

1. 研究課題・受託者・研究開発期間・研究開発予算

- ◆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
- ◆受託者 (大) 東京科学大学、(大) 東京大学、(大) 愛媛大学、(大) 一橋大学、日本放送協会、(株) 時事通信社
- ◆研究開発期間 令和3年度～令和7年度 (5年間)
- ◆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7年度100百万円)

2. 研究開発の目標

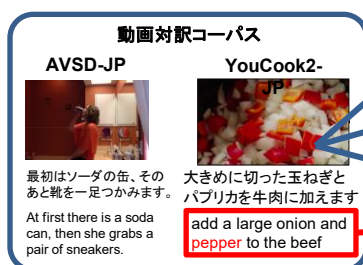
人間同士および人間とコンピュータ間の円滑なコミュニケーションの実現に向けて、コンピュータが文脈・状況を考慮しながら、言語、音声、画像、動画、表などによるマルチモーダルな情報を統合的に理解する手法を探索する(研究開発項目1)。また、長さや焦点、スタイル、難易度などを指定したり、外部知識を与えることによって、コンピュータが人間のニーズに合わせてテキストを生成する技術を確認する(研究開発項目2)。さらに、これらの成果を統合することで、自然言語生成の新たな応用を開拓するとともに、自動翻訳や同時通訳の研究開発の高度化につなげる。

3. 研究開発の成果

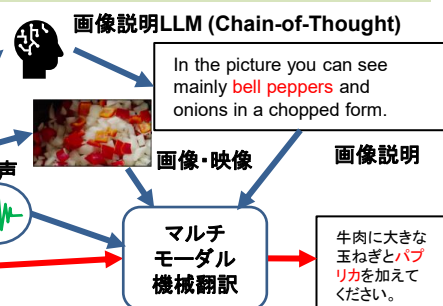
研究開発項目1: マルチモーダル情報理解技術の研究開発

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索する

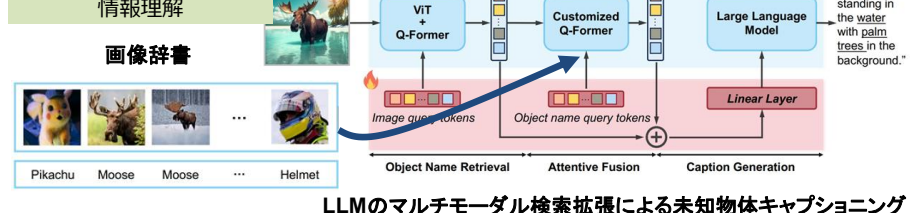
項目1-1 マルチモーダル動画対訳コーパスの作成



項目1-2 マルチモーダル機械翻訳



項目1-3 マルチモーダル情報理解



項目1-1成果: マルチモーダル動画対訳コーパス

- 既存の動画画像データセットの人手翻訳により、動画長約75時間・1万3千文対からなるプレゼンテーション動画対訳コーパス(YouCook2-JP) および動画長約8.7時間・1万9千文対からなるインタラクション動画対訳コーパス(AVSD-JP)を開発・公開した
- 開発したYouCook2-JPを用いた国際技術コンペティションをWAT2025で実施した

項目1-2成果: マルチモーダル機械翻訳

- 音声認識誤り訂正を用いた講演動画のためのマルチモーダル機械翻訳を開発した
- 拡散モデルを用い、入力画像を原言語文の説明に沿った画像に変換するマルチモーダル機械翻訳手法を開発した
- 項目1-1で開発した料理動画英日対訳コーパス(YouCook2-JP)に対し、LLMおよび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
- 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、コマを超えた文脈情報を扱う機械翻訳手法および書誌情報をを用いた機械翻訳手法を開発した

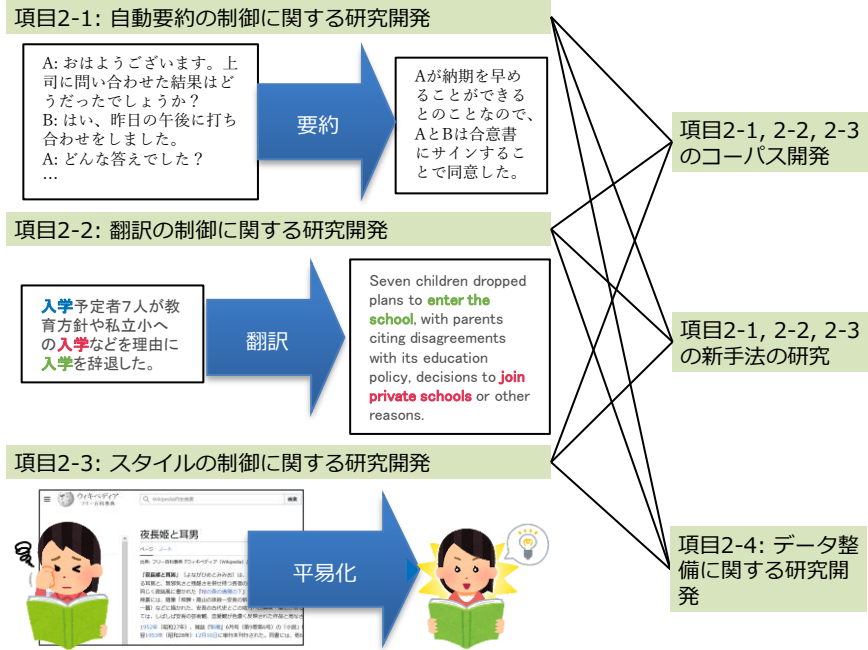
項目1-3成果: マルチモーダル情報理解

- 画像識別やシーングラフ生成のためのデータ拡張手法を複数開発した
- シーングラフに基づく画像・テキスト情報の統合方法や、外部知識の接続方法を開発した。また、これを応用した予測画像キャプションングを世界で初めて実現した
- LLMのマルチモーダル検索拡張による未知物体キャプションング手法を開発した
- 知識グラフに基づくLLM対話エージェントの自己検証・修正手法を開発した

3. 研究開発の成果

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

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り、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



項目2-1成果: 自動要約の制御に関する研究開発

- ビジネスシーン対話要約コーパス(2万文規模)を構築・公開した
- 対話によりLLMの制御と性能向上を達成できることを実証した
- LLMで生成された文の検出やその自動評価など、新しい研究を進めた
- 項目1との融合研究として図や表のキャプション生成のコーパス・手法を構築した

項目2-2成果: 翻訳の制御に関する研究開発

- ニュースの書式を考慮した日英ニュース翻訳用のテストデータを3種類構築し、WAT2023, WMT2024, WAT2025のshared taskとして国際技術コンペを実施した
- 複数の大規模言語モデルを活用した翻訳モデルと、大規模言語モデルを活用した翻訳モデルの学習データのデータ拡張手法を提案し、有効性を実証した

項目2-3成果: スタイルの制御に関する研究開発

- 文書単位の日本語テキスト平易化に向け、一般向け新聞記事と小学生向け新聞記事およびWikipedia記事から文書単位の編集タグ付きコーパスを構築し、その詳細および性能について国際会議で発表するとともに、コンペのサイトを公開した
- 文書単位の英語のテキスト平易化において難易度が制御可能な手法について、LLMを用いたモデルを提案し、有効性を実証した

項目2-4成果: データ整備に関する研究開発

- 編集・配信システムより日英記事、写真、子供ニュース等のデータを配布した
- 記事データを活用し、要配慮情報を含む記事を分類するモデルを構築した

4. 特許出願、論文発表等、及びトピックス

国内出願	外国出願	研究論文	その他研究発表	標準化提案 標準化採択	プレスリリース 報道	展示会	受賞・表彰
0 (0)	0 (0)	13 (3)	154 (39)	0 (0)	4 (1)	1 (1)	19 (6)

※成果数は累計件数、()内は当該年度の件数です。

5. 研究開発成果の展開・普及等に向けた計画・展望

本研究では、動画・音声・文章など複数の情報を統合的に理解し、目的に応じた文章を生成できる新しいAI技術を開発した。今後は、本技術をマルチモーダルAIやエージェント型AIなどの研究基盤として展開するとともに、行政文書処理や多言語対応を想定した実証・試作を通じて、実用化に向けた可能性を高める。研究成果はデータやソフトウェアとして公開し、地方や中小企業も含めた幅広い活用を促すことで、国産AI基盤の確立と人材育成に貢献する。