

申請者	学科名	情報通信工学科	職名	教授	氏名	岩橋 直人
調査研究課題	雑談コミュニケーションを通じたロボットの実世界理解を伴うマルチモーダル対話能力の機械学習の研究					
調査研究組織	氏名	所属・職		専門分野	役割分担	
	代表	岩橋 直人	情報通信工学科・教授	機械学習	マルチモーダル対話学習・研究全体	
	分担者	高瀬 健太 石田 卓也	情報系工学研究科・大学院生（博士前期課程）	マルチモーダル情報処理 音声言語情報処理	物体概念学習・情報統合 対話処理・音声インターフェース	
調査研究実績の概要	<p>1.背景と目的</p> <p>近年、雑談システムの研究が盛んである。雑談を行うコミュニケーションロボットが数多く開発されている。これらのロボットは、雑談の他、人間の要求に応じて、天気予報を調べたり、電話をかけたりと、さまざまなドメインの対話ができる。従来研究のほとんどは、眼前の事物とは関係のない、たとえば「どんな映画が好きですか」といった発話（以降、非接地発話と呼ぶ）によるコミュニケーションをいかに実現するかという問題に焦点を当てたものである。</p> <p>一方で、眼前の事物に関する発話（以降、接地発話と呼ぶ）の言語表現および概念をロボットに学習させる、言語獲得、記号接地、記号創発に関する研究が盛んである。従来研究のほとんどは、記号的知識をまったく持たないロボットがいかに記号的知識を学習するかという問題に焦点を当てたものである。</p> <p>こうした研究の二極化の流れの中で、人間とロボットの間で、非接地発話によるコミュニケーションが既に成り立っている状態から、接地発話により、言語と眼前の事物との対応付けをいかに学習するかという問題は、ほとんど取り扱われてこなかった。実際、従来の雑談システムでは、言葉と眼前の事物との結び付きを学習する対話をすることはできない。たとえば、人間がリモコンをロボットに見せながら「これはリモコンだよ」と教えて、「リモコン」という単語と‘リモコン’という物体を結び付けて学習することができない。人間が発した接地発話を、ロボットが適切に理解するためには、言語情報と実環境情報を統合して処理する必要がある。雑談の中に接地発話と非接地発話が混在する可能性を考える必要がある。雑談システム研究において、接地発話をいかに取り扱うかは重要な課題である。このような領域の研究が希求されている。</p> <p>そこで我々は、ロボットが人間との雑談を通じた自然なインタラクションにより、言語と実世界事物の対応付けを学習できるロボットの実現を目指している。本研究では、接地発話として物体教示発話のみを対象にして、ロボットが人間との雑談の中から、人間の物体教示発話を検出し、その発話の対象としている物体とそれを指示する単語を学習する手法を提案した。</p>					



図 1 人間-ロボットインタラクション

調査研究実績の概要

2.提案手法

提案手法は、音声情報とシーン画像情報の組の集合を入力データとし、物体と単語の組の集合を出力するものである。本手法は、図2に示すように連続する二つのプロセスから構成される。まず、物体教示発話検出プロセスである。人間の発話情報と身体動作情報から、発話が物体教示発話であるか否かを判別する。次に、物体と単語の学習プロセスである。発話がシーン画像情報に含まれる複数の物体の内のどれを指し示すものであるか、同時に、発話中のどの単語がその物体を指し示すものであるかを推測する。

3.実験

提案手法におけるこれら二つのプロセスそれぞれの有効性を実験により評価した。まず、提案手法による物体教示発話の検出性能を10分割交差検定により評価した結果を表1に示す。2000発話の内、物体教示発話であると分類されたのは196発話であった。提案手法により、高い性能で物体教示発話の検出が可能であることが示された。次に、物体と単語の学習の性能を、物体教示発話と分類された文の中で、各発話において正しい単語が選択された確率 P_w 、各発話において正しい画像クラスが選択された確率 P_c 、各発話において正しい単語と画像クラスが選択された確率 P_{wc} によって評価した。学習法A, B, Cの評価結果を表2に示す。性能は、 $A < B < C$ の順で高くなっていることが分かる。さらに、CRFによる物体教示発話の教師無し判別の評価結果を表3に示す。Precisionが高いことから、線形SVMでの記号接地発話の誤検出を、ある程度排除できていることが見て取れる。

4.まとめ

雑談を通した物体と単語の学習という、従来まったくなかった新しい発想の研究アプローチを提案した。本アプローチに基づいた具体的なアルゴリズムを提示、実装し、実験により評価した。今後は、より詳細な評価と改良を行ってゆく予定である。

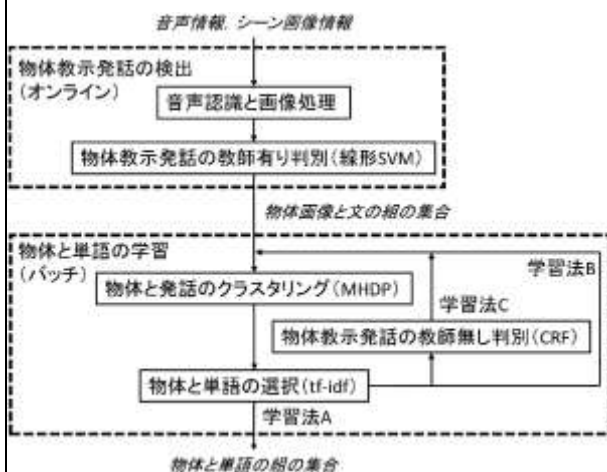


図 2 提案手法の処理の流れ

表 1: 物体教示発話の検出性能 (%)

Precision	Recall	F-measure
87.3	88.2	87.6

表 2: 学習された物体と単語の精度 (%)

学習法	P_w	P_c	P_{wc}
A	31 (61/196)	30 (59/196)	10 (19/196)
B	35 (69/196)	57 (112/196)	28 (54/196)
C	46 (66/144)	63 (90/144)	34 (49/144)

表 3: 物体教示発話の教師無し判別性能 (%)

Precision	Recall	F-measure
95 (137/144)	78 (137/175)	86

成果資料目録

- [1] 山本一馬, 石田卓也, 岩橋直人, Ye Kyaw Thu, 国島文生. “人とロボットによる雑談から眼前の物体に記号接地している発話の検出法,” HAIシンポジウム, Dec. 2016.
- [2] 石田卓也, 山本一馬, 岩橋直人, イェチョウトウ, 中村友昭, 長井隆行, 国島文生. “雑談を通した物体と単語の学習,” 人工知能学会全国大会 (5月発表予定)
- [3] Ye Kyaw Thu, T. Ishida, N. Iwahashi, T. Nakamura, T. Nagai. “Symbol Grounding from Natural Conversation for Human-Robot Communication,” INTERSPEECH (投稿済)