

論文 / 著書情報
Article / Book Information

題目(和文)	統計的機械翻訳における翻訳モデルの学習手法に関する研究
Title(English)	
著者(和文)	上垣外英剛
Author(English)	Hidetaka Kamigaito
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第10542号, 授与年月日:2017年3月26日, 学位の種別:課程博士, 審査員:高村 大也,新田 克己,中本 高道,奥村 学,長谷川 晶一
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第10542号, Conferred date:2017/3/26, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

論文審査の要旨及び審査員

報告番号	甲第	号	学位申請者氏名	上垣外 英剛	
論文審査 審査員		氏名	職名	氏名	職名
	主査	高村 大也	准教授	長谷川 晶一	准教授
	審査員	奥村 学	教授		
		中本 高道	教授		
新田 克己		教授			

論文審査の要旨 (2000 字程度)

機械翻訳は入力された文書を目的の言語に自動で翻訳するという課題である。電子計算機が発明されて以降、機械翻訳に関して様々な研究が行われ、数多くの手法が開発されてきた。本研究ではその中でも統計的機械翻訳の枠組みに注目し、その構成要素の一つである翻訳モデルに残された課題の解決を行う。本研究で扱う翻訳モデルに残された課題は次の二つである。

一つ目は、翻訳モデルを学習する際に用いられる単語アライメントの生成において、明示的に機能語と内容語の差異を考慮した学習ができないことによって日本語と英語のような言語学的に遠い言語対における翻訳精度が低下してしまう問題である。二つ目は、統計的機械翻訳の翻訳手法の一つである、階層的フレーズベース翻訳において、適切な長さのフレーズ対や書き換え規則を残して、翻訳精度を低下させることなく翻訳モデルのサイズを削減することができないという問題である。

本論文では、上記二つの問題に、それぞれ人手によるデータの作成を必要としない教師なし学習手法により対処することで、翻訳モデルの学習を改善することを目的としており、「統計的機械翻訳における翻訳モデルの学習に関する研究」と題し、全5章により構成されている。

第1章「序論」では、最初に機械翻訳の歴史と統計的機械翻訳における翻訳モデルの概要を述べ、次に翻訳モデルに存在する問題として、翻訳モデルの学習時に必要な単語アライメントの学習において機能語と内容語の差異を明示的に考慮できないという問題と学習された翻訳モデルのサイズが増大してしまうという問題が存在することを説明し、それぞれの問題に対する解決方針について説明し、最後に本論文の構成を紹介している。

第2章「関連研究」では、まず統計的機械翻訳の概要について説明している。次に本論文の一つ目の課題において対象とするフレーズベース翻訳と、二つ目の課題において対象とする階層的フレーズベース翻訳についての説明を行っている。そして一つ目の課題に関連する教師なし単語アライメントモデルについての説明を行っている。さらに二つ目の課題に関連する翻訳モデルのサイズの削減手法の研究についての説明を行っている。

第3章「頻度制約を考慮した制約付き EM に基づく教師なし単語アライメント」では、本研究で取り組んだ翻訳モデルにおける一つ目の課題について述べている。そして事後確率正則化学習法に基づく教師なし単語アライメントの学習に、機能語と内容語を識別可能な制約を用いる手法を提案している。提案手法と従来用いられていた対称化制約に対して比較評価を実施し、日本語と英語の言語対において、単語アライメントの精度指標である AER と F 値において単語アライメント精度が向上し、自動翻訳指標である BLEU においても翻訳精度が向上することを確認している。

第4章「階層的バックオフ過程に基づく Hiero 文法の学習」では、本研究で取り組んだ翻訳モデルにおける二つ目の課題について述べている。そして従来のベイズ的手法に基づく同期文脈自由文法の学習手法に階層的なバックオフ過程を導入することで Hiero 文法を学習する手法を提案している。提案手法と従来のベイズ的手法を比較するためにドイツ語/フランス語/スペイン語/日本語から英語へ翻訳する実験を実施した結果、同等のルールテーブルサイズの元で、提案手法は BLEU において同等以上の翻訳精度を示すことを確認している。また、従来のヒューリスティックな手法との比較評価についてもドイツ語から英語への翻訳を行う実験を実施し、提案手法は

10分の1以下のルールテーブルサイズで、BLEUにおいてより高い翻訳精度を示すことを確認している。さらに、ヒューリスティックな手法に統計的な検定に基づくルールテーブル削減手法を適用したものとの比較評価についても、ドイツ語から英語への翻訳を対象に実施し、その結果、提案手法はより小さいルールテーブルサイズで、同等の翻訳精度を示すことについても確認している。

第5章「結論と今後の課題」では、本研究における課題に対する提案手法の有効性について実験結果に基づく考察を実施し、その結論を述べている。さらに本研究において残された課題について統計的機械翻訳とニューラルネットワークを用いた翻訳手法の両面から説明し、その解決のために必要な方針について論じている。

以上を要するに、本論文は、統計的機械翻訳において翻訳モデルを学習する際に生じる、日本語と英語のような言語学的に遠い言語対における翻訳精度が低下してしまう問題と、階層的フレーズベース翻訳において翻訳精度を低下させることなく翻訳モデルのサイズを削減することができないという問題に焦点を当て、前者においては頻度に基づいて機能語と内容語を識別する制約を用いた学習手法を提案することで、後者においては従来のベイズ的手法に基づく同期文脈自由文法の学習手法に階層的なバックオフ過程を導入することで翻訳モデルの性能向上を行っている。また、提案された両手法は共に教師なし学習に基づいており、手法を適用する際に人手により新たに学習データを作成する必要がなく、従来手法に基づいて作成された自動翻訳システムに対して簡易に適用することが可能であり、工学上貢献するところが大きい。よって、博士(工学)の学位を授与するに十分な価値をもつものと認められる。