

令和 5 年 6 月 14 日現在

機関番号：34304

研究種目：基盤研究(C)（一般）

研究期間：2018～2022

課題番号：18K11557

研究課題名（和文）言語運用能力向上のためのAIのマルチメディアデータ読解力に関する研究

研究課題名（英文）Research on AI's reading comprehension of multimedia data for improving language proficiency

研究代表者

宮森 恒（Miyamori, Hisashi）

京都産業大学・情報理工学部・教授

研究者番号：90287988

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：本研究の目的は、画像や図表等のマルチメディアデータを活用したモデルを構築することにより、従来の自然言語処理だけでは得られない、より人間に近い確な読解力を実現する基盤技術を確立することである。

主な成果として、テキストと、テキストから得られる画像の特徴表現を融合的に利用した質問応答手法、ユーザが調べたいテキストから、表形式で記載された統計データを対象としてランキングし関連づける手法、文エンコーダによるクエリ指向要約を行い、把握した内容を説明する手法が挙げられる。

研究成果の学術的意義や社会的意義

本研究の学術的独自性は、文章を読みながら頭の中に視覚的要素を思い描いたり、文章を読む際に様々な知識や文脈と関係付けながらその内容理解を促進するといった、より人間らしい読解力を、画像や図表といったマルチメディアデータを活用したモデル構築により実現することに着眼している点である。本研究は、人とのよりの確なインタラクションが求められるアプリケーションに広く応用することが可能であり、あらゆる場面で快適で豊かな生活を実現する超スマート社会の根幹を担う技術の一つとして貢献することが期待できる。

研究成果の概要（英文）：This research aims to establish an essential technology to realize more human-like accurate reading comprehension, which cannot be achieved by conventional natural language processing alone, by building a model that utilizes multimedia data such as images and charts. The main achievements include a question-answering method that jointly utilizes text and feature expressions of images derived from text, a method for ranking and relating statistical data described in tabular form from text that the user wants to examine, and a method for performing query-oriented summarization using sentence encoders to explain what has been grasped.

研究分野：マルチメディアデータ工学

キーワード：マルチモーダル 読解力 統計データ リランキング 質問応答 画像特徴量 深層学習 人工知能

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

AIの自然言語運用能力には、「読む」「話す」「聞く」「書く」があると考えられるが、本研究では、読む力、すなわち、AIの読解力に焦点を当てる。読解力は、文章に様々な意味を見出す能力であり、人間らしい創造力の基盤となる能力と考えられる。

従来の自然言語処理研究においては、語義曖昧性解消や照応解析、談話構造解析をはじめ、読解力に関係する様々な要素技術が研究されてきたが、読解力そのものに直接着目した研究としては大きく2つの例を挙げることができる。一つは、文章の重要な箇所を制限字数でまとめる自動要約の研究である。これは、読解力の中でも、文章を読み「要点を的確に取捨選択する能力」と捉えることができる。もう一方は、大学入試センター試験における国語や英語での長文読解問題に対する自動解答についての研究である。これは、文章を読んで「理解した内容を選択肢に結びつける能力」と捉えることができる。

これらの研究はいずれも、基本的に自然言語が書かれたテキストデータのみを用いて研究がなされてきた。しかし、人間の読解力というものは、テキストデータのみでの処理によって構成されているのだろうか。

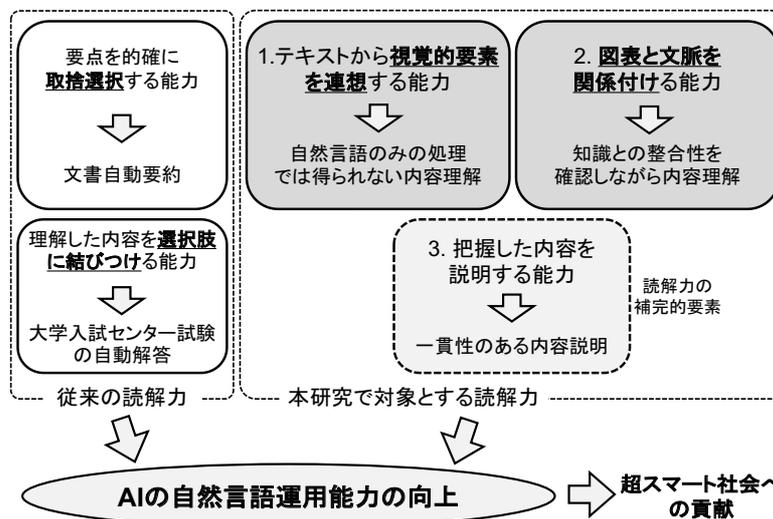


図1. 本研究の概要

(1) 人間は文章を読む際、しばしば頭の中に何らかの視覚的要素を思い描く。例えば、「オリンピック選手たちは大きく手を振りながら並んで入場してきた」という文と「小学生たちは大きく手を振りながら並んで入場してきた」という文は、主語が異なるだけだが、頭の中で想像する手の振り方は、それぞれ「左右」と「前後」というように、大きく異なるのではないだろうか。これは「オリンピック」と「運動会」では入場の様子が異なることを視覚的に思い浮かべた結果であると考えられる。視覚的要素を思い描くことにより、自然言語のみの処理では得られない内容理解に繋がるといえる。

(2) 人間は文章を読む際、自身の知識や経験と照らしてその内容が整合しているかどうかを自問しながら読み進める。例えば、「少年犯罪が急増している」という文の場合、法務省の犯罪白書の統計データと照らし合わせると、少年犯罪が急増しているとまでは言えず、この一文の内容はおかしいという違和感を覚える。このような違和感は、文章全体の説得力に影響を及ぼす要素であり、自然言語のみの処理では得られない内容理解に繋がると考えられる。

2. 研究の目的

(1) 本研究の目的は、画像や図表等のマルチメディアデータを活用したモデルを構築することにより、従来の自然言語処理だけでは得られない、より人間に近い的確な読解力を実現する基盤技術を確立することである。

(2) 本研究の学術的独自性は、文章を読みながら頭の中に視覚的要素を思い描いたり、文章を読む際に様々な知識や文脈と関係付けながらその内容理解を促進するといった、より人間らしい読解力を、画像や図表といったマルチメディアデータを活用したモデル構築により実現することに着眼している点である。本研究は、人とのよりの的確なインタラクションが求められるアプリケーションに広く応用することが可能であり、あらゆる場面で快適で豊かな生活を実現するスマート社会の根幹を担う技術の一つとして貢献することが期待できる。

3. 研究の方法

本研究では、画像や図表等のマルチメディアデータを活用したモデルを構築することにより、従来の自然言語処理だけでは得られない、より人間に近い的確な読解力について研究する。具体

的には、以下の3項目について研究開発を行う。

(1) テキストから視覚的要素を連想する能力

テキストと画像内容に一定の対応関係が認められるデータを用いて、テキストの分散表現、画像の分散表現を各々獲得する Encoder と、テキストから画像の分散表現を連想する Encoder を構築する。これにより、テキストから、テキストの分散表現だけでなく、視覚的要素の分散表現を得ることができる。これら2つの分散表現を用いて文生成を行う Decoder を構築する。視覚的要素の連想や生成された文を分析し、明らかにした課題をもとに、性能改善を図る。

(2) 図表と文脈を関係付ける能力

政府統計や業界団体等が公開している統計表、放送番組で用いられる図表画像を検出し、知識源とする。文章と知識源を関連づけるため、文章から適切なクエリを生成し、関連度や重要度等の観点でスコアを計算する。また、得られた図表画像の傾向を表現するテキストを生成するモデルを構築する。図表の種類に応じて用いられる様々な表現を推定し、必要に応じて類型化された変換処理を施すことで図表の傾向を表現するテキストを生成する。

(3) 把握した内容を説明する能力

与えられた文章とそれに関連する問いかけの入力に対し、(1)と(2)のモデルで把握した内容を用いて回答説明を自動生成する手法を研究開発する。質問及び文章から得られる分散表現と、(1)、(2)のモデルを利用し、回答となる説明を生成する Decoder を構築する。比較的長文の説明を生成するために、一貫性を保つためのモデルを構築する。

4. 研究成果

(1) テキストから視覚的要素を連想する能力-1

まず、テキストから視覚情報を連想し活用した場合の効果を検証するためのデータセット iNatVQA を、動植物の画像と学名のデータセットである iNaturalist から作成した。データ数は、訓練データが2.7M、開発用とテスト用データが各0.1Mである。

次に、質問文から画像検索して得られる画像の特徴表現を融合的に利用することでどの程度質問応答に効果があるかを調べた。その結果、テキストのみで学習するモデル (Seq2seq) よりも、画像の特徴表現を融合的に利用した手法 (提案手法) の方が、テストデータに対し高い正答率を出すことがわかった。

表1. 各手法の精度比較

Model	Accuracy (%)
seq2seq	58.78
VQA	57.54
提案手法	60.10

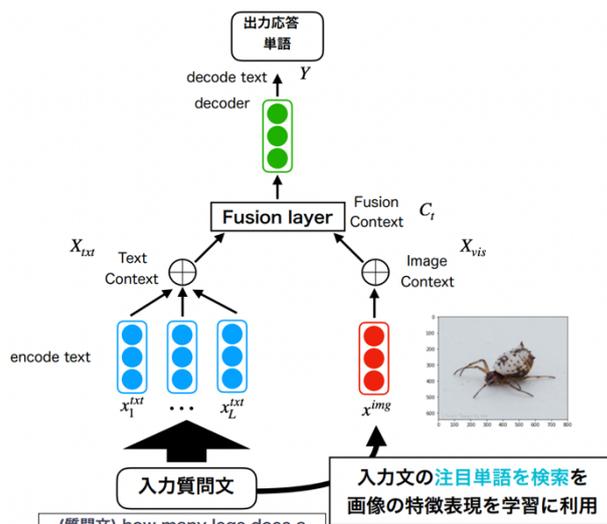


図2. 画像の特徴表現を融合的に利用したモデル

(2) テキストから視覚的要素を連想する能力-2

テキストから視覚情報を連想して質問応答する従来法で必要だった3段階の学習ステップ(テキスト特徴と視覚特徴の抽出、視覚特徴の連想、連想を用いた応答生成)を、1ステップで学習する手法(手法1)、及び、これにテキスト特徴と視覚特徴を各々どのような重要度で利用するかを適応的に変化させるゲート機構を追加した手法(手法2)について検討した。手法1では、色を尋ねる質問には適切に対処できる一方、上位カテゴリを尋ねる質問に誤りが目立ったが、手法2では、ゲート機構の導入により、色や生息地を除くほとんどのカテゴリでより誤答を防ぎ、多様な質問により適切に回答できるようになっていることを確認した。

(3) 図表と文脈を関係付ける能力-1

与えられた画像に関する説明文を生成する従来手法が、図表画像の内容をどの程度読み取り、テキストとして表現する能力があるのかについて調べた。その結果、生成結果が正しいと評価された割合は、文法的観点で98.8%、内容的観点で39.8%、関係性観点で48.2%であった。文法的観点については概ね正しい結果が生成できていたものの、内容的観点や関係性観点では、類似した色を混同したり、縦書きの軸ラベルを適切に説明文に反映しにくいといった傾向があることがわかった。

(4) 図表と文脈を関係付ける能力-2

ユーザが調べたい内容をテキストクエリとして入力し、政府などが公開している統計データ（一般に表形式で記載され、従来のテキスト文書とは異なる特徴を持つ）を対象としてランキングを行う統計データ検索タスクに対する検討を進めた。具体的には、被検索文書とクエリをカテゴリ分類し、候補となる被検索文書を絞り込む手法、統計データのメタデータにはない情報を統計表本体から抽出し、被検索文書を補強する手法、および、クエリに類似した拡張語を用いる手法で構成されるランキング手法について検討した。評価実験の結果、カテゴリ検索とデータ補強は、それぞれ、nDCG@10 の値を改善し、特に、データ補強のみの手法では nDCG@10 で 0.567 と最も良好な値を示すことを確認した。

表 2. ランキング結果の評価結果

Method	NTCIR-15 Data Search タスクにおける RUN-ID	nDCG@10
BM25		0.351
BM25+PRF+BERT [15]	NII-J-EX-10	0.406
BM25+QM [16]	uhai-J-10	0.392
RUN-1	KSU-J-EX-3	0.353
RUN-2		0.567
RUN-3	KSU-J-EX-1	0.426
RUN-4		0.534

(5) 把握した内容を説明する能力-1

原文の単語単位のベクトルに加え、文単位のベクトルを導入し、要約を生成する手法を提案した。また、文単位の注意機構の出力ベクトルを適応的に用いる Sentence Adaptive Attention を導入したモデルを提案した。本モデルにより、文長が長い原文が入力された場合でも、モデルが頑健に働くことが期待される。

実験により、提案手法によって、ROUGE における精度が改善したことを確認できた。特に、デコード時に、アテンションした文単位のベクトルを、適応的に用いるモデルの精度が安定して高いことがわかった。

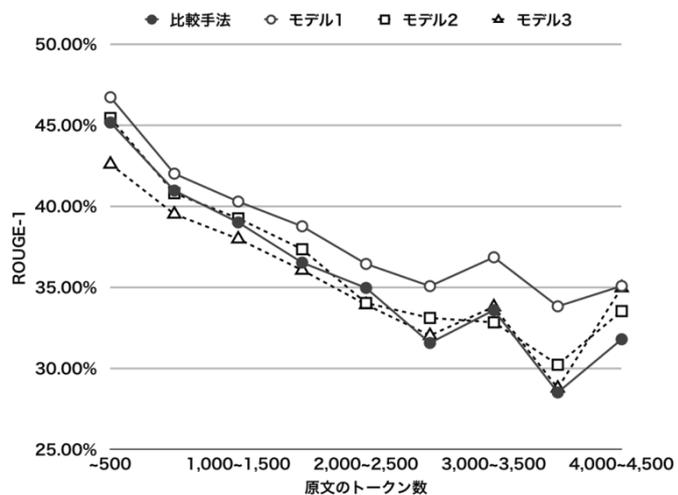


図 4. 原文のトークン数に対する ROUGE-1 による各モデルの比較

(6) 把握した内容を説明する能力-2

質問応答や対話システムにおいて、開発者の意図した人物らしさを、話者情報（話者用例とその話者のプロフィール情報）を与えることで比較的容易に反映させることが可能な応答生成手法について検討した。評価実験では、複数の被験者がそれぞれ思い描いた人物らしさが提案手法による出力応答へ反映されることを確認した。一方、特殊な語尾や口癖などについては必ずしも適切に反映できていない点があり、改善すべき課題であることがわかった。

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 4件/うち国際共著 0件/うちオープンアクセス 4件）

1. 著者名 岡本 卓, 宮森 恒	4. 巻 14
2. 論文標題 被検索文書の絞り込みと補強, クエリ拡張に基づく統計データ向けアドホック検索	5. 発行年 2021年
3. 雑誌名 情報処理学会論文誌データベース(TOD)	6. 最初と最後の頁 1--13
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -
1. 著者名 木村 輔, 田上 諒, 宮森 恒	4. 巻 12
2. 論文標題 文エンコーダによるクエリ指向要約モデルの強化	5. 発行年 2019年
3. 雑誌名 情報処学会論文誌データベース(TOD)	6. 最初と最後の頁 66-79
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -
1. 著者名 杉本 翔, 岡部 稜, 喜田 将生, 宮森 恒	4. 巻 60
2. 論文標題 幸福感を提供するVR画像閲覧システムGaZone	5. 発行年 2019年
3. 雑誌名 情報処理学会論文誌	6. 最初と最後の頁 1970-1977
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -
1. 著者名 奥井 優太, 宮森 恒	4. 巻 1
2. 論文標題 プログレッシブトランスフォーマーを用いた日本手話SLPモデルの性能評価	5. 発行年 2023年
3. 雑誌名 日本データベース学会データドリブンスタディーズ	6. 最初と最後の頁 1-8
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

〔学会発表〕 計21件（うち招待講演 0件 / うち国際学会 5件）

1. 発表者名 Okamoto Taku, Miyamori Hisashi
2. 発表標題 Ad Hoc Search On Statistical Data Based On Categorization And Metadata Augmentation
3. 学会等名 IEEE 4th International Conference on Multimedia Information Processing and Retrieval (IEEE MIPR 2021) (国際学会)
4. 発表年 2021年

1. 発表者名 岡本 卓, 宮森 恒
2. 発表標題 統計データ検索のためのFT-Transformerによるランキング学習
3. 学会等名 第14回データ工学と情報マネジメントに関するフォーラム DEIM2022
4. 発表年 2022年

1. 発表者名 奥井優太, 宮森恒
2. 発表標題 プログレッシブトランスフォーマーを用いた日本手話SLPモデルの性能評価
3. 学会等名 第14回データ工学と情報マネジメントに関するフォーラム DEIM2022
4. 発表年 2022年

1. 発表者名 岡本 卓, 宮森 恒
2. 発表標題 KSU Systems at the NTCIR-15 Data Search Task
3. 学会等名 The 15th NTCIR Conference (国際学会)
4. 発表年 2020年

1. 発表者名 岡本 卓、宮森 恒
2. 発表標題 カテゴリ分類とメタデータ補強に基づく統計データに対するアドホック検索
3. 学会等名 第13回データ工学と情報マネジメントに関するフォーラム DEIM2021
4. 発表年 2021年

1. 発表者名 木村 輔、田上 諒、勝山光、杉本 翔、宮森 恒
2. 発表標題 KSU Systems at the NTCIR-14 QA Lab--PolInfo Task
3. 学会等名 The 14th NTCIR Conference (国際学会)
4. 発表年 2019年

1. 発表者名 杉本 翔、宮森 恒
2. 発表標題 GaZone: VR Image Browsing System Providing Feelings of Happiness
3. 学会等名 Entertainment Computing and Serious Games. ICEC-JCSG 2019. LNCS11863 (国際学会)
4. 発表年 2019年

1. 発表者名 杉本 翔、岡部 稜、喜田 将生、宮森 恒
2. 発表標題 幸福感を提供するVR画像検索システムGaZone
3. 学会等名 エンタテインメントコンピューティングシンポジウム2018論文集
4. 発表年 2018年

1. 発表者名 木村 輔、田上 諒、宮森 恒
2. 発表標題 Query-focused Summarization Enhanced with Sentence Attention Mechanism
3. 学会等名 6th IEEE International Conference on Big Data and Smart Computing (BigComp 2019) (国際学会)
4. 発表年 2019年

1. 発表者名 木村 輔、田上 諒、宮森 恒
2. 発表標題 文エンコードによるクエリ指向要約モデルの強化
3. 学会等名 第11回データ工学と情報マネジメントに関するフォーラム DEIM2019
4. 発表年 2019年

1. 発表者名 田上 諒、木村 輔、宮森 恒
2. 発表標題 前提知識を考慮した根拠の妥当性判断による自動回答手法
3. 学会等名 第11回データ工学と情報マネジメントに関するフォーラム DEIM2019
4. 発表年 2019年

1. 発表者名 杉本 翔、宮森 恒
2. 発表標題 話者用例に基づく人物らしさを反映した対話応答文生成
3. 学会等名 第11回データ工学と情報マネジメントに関するフォーラム DEIM2019
4. 発表年 2019年

1. 発表者名 石橋陽一、森泰、木村輔、宮森恒
2. 発表標題 質問文から連想した画像特徴量を用いた質問応答モデル
3. 学会等名 言語処理学会第25回年次大会(NLP2019)
4. 発表年 2019年

1. 発表者名 森泰、石橋陽一、宮森恒
2. 発表標題 テキストと関連画像の視覚的要素を活用した質問応答
3. 学会等名 言語処理学会第25回年次大会(NLP2019)
4. 発表年 2019年

1. 発表者名 田上諒、木村輔、杉本翔、勝山光、宮森恒
2. 発表標題 地方議会における事実確認のための会議録分割
3. 学会等名 言語処理学会第25回年次大会(NLP2019)
4. 発表年 2019年

1. 発表者名 勝山光、杉本翔、田上諒、木村輔、宮森恒
2. 発表標題 地方議会特有の表現に基づく政策に対する発言の賛否分類
3. 学会等名 言語処理学会第25回年次大会(NLP2019)
4. 発表年 2019年

1. 発表者名 小川 志龍, 林 寛治, 宮森 恒
2. 発表標題 Transformerモデルに関する順序数の的確な把握と活用能力の調査
3. 学会等名 第15回データ工学と情報マネジメントに関するフォーラム DEIM2023
4. 発表年 2023年

1. 発表者名 王 墻, 宮森 恒
2. 発表標題 外部知識参照型拡散モデルによるテキストからのビデオ生成
3. 学会等名 第15回データ工学と情報マネジメントに関するフォーラム DEIM2023
4. 発表年 2023年

1. 発表者名 戸崎 友輔, 宮森 恒
2. 発表標題 語彙習得段階を考慮した英文法多肢選択問題の誤答選択肢自動生成
3. 学会等名 第15回データ工学と情報マネジメントに関するフォーラム DEIM2023
4. 発表年 2023年

1. 発表者名 増田 琉斗, 宮森 恒
2. 発表標題 画像・言語モデルに関する順序数の的確な把握と活用能力の調査
3. 学会等名 第15回データ工学と情報マネジメントに関するフォーラム DEIM2023
4. 発表年 2023年

1. 発表者名 今西 智也, 宮森 恒
2. 発表標題 スマートフォンを用いた軽量な道路損傷の自動検出
3. 学会等名 第15回データ工学と情報マネジメントに関するフォーラム DEIM2023
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

京都産業大学 情報理工学部 宮森研究室 (メディア知能研究室)
<http://milab.kyoto-su.ac.jp/>

6. 研究組織		
氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関