

Title	市街地での移動指示理解タスクにおける UNITER Regressor による目標位置予測
Sub Title	
Author	畑中, 駿平(Hatanaka, Shunpei)
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.32- 32
JaLC DOI	
Abstract	Recent advancement of vehicle automation technology is expected to improve the interaction between human and mobility modes. As the promising means, language-based/vision-and-language navigation and referring expression comprehension (REC) approaches have been studied. While most existing approaches identifies reference target, almost no methods can predict target position (destination) for the mobility. In this work, we extend the UNITER architecture to a multimodal regressor and train the model on the extended Talk2Car-STOP dataset with target position annotations. The experimental results show the effectiveness of the proposed method.
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-7
Genre	Conference Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO11003001-20230304-0032

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.

市街地での移動指示理解タスクにおける UNITER Regressor による目標位置予測

畑中駿平

慶應義塾大学大学院理工学研究科開放環境科学専攻

Abstract: Recent advancement of vehicle automation technology is expected to improve the interaction between human and mobility modes. As the promising means, language-based/vision-and-language navigation and referring expression comprehension (REC) approaches have been studied. While most existing approaches identifies reference target, almost no methods can predict target position (destination) for the mobility. In this work, we extend the UNITER architecture to a multimodal regressor and train the model on the extended Talk2Car-STOP dataset with target position annotations. The experimental results show the effectiveness of the proposed method.

Keywords: Understanding Navigation Instructions, Target Position, UNITER Regressor

1. 研究背景・目的

車などの自動化技術の進展とともに、人とモビリティのインタラクション方法が変化することが期待されている [3]。モビリティに対して、移動指示を自然言語で与えることができれば便利であるが、現状では自然な発話に対する言語理解性能は十分でなく人が操作する必要がある。将来的には人の役割はモビリティに指示を与え、自動化技術が命令を遂行するような関係になる可能性も指摘されている。このような状況を対象に、本研究では移動指示をモビリティが理解することを目指す。

2. 方法

本論文では、モビリティに対する移動指示、画像全体、ランドマーク領域を入力として目標位置の絶対座標を出力するタスクを扱う。具体的には UNITER [1] を拡張し、回帰問題として目標位置（2次元座標）を予測する手法 UNITER Regressor を提案する。

本研究の独自性は以下である。

- 移動指示文およびランドマーク（物体の領域）と画像全体入力とした目標位置回帰モデルを構築する。

3. 実験設定

本論文では、Talk2Car データセット [2] を拡張した Talk2Car-STOP データセットを新たに作成した。本研究では目標位置を予測するため、Talk2Car データセットから移動指示およびランドマークに対応した目標位置をアノテーションして新たに Talk2Car-STOP データセットを作成した。

4. 結果

図1に定性的結果を示す。図1に、予測成功の例を示す。図1 与えられた移動指示文 “I think this is where we are supposed to meet everyone, pull up **behind this bus** and park.” であり、ランドマークは右手のバスである。この例において提案手法はバスの後ろを目標位置として予測することに成功している。



移動指示文：“I think this is where we are supposed to meet everyone, pull up behind this bus and park.”

図1 本手法による成功例。緑色の矩形領域、赤色の点、青色の点はそれぞれランドマーク、正解の目標位置、予測した目標位置を表す。

5. 結論

本研究では、モビリティに対する視覚言語情報を入力とする目標位置の予測タスクを扱った。提案手法による貢献は以下である。

- 移動指示理解タスクにおいて、UNITER を拡張し、移動指示文およびランドマーク領域を入力とする目標位置を推定する回帰モデル UNITER Regressor を導入した。
- Talk2Car-STOP データセット上の有効性を示した。

将来研究としては、モビリティが使用される状況を想定した新たなシミュレーションデータセットの作成およびモデル構造の改良による性能改善が挙げられる。

参考文献

- [1] Yen Chen, Linjie Li, Licheng Yu, Ahmed El Kholy, Faisal Ahmed, Zhe Gan, et al. Uniter: Universal image-text representation learning. In *ECCV*, pages 104–120, 2020.
- [2] Thierry Deruyttere, Simon Vandenhende, Dusan Grujicic, et al. Talk2car: Taking control of your self-driving car. In *EMNLP-IJCNLP*, pages 2088–2098, 2019.
- [3] Arun Vasudevan et al. Talk2nav: Long-range vision-and-language navigation with dual attention and spatial memory. *IJCV*, 129(1):246–266, 2021.