Osaka University Knowledge Archive

Title	ロボットの言語獲得に向けたマルチモーダル世界モデルの構築
Author(s)	宮澤, 和貴
Citation	大阪大学, 2022, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/91759
rights	
Note	やむを得ない事由があると学位審査研究科が承認したため、全文に代えてその内容の要約を公開しています。全文のご利用をご希望の場合は、大阪大学の博士論文についてをご参照ください。

Osaka University Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

Osaka University

様式３

論　文　内　容　の　要　旨

氏　名　（　宮澤　和貴　）	
論文題名	ロボットの言語獲得に向けたマルチモーダル世界モデルの構築

論文内容の要旨

本論文では，実世界で言語を獲得するロボットの実現を目指す．人間は複数の感覚器官から取得したマルチモーダル情報をもとに心の中に世界のモデルを形成する．日常の中で経験する様々な情報を抽象化した世界モデルを形成することで，より効率的により正確に物事を理解して行動や決断をする．この世界モデルの中に言語を内包することで，マルチモーダル情報と言語情報の関係を捉え，自身の経験を言葉として他者に伝えることや，他者の言葉を自身の世界モデルを介して解釈することができる．ロボットも人間のように世界を理解して言語を獲得することで，ロボットの適応可能な環境や状況は増え，ロボットの価値が向上する．そこで本論文では，ロボットの言語獲得に向けて，確率的生成モデルとtransformerによる２種類の機械学習手法を採用して，マルチモーダル世界モデルを構築する．

初めに確率的生成モデルによる統合モデルを提案する．マルチモーダルカテゴリゼーションに基づく情報の抽象化を核として，複数の個別の機能をもつモジュールを統合することで，ロボットの自律的な言語と行動の学習を行う．ロボットは，人や環境との相互作用を通して学習した結果，簡単な言語の理解や文章の生成を実現した．確率的生成モデルが有用である一方で，近年，深層学習モデルであるtransformerによる言語学習やマルチモーダル学習が優れた性能を示している．そこで，transformerのロボットでの利用に関する調査を行い，その後，transformerによるモデルを２つ提案する．１つ目のモデルとして，事前学習済みtransformerモデルを用いたモデルを提案する．深層学習では十分な量の学習データを必要とするが，マルチモーダル情報は単モーダル情報と比較して大量にデータを収集することが難しい．そこで，単モーダルデータにより事前に学習されたtransformerモデルを利用した，データ効率の良いマルチモーダル学習を実現する．２種類のマルチモーダルデータセットを用いた分類タスクの結果，各データセットに対して高い分類精度を示すとともに，学習データ量の減少に対しても頑健であることを示した．２つ目のモデルとして，言語モデルであるBERTをマルチモーダルに拡張したモデルを提案する．このモデルは自己教師有り学習により，ロボットが取得したマルチモーダル情報と言語情報の関係を学習する．マルチモーダル情報を入力とした分類タスク及び単語予測タスクの結果，高い分類精度を示し，語順を考慮した単語予測を実現した．

様式７

論文審査の結果の要旨及び担当者

氏　　名　　（　　宮澤　和貴　　）		
論文審査担当者	（職）	氏　　名
	主 査　　教　授	長井　隆行
	副 査　　教　授	石黒　浩
	副 査　　教　授	原田　研介

論文審査の結果の要旨

本論文は，実世界で言語を獲得するロボットの実現を目指し，必要となる計算メカニズムをマルチモーダル統合の視点を中心に検討したものである．特に，確率的生成モデルとtransformerの2種類の機械学習手法に着目し，マルチモーダル情報を統合しつつ言語を包含するロボットの「世界モデル」を構築する手法を提案している．

本論文で示されている具体的な成果は以下の通りである．まず，確率的生成モデルによる統合認知モデルを提案した．マルチモーダルカテゴリゼーションに基づく情報の構造化を核として，個別の機能をもつ複数のモジュールを統合することで，ロボットの自律的な言語と行動の学習を行うモデルである．ロボットが人や環境との相互作用を通して学習することで，言語の理解や文章の生成ができるようになることを実験により示した．

次に，近年自然言語処理で注目されているtransformerを基盤としたモデルを検討した．transformerを用いる場合，一般に大量の学習データを必要とする．ロボットに関連するマルチモーダルデータを大量に収集することは困難であるため，この問題を解決する必要がある．そこで，単モーダルデータにより事前に学習されたtransformerモデルを利用したデータ効率の良いマルチモーダル学習を実現した．さらに，transformerと同様の注意機構を利用した言語モデルであるBERTをマルチモーダルに拡張したモデルを提案した．このモデルは自己教師有り学習により，ロボットが取得したマルチモーダル情報と言語情報の関係を学習するものである．実際，マルチモーダル情報を入力とした分類タスク及び単語予測タスクの結果，高い分類精度を示し，語順を考慮した単語予測が実現できることを示した．

以上のように，本論文はロボットの言語能力を含む知能を実現するための学習モデルを複数提案し，その可能性を実験的に検証したものであり，今後の知能ロボットの発展に大きく寄与すると考えられる．また，実際のロボットを用いた学習実験の結果を詳細に解析することで得た知見は，知能の構成的理解という科学的貢献の側面も有している．よって，本論文は博士（工学）の学位論文として価値のあるものと認める．