

採択番号 19701
研究開発課題名 多言語音声翻訳高度化のためのディープラーニング技術の研究開発
副題 深層学習によるマルチモーダル文脈理解と機械翻訳の高度化

(1) 研究開発の目的

ビジネスシーンにおける外国人と日本人のコミュニケーションを円滑にするため、会議や社交における対話・雑談、およびニュース記事に対応した機械翻訳技術を研究・開発する。また、より日常生活に密着した機械翻訳に向け、マルチモーダル機械翻訳に取り組み、テキストと画像の統合的な理解に向けた研究を進める。全ての研究・開発において、深層学習を共通基盤として採用し、タスクやメディア、モーダルの垣根を超えた人工知能の実現を目指す。また、自然言語処理、画像処理、深層学習、報道などの分野で最先端の取り組みを継続しているグループでチームを結成し、人材やデータ、技術の交流を促進する。これにより、日本語に関する技術で世界トップレベルを維持するとともに、マルチモーダルな文脈理解など、世界に先駆けた研究に挑戦する。

深層学習に基づくアプローチでは、学習データの質と量が性能を左右する。そこで、本提案では研究資源（対話や翻訳、多言語画像技術文のコーパスなど）の開発にも注力する。本研究で得られた成果は、プロジェクト終了後も活用できるように検討し、研究コミュニティや社会に還元する。また、ニュース記事に対応した機械翻訳技術および自動要約技術などの成果は、ニュースの配信の実務と照らし合わせながら評価し、研究成果の社会実装に向けた実行可能性調査を行う。また、対話翻訳技術はNICTが開発しているVoiceTraの翻訳エンジンや、スマートスピーカーへの応用を模索し、社会実装につなげていく。

(2) 研究開発期間

平成30年度から令和2年度（3年間）

(3) 実施機関

国立大学法人東京工業大学 <代表研究者>
国立大学法人東京大学
国立大学法人愛媛大学
日本放送協会
一般財団法人NHKエンジニアリングシステム
株式会社時事通信社

(4) 研究開発予算（契約額）

総額 780 百万円（令和2年度 240 百万円）
※百万円未満切り上げ

(5) 研究開発項目と担当

研究開発項目1：高度な文脈理解技術（インテリジェント翻訳技術）の研究開発

1. 対話単言語コーパスの構築（東京工業大学）
2. 日英対話対訳コーパスの構築（東京大学）
3. 文脈に応じて適切な訳を選択する技術の開発（東京大学）
4. 訳語の一貫性を保持する技術の開発（東京大学）
5. 翻訳に必要な情報を補完する技術の開発（東京大学）

研究開発項目2：新語・新トピックへの即時対応技術（ニュース対応翻訳技術）

1. ニュース記事コーパスの整備（時事通信社）
2. 人手によるニュース対訳コーパスの作成（NHKエンジニアリングシステム）
3. 新語や新話題に対応した機械翻訳技術の構築（東京大学）

- 4. ニュース翻訳システムの研究開発 (日本放送協会)
- 5. ニュース記事からの要約文生成の研究開発 (東京工業大学)
- 6. ニュースの長文翻訳および要約文生成の評価 (時事通信社)

研究開発項目3：マルチモーダル翻訳技術の研究開発

- 1. マルチモーダル日英対訳コーパスの作成 (東京大学)
- 2. 機械翻訳のためのシーン画像認識手法の開発 (東京大学)
- 3. マルチモーダル翻訳手法の開発 (愛媛大学)

(6) 特許出願、外部発表等

		累計 (件)	当該年度 (件)
特許出願	国内出願	1	0
	外国出願	0	0
外部発表等	研究論文	4	3
	その他研究発表	90	34
	標準化提案・採択	0	0
	プレスリリース・報道	1	0
	展示会	1	0
	受賞・表彰	16	8

(7) 具体的な実施内容と成果

研究開発項目1：

1-1、1-2. ビジネスシーン日英対話対訳コーパスの構築

引き続きビジネスシーンにおける対話のシナリオを日本語および英語で創作し、それぞれ英語および日本語に翻訳することで、日英対話対訳コーパスを拡充した。令和2年度はこれまでに構築した84,800対訳文に加えて、新たに27,000万対訳文を追加した。さらに令和2年度は既存の単言語対話コーパスである名大会話コーパスの英訳89,008文およびSanta Barbara Corpus of Spoken American English (SBCSAE) の和訳69,993文も実施し、対訳コーパスの拡充を行なった。

最終的にシナリオ作成とその翻訳により11.2万対訳文を構築し、既存の英語対話コーパスの翻訳により約31万対訳文を構築し、全体として約42万対訳文からなる対話対訳コーパスを構築した。なおビジネスシーン対話対訳コーパスのうちの24,000対訳文とAMI Meeting Corpusのうち24,000対訳文は対話翻訳研究を広く促進するために一般公開した。

1-3 文脈に応じて適切な訳を選択する技術の開発

Multi-source Transformer モデルを用いて文脈文を適切に利用することにより翻訳精度が向上することを示した。Multi-source Transformer は Transformer のエンコーダが2つ分あり、一方が文脈文をエンコードするもので、もう一方が翻訳対象文をエンコードするものである。実験には項目1-1および1-2で構築した対訳コーパスを用いた。

単一エンコーダと空文脈を入力したマルチエンコーダの結果の間に有意に差があることから、モデルサイズによる正則化の効果が存在することがわかる。また文脈文を増やしていくと精度が徐々に向上していき、日英では直前3文、英日では直前2文を入力した場合に最も精度が高くなることがわかった。対話文においては直前の文は相槌などさほど重要ではない文も多く、それより前の情報を用いることで翻訳精度が向上するためであると考えられる。一方で文脈文を多くし過ぎると翻訳精度が低下しているが、これは不要な情報がノイズとなって混入してしまう可能性が増えるためであると考えられる。

文脈文として入力する文をランダムに選択した場合、文脈なしの場合よりも精度が高いという結果となった。これについてもある種の正則化の効果であると考えられる。一方で正しい文脈文

を与えた場合と比較すると精度は低くなっており、ここから文脈を考慮した翻訳モデルは正しい翻訳を生成するために必要な情報を適切に文脈文から補っていると結論づけることができる。

1-4 訳語の一貫性を保持する技術の開発

訳語の一貫性を保つ目的や未知語対策として一般的に利用される方法は対訳辞書と placeholder を組み合わせるものである。この方法ならば辞書登録されている語はほぼ正確に翻訳することができるが、欠点として前後の文脈に合わせて適切に語形を変化させることができないことが挙げられる。例えば活用（単数/複数、時制、三単現など）や品詞（形容詞・副詞、サ変名詞に「する」を付与など）などは placeholder 前後の文脈によって適切に変化させる必要がある。そこで placeholder への埋め込みを部分的な翻訳問題として解く手法を開発した。提案手法は通常の NMT の decoder の後段に、placeholder の翻訳を行う placeholder decoder を接続するものである。

科学技術論文ドメインの対訳コーパスである ASPEC の日英対訳コーパスのうち 100 万対訳文を用いた日英翻訳実験において提案手法の有効性を検証した。評価は自動評価尺度 BLEU と placeholder の翻訳精度（placeholder の翻訳結果が正解に含まれるかどうか）で測定した。比較として原形に戻した辞書訳語を代入する方法(lower boundに相当)と、原形に戻す前の辞書訳語代入する方法（upper boundに相当）についても評価を行った。結果として提案手法である placeholder decoder を用いた手法では lower bound と比較して BLEU 値には変化がなかったが、placeholder 翻訳精度は 82.2%から 91.2%へと 9%の改善が見られた。このことから提案手法を用いることで placeholder に埋め込むべき訳語を適切に変形させることができていることがわかる。

1-5 翻訳に必要な情報を補完する技術の開発

日→英に翻訳においては、代名詞の翻訳は典型的な翻訳誤りの一つである。特に省略された代名詞（ゼロ代名詞）の翻訳誤りは深刻で、特に日本語会話文においては一人称や二人称はほとんどが省略される傾向にあり、英語に翻訳する際には適切な代名詞を補う必要がある。事前調査では基本的に英語文で高頻度の代名詞ほど、日本語ではゼロ代名詞になりやすいことがわかった。ゼロ代名詞の翻訳誤りを改善するためには、文脈情報が必須である場合や話者情報が必要となる場合なども存在するが、単文でもゼロ代名詞が推定可能なものも存在する。そこでこのような単文でも推定可能なゼロ代名詞の翻訳誤りを低減するために、ゼロ代名詞に関する訓練データのデータ拡張を行う手法を開発した。

提案手法では対訳文の日本語側から代名詞および後続する機能語列を削除したものを、追加の訓練データとして利用する。代名詞および機能語列の削除は、日本語文の形態素解析結果から人手で列挙したパターンに基づいて行う。提案手法の有効性を示すために、本プロジェクトで構築した対話対訳コーパス 246,541 文を訓練データとして用いた実験を行った。代名詞削除によるデータ拡張により新たに 36,411 文が訓練データとして追加され、合計 282,952 文となった。実験の結果、自動評価尺度である BLEU の値はほとんど変化しなかったが、ゼロ代名詞精度はベースラインと比較して 10 ポイント以上向上し、提案手法によりゼロ代名詞が適切に訳出されるようになったことがわかった。

研究開発項目 2 :

2-1、2-2 ニュース記事コーパスの整備、人手によるニュース対訳コーパスの作成

令和元年度に引き続き、日英記事コーパス、写真データなどの整備を実施、記事データベース、写真データベースから各種データを抽出した。令和 2 年度は日英記事 7,147 ペア、和文記事 171,913 本、AFP 英文記事 262,941 本を抽出した。写真データは、時事通信、AFP 通信、EPA 通信を合わせて 1,702,415 枚を取得。また、これとは別に記事データベースから 28,996

枚の記事中写真データを抽出した。

また、時事通信社日本語ニュースの日英翻訳と、日本語ニュースを英語ニュースの内容に合わせて修正する方式を検討して業者に発注し、人手翻訳による高品質な対訳として 20 万文対を構築。昨年度までに構築した人手翻訳による対訳データや、日英のニュース自動対応付けデータなどと合わせて合計 241 万文対の日英対訳コーパスを構築した。

2-3 新語や新話題に対応した機械翻訳技術の構築

新語・低頻度語に対応した機械翻訳のため、辞書を利用した翻訳と、単言語データによるデータ拡張を実施した翻訳の2つの手法を提案した。辞書を利用した翻訳手法では、検出した新語を変数に置き換え機械翻訳し、出力文中の変数を新語の訳語に置換する。この際、学習データに数回程度しか出現しない語に対して、挿入した英語が出力されないケースが発生する。そこで、学習データにおける単語の出現頻度が低い単語は変数方式、それ以外は入力に訳を挿入して出力を制御する翻訳ガイド方式で学習、翻訳する手法を提案した。実験では、変数方式と翻訳ガイド方式を併用する提案手法は変数方式と比較して、新語・低頻度語の正解率で微減しているが BLEU は向上しており、一定の効果があることが分かった。また、辞書が無くても一定の訳質で翻訳するために、単言語データを拡張して作った対訳がどの程度新語や低頻度語の翻訳に有効化を検証した。その結果、目的言語である英語の単言語データを利用することにより、BLEU 値、新語・低頻度語の正解率の改善を確認した。

2-4 ニュース翻訳システムの研究開発

翻訳品質向上を目指し、翻訳の原言語側の文脈を利用する手法と、目的言語側の文脈を利用する手法を提案した。翻訳の原言語側の文脈を利用する手法では、まず、翻訳対象とする文の前の文脈を述語項構造解析し、主語・主題を抽出する。この抽出が成功した対象文にもっとも近い語を、翻訳元文に追加して、機械翻訳への入力とする。この結果、文脈を利用する手法では JPO Adequacy の基準による主観評価の向上を確認した。また、目的言語側の文脈の利用も、翻訳結果の表現に揺れがあるような場合は表現を揃えることができるなど機械翻訳では有効と考えられる。しかし、目的言語側の文脈を使う場合、学習時と翻訳時のデータに差が生じるため問題となっていた。そこで、目的言語側の文脈となる学習データと翻訳結果を学習の進行に応じて段階的に切り替えながら学習する手法を提案した。実験により、BLEU 値の向上を確認した。さらに、項目 2-1 と項目 2-2 で作成した日英対訳データを利用して、日英機械翻訳システムを構築し、内容の伝達レベル (JPO Adequacy Evaluation の基準) による人手の主観評価を実施した。その結果、JPO Adequacy の値が 4 を超え、「ほとんどの重要情報は正確に伝達されている (80% ~)」以上の評価を得た。構築した日英ニュース翻訳システムは、目標である 8 割の翻訳成功率を達成したと判断できる。

2-5 ニュース記事からの要約文生成の研究開発

記事に対する見出しの正解は一通りではない。特に、記者が伝えなかった内容に焦点を当てた見出しを生成したい場合や、記事のアクセス数向上のために見出しの内容を調整したい状況では、見出しの内容に対する制御が必要である。そこで、原文書中の着目したい語をキーワードとして指定することで、内容レベルで生成要約を制御する研究を進めた。このタスクに対して2つの手法を提案した。1つは、原文書中のキーワードの前後に特殊なトークンを埋め込むインラインアノテーション手法である。もう1つは、単語分散表現に対して、単語がキーワードかどうかを示す分散表現を連結ないし加算して利用する手法である。

手法の評価のため、日本語の新聞記事と見出しからなるデータセット JApAnese MUlti-length Headline Corpus (JAMUL) を利用し、1つの記事に対して複数のキーワードセットとそれに対応する人手で作成された見出しからなる新たなデータセットを構築した。提案モデルを時事通信社の日本語新聞記事のデータセットによって学習し、新たに構築したデータセットを用いて評価した。提案モデルが、キーワード指定によって要約を内容レベルで制御できることを示し、適切なキーワード指定によって ROUGE スコアが向上することを確認した。

2-6 ニュースの長文翻訳および要約文生成の評価

令和2年度は、最終的な評価に向けこれまで作成した「JJIQM 翻訳品質評価ガイドライン」のエラー項目等の改訂を行い、より利用しやすい基準とした。この改訂版 JJIQM に基づき、項目 2-1 から項目 2-4 で構築した日英、英日翻訳エンジンの評価を実施した。翻訳上のミスを深刻（-100点）、重度（-10点）、軽度（-1点）の3種類に分類し、エラー項目表に基づいて、専用ソフトを使って減点法で採点し、日本語1000文字あたり減点35点までを合格とした。合格点を100点に正規化した結果、日英評価では平均得点は17.7点、合格数は240記事中9記事、配信数は90記事中19記事、英日評価では平均得点は6.78点、合格数は127記事中0記事と、厳しい評価となった。今後は、ニュースらしい翻訳の生成などの研究課題の解決が必要と考えられる。

研究開発項目3：

3-1. マルチモーダル日英対訳コーパスの開発

昨年度に引き続き、Flickr30K Entities データセットの英文キャプションを日本語翻訳し、新たに31,783文の日本語文を作成した。これにより Flickr30K Entities に含まれる全ての英文キャプションの翻訳が完了し、合計で158,915文対からなるマルチモーダル日英対訳コーパスが完成した。これは、日英では現時点で最大のマルチモーダル対訳コーパスであり、語句と画像領域の対応関係を日本語訳においても保つように単語アラインメントを行っていることが大きな特徴である。本コーパスはLREC2020で発表し、データの公開を行った。さらに、本コーパスを用いて、WAT2020においてマルチモーダル日英翻訳タスクを新規に実施することで、研究成果の社会還元を行い、研究コミュニティをリードする活動を行った。本タスクはWAT2021においても引き続き実施中である。

3-2. シーン画像認識手法の開発

前年度に開発した自動データ拡張法を論文化し、コンピュータビジョンのトップ会議であるECCV2020で発表を行った。この手法に加え、14種類のCNNモデルによるネットワークアンサンブルを用いることで、100種類のシーン画像認識精度が80.7%となり、目標であった80%を超える精度を達成した。

さらに、複数の異なるシーン認識モデルによって得られる特徴をマルチモーダル機械翻訳において活用する手法の開発を行った。本手法は、動画像付き英中翻訳の国際コンペティションであるVideo-guided Machine Translation Challenge (ACL2020の併設WSにおいて開催)において、主催者提供の画像特徴のみを用いたベースラインと比較してBLEU値を1.34%向上させ、参加21チーム中の第1位となる成果を得た。

また、画像中のシーンのより詳細な認識理解を目指し、物体関係性認識の教師なし学習を世界で初めて実現した。また、本手法を半教師付き学習へ拡張したところ、教師付き学習を上回る精度を達成した。一般に、物体関係性認識の教師付きデータの作成コストは非常に高いため、本手法は有望な方向性の一つを示したと言える。

3-3. マルチモーダル翻訳手法の開発

昨年度まで継続して行ってきた教師付きアテンション制約を用いたマルチモーダル機械翻訳の研究を論文化し、自然言語処理分野のメジャー会議であるCOLING2020において発表を行った。この研究成果は自然言語処理分野のジャーナル「自然言語処理」に掲載され、2021年6月に掲載予定である。

また、項目2-2で開発した日英対訳コーパスに付随する記事画像を用いて画像付き日英ニュース記事対訳コーパスを開発した。人手で翻訳した約50万文対の時事通信ニュース記事対訳コーパスを対象として、それらの記事に付随する記事画像を記事画像データベースから抽出することにより、9,515文対(9,515枚画像)の画像付き英日対訳コーパスを実現した。50万文対のテキストだけから学習する場合に比べ、画像付き対訳コーパスを追加して学習することで翻訳精度(BLEU)が日英で+3.64%向上し、英日で+1.46%向上することを確認した。ニュース記事機

械翻訳において記事画像を用いることの有効性を示す結果であり、マルチモーダル機械翻訳の方向性の一つを示したと言える。

さらに、対訳コーパスと画像キャプションデータだけを用いて学習するマルチモーダル機械翻訳の研究開発を行った。マルチモーダル機械翻訳のためのリソース(画像-原言語文-目的言語文の3つ組)は希少なデータであるため、これらから十分な学習を行うことは難しい。本研究では、これらのマルチモーダル対訳コーパスを用いず、大量に存する対訳コーパスと画像キャプションデータだけを用いて学習するマルチモーダル機械翻訳手法を開発した。実験を行い、本手法により翻訳精度(BLEU)が+0.87%向上することを確認した。本手法では大量に存する対訳コーパスと画像キャプションデータだけを用いて学習することができるため、将来のマルチモーダル機械翻訳の利用可能性を大きく高めたと言える。

(8) 研究開発成果の展開・普及等に向けた計画・展望

- ニュースの日英機械翻訳モデルを翻訳業務フローに取り込み、業務効率化を目指すとともに、翻訳過程で生成される日英の記事ペアを蓄積してコーパス化する
- 取材映像や記者会見などの音声を文字化したのち、機械翻訳システムで英訳することで、英語の放送原稿や字幕を短時間で作成する字起こしシステムを開発する
- インバウンド向け英語ニュースの配信システムに、本委託研究で開発したニュースの日英機械翻訳モデルや日本語・英語の記事からの見出しの自動生成技術を活用する
- 本委託研究で新たな研究を開拓した文脈依存型の機械翻訳やマルチモーダル機械翻訳の研究開発を継続する
- Workshop on Asian Translation (WAT) などの共通タスクを通して、本委託研究で開発したコーパスの活用・普及を促進する