた出力文を書き換えるしくみを新たに導入して、その効果を検証した。

本論文は以下の6章から構成される。

第1章では、文圧縮の概念を導入し研究の動機を述べた上で、現状のアプローチにおける問題を提起し、解決すべき課題を設定している。また、これに基づき論文の貢献をまとめている。

第2章では、文圧縮の関連手法を紹介し、文書要約に関する研究全体の中での文圧縮

の位置づけを述べている。文圧縮は出力文として単一の文を想定するものであるが、入力文が単一の文である場合と、複数の文集である場合の2種類の問題設定があることを紹介し、それぞれについて、従来研究の歴史や現状を詳細に説明している。論文中では、第3章および第4章で単一入力文、第5章で複数入力文の問題を論じている。

第3章では、従来のニューラル文圧縮では言語的な知識が十分に活用できていないことを指摘し、それを解決するためのゲート付きニューラルネットワークモデルを新たに提案している。ここでの言語的知識は、品詞、係り受けラベル、固有表現タグの3種類で構成される単語レベルの情報である。実験では、提案するモデルの優位性を数値的に示すとともに、言語的知識を用いる効果を分析の上、考察している。

第4章では、文の統語構造に関する知識を活用するためのニューラル文圧縮モデルについて論じている。係り受けに基づく文構造情報を扱うために注意機構を導入するとともに、ニューラル言語モデルを用いて出力文の可読性を数量的に評価するしくみを提案し、強化学習の適用による最適化を可能にしている。実験では、評価に用いるニューラル言語モデルの構築時に言語的な知識を導入することで、ニューラル言語モデルの可読性評価指標としての性能が改善し、さらにニューラル文圧縮モデル自体の性能も向上することを示している。また、有用性の評価指標については、エントロピーに基づく手法を検討し、人間による評価との整合性に基つき有効性を示している。

第5章では、複数の文を入力として、重要な情報を残しつつ可読性の点でも問題がない単一要約文を生成するための手法について論じている。既存手法では入力文中に存在する表現だけで出力文を構成する点が、可読性を確保する上での妨げとなっている点を指摘し、いったん得られた出力文をより自然な形に言い換える機能を導入することで、可読性が向上できることを論じている。さらに訓練用データの不足を補うために、既存の言い換え辞書を言語知識として活用する手法を提案し、実験を通して有用性を示している。

第6章では、第5章までの手法や問題について総括するとともに、将来の展望に関して議論を行っている。

このように本論文は、ニューラル文圧縮において言語的知識を活用する方法について体系的に論じたものであり、文圧縮の研究に新たな方向付けを与えるものとして、当該分野の今後の発展に大きくつながることが期待できる。

よって本論文は博士（情報理工学）の学位請求論文として合格と認められる。