

## 審査の結果の要旨

氏 名 朱 中元

本論文は「Structure-Aware Latent-Variable Models for Neural Machine Translation (ニューラル機械翻訳のための文構造を考慮した潜在変数モデル)」と題し、英文で記されており、7章から成る。機械翻訳は、社会的な需要や期待が非常に大きい重要な技術であり、近年はニューラルネットワークに基づくニューラル機械翻訳が非常に優れた翻訳精度を達成するようになってきている。しかしながら、ニューラル機械翻訳には依然として、翻訳速度が遅いこと、モデルサイズが大きいこと、出力文に多様性がないこと、などの課題があり、特にエッジデバイスなど限られた計算資源上での実用を妨げる要因となっている。本論文では、これらの課題を改善するための一貫した方法論として、文構造を考慮した潜在変数をニューラル機械翻訳のモデルに組み込むアプローチを提案し、実際に各課題を改善するニューラル機械翻訳手法を複数提案している。実験では、機械翻訳の研究分野における標準的なベンチマークデータセットを用い、各提案手法の効果を定量的に確認している。また、それぞれの手法における潜在変数を分析し、いずれも妥当性のある言語的構造を反映していることを定性的に評価している。

第1章「Introduction and Background」では、ニューラルネットワークを用いた機械翻訳の発展について述べるとともに、現状の機械翻訳においては、翻訳速度の向上・モデルの軽量化・多様性のある翻訳の実現などが未解決の重要課題であることを指摘している。本研究では、これらを実現するための一貫した方法論として、文の言語的構造を適切に反映した潜在変数を用いてニューラル機械翻訳をモデル化するアプローチを提案し、それぞれの課題解決を行う手法を実現していくことを述べている。

第2章「Prerequisite Knowledge for Neural Machine Translation」では、ニューラル機械翻訳を構成するために必要となる、ニューラルネットワークの中核技術について解説している。

第3章「Prerequisite Knowledge for Latent-Variable Likelihood Models」では、潜在変数に基づく生成モデルや、それらのニューラルネットワークへの組み込み方法に関する背景技術を解説している。また、勾配法によって離散潜在変数を学習するための幾つかの技術について述べている。

第4章「Latent-Variable Non-autoregressive Neural Machine Translation」では、潜在変数に基づく非自己回帰型ニューラル機械翻訳手法を提案し、翻訳速度の大幅な向上を実現している。一般的に、ニューラル機械翻訳では自己回帰型モデルが翻訳文の出力に用いられてきたが、自己回帰型モデルは翻訳文中の各単語を逐次的に予測するため

並列化が困難であり、翻訳速度向上におけるボトルネックとなっていた。提案手法では、機械翻訳のプロセスを条件付き生成モデルとして定式化し、各入力単語の翻訳内容に対応する複数の潜在変数を自動的に学習する。最終的な翻訳文は各単語に対応する潜在変数からのフィードフォワードによって得られるため、並列化を活用した大幅な翻訳速度の向上が可能となる。さらに、翻訳精度を向上させるため、潜在変数の推定値を動的に更新するアルゴリズムを導出している。実験では、WMT14英独翻訳タスクにおいて、ベースラインの自己回帰型モデルに対して、翻訳精度はBLEU値で1%程度劣るものの、約8倍の速度で翻訳が実行可能であることを示している。

第5章「Discrete Representation Learning for Model Compression」では、単語ベクトルを離散潜在表現へ圧縮することによって、ニューラル機械翻訳モデルの軽量化を行う手法を提案している。提案手法はベクトル量子化に基づく圧縮を行うが、その際に離散コードと基底ベクトルをニューラルモデルの中で同時に学習するという特徴を有している。これを実現するために、Gumbel-Softmax法を利用した離散ボトルネックをモデルに組み込む工夫を行っている。実験では、提案手法を適用したニューラル機械翻訳モデルの圧縮率と翻訳精度の関係について検証し、翻訳精度を落とすことなく単語ベクトルに必要なモデル容量を94%~99%圧縮できることを示している。

第6章「Learning Syntactic Latent Variables for Diverse Translation」では、ある入力文に対して、多様な文法的構造を持つ複数の翻訳候補文を生成できるニューラル機械翻訳手法を提案している。提案手法は、文の文法的構造を捉える離散潜在変数を学習し、得られた潜在変数によってニューラル機械翻訳モデルの出力を条件づけることにより、出力文の構造を制御するメカニズムとなっている。実験では、提案手法による離散潜在変数を用いた条件付け構造を組み込むと、翻訳の品質（精度）に悪影響を与えずに、出力文の文法的多様性を大きく向上できることを示している。

第7章「Summary」では、本論文の成果をまとめるとともに、今後の研究や関連技術の展望について述べている。

以上を要するに、本論文はニューラル機械翻訳において、学習に基づき文構造を捉えた潜在変数を活用する方法論および一連の手法を提案し分析を行ったものであり、機械翻訳のみならず周辺分野に広く影響を与えうる大きな研究的価値を有する。さらに、機械翻訳の実用面における数々の未解決課題においてそれぞれ顕著な効果を挙げており、情報理工学における創造的実践の観点からの価値が認められる。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。