

# ユーザとのインタラクションに基づく学習を利用した ロボットのタスクプログラミング

Natural Language-Based Task Programming with Interactive Learning Ability of Robots

板谷 純希\*<sup>1</sup>      中村 友昭\*<sup>2</sup>      長井 隆行\*<sup>1</sup>  
Junki Itaya      Tomoaki Nakamura      Takayuki Nagai

\*<sup>1</sup>電気通信大学大学院情報理工学研究所

Faculty of Informatics and Engineering, The University of Electro-Communications

\*<sup>2</sup>電気通信大学大学院電気通信学研究所

Faculty of Electro-Communications, The University of Electro-Communications

Robots can execute only pre-programmed sets of actions in general. In order to make the robot truly useful in our living environments, robots are expected to carry out complex tasks in a flexible manner. Furthermore, the number of required actions can be huge, and it is not plausible that all of these actions are programmed in advance. Therefore, robots are required to learn novel actions through the interaction with users. In this paper, we examine the natural language-based task programming with the learning ability of the robot. The most important point in this paper is that the acquired new actions can be utilized within the natural language-based task programming framework. The proposed framework is implemented on a mobile robot platform and some experiments are carried out to validate the proposed framework.

## 1. はじめに

一般に、ロボットの行動はあらかじめプログラムされている必要があり、プログラムされていない行動を行うことはできない。全ての行動を事前にプログラムしておくことは困難であるため、実際の動作環境に応じてプログラミングを行う必要がある。しかし、ロボットのユーザーが必ずしもプログラミングの知識を持っているとは限らず、特に家庭用ロボットの普及を想定した場合、ユーザーが一般のコンピュータ言語を用いてプログラミングすることは困難である。従って、ロボットが人間からの自然言語による命令や説明を柔軟に理解し、自らが実行できるプログラムへと変換する能力が必要である。

こうしたロボットへの要求は、様々な場面で見ることができる。例えば、ロボカップ@ホームリーグでは、General Purpose Service Robot(GPSR)と呼ばれるタスクが2010年の大会より追加された[RoboCup 10]。このタスクでは、行うタスクの具体的な内容は事前には知らされず、ユーザーからの自然言語をロボットが解釈して動作する必要がある。ロボカップ@ホームリーグは、ロボカップの新しいリーグであり、家庭用ロボットのための技術を競技形式で競うものである。多くの競技が、予め決められたタスクをいかに正確に行うのかを競うのに対し、GPSRタスクでは、ロボットのフレキシブルな行動生成能力が問われることになる。このように、家庭内でのタスクをユーザーの命令に応じて適応的に行うことができる能力は非常に重要である。

こうした背景のもと、ユーザーの言語による命令を解析し、それをロボットが実行できる形へ自動的に変換する枠組みを提案する。著者らは、[板谷 10]において中核となる構文解析に基づくプログラムの自動生成について、固定ロボットを用いた評価を行っている。文献[板谷 10]において提案したシステムの基本的な考え方は、ロボットの行動が、基本的な動作の組み合わせで表現することができるというものである。例えば“掴

む”のような動作の場合、“手を開く”、“物体の位置に腕を動かす”、“手を閉じる”という基本的な動作を組み合わせることによって実現することができる。そこで、これらの基本的な動作プログラムをあらかじめ用意しておき、これらを組み合わせることによって複雑なロボットの行動を実現する。また、基本的な動作プログラムを、特定の単語と結びつけることによって、自然言語によるプログラミングを実現するものである。それらの動作を[板谷 10-06]において自律移動ロボットのタスクプログラミングへと拡張し、またユーザーとロボットとの対話を用いて命令の曖昧性を一部解消する機能を実現した。本稿では意味辞書を用いた類義語の判断を利用し、命令可能な動詞の幅を拡げ、未知語の学習を用いてロボットの動作に必要な情報の獲得を実現する。

関連研究としては、文献[青山 09]の走行プログラムの自動生成が挙げられる。しかしこれは、GUIによる操作のみを対象としており、入力された走行経路図からの走行プログラム生成のみを対象としている。文献[Knoop 07, Pardowitz 05]では、ヒューマノイドロボットの自動プログラミングを扱っているが、基本的にはユーザーの実演をいかにロボットで実行するかが主眼である。また、ロボットによる言語理解の研究も多く存在するが、対話システムにおける応答生成など、コミュニケーションが主眼であり、複雑なタスクを行うようなものは少ない。一方、ロボットによる言語獲得[Roy 02, Iwahashi 07]や模倣学習の研究も進んでおり、本質的な問題はこうしたアプローチにより解決される可能性がある。しかし、現段階ではまだ実用的であるとはいえず、また実用上はある程度の作りこみと学習のハイブリッドな手法が現実的であると思われる。また、ロボットと人間とのコミュニケーションの特徴を分析することでロボットに対する人間の認知特性を探るなどのような、人間とロボット間のコミュニケーションに対して人間視点から焦点を当てている研究も挙げられる[松本 07]。さらに、CGを自然言語から自動生成する研究も行われている[Oshita 09]。アプローチとしては非常に近いものの、実世界における物理的な動作とCGとは本質的な違いがある。

連絡先: 板谷 純希, 電気通信大学大学院情報理工学研究所, 東京都調布市調布ヶ丘 1-5-1, j-itaya@apple.ee.uec.ac.jp

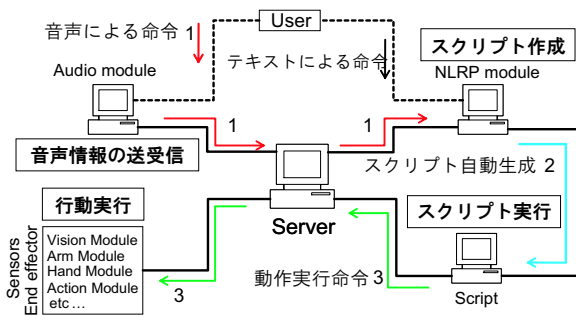


図 1: ロボットシステムの概要

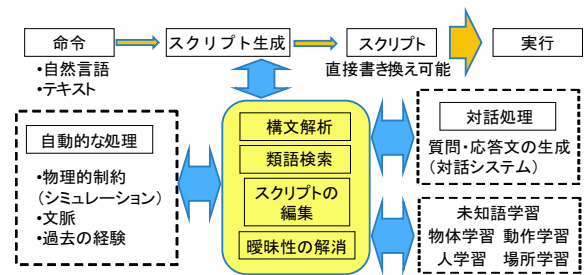


図 3: 自然言語によるロボットプログラミングの全体像

表 1: ロボットからの質問 (...; 動詞の原型 ~; 場所名 ; 物体名・人物名)

必要な情報	文脈予測	質問文
物体の場所	不可	はどこにありますか
人物の場所	不可	さんはどこにいますか
物体の場所	可	は~にありますか
人物の場所	可	さんは~にいますか
移動場所	不可	どこに...のですか
対象物体名	不可	何を...のですか
対象人物名	不可	誰を...のですか
種類	不可	どの ですか

```

from RobotMovePositionName import *      #モジュールの読みこみ
from ActionGrabObject import *

data.PositionName.append("キッチンテーブル") #移動場所の名前を引数へ追加
MovePositionName(data)                    #移動

data.ObjectName.append("ペットボトル")     #把持物体の名前を引数へ追加
GrabObject(data)                          #物体の検索・把持
    
```

図 2: スクリプトの例

## 2. ロボットシステム

### 2.1 モジュール構造

本稿で想定するロボットのソフトウェアは、モジュール群として実装される。これは、RTM などのミドルウェアを用いたネットワーク構造で実現可能である。本稿で用いるロボットでは、DiGOROnet と呼ぶ独自のミドルウェアを用いる。これは黒板モデルをベースとしており、モジュールの接続・切断が比較的柔軟に行えるのが特徴である。ロボットをプログラミングするためのソフトウェアも、一つのモジュールとして実現される。図 1 に、ロボット内のモジュール構造の概要を示す。Natural Language Robot Programming(NLRP) モジュールにおいてユーザの指示に基づいたスクリプト型のプログラムが自動生成され、実行する。

### 2.2 スクリプト実行

ロボットの行動は、基本的な動作をあらかじめモジュールとして用意しておき、それを組み合わせて順に実行することで実現する。タスクプログラムはスクリプト言語である IronPython によって作成し、図 2 のように基本動作モジュールへ必要な情報を渡し、順番に呼び出すことで実行していく。基本的な動作のモジュールは、設計者が事前に設計することになるが、動作学習によって新たに学習させることも可能である。

## 3. 言語からのスクリプト自動生成

### 3.1 提案システムの概要

図 1 における NLRP モジュールが自然言語からスクリプトを自動的に作成する仕事を行うことになるが、基本的な考え方は、用意した基本動作や上位行動に単語(動詞)を割り当てておき、ユーザーからの命令文を構文解析することで、動詞に対応したモジュールを並べていくというものである。この際、主語や述語となる名詞は動作モジュールの引数となるため、センサ情報処理結果を参照することで、その存在の有無や位置などの情報に変換する。自然言語を用いて命令する場合の大きな問題は、命令文の曖昧性である。特に、移動ロボットにおいてはロボットが行動するために必要となる情報の種類が多く、より命令に曖昧性が含まれる可能性が高くなる。図 3 に構想の全体像を示す。最終的には、背景知識の獲得や経験からの学習、といったロボットが自動的に判断する機能を実現する必要がある

が、本稿では基本的な処理に加え、音声辞書に未登録な単語を用いた物体学習や、意味辞書を用いた類義語の判断を利用し、命令可能な動詞の幅を広げる。

### 3.2 基本処理

#### 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。しかし、名詞には物体位置を表わすものやロボットの移動座標を表わすものなどの種類があるため、それらの名詞にあらかじめ対象の位置特定に必要な情報を結びつけて学習しておくことで実現する。例えばロボットの移動先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るが、物体を表わす名詞では結びついている情報と、物体認識などセンサ情報処理結果を比較することによって得られた結果を最終的に出力する。しかし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するためには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。そのため物体名詞にはその位置を表わす場所名詞が修飾語として係っている必要があるとして考え、物体名詞に係る場所名詞が存在する場合はあらかじめその場所に移動してから物体の検索を開始し、名詞が存在しなかった場合は対話処理により補う。

#### 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 [NICT 10] 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問題があったが、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、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。また、各動作では必要となる情報があらかじめ定義されているため、命令が曖昧な時に不足している情報の判断も動作自体で行うことができる。



図 4: 新規動作”捨てる”の学習風景

### 3.2.3 対話処理

命令文の曖昧性については前述したように動詞について処理した後に、その動詞に必要な名詞が不足していないかどうかで判断を行っており、それを補うようにロボットが人間に質問を行い、情報の補間を行う。また動詞に必要な名詞の他にも、名詞の処理の項で述べたように物体名詞に場所名詞が係っているかどうかの判断も行って質問をする。ロボットから行われる質問の一覧を表 1 に示す。表のように不足している情報（名詞）の種類とすでに獲得している情報の種類によってロボットからの質問を決定しており、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が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。さらにコップを持っていく場所についての情報も不足しており、この命令の動詞の原型は「持つてくる」であるため表の...部分にこれを当てはめ「どこに持つてくるのですか?」という質問が生成される。しかし、「キッチンテーブルに行き、林檎を持ってきて」のように文章内に物体の位置情報と予想される場所名詞が含まれていた場合、ロボットからの質問は「リングはキッチンテーブルにありますか?」という確認の内容となり、これが否定された場合によく「リングはどこにありますか?」と質問するようになっており、文脈から簡単な予測処理を行うことで対話処理の向上をはかっている。ロボットの質問は 1 つ行うごとにユーザーからの返答を待ち、質問を行った後の返答に含まれる名詞とその名詞に係る修飾語（形容詞など）に注目することで情報の取得を行う。

### 3.3 未知語の学習

動詞と基本動作の組を全て事前に用意することは難しいため、ロボットに動作を見せ、動詞を結びつけることで未知動作を学習させることが可能である。これはトラジェクタの軌道とランドマークの位置関係を何通りも学習し、HMM (Hidden Markov Model) を用いて汎化をすることで新たな動作として獲得している [Sugiura 11]。図 4 は実際に「捨てる」という動作を学習している様子である。この場合ランドマークが赤いごみ箱であり、赤いラインがトラジェクタであるぬいぐるみの軌道である。さらにトラジェクタの高さや左右を入れ替えるなど物体の位置関係を変えて学習させることで「捨てる」を学習する。ただし、この時結びつける単語自体は音声認識の辞書に登録されている必要がある。従ってこれは厳密な意味での未知語の学習ではない。しかし、物体の学習や場所の学習は未知語を用いて行うことが可能である。この手法に関して著者らは、[Attamimi 10] において厳密な意味での未知語の学習として議論しており、本稿での枠組みに組み込まれている。

## 4. 実験

実験に用いたロボットと動作環境を、図 5 と図 6 に示す。ロボットは、全方位移動台車をベースとして、4 台のオンボード

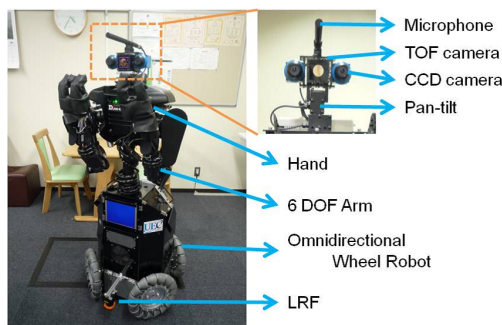


図 5: 実験で用いたロボット DiGORO

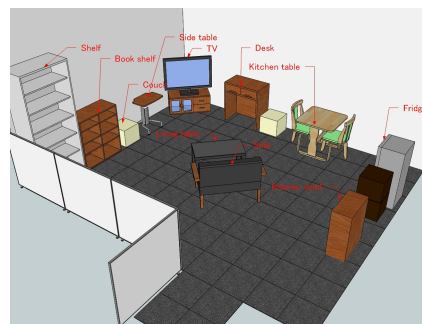


図 6: 実験で用いた動作環境

PC, LRF, 6 自由度のアーム, 赤外線カメラ, CCD カメラ, マイクフォンから構成されている。カメラはパンチルト台に載っており、周囲を見まわし物体の位置情報の取得や人の顔検出などを行うことができる。また、全方位移動台車で移動することができ、LRF を用いた自己位置の推定や、カメラで検出した物体をアームにより把持することが可能である。

実際に命令から生成したスクリプトが実行可能であるかの評価を行った。命令はロボカップ@ホームリーグの 2010 年度版ルールブック [RoboCup 10] の競技内容の文章から 25 種類抽出、模倣学習により獲得した動作を含めた命令を 5 種類使用し、全 30 種類の命令を用いた。実験で使用する物体は 10 種類、そのうち 5 種類の物体の名前と家具や人物の名前も予め定義しておき、残り 5 種類の物体の名前は未知語を用いて学習しておく。命令は名詞のパターンを変えそれぞれ 3 回ずつ行い、スクリプトの生成率を評価した。また、命令の動詞を理解できる確率を、意味辞書による類義語検索を行わない場合と、日本語の類義語検索のみを用いた場合、日本語と英語の類義語検索を両方用いた場合とでそれぞれ比較し、理解可能な動詞のカバー率についても評価する。

結果は図 7 に示す。図の (a) は実際行った動作の成功失敗にかかわらず、命令の内容に沿ったスクリプトが生成された割合を示し、全命令のうち 87% の割合でスクリプトが生成されたことを示す。しかし、スクリプト生成に成功した命令のうち 14% は命令内の動詞が解析できず、ロボットがユーザーへ命令の言い換えを要求することで最終的にスクリプトの生成に成功している。例えば「物体を回収しろ」という命令では「回収」という動詞に類義語が存在せず解析ができないため、ロボットは「回収するが実行できません。別の言葉で命令して下さい。」と発話する。また、「手を振っている人に近付いて」という命令では人の状態を認識することができないためスクリプトの生成に失敗してしまうので、今後は人や物体の名詞に対して位置情報だけでなく状態を認識するための処理が必要になってくる。残りの 73% が動詞の解析に成功した割合であり、これが動詞の認識率となる。命令の解析に日本語の類義語検索しか行わな

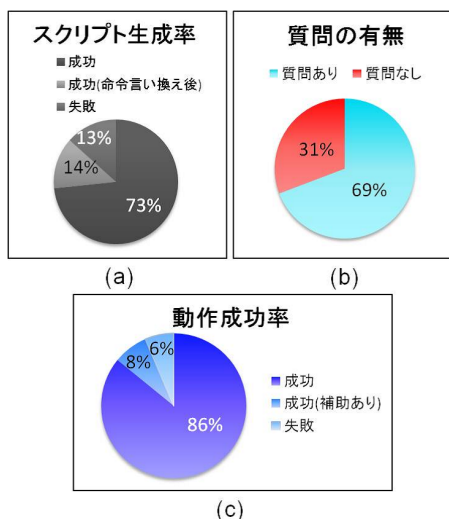


図 7: 実験結果 (a); スクリプト生成結果 (b); ロボットが質問を行う割合 (c); 動作実行結果

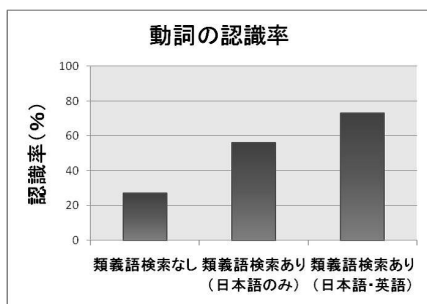


図 8: 動詞の認識率の比較

かった場合この認識率は 56%となり、意味辞書による類義語検索を行わなかった場合この割合は 27%となる。(図 8)したがって類義語検索によって動作可能な動詞数を大きく拡大することができるといえる。さらに、言い換えた命令を含めてユーザーへ何らかの質問を行い情報の補間を行いながらスクリプトの生成を行った確率は 69%であった。(図 7(b)) 従ってロボットから人へ指示や質問を行うことでスクリプト生成率の向上が達成できた。また、生成されたスクリプトを実行したときの動作結果は図 7(c) に示す通りで、86%の割合で動作に成功した。8%はハードウェア的にロボットが実行できない動作であり、ロボットがユーザーに補助の依頼を行うことで実行している。例えば「ドアを開けて退場しろ」という命令では「ドアが開けられないので開けてください」と発話する。残りの 6%は物体認識の失敗や移動の失敗により動作が失敗してしまうが、動作自体の安定性も高いと言える。また、今回用いた物体と家具の半数は音声辞書に名前が未登録のものを用い、実験直前にそれらの未知語学習を物体学習や家具の位置学習と同時にを行い実験に使用したが、それらの情報がロボットの動作に与える影響は少なく、使用する物体すべてがあらかじめ定義されている場合、その動作が失敗する割合は 5%であり、ほとんど変化は見られず、未知語を用いた情報の獲得が有効であることがわかる。

## 5. まとめ

本稿では文献 [板谷 11] にて提案した日本語意味辞書による類義語検索を利用した動作可能な動詞数の拡大に、英語の類義語検索を加えることでさらなる動詞数の拡大に成功した。さら

に人とのインタラクションによる動作の学習を取り入れ、さらなる動詞数の拡大を実現した。また、音声辞書に未登録な単語を用いた物体や場所の学習を行うことで動作実行のための情報をロボットが柔軟に獲得することが可能となった。今後はロボットが動作する環境が複雑化しても行動が可能かどうかの検証が必要であり、物理シミュレーション上での動作実験も予定している。将来的にはシミュレーション上で動作の可否を自動的に判断したうえで、動作を行うことも視野に入れている。また、経験や文脈などからロボットが自動的に曖昧性を解決するなどの機能を検討する予定である。

## 参考文献

- [青山 09] 青山ほか: “次世代ロボット知能化技術開発プロジェクト - RTM 化におけるプログラム開発効率と品質の向上”, 第 27 回日本ロボット学会学術講演会, 2D2-06, 2009
- [Knoop 07] Knoop, S. *et al.*: “Automatic robot programming from learned abstract task knowledge”, in Proc. of IROS07, 2007
- [Pardowitz 05] Pardowitz, M. *et al.*: “Learning Sequential Constraints of Tasks from User Demonstrations”, in Proc. of HUMANOIDS2005, 2005
- [松本 07] 松本ほか: “人間とロボットの対話の認知特性 ホームロボットを用いた生活実験の対話ログの解析”, IEICE, MVE2007-16, 2007.6
- [Oshita 09] Oshita, M.: “Generating Animation from Natural Language Texts and Framework of Motion Database”, in Proc. of CW2009, 2009
- [板谷 10] 板谷, 中村, 長井: “自然言語によるロボットプログラミング”, IPSJ 全国大会, 2010.03
- [板谷 10-06] 板谷, 中村, 長井: “自律移動ロボットのための自然言語タスクプログラミング”, JSAI 全国大会, 2010.06
- [板谷 11] 板谷, 中村, 長井: “自然言語を用いた自律移動ロボットのタスクプログラミング”, IEICE 全国大会, 2011.03
- [Roy 02] Roy, D. and Pentland, A.: “Learning Words from Sights and Sounds: A Computational Model”, Cognitive Science, Vol.26, No.1, pp.113-146, 2002
- [Iwahashi 07] Iwahashi, N.: “Robots That Learn Language: A Developmental Approach to Situated Human-Robot Conversations”, In N.Sankar ed. Human-Robot Interaction, pp.95-118, I-Tech Education and Publishing, 2007
- [Attamimi 10] Attamimi, M. *et al.*: “Learning Novel Objects Using Out-of-Vocabulary Word Segmentation and Object Extraction for Home Assistant Robots”, in Proc. of ICRA 2010, 2010
- [NICT 10] 日本語ワードネット (XX 版) (c) 情報通信研究機構, 2009-2010 available at <http://nlpwww.nict.go.jp/wn-ja/>
- [RoboCup 10] Robocup@home2010 rulebook available at [http://www.ai.rug.nl/robocupathome/documents/rulebook2010\\_FINAL\\_VERSION.pdf](http://www.ai.rug.nl/robocupathome/documents/rulebook2010_FINAL_VERSION.pdf)
- [Sugiura 11] Komei Sugiura, Naoto Iwahashi, Hideki Kashioaka, and Satoshi Nakamura: “Learning, Generation, and Recognition of Motions by Reference-Point-Dependent Probabilistic Models”, Advanced Robotics, Vol. 25, No. 6-7, pp. 825-848, 2011