

2023 年度年次報告書
AI 活用で挑む学問の革新と創成
2022 年度採択研究代表者

八木 拓真

産業技術総合研究所 人工知能研究センター
研究員

大規模言語モデルからの知識抽出に基づく視覚スクリプトの創成

研究成果の概要

画像からシナリオ(主役およびゴール)とスクリプト(主役がゴールを達成するための典型的なイベント系列)を予測するタスクについて、(i) 評価用データセットの作成 (ii) 大規模言語モデル(LLM)およびマルチモーダル大規模言語モデル(MLLM)を用いたモデルの構築と精密な評価を実施した。

評価用データセットの作成では、画像から想起されるスクリプトについてこれまで定量評価に耐えうる量のデータセットは存在しなかった。そこで今回新たに飲食・宿泊・仕事場・パーティーなどの様々な屋内外の場面を含む 600 枚の画像に対して 1,200 件のシナリオおよびスクリプトのアノテーションを実施した。

モデル構築では、画像からのシナリオ生成および画像およびシナリオからのスクリプト生成の2種類のタスクについて、LLMがシナリオおよびスクリプトの想起に必要な知識を含んでいるという仮説の元、シナリオ生成・スクリプト生成タスクそれぞれについて画像説明文生成モデルと LLM を組み合わせる手法および画像とテキストを同時に受け付ける MLLM を使用する手法を提案しその評価を行った。

提案手法は両タスク共に定量・定性双方優れた結果を生成し、シナリオ生成・スクリプト生成双方について、自動評価・人手評価共に MLLM が上回る性能を示した。一方、自動評価と人手評価との相関はわずかに認められるものの低い水準に留まり、多様な場所における常識的知識を評価する方法に課題が見いだされた。

今後の展望としては、刻一刻と変化する状況において正しい行動を生成するためのスクリプト知識の活用などが考えられる。