

Title	Switching Head-Tail Funnel UNITER による対象物体および配置目標に関する指示文理解と物体操作
Sub Title	
Author	是方, 諒介(Korekata, Ryōsuke)
Publisher	慶應義塾大学AI・高度プログラミングコンソーシアム
Publication year	2023
Jtitle	AICカンファレンス予稿集 (2023.) ,p.30- 30
JaLC DOI	
Abstract	本研究では, 生活支援ロボットが自然言語による指示文に基づいて日用品を家具へ運搬する手法の構築を目的とする. 本論文では, 対象物体および配置目標に関する予測を単一モデルで個別に行う方法でタスクを解くことが可能なSwitchingHead-Tail Funnel UNITERを提案する. Switching Head-Tail機構の導入により, 対象物体候補および配置目標候補がそれぞれMおよびN個存在する状況において, 推論回数を $O(M \times N)$ ではなく $O(M + N)$ とすることが可能になる. 結果として, 標準的なシミュレータにおいて収集された新規データセットにおいて, 提案手法がベースライン手法を言語理解精度で上回った. さらに, 標準化された家庭内環境で実機実験を行い, 物体把持および配置動作についてそれぞれ90%以上の成功率を達成した.
Notes	会議名: AICカンファレンス2023 開催地: 慶應義塾大学日吉キャンパス 日時: 2023年3月4日 第2章ポスター発表要旨 ポスター要旨-5
Genre	Conference Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO11003001-20230304-0030

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

Switching Head-Tail Funnel UNITER による 対象物体および配置目標に関する指示文理解と物体操作

是方諒介

慶應義塾大学理工学部情報工学科

Abstract: 本研究では、生活支援ロボットが自然言語による指示文に基づいて日用品を家具へ運搬する手法の構築を目的とする。本論文では、対象物体および配置目標に関する予測を単一モデルで個別に行う方法でタスクを解くことが可能な Switching Head-Tail Funnel UNITER を提案する。Switching Head-Tail 機構の導入により、対象物体候補および配置目標候補がそれぞれ M および N 個存在する状況において、推論回数を $O(M \times N)$ ではなく $O(M + N)$ とすることが可能になる。結果として、標準的なシミュレータにおいて収集された新規データセットにおいて、提案手法がベースライン手法を言語理解精度で上回った。さらに、標準化された家庭内環境で実験を行い、物体把持および配置動作についてそれぞれ 90% 以上の成功率を達成した。

Keywords: Multimodal Language Understanding, fetch-and-carry, Object Manipulation, Domestic Service Robot

1. 研究背景・目的

高齢化が進行する現代社会において、在宅介護者不足に対する一つの解決策として被介護者を物理的に支援可能な生活支援ロボットが注目されている [4]。しかし、人間からの自然言語による指示をロボットが理解する能力についてはいまだ不十分である。本研究では、複数の日用品および家具から、参照表現を含む指示文の対象物体および配置目標を特定し、運搬する Dual Referring Expression Comprehension with fetch-and-carry (DREC-fc) タスクを扱う。

2. 方法

本論文では、対象物体および配置目標に関する予測を単一モデルで個別に行う方法でタスクを解くことが可能な Switching Head-Tail Funnel UNITER (SHeFU) を提案する。これにより、対象物体候補および配置目標候補がそれぞれ M および N 個存在する状況において、言語理解に必要な推論回数を $O(M \times N)$ から $O(M + N)$ に削減可能となる。

既存手法 (e.g., [1]) と異なる点は、Switching Head-Tail 機構を導入することで、単一モデルで対象物体候補および配置目標候補のどちらも入力として扱い、効率的に推論することが可能な点である。Switching Head 機構では、対象物体および配置目標を予測するためのパラメータを暗黙的に共有するモデルの条件付けを行う。一方で、Switching Tail 機構ではマルチタスク学習を可能にする。これらの機構により、対象物体を予測する際に配置目標に関する視覚および言語の情報を活用することが期待され、その逆もまた然りである。また、単一モデルでの学習が可能になるため、別々のモデルを用意する必要がなくなる。

3. 結果

シミュレーション環境において新たに収集した ALFRED-fc データセットおよび実機による評価を行った。前者は ALFRED [3] データセットを基に作成し、後者は World Robot Summit [2] で標準化された家庭内環境を用いた。

図 1 に、実機における定性的結果として成功例を示す。赤色、橙色、および青色の矩形領域はそれぞれ対象物体の ground truth、配置目標の ground truth、および対象物体候補または配置目標候補を表す。対象物体および配置目標はそれぞれポテトチップス缶およびサッカーボールの乗ったテーブルであり、提案手法は正しい予測を基に運搬に成功した。



図 1 実機における定性的結果。(a) 対象物体候補および配置目標候補。(b) 把持および配置動作。指示文：“Put the red chips can on the white table with the soccer ball on it.”

4. 結論

本研究の貢献を以下に示す。

- 言語理解における対象物体および配置目標の探索に必要な推論回数を削減可能な SHeFU を提案した。
- Switching Head-Tail 機構の導入により、対象物体および配置目標を単一モデルで個別に予測可能にした。
- ALFRED [3] を基にした DREC-fc タスクのための ALFRED-fc データセットにおいて、SHeFU がベースライン手法 [1] を言語理解精度で上回った。
- 実機を用いて実環境で実験を行い、SHeFU がベースライン手法を言語理解精度で上回った。さらに、把持および配置動作を高い成功率で実行した。

参考文献

- [1] Shintaro Ishikawa and Komei Sugiura. Target-dependent UNITER: A Transformer-Based Multimodal Language Comprehension Model for Domestic Service Robots. *IEEE RA-L*, 6(4):8401–8408, 2021.
- [2] Hiroyuki Okada et al. What competitions were conducted in the service categories of the World Robot Summit? *Advanced Robotics*, 33(17):900–910, 2019.
- [3] Mohit Shridhar et al. ALFRED: A Benchmark for Interpreting Grounded Instructions for Everyday Tasks. In *CVPR*, pages 10740–10749, 2020.
- [4] Takashi Yamamoto, Koji Terada, Akiyoshi Ochiai, et al. Development of Human Support Robot as the research platform of a domestic mobile manipulator. *ROBOMECH Journal*, 6(1):1–15, 2019.