審 査 の 結 果 の 要 旨

氏　　名　　岩月　憲一

本論文は、学術論文で用いられる定型表現を、その表層の多様性を確保しつつ検索するための枠組みを提案している。これまで定型表現検索はキーワードマッチングによるものが主流であったが、ユーザの入力したキーワードと文字列の一致度が高い定型表現しか得られないという問題があった。論文執筆支援においては、用途が同じでありながらユーザの入力とは表層が異なる定型表現を提示することが重要である。本論文で提案されている枠組みでは、定型表現の伝達機能を利用することでこれを解決する。本枠組みでは、伝達機能のラベルが付与された定型表現データベースを予め構築しておき、ユーザ入力と伝達機能が同じ定型表現を検索することで、文字列の一致度が低いにもかかわらず伝達機能が同じである定型表現を提示することができる。これを実現するためには、伝達機能ラベル付き定型表現データベースが不可欠である。データベース構築のためには、伝達機能に基づく文分類と、定型表現抽出を行う必要があるが、いずれもその手法は明らかでなかった。本論文は、正解データを構築した上で教師あり学習によって伝達機能に基づく文分類が可能であることを示し、更に固有表現抽出と依存構造解析を組み合わせた定型表現抽出手法を提案している。更に、これらの手法を組み合わせ、伝達機能ラベル付き定型表現データベースを構築している。最終的に、従来のキーワードマッチングによる検索と、提案された枠組みを比較し、伝達機能が同じ多様な定型表現を検索できることを示している。

本論文は、全8章からなる。

第1章では、計算機を用いた英語論文の執筆支援において、定型表現の検索に課題があることを指摘している。続いて、伝達機能に基づいた定型表現検索の枠組みを提案している。

第2章では、学術論文を対象とした自然言語処理分野の研究を俯瞰し、学術論文における定型表現及び伝達機能に関する既存の言語学的研究と自然言語処理的研究を網羅的に説明している。さらに、本論文で用いた各手法の背景について述べている。

第3章では、伝達機能に基づく文分類の訓練・評価用データセット及び定型表現抽出評価用データセットを構築している。また、構築したデータセットの質的評価を行い、伝達機能ラベルが正しく付与されていることを示している。

第4章では、文に対する伝達機能ラベルの付与手法を提案している。第3章で構築した

データセットを用い、学術論文向け言語モデル SciBERT を利用することによって、高精度な文分類が可能であることを実証している。さらに、汎用言語モデル BERT と SciBERT の比較、また異なる分野の論文データを用いた比較を行うことによって、訓練データと分類対象データの論文の分野が異なる場合にも分類精度が保たれることを示している。

第5章では、定型表現抽出手法及び評価手法を提案している。これまで定型表現の評価手法は確立されていなかったが、伝達機能を表現できているかどうかによって評価する手法を提案している。また、固有表現抽出及び依存構造解析を利用した定型表現抽出手法を提案している。既存の定型表現抽出手法との比較実験を行い、提案手法が最も高精度で定型表現抽出を行うことを示している。

第6章では、第4章及び第5章で提案した手法を組み合わせ、大規模な伝達機能ラベル付き定型表現データベースを構築している。次に、このデータベースを用い、既存のキーワードマッチングによる検索と、伝達機能ラベルを用いた検索を比較している。比較の結果、伝達機能ラベルが付与されていることによって、伝達機能を維持したまま多様な定型表現を検索することが可能であることを示している。

第7章では、第6章までの結果をもとに、今後の課題、特に伝達機能体系における伝達機能の粒度と、伝達機能を体現するテキストの単位について議論している。

第8章では、本論文の貢献をまとめ、結論を述べている。

以上のように、本論文は、多様な定型表現を検索可能にするために必要な伝達機能ラベル付き定型表現データベースを、伝達機能ラベル付与手法と定型表現抽出手法を提案することによって構築し、実際に伝達機能に基づく定型表現検索が可能であることを示したものであり、自然言語処理分野における定型表現の研究に新たな方向性を示すものであると評価できる。

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。