

令和 3 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 3 年度 30 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	0	0
	その他研究発表	1	1
	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	0	0

(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仕様設計を開始した。まず、既存の動画対訳コーパスを調査した結果、基本的に動画キャプションングタスクのテキストに基づいていることを確認した。これに対し、本研究項目ではプレゼンテーション・インタラクションのように、単に動画の状況説明ではない発展的なコミュニケーションのためのテキストをベースとすることで、新規性が高く価値あるコーパスの実現が見込めるとの感触を得た。また、具体的なコーパス作成のアプローチについても、費用対効果・公開可能性等を念頭に置きながら、次のように検討を進めた。本研究において開発したコーパスは速やかに公開する必要があるため、権利関係やプライバシー面で懸念のあるデータを用いることは適切ではない。このため、これらの問題がクリアされている既存の公開動画データセットの英語テキストを日本語翻訳し、その日本語テキストのみを公開することが効率的なアプローチであると判断した。ただし、既存の公開動画データセットも動画自体の配布は行っていないことがほとんどであるため、再現性の観点から、特徴量データを公開している YouCook2 や HowTo100M などのデータセットを利用することが、現段階の最も有望な方針である。

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、既存のスピーチ機械翻訳の調査と音声・映像付きマルチモーダル対訳コーパスの準備を行った。既存のスピーチ機械翻訳ソフトを実際の会議に用いて、音声認識の問題および機械翻訳の問題を調査した。音声認識に問題があることが判明し、より精度の高い音声認識ソフトを用いることが必要であることがわかった。マルチモーダル対訳コーパスの準備について、TED Talks の対訳データである MSLT17 対訳コーパス(英日 223 万文対)と原ビデオの紐付けを行い、マルチモーダル対訳コーパスデータの準備を行った。音声と対応していない箇所があるなど、いくつかの問題点があることが判明した。

1-3 マルチモーダル情報理解に関する研究開発

マルチモーダル情報理解に関する既存手法の調査と問題点の整理を行った。現在のマルチモーダル情報処理では、大規模な画像・テキストペアデータを用いた教師なし学習により事前学習済みモデルを得るアプローチがトレンドであるが、このようなモデルでは説明性の欠如や、学習データに起因する不公正なバイアスが大きな未解決課題である。このことから、本研究項目で提案する明示的なグラフベースの情報理解が有望であることが示唆された。また、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の検討に着手した。具体的には、画像認識におけるニューラルモデルの選択を検証データなしで行う手法や、自然言語処理における単語埋め込みに複数の双曲空間の積空間を用いる手法を開発し、論文投稿を行った。

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

近年の言語生成モデルの発展により、見出し生成（記事に対して見出しを作文する）タスクの性能が向上し、人間が作成した見出しに近い品質の見出しを自動で生成できるようになった。しかし、自動見出し生成は記事内容から逸脱した見出しを生成することがある、という問題が報告されている。従来研究では、訓練データ中で忠実性が低い見出しを取り除くアプローチが提案されたが、見出し生成モデルの学習に用いる訓練事例数が減少するため、生成される見出しの品質が低下するという問題があった。そこで、マスク付き言語モデルを用いて忠実性の低い訓練事例を忠実性の高い事例に書き換える手法、タグにより見出しの忠実性を制御する手法を検討した。自動評価と人手評価の結果から、提案手法は既存手法と同程度の忠実性を保ちながら品質の高い見出しを生成できることを確認した。

また、DialogSum と呼ばれる英語の対話の要約コーパスを日本語に翻訳し、対話の書き起しを要約する研究について検討を進めた。

2-2 翻訳の制御に関する研究開発

翻訳の制御対象とする日本語と英語のニュース記事に対して、その差分を抽出するためのアノテーションの指針を策定した。まず、複数文で構成される日本語と英語のニュースに対して、文のアライメントを実施する。この際、複数文対複数文の対応も許可している。次に、対応付けられた文対において、相手言語側に対応する内容が無い表現、および対応する内容は存在するが、日付と曜日または代名詞と固有名詞など情報の内容が大きく違っている表現に対してタグを付与する。この指針に沿って、少量の時事通信社のニュースに対して差分を抽出するためのタグ付与を実施し、タグ付与作業における問題点を洗い出した。

2-3 スタイルの制御に関する研究開発

日本語のテキスト平易化に関する変換器構築に必要なコーパス（一般向けの新聞記事のテキストと、同じ時期の小学生新聞の記事）を取得した。テキスト平易化に用いることのできる言語資源を整理し、レベル別に統制した語彙による語彙平易化を行うシステムのベースラインを試作した。また、次年度以降の平易化の人手タグ付け対象とするコーパスについて、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検討を進めた。

2-4 データ整備に関する研究開発

本研究に資するデータの整備に向け、時事通信社の社内システムの確認、必要な改修を行うとともに、個人名の匿名化処理に関する検討などを進めた。一般ニュースについては、日本語記事、日本語と英語が対になった記事、記事中の写真やキャプチャを取得する方向でデータベースを確認し、異なるフォーマットを統一するための準備作業を実施した。また、子供ニュースを一括取得するための仕組みを新たに構築した。映像・音声については、ストックフォトサービス内のデータベースから書誌情報を確認した。特に公の場での記者会見の動画などに関し、抽出対象となり得るかの洗い出し等を行った。個人名等の匿名化について、既存サービス提供企業からヒアリングを行うなど、公開を前提としたコーパス作成にあたり、匿名化処理の検討を開始した。

(8) 今後の研究開発計画

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設計を完了し、翻訳作業を外注するための詳細な仕様を決定する。また、翻訳作業ではテキストと同時に動画内容を確認する必要があるため、作業を効率化するためのインターフェースの開発も行う。2022年度中に一部データの発注を行う。

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。MSLT17 対訳コーパスに対し原ビデオとの対応が今年度完了したので、今後は、ビデオ内の音声と対訳文とのアライメントを行う。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

引き続き、各モダリティにおける情報抽出や認識理解の精度向上を目指した技術開発を行うと共に、シーングラフに基づくマルチモーダル情報の統合手法についても検討を進める。

2-1 自動要約の制御に関する研究開発

新聞記事、対話のドメインにおいて、自動要約の制御に関するコーパスの構築および新しい手法の開発を進める。2021 年度に試作したコーパスの分析や、コーパス構築のコストパフォーマンスの検討を進めながら、2022 年度に構築するコーパスの仕様策定を進める。また、既存のコーパスを用い、自動要約の制御に関する予備的な研究開発を進める。

2-2 翻訳の制御に関する研究開発

日英機械翻訳における訳語補完の制御に関する研究開発を進める。2021 年度に作成した日英ニュース記事の差分を抽出するためのアノテーションの指針に従い、大量の日英ニュースに対して差分のアノテーションを実施する。この日英の差分をアノテーションしたコーパスから訳語補完されやすい特徴を持つ語を判別し、その語に関する一文を超えた情報（Knowledge-graph、文脈、直前までのニュース、Web など）から獲得して訳語補完を制御する手法を研究開発する。

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。統計的手法で一般向けの新聞記事と小学生向けの平易な記事の対応づけを行い。また、対応づけを行った記事を基に文の対応づけを行い、文単位で対応のついたテキスト平易化コーパスを作成して、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を構築する。構築した平易化器を用いて多様な平易文を生成し、平易化の種類や平易度などのメタ情報を人手で付与することによってテキスト平易化コーパスを構築する。

2-4 データ整備に関する研究開発

時事通信社の社内システムから、必要なデータを抽出し、各機関に提供する。個人名等の匿名化処理について引き続き検討を進めるほか、個人情報保護法、著作権法など各種法令、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。