

【論文審査の要旨】

テキスト平易化は、難解なテキストの意味を保持したまま平易に書き換えるタスクである。テキスト平易化システムは、言語学習者や子どもをはじめとする人々の文章読解を支援し、他の自然言語処理応用タスクの性能改善にも寄与する。このタスクは、機械翻訳や文書要約などと同じく、テキストからのテキスト生成タスクとして位置づけられる。

テキスト平易化の先行研究には語彙的換言と機械翻訳の2つのアプローチがある。語彙的換言アプローチでは「難解語検出・換言生成・ランキング」のパイプラインで平易な同義文を生成する。機械翻訳アプローチでは機械翻訳器を用いて平易な同義文を生成する。どちらのアプローチでも大規模なパラレルコーパスから平易化規則を獲得する手法が主流であるため、これまでは言語資源の豊富な英語を中心に研究されてきた。しかし、英語以外の多くの言語では平易に書かれた大規模コーパスを利用できない問題がある。

そこで、本研究では、平易な大規模コーパスを利用できない言語、特に日本語を対象として、語彙的換言アプローチと機械翻訳アプローチによるテキスト平易化を実現する。まず、平易なコーパスを用いない語彙的換言アプローチとして、本研究では日本語のための新しい言い換え知識獲得、意味的等価性フィルタリング、平易性フィルタリング、文法性ランキングの各手法を提案する。また、平易なコーパスを用いない機械翻訳アプローチとして、本研究では文の難易度推定と文間類似度推定を組み合わせるコーパスから擬似的なテキスト平易化のためのコーパスを構築し、任意の言語での機械翻訳器を用いたテキスト平易化を可能にする。

実験結果によって、本研究の語彙的換言アプローチが既存の言語非依存な教師なしテキスト平易化手法を上回ることを示す。また、機械翻訳アプローチにおいては、本研究の疑似的な平易化コーパスが既存のテキスト平易化コーパスと同等に機械翻訳アプローチによる平易化に用いることができることを示す。

また、本研究では上記の主要な貢献に加え、さらにテキスト平易化を改善する。まず、言い換え獲得のタスクに取り組む。言い換え獲得は語彙選択アプローチの改善のために重要である。次に、文間類似度タスクに取り組む。文間類似度は機械翻訳アプローチの改善のために重要である。最後に、品質推定タスクに取り組む。品質推定は、テキスト平易化の自動評価尺度の改善のために重要である。これらのタスクの全てにおいて、本研究は世界最高精度を達成した。

本研究の主要な貢献は以下である。

1. 世界最高精度の日本語の語彙平易化手法を提案した。
2. 世界初の日本語の語彙平易化データセットを構築した。
3. 単語の分散表現のアライメントに基づく新しい文間類似度を提案し、平易な文と難解な文のアライメントタスクで先行研究より性能が高いことを示した。
4. 上記の文間類似度を用いて文アライメントを改善し、統計的機械翻訳手法と組み合わせ

せることで、英語のテキスト平易化タスクで世界最高精度を達成した。

5. 上記の文間類似度を用いれば、平易化コーパスを用いない統計的機械翻訳によるテキスト平易化手法でも、平易化コーパスを用いたテキスト平易化手法と遜色ない精度でテキスト平易化が可能であることを示した。
6. 広く用いられている言い換え候補獲得手法に対し、情報理論的な解釈を与え、かつ性能を改善した。

以上のように、本論文では、平易な大規模コーパスが利用できない言語において、生コーパスからテキスト平易化のための知識を獲得し、それを用いた教師なしテキスト平易化手法が教師ありテキスト平易化手法と同等の性能であることを実験により示し、有効性を評価している。提案手法である平易なコーパスを用いないテキスト平易化手法は、様々な言語に低コストで適用可能であり、テキスト平易化の多言語展開の基礎技術となることが期待されるため、工学的に重要な意義があると考えられる。よって、本論文は博士（工学）の学位を授与するに十分な価値があるものと認められる。

（最終試験又は試験の結果）

本学の学位規則に従い、最終試験を行った。公開の席上で論文発表を行い、学内外の教員による質疑応答を行った。また、論文審査委員により本論文及び関連分野に関する試問を行った。これらの結果を総合的に判断した結果、専門科目についても十分な学力があるものと認め、合格と判定した。