

氏名	龍 梓			
学位の種類	博士（工学）			
学位記番号	博 甲 第 8 5 3 3 号			
学位授与年月日	平成 3 0 年 3 月 2 3 日			
学位授与の要件	学位規則第 4 条第 1 項該当			
審査研究科	システム情報工学研究科			
学位論文題目	機械翻訳における大規模フレーズ翻訳知識の獲得と 利用に関する研究			
主査	筑波大学 教授	博士(工学)	宇津呂 武仁	
副査	筑波大学 教授	工学博士	丸山 勉	
副査	筑波大学 教授	博士(工学)	矢野博明	
副査	筑波大学 教授	博士(工学)	古賀弘樹	
副査	筑波大学 教授	博士(工学)	山本幹雄	

論 文 の 要 旨

本論文では、機械翻訳において翻訳対象となる文書ジャンルの中でも、最もその重要性が大きいジャンルの一つである特許文書の翻訳をとりあげている。大規模対訳特許文を情報源として、機械翻訳における大規模フレーズ翻訳知識の獲得と利用の観点から、「対訳専門用語の同定」、「同義対訳専門用語の同定」、「統計的機械翻訳による大語彙フレーズ翻訳との併用によるニューラル機械翻訳」の課題について取り組んでいる。

「対訳専門用語の同定」、および、「同義対訳専門用語の同定」の課題においては、機械翻訳、特に、RBMT(人手によって作成された規則に基づく機械翻訳)のパラダイムにおいて、利用するための大規模フレーズ翻訳知識の獲得手法に取り組んでいる。「対訳専門用語の同定」の課題においては、SMT(統計的機械翻訳)が持つ特異な機能として、大規模フレーズ翻訳テーブルの自動生成機能に着目し、これを利用して、大規模フレーズ翻訳辞書作成支援方式を考案している。「同義対訳専門用語の同定」の課題においては、「対訳専門用語の同定」の課題において、異なる対訳文から得られた専門用語対訳対の関係性に着目し、同義関係にある対訳専門用語を同定している。

一方、「統計的機械翻訳による大語彙フレーズ翻訳との併用によるニューラル機械翻訳」の課題においては、NMT(ニューラルネットワーク機械翻訳)のパラダイムにおいて大規模フレーズ翻訳知識を利用する方式について取り組んでいる。この方式においては、標準的なNMTモデルにおいて扱える語彙のサイズに限界があり、モデル中に低頻度語彙を含むことができない点に着目し、SMTの長所である大語彙フレーズ翻訳、および、NMTの長所である自然かつ流暢な訳文生成の両方の長所を併せ持ったNMT方式を実現している。

審査の要旨

【批評】

本論文で明らかにした事項は以下の通りである。

(1)「対訳専門用語の同定」の課題においては、まず、基盤的技術として、対訳文およびフレーズ翻訳テーブルを用いた訳語推定を行い、フレーズ翻訳テーブルに含まれる大部分の誤訳・ノイズを排除して対訳専門用語の候補を獲得する方式を確立している。そして、複数の対訳文から得られた素性を用いた分類器学習方式を適用することによって、対訳専門用語の同定を行う方式を考案している。この方式の評価を行い、適合率を最大化する調整を行った場合に 90%以上の適合率を、一方、F 値を最大化する調整を行った場合に 80%以上の F 値を、それぞれ達成している。

(2)「同義対訳専門用語の同定」の課題においては、原言語・目的言語方向、および、目的言語・原言語方向、の両方向において専門用語の訳語推定を繰り返すことによって、人手で選定した中心的対訳対に対して同義の可能性のある対訳対を網羅する同義対訳専門用語候補集合を生成する技術を確立している。そして、分類器学習方式を適用することによって同義集合および異義集合を同定する方式を提案している。評価実験においては、日中パテントファミリーから抽出した 360 万対の日中対訳文に対して提案手法を適用し、同義関係にある日中対訳専門用語の同定において、再現率が 25%以上という条件のもとで、約 90%の適合率を達成している。

(3)「統計的機械翻訳による大語彙フレーズ翻訳との併用によるニューラル機械翻訳」の課題においては、標準的な NMT モデルにおいて扱える語彙のサイズに限界があり、いかに流暢な訳文を生成できたとしても、特許文書の正確な翻訳において不可欠である新語・固有表現が未知語となって訳出されないという問題に取り組んでいる。この問題に対して、訓練用対訳文からフレーズ間の二言語対応の情報を収集し、二言語間で対応済みのフレーズ対訳対を同一のトークンに置き換えた後、NMT モデルの訓練を行う方式を考案している。そして、翻訳時において、NMT モデルの語彙集合によって対応可能な部分については NMT モデルによる訳文生成を行い、一方、その他のフレーズまたは単語語彙部分については SMT モデルによって翻訳を行う方式を確立した。日中、中日、日英、英日の各方向の翻訳において評価を行い、提案手法の有効性を検証している。結果として、ベースラインである標準的 SMT モデル、および、提案手法が適用されていない NMT モデルを上回る性能を達成した。

以上の議論に基づき、本論文は博士論文に値するものと認められる。

【最終試験の結果】

平成 30 年 1 月 31 日、システム情報工学研究科において、学位論文審査委員の全員出席のもと、著者に論文について説明を求め、関連事項につき質疑応答を行った。その結果、学位論文審査委員全員によって、合格と判定された。

【結論】

上記の学位論文審査ならびに最終試験の結果に基づき、著者は博士（工学）の学位を受けるに十分な資格を有するものと認める。