

先端科学技術研究科 修士論文要旨

所属研究室 (主指導教員)	サイバネティクス・リアリティ工学 (清川 清 (教授))					
学籍番号	2411261	提出日	令和 8年 1月 19日			
学生氏名	松崎 優陽					
論文題目	LLM は野球を理解できるか？長文文脈ルール推論ベンチマークの構築と評価					
要旨						
<p>人工知能の進展はスポーツ分野にも波及しており、パフォーマンス分析や戦術最適化など、多様な応用が進められている。中でも、大規模言語モデル (Large Language Models; LLMs) の発展は、選手評価や戦略分析といった言語的タスクの自動化を加速させていている。こうした背景のもと、スポーツ領域における知識や理解を評価するための質問応答型ベンチマークが提案してきた。</p> <p>しかし、既存のベンチマークの多くは、定義や一般的な知識の理解に主眼を置いており、競技中に発生する複雑なルール判断を十分に評価できていない。特に野球は、星状況、走者の動き、プレイの順序といった複数の要因が相互に作用し、得点やアウトの成立条件が状況に応じて変化する競技である。そのため、試合状況を正確に把握した上でルールを適用する高度な推論能力が求められるが、このような能力を体系的に検証するデータセットは限られている。</p> <p>本研究では、野球における複雑なルール判断に焦点を当てた新たな質問応答ベンチマークを構築し、LLM の推論精度を評価する。提案ベンチマークでは、試合状況を模した詳細なプレイ記述に基づき、「得点の有無」や「アウトの成立」といったルール適用判断を問う設問を用いる。設問の難易度に応じて二種類のデータセットを構築し、複数の LLM を対象に、zero-shot, few-shot、および knowledge-grounded prompting (KGP) におけるプロンプト構造の違いが回答精度に与える影響を検証した。</p> <p>実験の結果、zero-shot における最高の正答率は 68.5% となり、より複雑な設問を含むデータセットでは正答率が低下する傾向が確認された。一方、few-shot では正答率の顕著な向上は限定的であったのに対し、KGP では精度の向上が確認され、設問に関連するルール情報を提示することが、LLM の推論を有効に補助する可能性が示された。本研究を通じて、LLM が野球の複雑なルールを含む設問に對して示す推論能力の実態を明らかにし、今後の応用可能性と改良の方向性を示す。</p>						