【学位論文審査の要旨】

（論文審査の要旨）

ニューラル機械翻訳には未知語の翻訳精度が非常に悪いという問題がある．そこで，単語をサブワードという単語より小さい文字列に分解して翻訳する手法が盛んに研究されている．しかしながら，既存のアルゴリズムは文字の種類が少なく単語の区切りが明確なアルファベットを用いる言語（表音文字言語）ではうまくいくが，日本語や中国語のように文字の種類が多く単語の区切りも明示しない言語（表意文字言語）ではうまくいかない．一方，漢字は文字より細かい単位である篇（へん）や旁（つくり）に分解することができる．

そこで，本研究ではサブキャラクタという文字より小さい単位でニューラル機械翻訳を行う手法を提案する．文字をさらに分解することで，ヨーロッパの言語間でサブワードが共有されて語彙サイズが小さくなり，翻訳精度が向上するように，日中のように漢字を共有する言語間でサブキャラクタが共有されて語彙サイズが小さくなり，翻訳精度が向上することが期待される．

まず，第 1 章ではサブキャラクタを用いたニューラル機械翻訳における一般的な問題について紹介する．また，論文の構成と貢献についても述べる．

次に，第 2 章では本研究で用いるモデルである教師ありニューラル機械翻訳と教師なしニューラル機械翻訳の基礎知識を説明する．また，表意文字の分解についても取り扱う.先行研究では表音文字を用いる言語と表意文字を用いる言語の違いに敏感ではなく，表意文字の特徴を有効に活用することで性能の向上が期待できる．

第 3 章は提案手法について詳細に述べる．まず表意文字言語に対する文字分解の方法について説明し，次にどのようにして分解したサブキャラクタ列を文字列に変換することができるかを説明する．そして本研究で用いる教師ありニューラル機械翻訳モデルの LSTM（Long short-term memory）モデルと Transformer モデル，そして Transformer モデルを用いた教師なしニューラル機械翻訳モデルについて紹介する．さらに，Transformer モデルに対しては，サブキャラクタレベルの系列に対する効果的な位置情報の符号化方法についても述べる．

そして，第 4 章では実験設定について詳述する．本研究では中国語・日本語，中国語・英語，そして日本語・英語という言語対を用い，それぞれ表意文字言語同士，そして表意文字言語と表音文字言語の言語対として，複数の粒度の文字分解によるサブキャラクタを比較する翻訳実験を行う．

第 5 章は実験結果を示す．実験結果から，サブキャラクタを用いて翻訳を行う提案手法は，文字レベルで翻訳を行うベースラインと比較して，教師ありニューラル機械翻訳においても教師なしニューラル機械翻訳においても高い翻訳精度であることを示した．

また，第 6 章では実験の結果について考察する．実験より，細かい粒度の文字分解にすればするほど，翻訳の性能向上が見られる．一方，文字分解をすることで，非常に長

いサブキャラクタ系列の訓練データでは性能が落ちることも確認された．これらの現象について考えられる原因について，様々な角度から検討した．

最後に，第 7 章では本研究についてまとめ，今後の研究の方向性について議論する．本研究では表音文字と表意文字の違いに注目し，文字をさらに分解することでニューラル機械翻訳の精度を向上させる手法を提案したが，機械翻訳以外の他のタスクでも文字分解が有効であることが期待できる．

以上のように，本論文で提案する手法は，表意文字を用いる言語におけるニューラル機械翻訳において翻訳精度を向上させることができ，工学的に重要な意義があると考えられる．よって，本論文は博士（工学）の学位を授与するに十分な価値があるものと認められる．

（最終試験又は試験の結果）

本学の学位規則に従い，最終試験を行った．公開の席上（オンライン）で論文発表を行い，学内外の教員による質疑応答を行った．また，論文審査委員により本論文及び関連分野に関する試問を行った．これらの結果を総合的に判断した結果，専門科目についても十分な学力があるものと認め，合格と判定した．