

(続紙 1)

京都大学	博士 (情報学)	氏名	白井 圭佑
論文題目	Interpreting Instructional Texts Towards Robot Execution (ロボット実行のための指示テキスト解釈)		
<p>(論文内容の要旨)</p> <p>本論文は、手順書を解釈し実行するロボットシステムの実現を目標に、手順書のワークフローの解釈、動作結果の視覚的予測、言語指示のロボット計画への変換のそれぞれのアプローチに関する研究成果をまとめたものであり、全5章から構成されている。</p> <p>第1章では、手順書ワークフローの表現の獲得、手順書上での物体の状態変化の予測、言語指示のロボットの動作表現への変換という研究方向に焦点を当て、自然言語処理、マルチメディア、ロボット工学における取り組みについて説明し、これらを統合することで、主に調理レシピを対象としたロボットシステムを提案している。その後、本博士論文における研究主題として、(i)調理レシピを対象とした物体の動作後の視覚状態予測、(ii)言語指示のロボットの計画への変換、(iii)本システムの調理以外の分野への拡張という3つを挙げ、それぞれの導入を行っている。</p> <p>第2章では、調理レシピにおける物体の動作後の視覚状態予測という課題に取り組んでいる。これは、手順書上の物体の状態変化の追跡という自然言語処理のタスクと、動作後の物体の視覚状態認識というコンピュータビジョンのタスクを組み合わせたものであり、手順書のワークフローを解釈しつつ動作による物体の視覚的状态変化を行う必要がある。ここでは、このタスクに取り組むために新たにVisual Recipe Flow (VRF)データセットを提案している。VRFデータセットでは、レシピフローグラフを拡張して調理レシピのワークフローを表現し、物体の視覚的な状態変化を動作前後に対応する画像の組を用いて表現している。データ収集方法、アノテーションの統計量について説明を行った後、複数者によるアノテーション間の一致率を計算することで、アノテーションの質を確認している。最後に、テキストと画像の両モダリティが動作後の視覚状態の予測に有用であることを示すため、マルチモーダルな検索の実験を行っている。</p> <p>第3章では、言語指示をロボットの動作計画に変換する課題に取り組んでいる。近年では、大規模言語モデル (LLM) を用いて言語指示を計画に直接変換するアプローチが注目を集めている。しかし、これらのアプローチは入力から出力への変換をend-to-endで行うため解釈性に欠け、加えてLLMの出力した計画が有効であるとは限らないという問題がある。一方で、古典計画においては、シンボリックプランナーを用いて問題記述に対応する計画の探索を行うのが主流である。ここで、シンボリックプランナーによって得られた計画は論理的に正しいことが保証されている。ここでは、LLMベースの手法の問題を解決するために、Vision Language Interpreter (ViLaIn)の提案を行っている。ViLaInは言語指示と環境の観察から対応する問題記述の生成を行い、シンボリックプランナーを用いることで有効な計画の探索を行う。ViLaInの評価のため、ここでは新たにProblem Description Generation (ProDG)データセットを併せて提案している。実験では、ViLaInがProDGデータセットにおいて、文法的に正しい問題記述の生成と有効な解の探索を高精度で行えることを示している。</p>			

第4章では、提案システムを調理以外の分野に拡張する課題に取り組んでいる。ここでは、手順書のワークフロー解釈に焦点を当て、先行研究で提案されたレシピフローグラフ表現を調理以外の分野を扱えるように拡張した、wikiHowフローグラフ表現を提案している。調理以外の分野におけるフローグラフ予測精度を調査するために、wikiHowから収集した4分野の手順書にアノテーションを行うことで、w-FGコーパスを構築している。実験では、低リソース設定を想定し、既存の調理分野のコーパスを用いて対象分野への分野適応を行うことで大幅に精度向上が実現できることを示している。

第5章では、本論文における貢献のまとめを行い、各章における今後の課題について議論している。

(続紙 2)

(論文審査の結果の要旨)

本論文は、調理分野を対象として、手順書を解釈し実行するロボットシステムの実現を目指して得られた研究成果をまとめたものであり、主な成果は次の通りである。

1. 調理レシピにおいて、物体の動作後の視覚的状态の予測を行うための新たなデータセットを提案した。データの収集方法、アノテーション手順について順に説明を行い、最後に、データセットの評価をアノテーション一致率の計算とアノテーションを用いた実験により行った。

2. 大規模言語モデルを用いて、人が解釈可能な方法で言語指示をロボットの計画に変換する新たなフレームワークを提案した。評価のための新たなデータセットを構築し、これにおいて提案フレームワークは言語指示から計画への変換を高精度で実現できることを確認した。

3. 提案システムの調理以外の分野への拡張を試みた。ここでは、手順書のワークフロー解釈に焦点を当て、先行研究で提案された調理レシピの表現を拡張した、新たな表現を提案した。調理以外の分野におけるフローグラフ予測精度を調査するため、新たなデータセットを構築し、実験を行った。実験では、既存の調理分野のコーパスを用いて分野適応を行うことで、低リソース下で精度向上が実現できることを確認した。

以上、本論文は、手順書を解釈し実行するロボットシステムの実現を目指し、物体の動作後の視覚状態予測、言語指示のロボットの計画への変換、提案システムの調理以外の分野への拡張というそれぞれの課題に対して、適切な解決方法を提案したものであり、学術上・実応用上寄与するところが少なくない。よって、本論文は博士(情報学)の学位論文として価値あるものと認める。また、令和6年2月19日に実施した論文内容とそれに関連した口頭試問の結果、合格と認めた。なお、インターネットでの全文公表を行うことについて支障がないことを確認した。