

Title	大規模言語モデルのリーズニング能力の評価：言語理解と推論の観点から
Sub Title	
Author	安東, 里沙子
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	2024
Jtitle	AICカンファレンスWinter Edition予稿集 (2024.)
JaLC DOI	
Abstract	近年、大規模言語モデル（LLM）は、人間に匹敵する自然な言語能力を獲得していると考えられている。しかし、それらがどの程度正確な論理推論能力を備えているかは、依然として明らかではない。また、推論のどの過程でエラーが生じやすいのか、さらには、そのようなエラーをどのように低減できるのかといった問いも重要である。 本発表では、言語理解と推論の観点によるLLMの評価に関する発表者らの研究成果を報告する。特に、自然言語から形式言語への翻訳を含む推論タスクや、図的表現を補助として用いる推論タスクに基づく実験結果を、簡単に紹介する。
Notes	会議名：AICカンファレンス2024 Winter Edition 開催地：慶應義塾大学協生館AICラウンジ 日時：2024年12月7日 講演セッション要旨
Genre	Conference Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO11003001-20241207-0009

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.

エンドエフェクタの軌道に基づく open-vocabulary 物体操作タスクにおける将来成否予測

神原 元就
慶應義塾大学大学院理工学研究科

Abstract:

open-vocabulary 物体操作タスクにおいて、実行前に失敗を予測し、適切な軌道を再生成することができれば、安全かつ効率的なタスクの実行が可能である。しかし、生成された軌道に基づいてタスクの成功や失敗を事前に予測する手法はほとんど存在しない。そこで本研究では、操作前の一人称視点画像、物体操作指示文、及び軌道に基づいて、物体操作タスクの将来的な成功または失敗を予測する方法を提案する。また、エンドエフェクタの軌道から効果的な特徴を抽出し、画像特徴とアラインメントを行うための Trajectory Encoder を導入する。

大規模言語モデルのリーズニング能力の評価: 言語理解と推論の観点から

安東 里沙子
慶應義塾大学大学院文学研究科

Abstract:

近年、大規模言語モデル (LLM) は、人間に匹敵する自然な言語能力を獲得していると考えられている。しかし、それらがどの程度正確な論理推論能力を備えているかは、依然として明らかではない。また、推論のどの過程でエラーが生じやすいのか、さらには、そのようなエラーをどのように低減できるのかといった問いも重要である。本発表では、言語理解と推論の観点による LLM の評価に関する発表者らの研究成果を報告する。特に、自然言語から形式言語への翻訳を含む推論タスクや、図的表現を補助として用いる推論タスクに基づく実験結果を、簡単に紹介する。

Image Captioning における自動評価の最前線

和田 唯我
慶應義塾大学大学院理工学研究科

Abstract:

Image Captioning は、マルチモーダル LLM に代表されるように、幅広い研究が行われている。研究に際して、円滑なモデル開発のためには、人手評価を代替しうる高性能な自動評価尺度が重要である。そこで、本発表では、Image Captioning における様々な自動評価を紹介する。具体的には、JaSPICE [Wada+, CoNLL23] のような古典的な評価尺度から Polos [Wada+, CVPR24] や DENEb [Matsuda+, ACCV24] といったデータ駆動型評価尺度まで、広範な自動評価尺度を紹介する。

生活支援ロボットにおけるマルチモーダル検索に基づく移動マニピュレーション

是方 諒介
慶應義塾大学大学院理工学研究科

Abstract:

本研究では、生活支援ロボットが自然言語指示文に基づき日常物体を指定された家具へ運搬する open-vocabulary mobile manipulation タスクを扱う。提案手法では、指示文および画像を基盤モデルを用いて複数粒度で分解および統合し、環境中の画像群から対象物体および配置目標を検索可能にする。大規模な屋内環境データセットにおいて実験を行い、標準的な評価指標において既存手法を上回る結果を得た。さらに、標準化された屋内環境において実機実験を行い、ゼロショット転移条件のもと 80% 以上のタスク成功率を達成した。