

平成30年度 情報工学コース卒業研究報告要旨

外山 研究室	氏 名	浅川 脩介
卒業研究題目	法令文のニューラル機械翻訳への ソフトテンプレートの導入	
<p>急速にグローバル化が進む今日において、国際取引の円滑化や法整備支援の推進などを目的として、日本法令の外国語訳の必要性が高まってきている。しかし、法令の量は膨大であり、それを翻訳するには多大な労力を要する。また、法令のような公文書の翻訳では、翻訳の前後で文の意味を正しく保持する翻訳が求められる。さらに、法令の翻訳では、訳文、訳語の統一を図る必要がある。そのため、翻訳者の負担はさらに大きくなる。よって、法令の翻訳を迅速にするには何らかの支援が要求される。</p> <p>このような背景から、法令文の翻訳に関して様々な研究がなされている。その中の1つにニューラル機械翻訳(NMT)がある。近年、NMTに関する研究は、高い性能を発揮しているが、不自然な訳、訳抜け、誤訳などの課題があり、妥当性という観点から、法令の機械翻訳の手法としてはまだ実用的ではない。</p> <p>NMTの性能を上げるため、ニューラルモデルを用いた自動要約にソフトテンプレートを導入した研究に着目した。ソフトテンプレートとは、コストのかかるテンプレート要約の代替案として考案されたもので、ここでは、原文と類似度の高い文に対する要約である。先行研究では、まず、全文検索ソフトウェアであるLUCENEを用いて、学習データから原文と類似度の高い30文を検索し、それらの要約文を要約のソフトテンプレート候補とした。次に、この30文に対して、ソフトテンプレートとしての有用度を計算し、1文を選んだ。このようにして選んだ1文と原文を入力として、オープンソースの機械翻訳システムOpenNMTを用いて要約している。有用度と翻訳モデルは、ニューラルを用いて同時に学習される。この研究では、ソフトテンプレートの一部を書き換えた要約が実現されており、従来手法よりも性能の良い結果を示している。こうしたソフトテンプレートの一部を書き換えるという手法は、訳文、訳語の統一が要求される法令翻訳においても有効であると考えられる。</p> <p>そこで本研究では、既存の翻訳手法よりも性能の高い翻訳の実現を目的として、NMTにソフトテンプレートを導入する手法を提案する。本手法では、まず、類似文字列検索ツールであるSimStringを用いて、学習データから原文と類似度の高い30文を検索し、それらの訳文を翻訳のソフトテンプレート候補とする。次に、この30文の中から、ソフトテンプレートとなる1文を決定する。決定の際、ソフトテンプレートの有用度を計算し、有用度が最も高くなると考えられる1文をソフトテンプレートとして採用する手法と、SimStringによるコサイン類似度が最も高い文をソフトテンプレートとして採用する手法を用いる。最後に、それぞれ決定したソフトテンプレートに対して、原文とソフトテンプレートを入力として、OpenNMTを用いて翻訳する。</p> <p>本手法の有効性を検証するため、法令705本(356,323文)を学習データとして用い、28本(10,307文)をテストデータとして用いて実験した。実験の結果、NMTの結果と比較して、翻訳性能が上がらなかった。提案手法では、ソフトテンプレートの一部が不必要に書き換えられたものが出力され、文としての意味が成立しないものがほとんどであった。ソフトテンプレートを導入した要約の研究では、ソフトテンプレートと入力文が同じ言語であったのに対し、本研究ではそれらが異なる言語であったため、学習が正しくされなかったと考えられる。今後、正しい学習をするために、異なる言語の分散表現を同じ空間にマップするなどの手法を取り入れることが必要である。</p>		