

審査の結果の要旨

氏 名 石渡 祥之佑

本論文は「Translation and Description Methods for Multilingual Text Understanding (多言語テキスト理解のための翻訳および語義説明手法)」と題し、英文6章から構成されている。多言語で記述された多領域に渡るウェブ上の文書の容易な理解を可能とすることを目的として、領域内コーパスを必要としない高精度で即時的な機械翻訳の領域適応手法、差異の大きな言語間における機械翻訳技術の高精度化手法、および未知の語句の語義を自動生成により提示する手法を提案し、未知語翻訳、機械翻訳、および定義文生成タスクにおける実験を行い、提案手法の有効性を論じている。

第1章は「Introduction (序章)」であり、本論文の背景、および取り組む課題について概観し、本論文の構成を述べている。

第2章は「Related Work (関連研究)」と題し、多言語で記述される多領域文書の理解を目的とする領域適応・ニューラル機械翻訳・ならびに定義文生成に関して、現時点までに提案されている手法の限界点を明らかにするとともに、本論文において提案する手法との違いを明らかにしている。

第3章は「Accurate and Instant Translation Model Adaptation for Statistical Machine Translation (高精度かつ即時的な統計的機械翻訳の領域適応)」と題し、高精度かつ即時的に統計的機械翻訳モデルを他の領域に適応させる手法を提案している。従来の機械翻訳の問題点として、翻訳モデルの訓練に用いた教師データと領域が異なる場合、未知語の出現頻度が増加し、文書の翻訳性能が著しく低下することが挙げられる。本章では、教師なし学習によって容易に得られる単語のベクトル表現の言語横断写像を用いる手法を提案し、これにより高精度な未知語翻訳が可能になることを示している。本手法は、対象とする領域の対訳辞書や対訳コーパスを必要としない点に特徴がある。また、この高精度な未知語翻訳の手法を統計的機械翻訳に導入し、実験により訓練データの領域と異なる領域の文書の翻訳性能向上 (BLEU +0.5~1.5) を確認している。

第4章は「Chunk-based Decoder for Neural Machine Translation (ニューラル機械翻訳のための文節に基づくデコーダ)」と題し、言語間の差異の大きさから特に翻訳が困難となる英日翻訳の性能

を向上させる手法を提案している．本手法は，2015 年以前の統計的機械翻訳では広く用いられていた Phrase/Chunk 単位の翻訳を，2016 年以降その性能の高さから広く用いられるようになったニューラル機械翻訳に導入したものであり，世界ではじめて日本語の Chunk（文節）単位の翻訳を可能とし，単語単位の翻訳では困難であった英日翻訳の性能を向上させた点に特徴がある．また，この文節に基づくデコーダを用いて科学技術論文領域での翻訳実験を行い，従来手法からの性能向上（BLEU +0.9, RIBES +0.6）を確認している．

第 5 章は「Learning to Describe Phrases with Local and Global Contexts (局所的・大局的な文脈を用いた句の語義説明)」と題し，未知の語句の語義を自然言語のかたちで自動生成することにより提示する手法を提案している．本章ではまず未知の語句を多く含む説明文生成のデータセットを構築するため，Wikipedia から（語句，説明文，局所文脈）の組を自動的に獲得する方法を提案している．また，多領域の文書から人間が情報を効率的に入手することを補助するため，RNN (リカレントニューラルネットワーク) により未知の語句の語義を説明する文を生成する方法を提案している．本手法は，説明文の生成に際してその句の局所文脈 (Local Context) と大規模なコーパスにおける大局的な文脈 (Global Context) を同時に考慮し，計算機が自律的に文脈情報の取捨選択を行う点に特徴がある．本章で構築したものを含む計 4 つのデータセットにおける実験により，提案手法が多様な領域の語句に対して正確 (BLEU +0.9, ならびに 5 段階の人手評価 で +0.3) な説明文の生成を可能とすることを確認している．

第 6 章は結論であり，本研究の成果と今後の研究課題について総括している．

以上，これを要するに，本論文は単語のベクトル表現の言語横断写像によって即時的な機械翻訳の領域適応を実現する手法と，従来の単語単位ではなく文節単位の英日翻訳を可能とする高精度なニューラル機械翻訳のデコーダの構成法，ならびに局所的・大局的な文脈を適応的に取捨選択することで未知の語義を高い精度で生成する語義説明手法を提案し，機械翻訳ならびに語義説明の性能向上により人間の文書理解を容易にしており，電子情報学上貢献するところが少なくない．よって本論文は博士（情報理工学）の学位請求論文として合格と認められる．