

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)	知能コミュニケーション (中村 哲 (教授))		
学籍番号	1911059	提出日	令和 4年 1月 19日
学生氏名	押田 祐治		
論文題目	字幕情報を考慮した英日ニューラル機械翻訳		
要旨			
<p>字幕翻訳は、文字数の制限がある、改行が存在する、句読点を使わないなどの特徴があり、一般的な翻訳とは異なる。近年、ニューラル機械翻訳の登場により、機械翻訳の翻訳精度が大幅に向上したものの、字幕翻訳に焦点を当てた機械翻訳の研究は少ない。既存の研究では字幕の対訳データを作る際、改行を無視し、文を連結することにより対訳データとして用いているが、それでは改行の情報を取り扱うことができないという問題がある。本研究では改行情報を用いてモデルを学習すれば、自動で改行を含んだ訳出をしてくれるという仮説をもとに検証を行った。またその際、句読点の有無や文字数について評価した。提案手法として字幕翻訳データの改行の位置に改行記号
を挿入し、文を連結した。そして改行記号を挿入したデータを学習に用いて実験を行った。実験の結果、提案手法を用いたデータで学習したモデルは、自動で改行記号を訳出できることがわかった。また句読点を使わない、制限された文字数以内で翻訳するなど、字幕翻訳に適応した訳出をすることがわかった。</p>			