

琉球大学学術リポジトリ

低資源言語処理への機械学習および統計手法の適用
に関する研究、事例：ダリ語とパシュート語

メタデータ	言語: 出版者: 琉球大学 公開日: 2021-11-17 キーワード (Ja): キーワード (En): 作成者: Dawodi, Mursal メールアドレス: 所属:
URL	http://hdl.handle.net/20.500.12000/50046

令和 3年 8月 4日

琉球大学大学院
理工学研究科長 殿

論文審査委員

主査 氏 名 和田 知久
副査 氏 名 名嘉村 盛和
副査 氏 名 岡崎 威生



学位（博士）論文審査及び最終試験の終了報告書

学位（博士）の申請に対し、学位論文の審査及び最終試験を終了したので、下記のとおり報告します。

記

申請者	専攻名 総合知能工学 氏名 Dawodi, Mursal 学籍番号 [REDACTED]	
指導教員名	和田 知久	
成績評価	学位論文 <input checked="" type="radio"/> 合格 <input type="radio"/> 不合格	最終試験 <input checked="" type="radio"/> 合格 <input type="radio"/> 不合格
論文題目	A Study on Applicability of Machine Learning and Statistical Approaches on Low Resources Language Processing: Case Dari and Pashto 低資源言語処理への機械学習および統計的手法の適用に関する研究、事例：ダリ語とパシュート語	
審査要旨（2000字以内）	情報技術が進展し、音声認識や、文書の分類など計算機を用いた言語処理が一般的に用いられるようになった。英語のような、メジャーな言語では辞書や語彙目録のようなデータベースが充実しており複雑な言語処理可能となり研究がなされている。 本研究では、アフガニスタンの公用言語であるダリ語とパシュート語という	

(次頁へ続く)

審査要旨

データベースが充実していない言語に対して、計算機による言語処理特に、低資源な計算処理での処理の可能性を検討したことにある。ダリ語は右から左に記載する言語で、また文語と口語で大きな違いがあり、アクセントも地域で大きく異なり、接頭語や接尾語での変化も大きく複雑な言語である。パシュート語もアフガニスタンの公用語であるが、イラン系の言語であり、発音が異なる文字や、地域によって綴りの変化など特徴のある言語である。

コンピュータを用いた言語処理の中で、本研究では統計的手法と、機械学習や深層学習アルゴリズムを以下の4つの応用分野に対して適用してその有効性や性能差の定量的比較を行った。4つの応用とは、①自動音声認識、②文書のカテゴリ分類、③文書のポジティブ・ネガティブなどの分類、④文書の要約生成システムである。

特にダリ語の自動音声認識では、90%を超える高い認識率を深層学習ディープラーニングニューラルネットでの実現に成功し、このような言語に対する自動応用処理の可能性を示した。ただ、パシュート語ではやや性能が低く、言語特性による今後の改善の示唆を与えている。

文書のカテゴリ分類では、歴史、ニュース、スポーツ、技術などの8程度のカテゴリ分類システムを構築し、従来の統計手法と機械学習での定量的比較結果を示している。処理本体の前に効率的なシステム構築のために、事前処理を導入し、その有効性を示している。メジャーでない言語、特徴の大きく異なる言語などに対しても今後益々コンピュータによる自動化は工学的に有効であり、今回ダリ語とパシュート語を事例とし、低資源言語処理の有効性を定量的に示し、機械学習および統計的手法の比較を明らかにしている。

以上の結果により、本研究成果は工学的に有用であり、提出された学位論文は博士の学位論文に相当するものと判断し学位論文の審査を合格とする。また、2021年8月2日15:00～16:00に開催した論文発表会における発表ならびに質疑応答において、申請者は専門分野および関連分野の十分な知識ならびに十分な研究能力を有していることが確認できたので最終試験を合格とする。