

論文審査の要旨および学識確認結果

報告番号	甲 第 号	氏 名	神原 元就
論文審査担当者：	主査	慶應義塾大学教授	博士（情報学） 杉浦 孔明
	副査	慶應義塾大学教授	博士（工学） 斎藤 英雄
		慶應義塾大学教授	工学博士 斎藤 博昭
		慶應義塾大学教授	博士（工学） 高橋 正樹
(論文審査の要旨) 学士（工学）、修士（工学）神原元就君提出の学位請求論文は「Multimodal Language Processing for Everyday Task Execution by Service Robots（サービスロボットによる日常タスク実行のためのマルチモーダル言語処理）」と題し、全6章からなる。 本論文は、自然言語によるコミュニケーションを伴う日常タスクをサービスロボットが実行できるようにすることを目的とし、言語情報に加えて視覚情報およびロボットの軌道情報を統合するマルチモーダル言語処理モデルを構築した研究である。本論文では、自然言語指示により指定されたタスクを適切なサブタスクに分解し実行する、あるいは分解された各サブタスクの成否判定を行うためのマルチモーダル言語理解モデルを提案する。 第1章では、研究の背景、目的、本研究の位置付け及び論文構成を示している。 第2章では、画像を入力として、物体操作指示文を生成する Fetching Instruction Generation タスクを目的として、対象物体、目標領域および画像中の物体群の位置に関する関係性をモデル化する Case Relation Transformer (CRT) を提案している。実験を通じて、CRT が位置情報を用いた注意機構によって、物体同士の空間的な関係性に関する参照表現を、必要に応じて適切に生成できることを明らかにしている。 第3章では、家庭環境内における自然言語指示に基づく Fetch-and-Carry タスクを対象とし、シミュレータ及びマルチモーダル言語理解・生成モデルを用いることでタスク生成からタスクの実行までを完全に自動化したパイプラインを開発している。このパイプラインによって多様な指示文で Fetch-and-Carry タスクの実行システムを評価可能であることを示している。さらに、指示文と画像の類似度に基づき操作対象の物体を特定するマルチモーダル言語理解モデルを提案している。この手法により、既存のルールベースシステムを上回る成功率を達成できることが示されている。 第4章では、ある時刻までの動画を入力として、それ以降に起き得る事象に関する説明文を生成する Future Captioning タスクを対象とし、動画に含まれる各事象の自己相関を計算し将来的に起き得る事象の推論およびそれに関する説明生成を行う Relational Future Captioning Model を提案している。実験を通じて、物体の配置方策および入力動画に基づいて、将来的にどのような物体同士の衝突やその連鎖が起き得るかについて、適切な説明を生成できたことを示している。 第5章では、マニピュレータの軌道、自然言語指示、および物体操作前の画像を入力として、それらのアラインメントを計算し、ひいては将来的な自然言語指示に基づく物体操作の成否に関する予測を行う Pre-Manipulation Alignment Prediction タスクを対象とし、環境と軌道、およびそれらと自然言語指示の交差注意を段階的に計算することで各モダリティ同士のアラインメントに関する特徴を計算する手法を提案している。実験により、提案手法は、2段階の交差注意機構により物体操作を実際に行うことなく将来的な物体操作の成否を適切に予測可能であることを明らかにしている。 第6章では、本論文の総括を行い、今後の展望について述べている。 以上要するに、本論文は、自然言語を用いた日常タスクの実行を目的として、マルチモーダル言語処理モデルに基づいた基盤技術を開発し、実験を通じてその有効性を明らかにしたものであり、工学上寄与するところが少なくない。よって、本論文の著者は博士（工学）の学位を受ける資格があるものと認める。			
学識確認結果	学位請求論文を中心にして関連学術について上記審査会委員で試問を行い、当該学術に関し広く深い学識を有することを確認した。 また、語学（英語）についても十分な学力を有することを確認した。		