

令和 6 年 7 月 2 日現在

機関番号：92668

研究種目：基盤研究(A)（一般）

研究期間：2021～2023

課題番号：21H04910

研究課題名（和文）自然言語指示に応じて多様な作業を行うロボット実現のための動作生成技術の開発

研究課題名（英文）Development of Motion Generation Technology to Realize Robots that Perform Various Tasks according to Natural Language Instructions

研究代表者

橋本 敦史（Hashimoto, Atsushi）

オムロンサイニックエクス株式会社・リサーチアドミニストレイティブディビジョン・シニアリサーチャー

研究者番号：80641753

交付決定額（研究期間全体）：（直接経費） 32,300,000円

研究成果の概要（和文）：本プロジェクトでは、言語指示に基づいてロボットに作業をさせるための汎用計算モデルを開発しました。Normanの行為の7段階モデルに基づいて設計し、サラダを対象に言語制御によるロボットの動作フローをほぼ完成させました。計画問題記述言語（Planning Domain Definition Language: PDDL）を使用することで、説明性や信頼性を向上させ、学習データなしでの制御を実現しました。これにより、従来のブラックボックス型システムよりも実用的なシステムを構築できました。

研究成果の学術的意義や社会的意義

本プロジェクトでは言語指示に基づきロボットを操作する技術を開発した。同時期に複数の研究機関も類似の研究を行ったが、これらの研究では「意図の推定」という段階が見落とされていた。そのため、これらのシステムはロボットの身体性に依存し、個々のロボットに特化したモデルを開発する必要があった。さらに、動作の安全性の確保や言語指示が正しく解釈されたかの判定が難しくなる問題も存在した。本プロジェクトでは、言語指示とシーン観測から物体の初期状態と目標状態を出力する方法を確立した。これにより、特定のロボットに依存せず、結果を解釈可能で、かつ、安全性を担保すること可能となった。

研究成果の概要（英文）：In this project, we developed a general-purpose computational model to make robots perform tasks based on verbal instructions. The design was based on Norman's seven stages of action model. We nearly completed the action flow for a robot controlled by verbal commands, using a salad as a test subject. By employing the Planning Domain Definition Language (PDDL), we enhanced the system's explainability and reliability, achieving control without the need for training data. This allowed us to create a more practical system compared to conventional black-box type systems.

研究分野：コンピュータビジョン

キーワード：自然言語処理 クロスモーダル処理 ロボティクス

1. 研究開始当初の背景

国立社会保障・人口問題研究所による「日本の将来推計人口 平成 29 年推計」によれば、生産年齢人口は今後 30 年で現在の約 71%程度に減少し、その後も減少傾向が続く。このため、生産性の向上に資する技術開発が喫緊の課題となっている。なお、この状況は研究期間が終了した現状でも変わっていない。このような社会情勢の中で、ロボットを活用して不足する労働力を補填する研究が活発化の兆しを見せていた。また、当時の技術動向として深層学習の開発に端を発した視覚言語融合技術が大きな発展の兆しを見せていた。

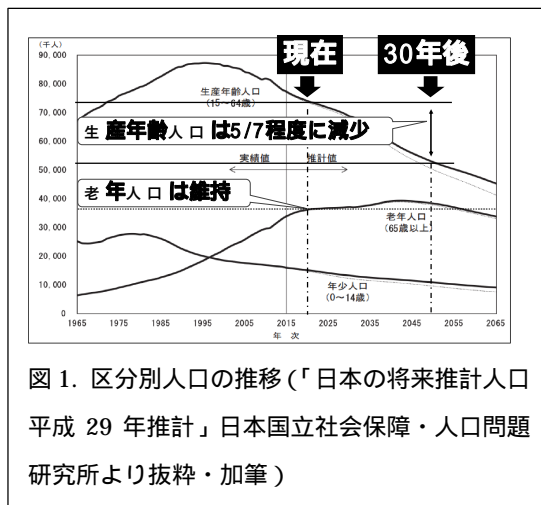


図 1. 区分別人口の推移(「日本の将来推計人口 平成 29 年推計」日本国立社会保障・人口問題研究所より抜粋・加筆)

当時、言語により動作するロボットの実例は Preferred Networks 社によるものだけであり、これは Object Picking タスクに特化して対話を通して物体の移動元と移動先を特定するというものであった。

2. 研究の目的

本プロジェクトの目的は、Object Picking 以外の複雑なタスクであっても言語指示で動作させられるような、言語指示によるロボット制御を行うための汎用の計算モデルの開発であった。従来システムにおいて、言語により特定すべき対象は「物体の移動元」と「移動先」に対応する、画像上の 2 つの矩形領域であった。これに対し、本研究ではこれらを物体の「初期状態」と「目標状態」へと一般化し、任意のタスクを表現できるようにした上で、それらを入力として受けとり、実際に動作を実行するロボットシステムの実現を目指した。また、コンセプト実証のための具体的な目標として、ごく簡単なサラダを言語指示により調理できるロボットの実現を目指した。サラダの調理では包丁を使った切断など、Object Picking に加えてより複雑なスキルが要求される。このような動作を言語指示により適切に呼び出すための計算モデルを構築することで、計算モデルの汎用性を示すこととした。

3. 研究の方法

人間は言語指示と状況の観測に基づいて多様な動作を実行できるという意味で、本研究が目標とする汎用性を持つロボットの一種とみなすことができる。このような視点に立ち、人間がどのような過程を経て行為の実行に移るのかを考えた。この手がかりとして、ヒューマンコンピュー

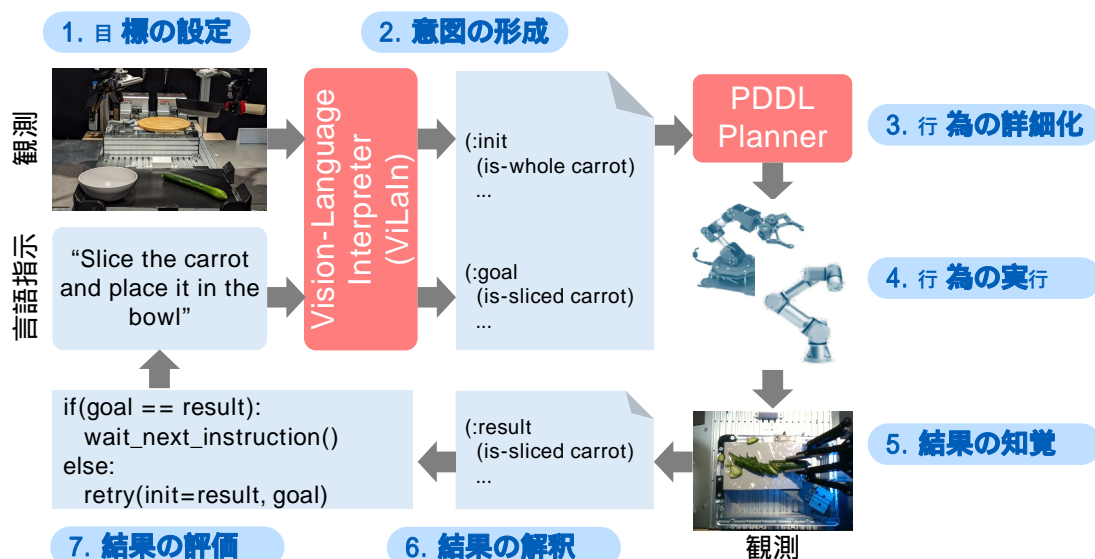


図 2. 行為の 7 段階モデルと言語指示ロボット制御の対応関係

タイインタラクション分野で古くから受け入れられて来た D.A.Norman の行為の 7 段階モデルを参考とした(図 2)。ロボットが行為を実行し、次の行為に移るまでの対応関係に基づき、ブラッ

クボックス部分を最小化して汎用性を担保する計算モデルを確立した。また、実際に既存の各言語モデル(Vision-Language Model: VLM)や大規模言語モデル(Large Language Model: LLM)を駆逐することにより、言語により人間が実行した「1.目標の設定」の段階から、次の「2.意図の形成」に至る部分を計算する最小限のブラックボックスモデルと、PDDL など信頼性や説明性が高い既存技術を繋げることで目的のシステムを構築した。

4. 研究成果

図 2 に示したフレームワークを実現する代表的な研究成果として、図 3 に示した ViLaIn [1](ICRA2024 に採択され、発表済み)と、その ViLaIn の出力と同じ定義の記述を受け取って野菜を切るロボット[2] (ICRA2024 の Workshop 「Cooking Robotics: Perception and motion planning」に採択され、Best Video Award を受賞)を実現した(図 4)。理論的にこれらは連結して動作するはずである。しかし、残念ながら研究計画段階の予算と実際に割り振られた予算との差分が想定よりも大きかったため、実装力の量的な確保ができず、研究期間内に[1]と[2]のシステムを実際に連結させて言語で動作させるデモを行うには至らなかった。しかし、野菜を切るロボットの中で使われている技術[3]も ICRA2024 に採択され、また、上述の Workshop において Best Poster Award を受賞するなど、当該領域においてトップレベルの成果を実現できた。この技術については [IEEE Spectrum](#) [にも取り上げられる](#)など、広い範囲のロボティクス研究者の注目を集めることができた。

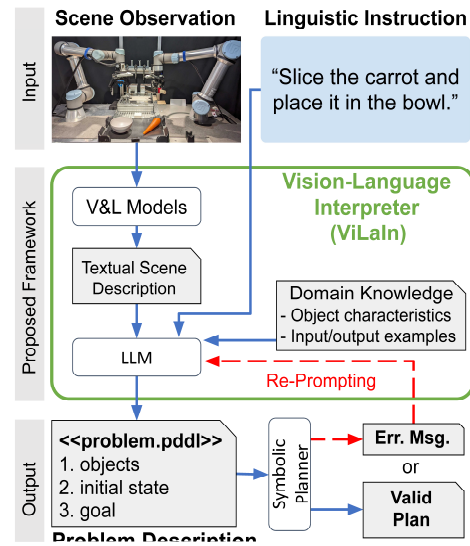


図 3. ViLaIn の概要



図 4. PDDL に従って食材の切断に関わる一連の作業を行う実際のロボットシステム(画像は [YouTube に投稿した動画](#)より引用)

上記[1-3]の個々の技術詳細はそれぞれの論文に譲るが、当該課題に関連した招待講演[4,5,9,10]や取材[7,8]、パネルディスカッション[6]への招待などロボティクスの専門家に留まらず、広い範囲の研究者・一般市民に広報できた点も大きな成果であると考えている。

今後は、図 1 のフレームワークに則り、さらに多様なロボットシステムでの汎用性の実証、および、状態認識能力がより高い VLM の実現により、幅広い分野で採用され得るシステムであることを証明していく。特にこのフレームワークを API として整備し、第三者がこのシステムに則って自前のシステムを動かそうとする場合の実装コストを極限まで低減させるなどして、他の研究グループを巻き込んだ実証を行っていきたい。

- [1] Keisuke Shirai¹, Cristian C. Beltran-Hernandez, Masashi Hamaya, Atsushi Hashimoto, Shohei Tanaka, Kento Kawaharazuka, Kazutoshi Tanaka, Yoshitaka Ushiku, Shinsuke Mori, “Vision-Language Interpreter for Robot Task Planning,” ICRA2024
ICRA 2024
- [2] Jeremy Siburian, Cristian Camilo Beltran-Hernandez, Masashi Hamaya, “Integrated Task and Motion Planning for Real-World Cooking Tasks,” ICRA2024 Workshop (Best Video Award)
- [3] Cristian Camilo Beltran-Hernandez, Nicolas Erbeti, Masashi Hamaya, “Slicelt! - A Dual Simulator Framework for Learning Robot Food Slicing,” ICRA2024 (Best Poster Award at a ICRA2024 Workshop)
- [4] 橋本敦史, “Vision & Language 技術とロボティクスの融合の先にある未来,” 計測自動制御学会制御部門 人の理解/誘導で強化される制御システム調査研究会 キックオフ, 2024 (招待講演)
- [5] 橋本敦史, “人と機械の融和における生成 AI の社会実装,” 日本鉄鋼協会 計測・制御・システム工学部会 シンポジウム, 2024
- [6] 橋本敦史 他 6 名, “社会実装と次世代技術基盤構築の両輪で推進する、これからのロボット研究開発 ~ロボットアクションプランの実現に向けて~, ” 国際ロボット展 NEDO ロボット・AI フォーラム, 2023, パネルディスカッション
- [7] Nikkei Tech Foresight, “ロボット制御にも ChatGPT 生成 AI で激変する製造業,” URL: <https://www.nikkei.com/prime/tech-foresight/article/DGXZQ0UC28A310Y3A220C2000000>, Nikkei Tech Foresight
- [8] TOKYO FM, “キャノンマーケティングジャパン presents Solution in my life ユージ - TOKYO FM 80.0MHz-, ” 2022 (ラジオ放送)
- [9] 橋本敦史, “自然言語に応じて多様な作業を行うロボット実現に向けたクロスモーダル機械学習の取り組み”, 日本ロボット学会 データ工学ロボティクス研究専門委員会主催 公開講演会, 2021
- [10] 橋本敦史, “クロスモーダル処理技術が統合する視覚・言語・ロボット制御技術の未来,” 日本機械学会年次大会, 2021

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 0件／うち国際共著 0件／うちオープンアクセス 0件）

1. 著者名 Ishikawa Reina, Hamaya Masashi, Von Drigalski Felix, Tanaka Kazutoshi, Hashimoto Atsushi	4. 巻 10
2. 論文標題 Learning by Breaking: Food Fracture Anticipation for Robotic Food Manipulation	5. 発行年 2022年
3. 雑誌名 IEEE Access	6. 最初と最後の頁 99321 ~ 99329
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ACCESS.2022.3207491	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Nishimura Taichi, Sakoda Kojiro, Ushiku Atsushi, Hashimoto Atsushi, Okuda Natsuko, Ono Fumihito, Kameko Hirotaka, Mori Shinsuke	4. 巻 29
2. 論文標題 BioVL2: An Egocentric Biochemical Video-and-Language Dataset	5. 発行年 2022年
3. 雑誌名 Journal of Natural Language Processing	6. 最初と最後の頁 1106 ~ 1137
掲載論文のDOI (デジタルオブジェクト識別子) 10.5715/jnlp.29.1106	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計8件（うち招待講演 2件／うち国際学会 4件）

1. 発表者名 宮本蒼太, 八木拓真, 牛久祥孝, 橋本敦史, 井上中順
2. 発表標題 手の軌道特徴を用いた一人称視点料理動画における詳細動作認識
3. 学会等名 電子情報通信学会PRMU研究会
4. 発表年 2023年

1. 発表者名 Keisuke Shirai, Atsushi Hashimoto, Taichi Nishimura, Hirotaka Kameko, Shuhei Kurita, Yoshitaka Ushiku, Shinsuke Mori
2. 発表標題 Visual Recipe Flow: A Dataset for Learning Visual State Changes of Objects with Recipe Flows
3. 学会等名 COLING2022 (国際学会)
4. 発表年 2022年

1. 発表者名 Atsushi Hashimoto, Taichi Nishimura, Yoshitaka Ushiku, Hirotaka Kameko, Shinsuke Mori
2. 発表標題 Cross-modal Representation Learning for Understanding Manufacturing Procedure
3. 学会等名 HCI12022 (国際学会)
4. 発表年 2022年

1. 発表者名 Nishimura Taichi, Sakoda Kojiro, Hashimoto Atsushi, Ushiku Yoshitaka, Tanaka Natsuko, Ono Fumihito, Kameko Hirotaka, Mori Shinsuke
2. 発表標題 Egocentric Biochemical Video-and-Language Dataset
3. 学会等名 The 4th Workshop on Closing the Loop Between Vision and Language in conjunction with ICCV2021 (国際学会)
4. 発表年 2021年

1. 発表者名 Nishimura Taichi, Hashimoto Atsushi, Ushiku Yoshitaka, Kameko Hirotaka, Mori Shinsuke
2. 発表標題 State-aware Video Procedural Captioning
3. 学会等名 The 29th ACM International Conference on Multimedia (国際学会)
4. 発表年 2021年

1. 発表者名 橋本敦史
2. 発表標題 自然言語に応じて多様な作業を行うロボット実現に向けたクロスモーダル機械学習の取り組み
3. 学会等名 日本ロボット学会データ工学ロボティクス研究専門委員会主催公開講演会 (招待講演)
4. 発表年 2021年

1. 発表者名 橋本敦史
2. 発表標題 クロスモーダル処理技術が統合する視覚・言語・ロボット制御技術の未来
3. 学会等名 日本機械学会年次大会（招待講演）
4. 発表年 2021年

1. 発表者名 白井圭佑, 橋本敦史, 牛久祥孝, 栗田修平, 亀甲博貴, 森信介
2. 発表標題 レシビ分野における動作対象の状態変化を考慮したデータセットの構築と検索モデルの提案
3. 学会等名 言語処理学会第28回年次大会
4. 発表年 2022年

〔図書〕 計0件

〔出願〕 計1件

産業財産権の名称 制御装置、制御方法、及び制御プログラム	発明者 濱屋政志、石川玲 奈、橋本敦史、田中 一敏	権利者 同左
産業財産権の種類、番号 特許、2022-032176	出願年 2022年	国内・外国の別 国内

〔取得〕 計0件

〔その他〕

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
研究分担者	井上 中順 (Inoue Nakamasa) (10733397)	東京工業大学・情報理工学院・准教授 (12608)	
研究分担者	牛久 祥孝 (Ushiku Yoshitaka) (10784142)	オムロンサイニックエックス株式会社・リサーチアドミニスト レイティブディビジョン・プリンシパルインベスティゲ ーター (92668)	

6. 研究組織（つづき）

	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
研究分担者	濱屋 政志 (Hamaya Masashi) (10869176)	オムロンサイニックエックス株式会社・リサーチアドミニストレイティブディビジョン・シニアリサーチャー (92668)	
研究分担者	松原 崇充 (Matsubara Takamitsu) (20508056)	奈良先端科学技術大学院大学・先端科学技術研究科・教授 (14603)	
研究分担者	森 信介 (Mori Shinsuke) (90456773)	京都大学・学術情報メディアセンター・教授 (14301)	
研究分担者	ベルトランエルナンデス クリスティアンカミロ (Beltran-Hernandez Cristian Camilo) (30984017)	オムロンサイニックエックス株式会社・リサーチアドミニストレイティブディビジョン・リサーチエンジニア (92668)	
研究分担者	VON・DRIGALSKI FELIX (Von Drigalski Felix) (90869215)	オムロンサイニックエックス株式会社・リサーチアドミニストレイティブディビジョン・シニアリサーチャー (92668)	削除：2021年11月30日

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