

自動生成した画像による 教師なしマルチモーダルニューラル機械翻訳

岩本裕司（愛媛大学）

概要

近年、機械翻訳の際に、原言語文（翻訳元言語の文）に加えて翻訳に関連する画像を利用することで、翻訳精度の向上を図るマルチモーダルニューラル機械翻訳（MNMT）が注目されている。しかし、MNMTモデルの学習には、原言語文、目的言語文（翻訳先言語の文）、関連画像で構成される3つ組データが必要となり、学習データ数不足が問題となっている。そこで、本研究では3つ組データを必要とせず、3つ組データより用意しやすい対訳テキストデータと原言語側の画像キャプションデータを用いてMNMTモデルを学習する新たな手法を提案する。提案手法では、まず対訳テキストデータからニューラル機械翻訳モデル（NMT）を学習し、学習したNMTモデルで画像キャプションデータの各文を翻訳することで、疑似3つ組データを作成する。そして、作成した疑似3つ組データを用いて、対訳文から画像を生成するtext-to-imageモデルとMNMTモデルを初期化する。その後、text-to-imageモデルとMNMTモデルを、逆翻訳形式のフレームワークで交互に繰り返し学習する。具体的には、MNMTモデルは、T2Iモデルによって生成された画像と対訳テキストデータによる疑似3つ組データで学習し、T2Iモデルは、MNMTモデルによって生成された目的言語文と画像キャプションデータによる疑似3つ組データで学習する。提案手法の有効性を、対訳テキストデータとしてMulti30kデータセット、画像キャプションデータとしてMSCOCOデータセットを用いた英独翻訳タスクで検証した。その結果、提案したMNMTモデルは画像入力なしのNMTモデルよりも優れており（+1.38 BLEUスコア）、また、提案した反復逆翻訳学習方式は初期のMNMTモデルの性能を向上させる（+2.8 BLEUスコア）ことを確認した。さらに、大規模なデータセットを用いた事前学習により、さらなる翻訳精度の向上が可能であることを実験的に示した（+1.07 BLEUスコア）。