

審査の結果の要旨

氏 名 羽鳥 潤

本論文は単語間の区切りに空白を用いない言語を対象とした自然言語処理において、品詞解析・構文解析・機械翻訳等を含むあらゆるタスクを行うための前提として必要不可欠なタスクである単語分割に対して、従来の研究が重要視してこなかった二つの問題に着目し、そのパラダイムを再検討し結合モデルを用いるアプローチにより問題の解決を試みている。第一の問題は、従来の多くの分割モデルにおいては品詞解析等の上位タスクとの関連を無視した分割基準が用いられている問題であり、第二の問題は、現在最高水準にある分割モデルのほとんどが局所的な情報のみに基づき、大域的な形態・統語論的情報を用いていない問題である。本論文では、それぞれの問題に対する深い洞察から、タスク間の結合モデルに基づく単語分割と上位タスクの相互作用を捉える新規手法を開発することで、第一の問題に対しては、上位タスクに基づく単語分割の自動最適化法を提案し、日本語の読み推定への応用例では様々な分割単位を同時に考慮し、読み推定モデル自身に最適な分割単位を選択させることにより、未知語と既知語の読み推定を単一のモデルで行うことに成功し、既存手法を上回る精度を記録している。第二の問題に対しては、単語分割・形態素解析・構文解析の3タスクを同時に解析する結合モデルを構築することでタスク間の大域的な相互作用を効果的に組み入れることに成功し、中国語ツリーバンクデータへの適用例では、各タスクの大幅な精度向上に成功している。

本論文は六章からなり、第一章では、問題の背景ならびに動機を明らかにし、また提案手法であるタスクに基づく単語分割単位の最適化および、単語分割タスクへの統語論的情報の統合の概要を述べている。第二章では、前提知識となる自然言語処理に関わるタスクの設定およびパラメータ推定に用いられる機械学習アルゴリズムについての説明をしている。第三章では、日本語読み推定タスクを適用例とした、結合モデルに基づく単語分割最適化法を提案し、様々なドメインの文章に対する実験の結果、特に未知語を多く含む文章に対して有意な精度向上が認められ、自由な単語単位による解析が柔軟で高精度な読み推定を可能にすることを明らかにしている。第四章では、品詞解析・構文解析結合モデルに対して、Shift-Reduce法を拡張した初めての漸進的アルゴリズムを提案している。そこでは漸進性を保ちつつ先の情報を用いる困難さを克服するために素性の遅延評価を導入し、漸進性を持たない関連研究に比べて計算量が有利なことを示し

ている。また中国語ツリーバンクデータに適用した結果、品詞解析・構文解析の精度が既存手法に対して大幅に向上することを明らかにしている。第五章では、第四章で提案した結合モデルを拡張して、単語分割・形態素解析・構文解析の3タスクを同時に解析する初めての結合モデルを提案している。中国語ツリーバンクデータに適用した実験の結果、統語的な依存関係を利用することは、単語分割・品詞解析の精度向上に大きく寄与することを明らかにし、特に未知語に対して顕著な効果があることを示している。また構文解析の精度自体も誤り伝播の問題が軽減することから大きく向上することを示している。第六章では、本論文で提案した手法の貢献を特に未知語の処理を中心に単語分割非依存タスク、単語分割依存タスクに分けて述べ、最後に全体としての総括を行い今後の発展性への展望を述べている。

本論文は、単語間区切り文字を用いない言語のあらゆる処理の基礎となる問題に対して大きな貢献をなすものであり評価に値する。

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。