

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年度から平成32年度（3年間）

(3) 実施機関

国立大学法人東京工業大学 <研究代表者>
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（平成30年度300百万円）
※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、論文発表等

		累計（件）	当該年度（件）
特許出願	国内出願	1	1
	外国出願	0	0
外部発表	研究論文	1	1
	その他研究発表	20	20
	プレスリリース・報道	1	1
	展示会	0	0
	標準化提案	0	0

(7) 具体的な実施内容と成果

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目標

- ・会議や社交における対話・雑談の日英対話対訳コーパスの設計および構築を行う。
- ・既存の対訳コーパスを分析し、翻訳に必要な文脈情報を調査する。
- ・予備実験により、文脈を考慮する既存の機械翻訳モデルの有効性を調査する。

②実施内容

- ・複数のビジネスシーンを設定し、シーンに沿ったシナリオとその翻訳を作成した。
- ・既存の英語対話コーパスを日本語へ翻訳した。
- ・文脈を考慮する上で、何文前までの情報を保持すべきか調査するため、照応のアノテーションが付与されている日本語および英語の既存データを分析した。
- ・文脈が必要となる翻訳事例のテストセットを構築した。
- ・1文前を考慮する既存の機械翻訳モデルを利用し、英日新聞翻訳や上記のテストセットでその効果を検証した。

③成果

- ・シナリオ作成と翻訳により、44,800文のビジネスシーン対話対訳コーパスを構築した
- ・既存の英語対話コーパスの和訳により、約6万文の対話対訳コーパスを構築した。
- ・文書中の共参照やゼロ照応の先行詞の2~4割は2文以上前に出現することが判明した。
- ・1文前のみを考慮する既存の文脈翻訳手法は、ある程度の翻訳精度向上を示すものの、その精度向上は不十分であり、より広い文脈を考慮できるモデルの研究が必要であることを確認した。

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目標

- ・ニュース対訳コーパスや要約コーパスの構築に注力する。時事通信社の記事管理システ

ムから記事データを抽出し、翻訳や要約の訓練データを半自動的に獲得する。人手翻訳の対象とするテキストを選び、コストパフォーマンスが最も良くなる方法を調査したうえで、その翻訳作業を進める。平成30年度中に構築できたコーパスを用い、ニュース機械翻訳、および要約文生成のベースライン手法を実装し、既存手法の問題点を洗い出す。

②実施内容

- ・時事通信社の記事管理システムから過去の記事テキストを抽出し、日本語記事と英語記事の対応付けを実施した。
- ・読売新聞社から日英対訳コーパスを購入した。
- ・時事通信社日本語ニュースの年度毎の翻訳計画を立案し、翻訳方式のコストを考慮しながら業者に翻訳を発注した。
- ・時事通信社と朝日新聞社のニュース記事から見出し生成用データを構築した。
- ・時事通信社の日本語ニュースにおける新語の出現割合を調査した。
- ・ニューラルネットワークを利用したニュース機械翻訳のベースライン手法を実装した。
- ・新聞記事から見出しと短文要約を生成するベースライン手法を構築した。
- ・翻訳結果の評価のために、翻訳業務と照らし合わせたガイドライン生成に着手した。

③成果

- ・日英ニュース対訳データ約161万文対と、要約コーパス約616万事例を生成した。
- ・予備調査により、記事中の14.3%が未知語（新語を含む）であることが分かり、辞書の併用が研究課題であることを確認した。
- ・日英ニュース翻訳システムを試作し、客観評価（BLEU）と主観評価（TAUS DQF-MQM）ともに、既存システムよりも良い評価結果であることを確認した。
- ・日英ニュース翻訳において、訳抜け、誤訳、重複訳、文脈による補完・用語統一などの研究課題を見出した。さらに、新聞社固有の記事のスタイルを考慮した機械翻訳、という新しい研究課題を見出した。
- ・ニュース要約システムを試作し、英語の従来研究と同程度の性能を達成した。
- ・ニュース要約において、要約長の制御、含めるべき情報の制御、矛盾の回避などの研究課題を見出した。

研究開発項目3：マルチモーダル翻訳技術の研究開発

①目標

- ・マルチモーダル日英対訳コーパスの仕様を定め、5万文程度の英語キャプション翻訳作業を行う。
- ・シーン画像認識のベースラインシステムの実装を行い、既存手法の評価・課題抽出および環境整備を行う。
- ・画像情報を用いるマルチモーダル機械翻訳システムを試作し、予備的な評価を行うことにより今後の課題を見出す。

②実施内容

- ・マルチモーダル日英対訳コーパスの仕様策定および発注作業を行った。具体的には、既存の公開データセットである Flickr30K Entities を用い、元々存在する英文キャプションを日本語翻訳することとした。この時、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うことが特徴である。
- ・シーン認識のベースラインシステムを実装し、評価および課題抽出を行った。具体的には、シーン画像認識の研究分野において標準的に用いられる大規模場所カテゴリデータセットである MIT Places2 を用いて、既存の最新手法の実装を行い、認識精度の評価を行った。また、分析結果を受け、新規手法の設計を行った。
- ・Transformer ベースのマルチモーダル機械翻訳を試作し、既存のマルチモーダルコーパスである Multi30k を用いてその性能評価および結果の分析を行った。

③成果

- ・マルチモーダル日英対訳コーパスについて、63,566文の対訳文作成作業を完了し、実

際に利用可能なデータセットの第一段階が完成した。これは、日英では世界初となるマルチモーダル対訳コーパスであり、既存データセット（独英）よりも大規模である。また、画像領域と語句の対応情報が付いた対訳コーパスとしても世界初のものである。

- シーン画像認識手法のベースライン（既存の最新手法）を実装し、評価を完了した。その結果、100種類のシーンの識別精度は、ランダムに抽出したシーンでは約75%程度と目標に近かったが、室内シーンのみ限定すると約65%となり、視覚的に類似した対象の識別が課題であることが知見として得られた。これを受け、詳細識別において優れた実績を有する双線形プーリングを発展させた新規手法を設計し、開発に着手した。
- マルチモーダル機械翻訳の試作と予備的評価を完了した。英独翻訳の性能評価を行い、画像情報を追加することで機械翻訳の精度(BLEU)が0.96向上することを確認できた(BLEU 34.30 → 35.26)。また、アテンションを可視化することで、単語と画像上のオブジェクトとの関連も学習できていることが確認できた。さらに、失敗例の分析を行い、今後の性能向上に向けての考察と課題抽出を行った。