

## 【学位論文審査の要旨】

(論文審査の要旨)

機械翻訳は深層学習の登場により、一気に発展した。しかしながら、ニューラル機械翻訳には大規模な対訳データ（文単位での対応の取れた対訳コーパス）の存在が不可欠であり、低リソースの言語対での課題となっている。

ほとんどの言語対では質の高い対訳コーパスが得られるわけではなく、自動的に生成された対訳コーパスは質が低いという問題がある。また、ニューラル機械翻訳モデルを訓練するための対訳コーパスが得られたとしても、実際に翻訳したい対象の分野の対訳コーパスは得られないか少量しか用意できないことも多く、分野外の対訳コーパスの活用方法も解決すべき問題である。そして、通常の機械翻訳の設定と異なり、同時機械翻訳も翻訳の際に全ての文を見ることができないという意味で低リソースの設定であり、いかにテキスト情報の不足を補うか、という問題を抱えている。

そこで、本研究では様々な追加リソースを用いて翻訳精度を向上させる研究に取り組む。1つ目はデータ選択とフィルタリングを組み合わせることで、機械翻訳で生成した擬似的な対訳文を活用する手法である。2つ目は分野適応と他言語のモデルの転移学習をすることで、分野外のデータを活用する手法である。3つ目はテキストの文脈情報が得られないときに画像の情報を活用する手法である。実験により、いずれの手法もニューラル機械翻訳の精度向上につながることを示した。

まず、本論文の第1章は低リソース下における機械翻訳とその課題についての概観を与える。本章では、様々なリソースから得られた追加の情報を活用することで効果的に低リソース言語対における機械翻訳の性能を向上させるという目的について述べる。

次に、第2章では逆翻訳と周遊翻訳という2つの方法で擬似的な対訳データを作成し、そしてフィルタリングを行う手法を提案する。フィルタリングされた擬似的な対訳コーパスを追加の訓練データとして用いることで、フィルタリングしない擬似的な対訳コーパスを逆翻訳および周遊翻訳の訓練に使うのと比較して、ニューラル機械翻訳の精度が向上することを示した。提案手法はロシア語から日本語への翻訳と日本語からロシア語への翻訳のそれぞれで、機械翻訳の標準的な自動評価尺度である BLEU スコアを 3.46 と 5.25 向上させた。

そして、第3章は低リソース言語対に対するニューラル機械翻訳の分野適応の課題について取り組む。この章では、分野外の対訳データを効率的に活用するため、多言語転移学習のアプローチを分野適応と組み合わせ、多段階でニューラル機械翻訳モデルの訓練を行う手法を提案する。提案手法はロシア語から日本語への翻訳と日本語からロシア語への翻訳のそれぞれで BLEU スコアを最大で 2.72 と 3.06 向上させた。

また、第4章では低リソース下での同時機械翻訳における追加のモダリティ（テキスト以外の情報）を活用する新しいアプローチについて述べる。これまでの同時機械翻訳では、文末まで発話が終わるまで待たずに翻訳するため、不十分なテキスト情報を用い

て翻訳するというタスクであったが、本章ではテキスト情報の不完全性を補うために画像の手がかりを活用する手法を提案する。実験結果により、提案手法は様々な言語対において、テキスト情報のみを使うベースラインと比較して有意に性能が向上しており、特に英語から日本語への翻訳のように語順の異なる言語対で性能向上が大きかったことを確認した。

最後に、第 5 章では本論文の結論をまとめ、得られた知見と提案手法の限界について議論し、低リソース機械翻訳の今後の課題について述べる。

以上のように、本論文で提案する手法は、低リソース環境における機械翻訳において、訓練データの拡張を行うことで性能を向上させることができ、リソースが十分ではない多くの状況に対して適用することが可能であるため、工学的に重要な意義があると考えられる。よって、本論文は博士（工学）の学位を授与するに十分な価値があるものと認められる。

（最終試験又は試験の結果）

本学の学位規則に従い、最終試験を行った。公開の席上（オンライン）で論文発表を行い、学内外の教員による質疑応答を行った。また、論文審査委員により本論文及び関連分野に関する試問を行った。これらの結果を総合的に判断した結果、専門科目についても十分な学力があるものと認め、合格と判定した。