

# ロボットによる環境の自律的探索と言語命令解釈の相互作用

板谷純希 † 伊東慶輔 † 荒木孝弥 † 中村友昭 † 長井隆行 † 船越孝太郎 ‡  
 長谷川雄二 ‡ 中野幹生 ‡ 岩橋直人 §

† 電気通信大学大学院 情報理工学専攻 知能機械工学専攻

‡ (株) ホンダ・リサーチ・インスティテュート・ジャパン

§ 独立行政法人 情報通信研究機構

## 1 はじめに

本稿では、ロボットがいかにかユーザーからの言語命令を解釈し、実行するかについて議論する。解釈の基本は認識文の構文解析であるが、環境への依存性や曖昧性、音声認識の誤りなど様々な要因でロボットがユーザーの自然発話による命令を正しく実行することは容易ではない。ユーザーの曖昧な言語命令をロボットが正しく解釈するための手がかりの一つは、環境情報である。しかし、ロボットが物や家具がどこにあるかといった環境情報の収集を完全に自律的に行うことは困難であり、環境情報の取得には、ユーザーとのインタラクションが有用である。例えば、未知の物体の名前を覚えるためには、ユーザーから名前を聞く必要があるが、この名前をユーザーの言語命令から推測することが可能となる場合がある。つまり、環境情報の収集と命令の解釈を相互に精緻化させることが可能であると思われる。そこで本稿ではまず、ロボットが室内を事前に探索し、環境（家具や物体の種類や位置）の情報を自律的に取得しデータベースとして保持することから始める。

これにより得られた情報は、ユーザーの曖昧な発話や音声認識の誤りを解消する手がかりとなる。さらに、ユーザーからの言語命令を実行するための可能な解釈の中で、取得できなかった情報を収集し、それを新たな命令の解釈に利用する手法を検討する。これにより、ユーザーの命令をロボットが実行できる可能性が向上することを実機の実験を通して示す。

実ロボットが、自然言語による命令を実行するための枠組みは、近年盛んに研究されている [1]-[3]。しかし、ヒューマノイドのレベルでマニピュレーションを含む複雑なことができるものはあまり多くない。また、言語理解と共にロボットが未知の物体や動作を学習できることは非常に重要であると考えられる [3]。本稿では、自然言語処理と物や人、動作など様々なレベルの学習を統合したシステムを考えるが、さらに環境の自律的な探索と推論を統合した上でこれらのインタラクションにより言語理解とロボットの知識を相互に精緻化することを検討する (図 1)。

## 2 ロボットシステム

### 2.1 モジュール構造

本稿で提案するシステムは、対話行動制御部構築ツールキット HRIME (HRI Intelligence platform based on Multiple Experts) [4] によって実装されることを想定する。これは、特定のタスクに特化した知識と内部状態を持つエキスパートと呼ぶモジュールを複数用いるモデルに

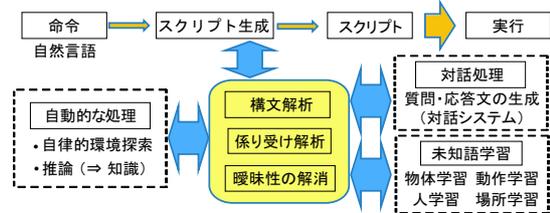


図 1: システムの概要

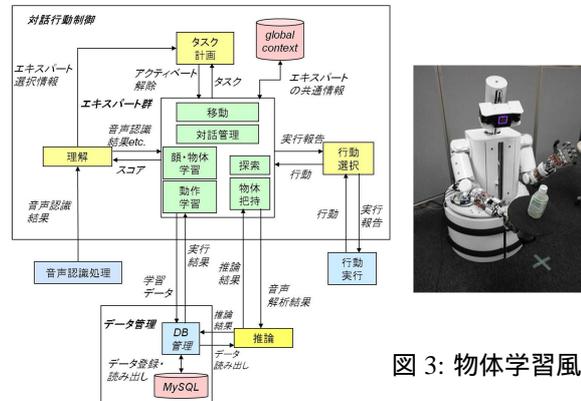


図 3: 物体学習風景

図 2: ロボットシステムの概要

基づいている。図 2 に、ロボット内のモジュール構造の概要を示す。データベースを管理するためのソフトウェアも、一つのモジュールとして実装される。ロボットが動作した結果はエキスパートからデータベース管理モジュールへと送られ、データベースへと保存される。それらのデータはユーザーの命令解釈や、動作を実行するために利用される。

## 3 提案方式

本稿で提案する手法の概要図を図 1 に示す。提案方式では命令文の構文解析だけでは解消できない命令の曖昧性をユーザーとの対話や対話を利用した未知語の学習、ロボットによる自律的な環境の学習や、それら学習で得られた知識を利用した推論により解消する。

### 3.1 環境の自律的探索

本稿では、室内の環境情報を取得するために 2 段階の学習を行う。まず室内に配置されている家具の位置情報を学習し、次に家具の上に置かれている物体の学習を行う。家具の学習時には室内の 2 次元マップを SLAM により作成し、そのマップ上での家具の範囲をデータベースに保存する。家具の学習終了後、ロボットはその家具を順番に回り、家具上の物体を学習する。物体の学習では、図 3 のようなテーブルを使って、ロボットが自律的にテーブルを回転させながら対象物体の様々な方向からの画像を取得する。学習後には物体を元の位置に戻し、物体の座標と物体が配置されている家具名をデータベースに登録する。同時にユーザーが音声により学習中の物体の名前を教えることで、その名詞と物体の画像情報を結びつけて学習することも可能であ

Interaction between autonomous environmental search and natural language understanding by robots

† Junki ITAYA, Keisuke ITO, Takaya ARAKI, Tomoaki NAKAMURA and Takayuki NAGAI

‡ Kotaro FUNAKOSHI, Yuji HASEGAWA and Mikio NAKANO

§ Naoto IWAHASHI

† The University of Electro-Communications

‡ Honda Research Institute Japan Co., Ltd.

§ National Institute of Information and Communications Technology

る。また、これらの情報は環境学習時以外であってもロボットが動作する毎に最新の情報に更新する(図2)。

### 3.2 命令文の解釈

#### 3.2.1 名詞の処理

名詞処理の基本は、その対象が空間的にどの位置に存在するのかを特定することである。環境学習により、室内の家具や物体の位置情報はデータベースに学習しておくためその情報を利用して動作を行う。例えば家具を表わす名詞の場合では結びついている座標情報が出力され、物体を表わす名詞では座標情報と、配置されている家具名が出力される。命令の中に物体の名詞が使われた場合、その物体の正確な座標位置を特定しなければいけないため、ロボットの視覚センサが物体を検出できる位置まで移動する必要がある。これはデータベースの座標情報をもとに実行が可能となるため、命令内に物体の位置を表す名詞が含まれる必要がない。

#### 3.2.2 動詞の処理

動詞は、対応する基本動作や上位行動を選択し、実行時の引数として名詞の情報を受け取る。動作と結びついていない動詞が使用された場合、意味辞書を利用した動詞の類義語検索により補う。この意味辞書による類義語検索はその単語が持つすべての意味で行われてしまうため、結びついている動作と違う意味の類義語が検索されてしまう可能性があるが、動作と結びついている動詞にあらかじめ同じ意味の英語の動詞を結びつけておき、日本語意味辞書と英語意味辞書を同時に使用して類義語検索を行うことで、より検索する単語の意味を限定することが可能になる。

#### 3.2.3 対話処理

命令文の曖昧性については、後に述べるようにデータベース内の情報を用いて推測することが可能な場合もあるが、それだけで解消できない時にはそれを補うようにロボットが人間に質問を行う。ロボットからの質問は、不足している情報(名詞・形容詞)の種類とすでに獲得している情報の種類によって決定しており、命令文を解析することで質問が自動的に生成される。例えばユーザーからの命令が「コップを持ってきて」というものであった場合、コップの場所とコップの種類の情報不足していることがわかるため、ロボットからの質問は「コップはどこにありますか?」と「どのコップですか?」となる。

### 3.3 インタラクティブな学習

環境の探索を事前に行うことで、家具の位置や物体の情報を得ることができるが、それ以外にも、必要に応じて物体や場所、人、動作学習などを行うことができる。物体や人の学習では、ユーザーが対象の名前や特徴を音声で教示すると同時にロボットが対象の画像を取得する。場所や家具を学習する際は、ロボットをその位置まで誘導した上で名前を教示することで、ロボットはデータベースに位置情報と名前を登録する。動作の学習に関しては、HMMに基づく見まね学習を実装している。

### 3.4 推論と知識の獲得

命令文の曖昧性は事前に行う環境学習で取得した情報をもとに推測することで解消し、新たな知識として再利用することが可能である。例えば、環境の探索時にユーザーとのインタラクションがなかった場合、ロボットは物体の位置情報と画像情報は取得できるが、物体名などを結びつけることができない。このような状態の時にユーザーから物体の把持命令があった場合、その物体の名前を知らないため実行することができない。しかし、命令文の中で物体が置かれている家具が指定された場合は、データベース内からその家具にある物

体を検索し、未知の物体があればその物体を対象として動作を実行すると同時に物体名をデータベースに登録することができる。具体的には次のようなシナリオを考案することができる。ロボットが、テーブル上に未知の物体があることを自律探索により知っており、ユーザーに「テーブルにあるコップを持ってきて」というような、未知物体の把持を命令をされたとする。その際ロボットは、テーブル上の未知物体がコップであろうと推測して行動し、またその時の解釈によって得られた「コップ」という名詞をその物体の名前としてデータベースへ登録する。

また、物体だけでなく場所や人物などについても同じ仕組みを適用できる。例えば「リビングにいるAさんにコップを渡して」というような命令を受けた際に、リビングに1人しか未知の人がいなければ、その人がAさんであると推測できる。さらにこれは、人や場所、物体、形容詞など様々なものに対して適用可能である。これらは全て、未知である(名前との対応付けがない)という情報を積極的に利用して対象を特定し、その情報を知識としてデータベースに登録することに相当する。ただしここでは、ユーザーの誤信念や嘘はないものと仮定している。また、物体や場所が複数の呼び名をもつ場合も扱うことができないため、これらへの対応は今後の課題である。

## 4 実験

一般ユーザーが実機のロボットに命令を行い、どの程度行動することができ、その過程で新たな知識を得ることができるかを確認する実験を行った。実験は、棚やテーブルなどの6種類の家具が配置されている5m四方ほどの部屋に10種類の物体、4人の人をランダムに配置して行った。ロボットは事前に部屋の環境探索を行うが、学習させる物体、家具、人それぞれの半数の名前は教示せずに、画像情報や位置情報のみ(人は画像情報のみ)を記憶させる。その後ユーザーがロボットに人・物体の探索、把持等の命令を複数行い、命令の成功率と共に、データベースになかった知識(名前)が動作後に得られているかどうかを検証した。

実行する命令は人と物体の探索、物体の把持、運搬などの動作で、推論によって解消できない場合は、物体学習や動作学習などのインタラクティブな学習を行う。また、実行に必要な物体や人はランダムに選ぶこととした。命令文は、システムを知らない被験者に自然言語で生成してもらい、合計30文をロボットに命令した。その結果、ロボットは87%の命令を実行することができた。また、最終的にすべての未知物体や人、場所の名前をデータベースに登録することができた。そのうち半数は、推論によって得られた知識であり、ユーザーへの質問やユーザーによる教示を行うことなく獲得することが可能であった。

## 5 まとめ

本稿ではユーザーの命令をロボットが実行するために、環境情報の収集と命令の解釈とを相互に利用する手法を提案し、その動作の確認を行った。その結果、命令の解釈の中から新たな情報が取得できることが確認できた。今後、大規模な評価実験を行う予定である。

## 参考文献

- [1] S.Tellex *et al.*, "Understanding Natural Language Commands for Robotic Navigation and Mobile Manipulation", Proc. of AAAI2011, 2011
- [2] M.Tenorth *et al.*, "Understanding and Executing Instructions for Everyday Manipulation Tasks from the World Wide Web", in Proc. of ICRA2010, pp.1486 - 1491, 2010
- [3] A.Kirsch, "Robot Learning Language-Integrating Programming and Learning for Cognitive Systems", Journal of Robotics and Autonomous Systems, Vol.57 Issue 9, pp.943-954, 2009
- [4] 中野ほか, "複数のエキスパートに基づくロボット対話行動制御モデル構築ツール", JSAI 全国大会, 2007