

先端科学技術研究科 修士論文要旨

所属研究室 (主指導教員)	自然言語処理学 (渡辺 太郎 (教授))		
学籍番号	2011121	提出日	令和 4年 1月 19日
学生氏名	坂井 優介		
論文題目	サブワード単位での知識グラフ埋め込みを用いたニューラル機械翻訳		
要旨			
<p>ニューラルモデルの登場以降、機械翻訳の精度は飛躍的に向上しているが、ニューラル機械翻訳では、新語や未知語、固有名詞など低頻度語については、翻訳精度が不十分であるという問題が知られている。この問題を解決するために、近年のニューラル機械翻訳はサブワード化という統計的情報に基づいた文の分割を機械的に行うことで、文章全体でのトークンを一定数し、トークン単位でのカバー率を拡大することによって対処しているが、単語の情報は一切考慮されていないため、意図しない翻訳結果となる場合がある。一方、サブワード化の弱点に対処するため、直接低頻度語に対して処理を行うことで翻訳精度を改善させる試みがある。特に知識グラフなど大規模な単言語データをニューラル機械翻訳に取り込むことで、翻訳精度を向上させる取り組みが行われている。しかし、大規模な単言語データを扱うため、知識グラフなどを効率的にニューラル機械翻訳に取り込むことは容易ではない。本研究では、サブワード化した巨大な知識グラフをニューラル機械翻訳に取り込む手法を提案する。提案手法では、知識グラフをニューラル機械翻訳に取り込む際に扱わなければならないトークン数を削減することができるため、知識グラフの情報を効率よくニューラル機械翻訳に注入することができる。提案手法により、既存の知識グラフを用いた手法より大幅な精度改善が示され、サブワード化のみ行ったニューラル機械翻訳よりも良い翻訳精度となることが示された。</p>			