

対話データを利用したキャラクター特徴を持つ会話の生成

山中 竜輔

本研究は、キャラクターの発話データを活用し、特定のキャラクター性を持つ対話システムの構築を目指すものである。従来のキャラクター対話システムの構築手法では、発話データの収集やアノテーションに多大な人手とコストが必要であり、効率性に課題があった。これに対し、本研究ではシナリオから自動的にキャラクター発話を抽出し、前発話を生成することで対話形式のデータセットを効率的に構築する手法を提案する。さらに、シナリオ構造の解析を通じてキャラクター間の関係性や背景情報を抽出し、これらを大規模言語モデル（LLM）である GPT に反映させることで、キャラクターの個性や一貫性を保持した自然な応答を生成することを目指す。

実験では、Rosebleu データセットを用い、3名の異なるキャラクター（ヴェル、ウルル、オペラ）に対して各 500 ペアの対話データを構築し、GPT-4 をファインチューニングしたモデルを評価した。自動評価および人手による主観評価の結果、提案手法はベースラインモデルおよび関係性データを考慮しないモデルと比較して、「口調のキャラらしさ」や「内容のキャラらしさ」において有意な改善を示した。また、「意味的自然さ」においても一定の向上が確認された。一方で、一部のキャラクターにおいては関係性データの影響が限定的であり、生成された応答の自然さに改善の余地があることが示唆された。

本研究の成果は、低コストかつ効率的にキャラクター性を反映した対話システムを構築するための有効なアプローチを提供するものである。今後の課題として、関係性データの精緻化や学習データの拡充、感情変化に対応した対話生成手法の導入が挙げられる。これらの課題に取り組むことで、より高度で魅力的なキャラクター対話システムの実現が期待される。

(指導教員 関 洋平)