

論文内容の要旨

博士論文題目 Automatic Summarization on Various Domains with Combinatorial Optimization and Machine Learning
(組み合わせ最適化と機械学習による多様な分野に対する自動要約)

氏名 西川 仁

(論文内容の要旨)

自動要約の対象となるテキストの種類、量が増加しつつある。これはインターネットの普及に伴い電子化されたテキストが増加したためであり、テキストの多様性が増すにつれて自動要約において新しい課題が生じている。課題は大きく分けて3つ存在する。第一に可読性の問題であり、第二に内容性の問題であり、最後に要約速度の問題である。本研究の目的は、自動要約において新しく生じたこれらの課題に対して、新しい自動要約の方法を提案し、これを解決することである。

本研究の主たる貢献は、これらの問題それぞれに対して解決の方法を示した点にある。まず、可読性の問題に対しては、文の抽出と順序付けを同時に行う要約モデルを新しく提案し、これを整数計画問題として表現し解くことによって、文の抽出と順序付けを別個に行う既存の手法に比べて可読性が改善された要約を生成できることを示す。次に、内容性の問題に対しては、転移学習を利用し異なる分野のテキストを教師事例として用いることで内容性が改善された要約を生成できることを示す。最後に、要約速度の問題に対しては、新しい要約モデルである冗長性制約付きナップサックモデルと、ラグランジュヒューリスティックに基づく高速なデコーディングアルゴリズムを提案し、良好な要約を高速に生成できることを示す。

本論文ではまず、本研究の目的、意義を明らかにする。次に、準備として、整数計画と構造学習について概観する。本論文中で扱われる自動要約の課題はすべて組み合わせ最適化問題の一種として表現されるため、組み合わせ最適化問題を表現しこれを解くための強力な枠組みである整数計画について簡単に説明する。また、自動要約の課題に含まれるパラメータを推定するための方法として、構造学習についても説明する。そののち、自動要約研究について広く概観し、本研究の位置づけを明らかにする。これらの準備ののち、上述した3つの問題について詳述する。第一の可読性に関する問題では、評価文書集合を要約の対象として取り上げ、提案する方法により要約の可読性が改善することを示す。第二の内容性に関する問題では、コンタクトセンタログを要約の対象として取り上げ、提案する方法により要約の内容性が改善することを示す。第三の要約速度に関する問題では、新聞記事集合と評価文書集合を要約の対象として取り上げ、提案する方法により要約速度が改善することを示す。最後に、これらの結果を踏まえ、更なる研究の方向性を示す。

氏名	西川 仁
----	------

(論文審査結果の要旨)

平成 25 年 8 月 2 日に開催した公聴会の結果を参考に平成 25 年 9 月 4 日に本博士論文の審査を行った。以下のとおり、本博士論文は、提案者が独立した研究者として、研究活動を続けていくための十分な素養を備えていることを示すものと認める。

西川 仁は、本博士論文において、可読性・内容性の高い自動要約を効率的に行う手法を提案し、性能評価を行った。本博士論文の貢献は次のようにまとめることができる。

1. 可読性の高い要約を得るため、要約文の抽出とその順序づけを同時に行う要約モデルを提案し、整数計画問題として解くことにより、文の抽出と順序づけをパイプライン的に行う手法よりも可読性の高い要約が生成できることを示した。評価文書集合の要約を題材として取り上げ、同手法の有効性を実験により評価した。
2. 内容性の高い要約を得るために、転移学習を利用し、異なる分野のテキストを教師事例として用いることで内容性が改善された要約を生成できることを示した。また、提案手法をコンタクトセンタログの要約を対象に適用することにより、要約の内容性が改善することを示した。
3. 要約速度の問題に対処するため、冗長性制約付きナップサックモデルと、ラグランジュヒューリスティックに基づく高速なデコーディングアルゴリズムを提案し、条件を満たす要約を高速に生成できることを示した。新聞記事集合と評価文書集合を要約の対象とし、提案する方法により要約速度が改善すること実験によりを示した。

自動要約研究における可読性、内容性、速度の問題を取り扱い、それぞれに対して有効な手法を提案した本研究は、独創性が高く、しかも実用的であり、自然言語処理の分野において高い貢献があると評価する。

よって、本論文は、博士（工学）の学位論文として価値あるものと認める。