様式 1-4-2 (2022-1)

令和６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

国立大学法人東京科学大学＜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
国立大学法人一橋大学
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７年度までの総額 430 百万円（令和６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　（国立大学法人東京科学大学）
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　（国立大学法人一橋大学）
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	9	4
	その他研究発表	114	30
	標準化提案・採択	0	0
	プレスリリース・報道	2	1
	展示会	0	0
	受賞・表彰	12	1

（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

前年度に引き続き、マルチモーダル動画対訳コーパスの拡張作業を行った。プレゼンテーション動画対訳コーパスについては、YouCook2 データセットの訓練データ・検証データの全ての日本語翻訳を完了し、約 75 時間・1万4千文対からなる YouCook2-JP を完成させた。また、インタラクション動画対訳コーパスについても、AVSD データセットをベースに拡張作業を進め、合計で約2時間・5千文対となった。さらに、画像をピボットとするマルチモーダル翻訳の訓練において、敵対的学習によるアラインメントを活用する手法を開発した。

1−2 マルチモーダル機械翻訳に関する研究開発

本年度は、①音声・画像付きスピーチ対訳を用いたマルチモーダル機械翻訳、②漫画を対象としたマルチモーダル機械翻訳を行った。①について、研究開発項目 1-1 で開発した料理動画英日対訳コーパス YouCook2-JP から画像を抽出することでマルチモーダル機械翻訳のための評価用データセット(150 文)を作成し、画像の説明を与える思考の連鎖（Chain-of-Thought）を用いたマルチモーダル機械翻訳手法を開発した。GPT-4o および Qwen2-VL を用いた実験を行い、思考の連鎖を用いることで評価用データセットに対する翻訳精度が大きく改善されることを確認した。②について、漫画を対象としたマルチモーダル機械翻訳の研究を行い、コマを超えた吹き出し間の文脈情報を扱う機械翻訳手法、および著者情報、出版社情報、ジャンル情報などの書誌情報を用いた機械翻訳手法を開発し、この研究成果論文は LREC-COLING 2024 で発表した。

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、画像情報を検索拡張の枠組みで大規模言語モデル(LLM)へ接続する手法の開発に注力した。まず、ごく少数の画像とラベルからなる画像辞書を外部知識として接続する、オープンワールド画像キャプショニング手法を開発した。本手法は二つの Q-Former を用い、検索のための画像埋め込み表現と、拡張された単語の埋め込み表現をそれぞれ学習することにより、マルチモーダルな検索拡張を実現している。実験により、提案手法は他の最新手法と比較して、数十分の一程度の小さなモデルサイズでありながら同等以上のキャプショニング精度を達成できることが示された。本成果は、コンピュータビジョンのトップ国際会議である CVPR 2024 で発表した。

さらに、前年度までに開発したシーングラフ認識手法の出力結果を用い、マルチモーダル LLM の出力結果を自己検証・修正させる手法の開発に着手した。GPT-4 による自動評価を用いた基礎的な検証の結果、提案手法により LLM のシーングラフ認識精度の改善が可能であることが確かめられた。

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

大規模言語モデルによる言語生成およびその制御に関する研究を進めた。大規模言語モデル（生徒役）から応答を引き出すときに、別の大規模言語モデル（教師役）からのフィードバックを与えることで、応答の質を向上させる研究が知られている。このとき、生徒が好ましくない応答を返した時だけにフィードバックをするのではなく、適切な応答を返した場合にも生徒を惑わすようなフィードバックを与える方が、生徒の応答の質が高まることを実証し、その研究を 27th European Conference on Artificial Intelligence (ECAI) で発表した。また、与えられたテキストが大規模言語モデルによって生成されたものであるか判定するタスク（LLM 検出タスク）において、生成内容を制御するプロンプトの有無が LLM 検出タスクの性能に大きな影響を与えることを示し、Findings of the Association for Computational Linguistics: EMNLP 2024 (EMNLP) で発表した。また、大規模言語モデルの生成に透かしを入れる従来手法では、透かしを入れることにより翻訳や自動要約などのタスクの性能が低下することが知られていた。そこで、LLM 検出器から与えられる報酬と、自動要約の評価器から与えられる報酬に基づく強化学習フレームワークを設計し、要約の性能を維持しながら LLM によって生成されたテキストであると検出されやすくする手法を提案し、言語処理学会第 31 回年次大会で発表した。

２－２ 翻訳の制御に関する研究開発

日本語と英語との間におけるニュースのライティングスタイルの違いを考慮した機械翻訳の研究を進めた。日本語ニュースと英語ニュースを比較して語句の繰り返しの傾向をいくつかのパターンに分類して識別実験を行った結果、大規模言語モデルにいくつかの見本を提示した上で予測を行う文脈内学習（in-context learning）を適用することで予測性能が伸びていくことが分かった。また、翻訳処理全体を扱う課題として、昨年度に引き続き、語句の繰り返しに関する訳出制御のシェアードタスクを The Ninth Conference on Machine Translation (WMT24) にて開催した。本年度は、評価用のテストセットの分量を拡大し、機械翻訳が出力した英文の表現の繰り返しの度合いを自動評価する手法を新たに導入した。

２－３ スタイルの制御に関する研究開発

日本語のスタイル変換に関する研究開発を進めた。2023 年度までに構築した日本語の Wikipedia 平易化コーパス（JADOS）のデータを用い、英語・ドイツ語の文書単位のテキスト平易化コーパスと合わせた共通タスクを提案・公開した。また、JADOS の開発データに対し、JaBART、Shallow、Gemma、GPT-4o の 4 モデルのテキスト平易化の出力を得て、人手による平易度ラベルの付与を行った。

２－４ データ整備に関する研究開発

時事通信社の編集・配信システムから日英記事データ、写真データ、子どもニュースデータなどを抽出、各研究機関に配布した。データ公開も視野に入れ、タグ構造を統一するなど、扱いやすさに留意した。これまでに抽出・配布した記事・写真データは、日英記事セット 10 万本、日本語記事 460 万本、記事中写真データ 57 万枚などである。個人名等の匿名化については、編集・配信システムから要配慮情報を含む記事を抽出し、そこから分類モデルを構築。再現率で 80－90%程度の性能を示した。さらに、日英対訳コーパス構築の効率化などを目指し、機械翻訳結果を LLM で自動校正するタスクに取り組んだ。

（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

完成されたプレゼンテーション動画対訳コーパス(YouCook2-JP)を用い、課題１－３と連携しながらマルチモーダル機械翻訳手法の実装を行うとともに、WAT 等で国際コンペティションを実施する。また、インタラクション動画対訳コーパスについて引き続き拡張作業を行い、年度前半に完成させる。二つのコーパスの最終版を年度終了までに公開する。

1−2 マルチモーダル機械翻訳に関する研究開発

引き続き 1-1 および 1-2 で開発されたマルチモーダル動画対訳コーパスを用いたマルチモーダル機械翻訳の研究開発を進める。特に LLM に対する思考の連鎖を用いたマルチモーダル機械翻訳手法の研究開発を進め、研究成果を論文にまとめて投稿する。

1−3 マルチモーダル情報理解に関する研究開発

引き続き、シーングラフ認識に基づくマルチモーダル LLM の自己検証・修正手法の開発に取り組み、画像に関する質問応答タスクなどでの精度改善を目指す。また、近年の LLM の著しい進歩を鑑み、マルチモーダル情報理解の達成度をより解像度高く評価するための新しいベンチマークデータセットの開発を進める。

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

昨年度は日本語に強くオープンな大規模言語モデルの研究開発が進んだ１年であった。今年度も最新の研究動向を踏まえながら、自動要約や自然言語生成の制御に関する研究開発を大規模言語モデル上で実施する。例えば、東京科学大学で開発している大規模言語モデル Swallow に基づく自動要約や自然言語生成の制御について研究を進める。

2−2 翻訳の制御に関する研究開発

引き続き、翻訳出力制御に関する研究開発を進める。複数文から構成される文書において、文と文のつながりや文書のスタイルを考慮して機械翻訳結果を制御する手法の研究を行い、統一感のある書式での翻訳性能の改善を行う。外部公開可能な翻訳制御のテストデータを新たに作成し、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。難易度の自動推定および難易度をコントロールできる日本語のテキスト平易化器を開発しつつ、文書生成時のハルシネーションに関するコーパスを構築し、大規模言語モデルの振る舞いについての分析を行う。

2-4　データ整備に関する研究開発

引き続き、時事通信社の社内システムから必要なデータの抽出作用を行う。また、各研究機関の要望なども聞きながら、必要な支援を行う。個人名の匿名化については、今年度取り組んだ手法や既存手法などを再検証し、可能な限り精度を高める方策を探り、実行する。