

令和 4 年度研究開発成果概要書

採 択 番 号 22501
研究開発課題名 自動翻訳の精度向上のための「マルチモーダル情報の外部制御可能なモデリング」の研究開発
副 題 マルチモーダル情報理解と制御可能なテキスト生成の研究開発

(1) 研究開発の目的

人間同士および人間とコンピュータ間の円滑なコミュニケーションの実現に向けて、コンピュータが文脈・状況を考慮しながら、言語、音声、画像、動画、表などによるマルチモーダルな情報を統合的に理解する手法を探求する。また、長さや焦点、スタイル、難易度などを指定したり、外部知識を与えることによって、コンピュータが人間のニーズに合わせてテキストを生成する技術を確立する。さらに、これらの成果を統合することで、自然言語生成の新たな応用を開拓するとともに、自動翻訳や同時通訳の研究開発の高度化につなげる。

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進め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。研究と並行して言語資源の開発に注力し、その成果物を研究コミュニティに還元する。これにより、日本語の自然言語処理に関する研究で世界トップレベルを維持するとともに、マルチモーダル情報理解や制御可能な自然言語生成などの研究分野で、世界に先駆けた研究を展開する。

(2) 研究開発期間

令和 3 年度から令和 7 年度 (5 年間)

(3) 受託者

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
東京都公立大学法人
日本放送協会
株式会社時事通信社

(4) 研究開発予算 (契約額)

令和 3 年度から令和 5 年度までの総額 230 百万円 (令和 4 年度 100 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 マルチモーダル情報理解技術の研究開発

- 1-1 マルチモーダル動画対訳コーパスに関する研究開発 (国立大学法人東京大学)
- 1-2 マルチモーダル機械翻訳に関する研究開発 (国立大学法人愛媛大学)
- 1-3 マルチモーダル情報理解に関する研究開発 (国立大学法人東京大学)

研究開発項目 2 制御可能なテキスト生成技術の研究開発

- 2-1 自動要約の制御に関する研究開発 (国立大学法人東京工業大学)
- 2-2 翻訳の制御に関する研究開発 (日本放送協会)
- 2-3 スタイルの制御に関する研究開発 (東京都公立大学法人)
- 2-4 データ整備に関する研究開発 (株式会社時事通信社)

(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	0	0
	外国出願	0	0
外部発表等	研究論文	1	1
	その他研究発表	44	43
	標準化提案・採択	0	0
	プレスリリース・報道	0	0
	展示会	0	0
	受賞・表彰	6	6

(7) 具体的な実施内容と成果

研究開発項目1 マルチモーダル情報理解技術の研究開発

1-1. マルチモーダル動画対訳コーパスに関する研究開発

マルチモーダル動画対訳コーパスの作成を開始し、一部データの試作と分析を行った。プレゼンテーション動画対訳コーパスは YouCook2、インタラクション動画対訳コーパスについては AVSD と呼ばれる既存の動画キャプションデータセットをベースに作成することを決定し、それぞれ動画を参照しながらオリジナルの英語キャプションを日本語翻訳することで、動画にグラウンディングされた対訳文対の作成を行った。この結果、プレゼンテーション動画対訳コーパスでは約 8 時間・1500 文対、インタラクション動画対訳コーパスでは約 1 時間・1300 文対からなる部分データをそれぞれ作成完了し、本年度の目標を達成した。

1-2. マルチモーダル機械翻訳に関する研究開発

マルチモーダル機械翻訳に関して、音声・映像付きマルチモーダル対訳コーパスの開発を行い、拡散モデルを用いたマルチモーダル機械翻訳の研究を行った。マルチモーダル対訳コーパスの開発について、TED Talks の対訳データである MSLT17 対訳コーパス(英日 22.3 万文対)と原ビデオの紐付けを行い、映像、画像、音声、自動書き起こし文、人手書き起こし文のアライメントを行うマルチモーダル対訳コーパスの試作を行い、マルチモーダル機械翻訳による性能評価を行った。マルチモーダル機械翻訳の研究においては、拡散モデルを用いて機械翻訳に不要な画像情報を除去する手法を考案した。マルチモーダル機械翻訳では、画像に含まれる翻訳に不要な情報がノイズとなり、却って性能が低下する場合がある。本研究では、拡散モデルを用いて、入力画像を原言語文の説明に沿った画像に変換することで、不要な画像情報を除去するマルチモーダル機械翻訳手法を考案した。実験によりマルチモーダル機械翻訳の性能が向上することを確認した。

1-3. マルチモーダル情報理解に関する研究開発

本年度は、各モダリティからの基本的な情報抽出の精度を向上させる技術や、複数モダリティのアライメントに関する基礎技術の開発を行った。まず、学習に基づくデータ拡張や入力量子化により、テスト時の未知ノイズに対し頑健な画像認識手法を複数開発した。また、カリキュラム学習により画像とテキストのグラウンディングを改善させる手法を開発し、Vision and Language タスクの精度向上が実現できることを示した。さらに、画像からのシーングラフ生成においてデータの頻度に関するバイアスを低減する手法や、外部知識グラフを用いて未来予測を行うキャプション生成手法を開発し、その有効性を示した。

研究開発項目2 制御可能なテキスト生成技術の研究開発

2-1. 自動要約の制御に関する研究開発

対話に関する自動要約および制御に関する研究開発を進めた。まず、要約対象の文書に対してクエリと要約の両方を自動生成する「クエリ推奨付き要約」を提案し、具体的なタスクと評価方法の設計を行った。文書の不要な部分を事前に予測して除去する手法を提案し、対話コーパス QMSum を対象とした実験により、その有効性を示した。また、特定の自然言語タスク(ここでは含意関係認識)においてシステムが予測を行う際に、その予測結果の説明を生成しながら人間と議論を行

い、人間とコンピュータと一緒にタスクを解決する研究に取り組んだ。実験により、提案システムが有益な議論ができること、問題に関する議論がシステムの性能改善につながることを示した。前者の研究は言語処理学会第29回年次大会のスポンサー賞（日立製作所）、後者の研究は言語処理学会第29回年次大会の委員特別賞を、それぞれ受賞した。

2-2. 翻訳の制御に関する研究開発

英語ニュースでは一般に一文内に同一の表現を繰り返し使うことを避け、言い換えや省略によって簡潔な文にすることが推奨される。一方で、日本語ニュースではこうした簡潔性が英語ニュースほどは考慮されず、一文内に同一の表現が複数回出てくることも珍しくない。このような日本語ニュースと英語ニュースのライティングスタイルの差分を分析し、日英のライティングスタイルを考慮した機械翻訳のための日英ニュース記事対のテストセットを構築した。また、機械翻訳の出力に対して編集量のパラメータを与えて生成結果を制御する手法を考案し、実験により有意に翻訳精度が向上することを確認した。

2-3. スタイルの制御に関する研究開発

日本語のテキストスタイル変換に関する研究開発を進めた。日本語のテキスト平易化に関する変換器構築に必要なコーパスとして、2013年～2020年の8年分の毎日新聞記事-毎日小学生新聞記事から統計的手法を用い対応する400記事ペア（各年50記事ペア）を取得し文間対応及び実施されている平易化操作ラベルを付与した。またテキスト平易化に用いることのできる言語資源として改変・再配布等が可能なライセンス及びジャンルの観点からWikipediaの秀逸な記事、良質な記事を取得し、各記事の概要欄の平易化テキストを作成した。新聞と同様、平易化前後の文間の対応付けと平易化操作ラベルの付与を行なった。

2-4. データ整備に関する研究開発

時事通信社の編集、配信システムから、コーパスの構築など本研究に資するデータ（日本語と英語の記事セット、写真と書誌情報のセット、こどもニュース）の抽出を開始し、各機関に提供した。また、子供ニュースに関しては、作成段階から一般記事、英文記事、写真、関連資料と紐付ける作業も実施。新開発のアプリを使い、年度内に66セットのデータを作成した。記事データからコーパスを構築する場合には、特に個人情報の扱いに難しさがある。個人名を自動で匿名化するツールを試験的に導入し、その有効性について検討を行った。

(8) 今後の研究開発計画

研究開発項目1 マルチモーダル情報理解技術の研究開発

1-1. マルチモーダル動画対訳コーパスに関する研究開発

引き続きマルチモーダル動画対訳コーパスの拡張作業を進めると共に、今年度作成したデータを用いてマルチモーダル機械翻訳の実験を行い、コーパスの有効性について検証する。また、作成したデータについて順次公開の準備を行い、コンペティションのセットアップについても検討を進める。

1-2. マルチモーダル機械翻訳に関する研究開発

音声を含むマルチモーダル機械翻訳のための研究開発を進める。IWSLT17 対訳コーパスに対し原ビデオとの文単位アライメント対応が今年度完了したが、不完全なアライメントが多く見受けられたため、これらの改善を今後行う。既存の音声付き対訳コーパスや本プロジェクトで開発した音声・映像付きマルチモーダル対訳コーパスを用いて、音声・画像・映像を用いたマルチモーダル機械翻訳の研究・開発を進める。

1-3. マルチモーダル情報理解に関する研究開発

画像からのシーングラフ出力の手法開発を進める。また、グラフによる知識表現を利用した応用タスクについても検討を行う。さらに、シーングラフに限らず、画像と言語を接地するための基礎技術について、最新の大規模マルチモーダル言語モデル等の動向を眺みながら多角的に検討・開発

を進める。

研究開発項目2 制御可能なテキスト生成技術の研究開発

2-1. 自動要約の制御に関する研究開発

2022年度は既存のコーパスを用いて自動要約や生成の制御に関する予備的な研究を進めた。一方で、OpenAI が発表した ChatGPT や GPT-4 など、大規模言語モデルを用いた生成型 AI が急速な進展を見せた年でもあった。今後は、日本語の要約タスクのコーパスを構築し、国際技術コンペを実施する予定であるが、最新の研究動向を踏まえながらタスクを設計し、自然言語生成に関する新しい手法の研究・開発を進める。

2-2. 翻訳の制御に関する研究開発

翻訳出力制御を評価するためデータを構築する。また、一文を超えた情報を利用して語彙、文法のきめ細かな制御を可能とする翻訳システムを試作し、予備的な評価を実施する。さらに、外部公開可能な翻訳制御のテストデータを作成し、国際技術コンペを企画する。

2-3. スタイルの制御に関する研究開発

日本語のテキスト平易化に関する研究開発を進める。テキスト平易化コーパス構築作業を継続しつつ、ニューラル系列変換モデルによる日本語のテキスト平易化器の開発を進める。構築したコーパスを使ったテキスト難易度制御のコンペティションを開催する。

2-4. データ整備に関する研究開発

引き続き時事通信社の編集・配信システムから、必要なデータの抽出作業を行う。個人名等の匿名化処理について検討を続けるほか、個人情報保護法、著作権法などの各種法令、その趣旨との関係に十分留意しながら、コーパス構築の作業を支援する。