

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

Symbol Grounding in Human-Robot Dialogs using Visual and Linguistic Cues

山本一馬 石田卓也 岩橋直人 Ye Kyaw Thu 國島丈生

Kazuma Yamamoto, Takuya Ishida, Naoto Iwahashi, Ye Kyaw Thu, Takeo Kunishima

岡山県立大学

Okayama Prefectural University

Abstract: Symbol grounding is the connection of words or symbols with their meanings through objects that exist in the real world. Our primary aim is to develop a robot that is capable of symbol grounding through human interaction. In this study, we have developed a method for detecting utterances that relate to symbol grounding based on SVM classification with data from both visual sensors, and the linguistic content of user's utterances acquired by automatic speech recognition.

1 はじめに

近年、人と雑談を行う対話システムの研究が盛んである [1][2][3][4]。さらに、人と雑談を行うことができるコミュニケーションロボットが数多く開発されている [5][6][7]。これらのロボットは人と雑談を行ったり、人の要求に応じて、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をするができる。しかしながら、言葉をロボットの周囲の実世界事物と結びつけること（記号接地）ができない。具体的な例として、ロボットに「お肉が食べたい」と言って、眼前の皿の上にある肉を食べさせてもらうことはできない。実際の雑談では、「どんな映画が好きですか」といった眼前の物体とは関係のない内容の発話（本稿では雑談的発話と呼ぶことにする）と、「これはコップだよ」「このワインを開けて」といった、眼前の物体に関する内容を述べている発話（本稿では記号接地発話と呼ぶことにする）とが混在していることがあるため、雑談対話システム研究において、記号接地発話をいかに取り扱うかは重要な課題である。

一方で、記号接地発話を理解・生成するための研究として、眼前の事物に関する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も盛んである [8][9][10]。従来の研究のほとんどは、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。前述の雑談対話システムのように記号的なコミュニケーションが既に成り立っている状態から、記号をいかに実世界の事物に接地させるかという問題は、少数の研究事例 [11] があるものの、ほとんど取り扱われてこなかった。雑談的発話と記号接地発話とが混在する雑談を扱った研究として、人同士の雑談からロボットに向けられた物体操作指示発話を検出する手法 [12] が提案されているが、このような領域の研究が希求されている。

そこで我々は、ロボットが人との雑談を通した自然なインタラクションにより、言葉を実世界に接地させることが可能なロボットの実現を目指している。そのための第一歩として、本研究では人との雑談から記号接地発話を検出する手法の検討を行った。

2 雑談システム

本研究では A.L.I.C.E. [13] などを用いられているルールベース型の雑談対話システムを構築した。できるだけ多くのユーザ発話がルールにマッチできるように、ユーザ発話の語尾にマッチするルールを最初に作成し、次に細かなルールを作成した。実際に作成したルールを、ルールがマッチするユーザの発話

表 1：雑談システムのルール例

ユーザ発話例	マッチするルール	システムの応答
こんにちは お腹が空きました おいしいお肉が食べたい おやつは何がいいですか テレビは何を見ますか 明日は雨が降るらしい 明日は学校休みですか これはコップだよ	* こんにちは * 腹*（空き 減り）* * おいしい* 食* べ* たい* * おやつ* 何* いい* か * テレビ* 何* 見* か * らしい * か * よ	こんにちは、今日は暑いですね 何か食べますか 私もおいしいお肉食べたいです あんばん 最近はバラエティーが多いです それは本当ですか 明日は学校休みですよ へーそうなんだ知らなかったよ

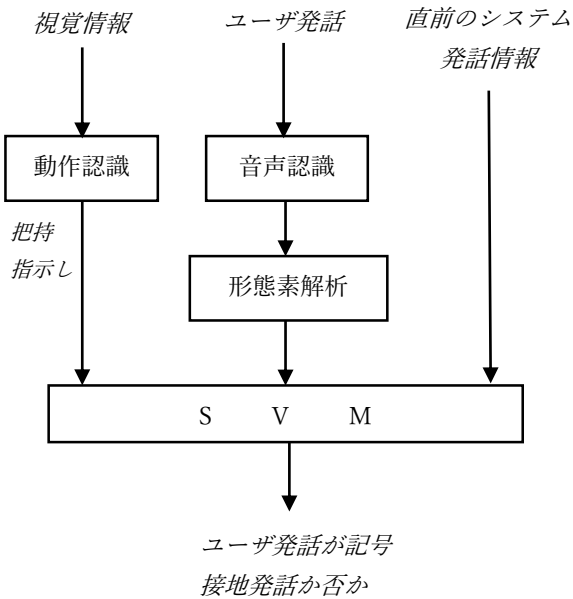


図 1：提案する手法の概要

例，ルールがマッチしたときのシステムの応答とともに表 1 に示す．表 1 の * 記号はあらゆる文字列にマッチすることを，| 記号は括弧内のいずれかの文字列にマッチすることを意味している．

本雑談システムではユーザ発話にマッチするルール 1 つに対して，複数のシステム応答を作成した．例えば「*ごめんなさい」というルールに対して，「気にしないでください」，「許しません」といった 2 つのシステム応答を用意し，システム内で生成した乱数からどちらを用いるかを決定した．作成したルールは全部で 248 個で，いずれのルールにもマッチしないユーザ発話に対しては「どういうことですか」，「ごめん，聞いてなかったからもう一回言って」といった応答をするようにした．

3 提案手法

本研究で提案する手法を図 1 に示す．本研究で構築したシステムは視覚情報とユーザの発話内容，直前のシステムの発話情報の 3 つの情報を入力として，Support Vector Machine(SVM) を用いてユーザ発話を記号接地発話と雑談的発話に分類する．

3.1 視覚情報

ユーザが「これはコップだよ」といった記号接地発話を行う際には，眼前にあるコップの把持や，指示しを同時にしていると考えることが自然である．そこで本研究ではユーザが眼前の物体を把持しているか，または指示しているかの 2 つの視覚情報を視覚センサで検出し，SVM の学習データとして用いる．視覚情報の獲得は Kinect を用いて行い，Kinect で検出されている眼前の物体の状態に変化がある度に，視覚情報を更新していく．

3.2 ユーザの発話内容

ユーザ発話は音声認識ツールキット *rospeech*[14] を用いて獲得する．音声認識によって得られたユーザ発話は形態素解析エンジン *MeCab*[15] を用いて形態素解析を行い，ユーザ発話を単語ごとに分割する．

3.3 直前のシステム発話の情報

ユーザとロボットが対話をする際に，ロボットからユーザに対して「これは何ですか?」，「ここにあるのは何?」といったような眼前の物体について問いかける発話があると予想される．そこで本システムでは，直前のシステム発話が疑問文であるか否かという情報も SVM の学習データとする．

4 評価実験

4.1 実験方法

本実験は、被験者は男性 1 名とし、ロボットと被験者の間に机を置いて行い (図 2)、机の上に置かれた物体を Kinect で検出させた。本実験では図 3 に示す 10 個の物体を用いた。被験者はこれらの物体を把持や指示しながらシステムとの対話を行った。ただし、一度に机の上に載せる物体の数は最大で 4 個とした。被験者の発話は全部で 2000 発話用意し、その内 1800 発話を雑談的発話、残りの 200 発話を記号接地発話とした。

本実験では表 2 に示すように、視覚情報と被験者の発話内容、直前のシステム発話の情報のうち、学習データとしてどれか 1 つを用いた場合と 2 つを用いた場合、そして 3 つ全てを用いた場合の合計 7 つの実験について 10 分割交差検定を行う。実験結果の評価尺度には、適合率 (precision), 再現率 (recall), F 値 (F-measure) を用いる。

4.2 実験結果

まず、本実験における音声認識の精度は表 3 の通りである。

次に、本実験で実際に被験者がシステムに対して発話した文章とその音声認識結果、被験者の発話に対するシステム応答の例を表 4 に示す。システム応答については、音声認識が誤った場合と正しく音声認識できた場合の応答例を示している。認識誤りは、「趣味」が「君」といった別の単語に置換された誤りに加え、「ここにあるのはぬいぐるみです」という被験者の発話の後半部分の「はぬいぐるみです」が脱落した誤りも存在している。音声認識が誤ることによって被験者の発話が別のルールにマッチしてシステムの応答が不自然になってしまったり、いずれのルールにもマッチせず「どういうことですか」、「ごめん、聞いてなかったからもう一回言って」といった応答が何度も繰り返される結果になった。

最後に、表 2 で示した 7 つの実験について、収集したデータを LIBSVM[16] を用いて 10 分割交差検定を行った結果を表 5 に示す。実験 3 の直前のシステム発話の情報のみを用いた実験では、2000 個全ての発話データが雑談的発話に分類されたので、実験 3 の結果については省略している。



図 2：実験の様子



図 3：実験で用いた 10 個の物体

表 2：本研究で行った 7 つの実験の一覧

実験	1	2	3	4	5	6	7
被験者の発話	○	-	-	○	○	-	○
視覚情報	-	○	-	○	-	○	○
直前のシステム発話	-	-	○	-	○	○	○

表 3：音声認識精度

文章	26.0%
単語	70.8%

表 4: 被験者の発話とその音声認識結果, システム応答の例

	被験者の発話	音声認識結果	システム応答	
雑談的発話	趣味を教えてください	君を教えてください	誤	うーん、どうしようかな
			正	映画を見ることと食べることです
	確かに	第二位	誤	どういうことですか
			正	やっぱりそうですね
	私もお寿司好きです	私もいいです	誤	気が合いますね
			正	お寿司のどんなところが好きなんですか
	京都は好きですか	今日はいいですか	誤	今日はいいですよ
			正	京都大好きです
	そうなんだ	今日は何名	誤	ごめん、聞いてなかったからもう一回言って
			正	そうですよ
	こんにちは、久しぶりだね	こんにちはがきれいだね	誤	そうですね
			正	久しぶりです。調子はどうですか
	知らないです	切らないで	誤	どういうことですか
			正	本当ですか
	何か暇だね	何か今だね	誤	そうですね
			正	どこかに行きませんか
	ありがとうございました	映画等ございました	誤	そうでしたか
			正	いえいえ
	今週三連休だって	今週連休だった	誤	そうでしたか
			正	それは本当ですか
記号接地発話	ここにあるのはぬいぐるみです	ここにあるの	誤	ごめん、聞いてなかったからもう一回言って
			正	へーそうなんですね
	この赤いコップは僕のです	この赤い石は僕のれる	誤	本当ですか
			正	分かった
	コップです	これって	誤	ごめん、聞いてなかったからもう一回言って
			正	分かった
	これはコップだよ	これはいくらよ	誤	どういうことですか
			正	わかった
	この黒い猫は僕のです	この黒い猫は送られる	誤	本当ですか
			正	へーそうなんですね

表 5：実験結果

実験	1	2	4	5	6	7
precision(%)	84.9	23.9	89.0	87.1	21.1	87.3
recall(%)	78.7	20.2	87.4	77.3	17.5	88.2
F-measure	0.814	0.217	0.879	0.813	0.191	0.876

5 考察

表 5 で示した実験結果のうち，学習データにユーザの発話を用いた実験 1, 4, 5, 7 では，F 値が全て 0.8 を超える結果が得られた．このことから記号接地発話の検出には，ユーザの発話内容の情報が最も有用であることが示された．

また，ユーザ発話のみを学習のデータとして用いた実験 1 とユーザ発話と視覚情報を学習のデータに用いた実験 4 の結果を比較してみると，実験 4 の方が評価尺度の値の全てが高くなっていることから，把持や指示といったユーザの動作情報も記号接地発話の検出に一定の効果があることが分かった．

直前のシステム発話が疑問文か否かといった情報を学習データに加えた実験 5, 6, 7 と加えなかった実験 1, 2, 4 では，前者の方で記号接地発話の検出精度が低下した．これは直前のシステム発話が疑問文となる学習のデータは記号接地発話と雑談的発話の両方に存在し，特に本実験では雑談的発話に多く含まれていたため，SVM が記号接地発話を誤って雑談的発話と分類してしまったことが原因だと考えられる．

本研究で行った実験では，文章における音声認識の精度が低い結果となっているにもかかわらず，ユーザの記号接地発話の検出精度は比較的高い結果が得られた．これは SVM の頑健性に依存しているためだと考えられる．また，ユーザ発話に対するシステム応答の中には不自然な応答が多く見られた．これは，雑談システムのルールが少なかったことと，音声認識の誤りによるものだと考えられる．

6 結論

本研究では，ロボットが物体認識や音声認識を用いてユーザとマルチモーダルなインタラクションを行いながら，ユーザの発話から記号接地発話を検出する手法を開発し，評価を行った．その結果として，ユーザの発話と動作情報が記号接地発話の検出に有効であることが示された．

我々はユーザの記号接地発話から実際に目前にあ

る物体に対してロボットが記号接地をすることが可能になると考えている．そのため，本研究で記号接地発話を検出する手法を開発できたことで，我々が目指しているロボットの実現に一步近づいたのではないかと考えられる．

今後は，雑談システムのルールの数を増やし，ユーザの発話に対して，より自然な応答ができるシステムになるように改良していくとともに，記号接地発話の検出精度を高めていきたいと考えている．さらに，記号接地発話に基づいて物体の概念を学習する手法を開発する予定である．

謝辞

本研究は，JSPS 科研費 15K00244，および，JST CREST 「記号創発ロボティクスによる人間機械コラボレーション基盤創成」の助成を受けたものです．

また，雑談システムを構築する際に協力していただいた，シナジーマーケティング株式会社の谷田泰郎氏と高椋琴美氏に感謝します．

参考文献

- [1] 前田和希, 宋鑫, 國政裕友樹, 豊田博之, 韓東力: 雑談システムにおける話題展開の性能向上, 言語処理学会第 16 回年次大会発表論文集, pp. 250-253, (2010)
- [2] 大西可奈子, 吉村健: コンピュータとの自然な会話を実現する雑談対話技術: NTT DOCOMO テクニカル・ジャーナル, Vol. 21, No. 4, pp. 17-21, (2014)
- [3] 目黒豊美, 杉山弘晃, 東中竜一郎, 南泰浩: ルールベース発話生成と統計的発話生成の融合に基づく対話システムの構築, 人工知能学会全国大会発表論文集, (2014)
- [4] 木村泰知, ジェブカラファウ, 高丸圭一: Radiobots 型対話システムの提案, 人工知能学会全国大会発表論文集, (2015)
- [5] <http://www.takaratomy.co.jp/products/omnibot/ohanas/>
- [6] <http://www.softbank.jp/robot/>
- [7] <https://www.unirobot.com/>
- [8] 岩橋直人: ロボットによる言語獲得 ―言語処理の新しいパラダイムを目指して―, 人工知能学会誌, Vol. 18, No. 1, pp. 49-58, (2003)
- [9] N.Iwahashi: Robots That Learn Language: Developmental Approach to Human-Machine Conversations, Human Robot Interaction (N.Sanker, Ed.), pp. 95-118, (2007)
- [10] T. Taniguchi, T. Nagai, T. Nakamura, N. Iwahashi, T. Ogata, and H. Asoh: Symbol Emergence in Robotics: A Survey, Advanced Robotics, Vol. 30, Issue 11-12, pp. 706-

728, (2016)

- [11] H. Holzapfel, D. Neubig, and A. Waibel: A dialogue approach to learning object descriptions and semantic categories, *Robotics and Autonomous Systems*, Vol. 56, Issue 11, pp. 1004-1013, (2008)
- [12] X. Zuo, N. Iwahashi, K. Funakoshi, M. Nakano, R. Taguchi, S. Matsuda, K. Sugiura, and N. Oka: Detecting Robot-Directed Speech by Situated Understanding in Physical Interaction, *人工知能学会論文誌*, Vol. 25, No. 6, pp. 670-682, (2010)
- [13] Richard S. Wallace: The Anatomy of A.L.I.C.E., Parsing the Turing Test, pp. 181-210, (2009)
- [14] 杉浦孔明, 堀智織, 是津耕司: rospeex : クラウド型音声コミュニケーションを実現する ROS 向けツールキット, *電子情報通信学会信学技報 CNR2013-10*, pp. 7-10
- [15] <http://taku910.github.io/mecab/>
- [16] C.-C. Chang, C.-J. Lin: LIBSVM: A library for support vector machines, *ACM Transactions on Intelligent Systems and Technology(TIST)*, Vol. 2 Issue 3, (2011)