

確率的空間記憶に基づくロボットのための 言語対話型インタフェースの実現

稲邑 哲也 園田 朋之* 岡田 慧* 稲葉 雅幸*

A stochastic space memory model for cooperative task achievement via human-robot interaction

Tetsunari INAMURA, Tomoyuki SONODA*, Kei OKADA* and Masayuki INABA*

Design and acquisition method of environmental model are important issues for robots which act in the real world such as a daily life environment. Robots have to follow up changes of environment and re-construct the model because of the conditions in the real world environment would be changed constantly. Also the model for unobserved part would be needed and should be inferred when users instruct the robot using symbolic expressions. In this paper, we have focused on stochastic representation of environmental memory to realize smooth communication between humans and robots, and realtime memory management with ambiguities in the real world. We also show that the representation is effective to construct a cooperative task achievement system on intelligent robots.

1. はじめに

近年、ロボットは人間の日常生活空間の中に進出し、人間と共に行動する機会が増えるにつれて多種多様な環境の中で人間の指示を理解することが求められるようになってきている。特に音声インタフェースの重要性は広く認知されており、音声言語によってロボットに指示を出し、組み立て作業や物の持ち運びタスク等を遂行させるためのヒューマンロボットインタラクションの研究例も多く発表されている^{1) 2) 3) 4) 5)}。しかしながらこれらの手法では、ユーザの発話に含まれる曖昧性や、環境をセンシングした結果生じる不確実に対応することが困難であった。小林らは人間と音声言語コミュニケーション可能なロボットにおいて、未知の語彙に対応する動作を獲得させるための手法を提案している⁶⁾。しかしながら、新規状況として対応可能なのは新規動作のみであり、環境変動への対応は考慮されていなかった。

ウィノグラードの指摘⁷⁾のように、自然言語処理によるロボットのユーザインタフェースを構築する際には、世界の表象モデルの果たす役割は非常に大きく、システムは環境の情報を積極的に収集するのみならず、過去の状態を記憶・保持しておく能力や、目の届かない領域に関する環境情報を推定する機能を構築しておく必要がある。特に日常生活環境下ではロボットは目の前の状況のみからではなく、過去に経験した記憶の中のみ存在し

ている情報を考慮したり、全く見たことが無い未知の環境に対応して状況判断を行う必要がある。特に、ロボット研究においては、環境モデルを事前に与えることを前提としていない状況がほとんどであり、モデルを自分自身の経験からリアルタイムに構築し、利用可能な表現形式でマネージメントして行く必要がある。

一方で、我々はこれまでに確率的情報処理の枠組みを用いてロボットと人間との間の適応的な対話制御システムを構築してきており、実環境の不確実性およびユーザの指示に含まれる曖昧性を解消する対話を実現してきた⁸⁾⁹⁾。しかし、状況の変動や物体、ロボットの位置関係などの把握、時系列的情報・文脈の利用などが課題として残されていた。そこで本年度の研究として、空間的・時間的な記憶を管理するシステムを提案し、変動していく環境の状況を考慮した上で、適切な発話行為と行動計画を行いながらタスクを遂行していくシステムの実現を目指す。

2. 確率的空間記憶モデル

2.1 環境と物体の空間記憶モデル

人間機械協調系の研究では、ユーザの曖昧な指示に対応するために確率的な計算・推論モデルが従来から提案されてきている。例えば若松らは確率的計算により感情・状況をモデリングし、Human-Robot 間の対話を円滑にさせる試みをしている¹⁰⁾。田中ら¹¹⁾、Stoppら¹²⁾、は人間との対話に基づいてロボットが行動をプランするための確率的計算モデルを提案し、それを用いて「このコ

2006年04月05日 受理

*東京大学大学院情報理工学系研究科

ップをその右側に置いて」というような、環境の物理的配置に関する指示の解釈を実現している。しかしながら、「さっきの物をここに運んで」というような、時間的な文脈表現を理解することは一般的に難しかった。文脈を追う会話が可能なロボットも存在する²⁾が、ロボットの注視点の管理やタスク知識の構造化など、与えられた文脈内で動作することが前提とされている場合が多く、一般的な拡張が難しい側面があった。

そこで、Human-Robot 間のインタラクションにおいて、時系列的情報・空間的情報の双方を確率的に表現する手法の確立を目指す。環境を記述するモデル、すなわち記憶モデルには、現在目の前に広がっている視野内の視覚情報のみならず、物体の背後に隠れている領域や、過去に経験した状況なども記憶にとどめておく機能が求められる。また、このモデルは、ロボットの行動決定や、人間との対話を推進するための発話行為のプランなどを行う際に有効的に機能できるような設計基準を設ける必要がある。

そこで本研究では、記憶モデル内で扱うべき情報として、確率表現と3次元情報の組み合わせが重要であると考へ、以下の3つの確率的記憶情報を導入する。

1. 存在確率: ロボットがある物体(人間やロボットを含む)を記憶モデルの中で、記憶オブジェクトとして保持しておくべき度合いを示す。同時に、記憶オブジェクトが参照された際に、その存在が行動計画に与える影響の強さを示す。
2. 物体の視覚属性に関する確信度: 物体が持つ視覚属性に関してセンシングの不確かさを表現する確率。具体的には色情報、3次元形状の種類、面積/ボリュームが該当する。物体のみでなく、対人行為のための人間やロボットも対象となる。
3. 3次元空間情報: 実際に行動の対象とする際に行動計画に利用できる形状の物理的情報。上記視覚処理で得られた物体の3次元空間内での位置情報はここで管理される。またロボット自身の身体情報も含み、ロボットと物体の相対的な位置関係の情報も生成する。

物体、人間、視覚情報に関する存在確率と視覚属性に関する確率、3次元的な空間情報の例を図に示す。

2.2 視覚に基づく空間記憶モデルの獲得と更新

空間記憶モデルを構築するには、まず最初にロボットの目の前の状況を記憶に反映させる。ステレオカメラに入力された画像情報から、色情報を元に領域分割を行い、それぞれの領域ごとに、3次元形状情報、3次元空間内での位置などの情報を取得する。記憶オブジェクトにはこれらの視覚特徴情報が付加された上で、空間記憶モデルに登録されて行く。

また実環境を対象とした画像処理では、左右の画像のマッチングミスによる距離情報の不確か性が生じやすい

ため、ステレオマッチング処理の内部段階で不確か性が生じた事を見出した場合には、処理結果に信頼性がないと判断し、物体の視覚プロパティの「確信度」を下げて

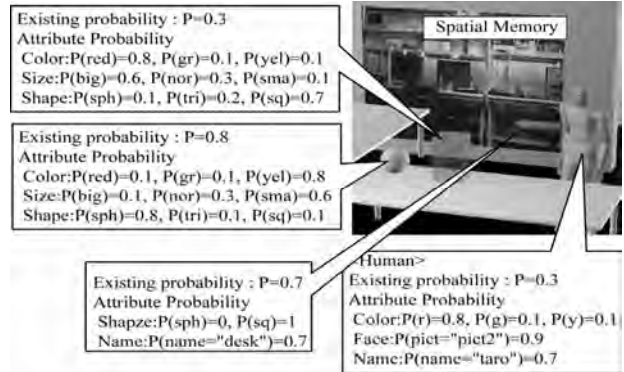


図1. 確率的空間記憶モデルによる記述例

登録をすることとする。この機能により、対象物体を自然言語を用いて参照する際、その視覚特徴量の持つ不確か性を考慮した上で、物体探索を行う事が可能となり、ユーザとの協調的な発話戦略にも活用される。詳細については第3章にて述べる。

2.3 存在確率の更新

次に、時間の経過とともに環境の変動が起こりうることを考慮するために、生成された記憶オブジェクトの持つ存在確率値を操作する。記憶に文脈情報すなわち時間的経過に関する情報を持たせるため、時刻 t における記憶オブジェクト o_i の存在確率 $p_t(o_i)$ を次の式(1)に従って更新する。

$$p_{t+1}(o_i) = p_t(o_i) \{1 - \omega + \alpha \sum_{j, j \neq i} p_t(o_j) R(o_i, o_j)\} \quad (1)$$

ただし ω は、記憶減衰係数 ($0 < \omega < 1$)、 $R(o_i, o_j)$ は、記憶オブジェクト o_i と o_j の間の記憶想起係数 ($0 < R(o_i, o_j) < 1$)、 α は正規化係数、をそれぞれ表す。

ここで、記憶想起係数 $R(o_i, o_j)$ は記憶オブジェクト o_i と o_j の関係性により定義され、一方の記憶オブジェクトが参照・更新された場合に、同時に他方の記憶オブジェクトも参照・更新される度合いを示すものである。定性的な側面を考えた場合、文脈内における相関係数のような役割を果たす。上式はロボットの記憶に対する存在の確信度を表し、忘却と想起・確認によって常に変動することを示している。これにより昔に見たものの確信度は薄れ、関係の深いものによって確信度を上げるような仕組みが実現可能となっている。

3. 協調的行動決定と対話制御への応用

3.1 タスクネットによるタスク遂行モデル

我々は従来までにユーザの曖昧な指示や環境のセンシングによる不確実性を考慮した上で、適切な発話戦略をとる手法を考案している⁹⁾。その手法ではベイジアンネットワークを利用してユーザの曖昧な指示の理解を実現した。一般的に人間が用いる自然言語では省略語が非常に多いが、ロボットにあるタスクを実行させるような状況では、使用される語彙は人名、部屋、物体の名前、タスクの種類などに集約されやすい。そこで、それらのタスクに関係する概念の関係を、個別のタスクごとにネットワーク構造の知識表現としてあらかじめ記述するアプローチを取る。概念間の関係は完全に固定して記述することはできないため、各概念間を確率的に表現しベイジアンネットワークで記述する。これをタスクネットと呼ぶ。図2はタスクネットの一例である。タスクネットの最下端のノードにはカメラから得られた視覚情報やタスクの指示に用いられた語彙が入力され、タスクの種別やタスクの対象となる目的物体などが指示に含まれていない場合、上層部のノードに対する推論を行って不足情報を補う。表1に推論対象となるノードとそのノードの命題値の例を示す。

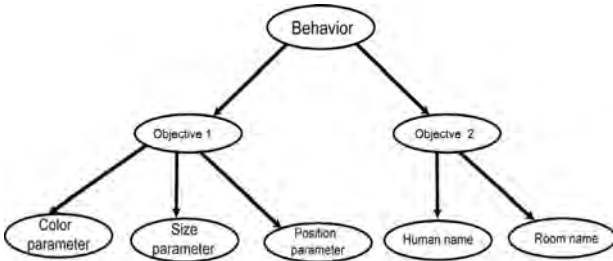


図2. タスク知識表現のためのベイジアンネットワーク

表1. タスクネットのノードに割り当てられる変数

Object 1	None, can, bottle, box,,
Object 2	Trash box, person, room,,
Behavior	Carry, discard, tidy,,

このベイジアンネットワークによるタスク表現により、ユーザが省略した概念を補間して遂行すべきタスクの種類や、そのタスクの遂行のために必要な情報を推測することが可能となる。また、従来研究の発話制御アルゴリズムで用いたベイジアンネットワークをここでも用いることにより、不足情報がある場合には質問を投げかけることで情報収集をするなど、タスクの知識を前提とした発話制御を実現できる。

この対話システムがもつ欠点の一つとして、実世界での物体の位置などの情報を有効に用いることができず、「目の前の情報」を用いるだけだった。そこで、前述した確率的な空間記憶モデルと統合することで、物体の空間内での分布状況、および時系列的な文脈を考慮することが可能となる。

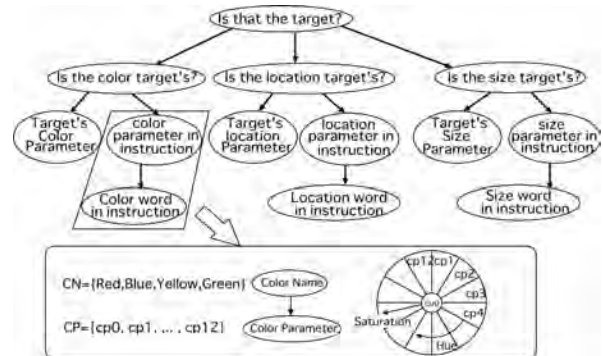


図3. 対象物指定のためのベイジアンネットワーク

3.2 ベイジアンネットワークによる発話制御

タスクネットにおいて不足情報が発生した場合に質問等の発話を行うことを述べたが、本節ではその具体的な手法について述べる。日常生活環境におけるタスクとして、最も頻繁に行われる行為が、物体の参照と同定である。図3は物体の参照行為に使用されるネットワークである。ネットワークを構成するノードは、注目している物体およびその視覚特徴量を意味している。視覚特徴量を表現するノードには、赤、青等の色の名前、大きい、小さい等のサイズの表現などの命題値が対応している。注目候補を表現するノードには、「注目している物体が参照すべき物体であるかどうか」という命題の値が対応する。他にも、「注目している物体の視覚特徴量は参照に用いられた表現と適合しているかどうか」という命題も存在し、推論を行う場合には、前者の注目候補を表現するノードに対する推論が行われ、その値は「物体特定の確信度」と呼ばれる。この値が高ければ、言語によって参照された物体と、現在カメラが注視している物体が同一のものであると判断する。

次に、物体の参照・同定の達成度を評価するパラメータとして、対象の局在性を導入する。対象の局在性とは、注目している対象が他の対象と混同される事なく存在している、すなわち局在している事を言う。この局在性を定式化する際に、この「対象同定の確信度」を用いる。 n 個の対象に対してそれぞれの同定確信度を CF_i とする。局在性が成立するためには、同定確信度が高い対象が唯一存在する必要がある。この判断基準を

$$L \equiv \frac{\max(CF)}{\text{second}(CF)} \geq 1.5 \quad (2)$$

とする。ここで $\max(CF)$ は存在する対象内で最大の同定確信度。 $\text{second}(CF)$ は2番目に大きい同定確信度の値とする。つまり、最大の同定確信度が、次点の確信度の1.5倍以上あれば、それが突出して高い同定確信度を持っていると判断する事にする。式(2)の L を、参照物同定タスクの達成度と呼ぶ事にする。この L が高い値になる

ように発話を計画することで、適切な発話を行う事が可能となる。

図4に、環境の観測モジュール、発話制御モジュール、行動決定モジュールの統合モデルを示す。このように、タスクネットによる、タスクに必要な概念間の関係性記述、対象物体の参照・同定確信度の導入によって、協調的にタスクを遂行する事が可能となる。

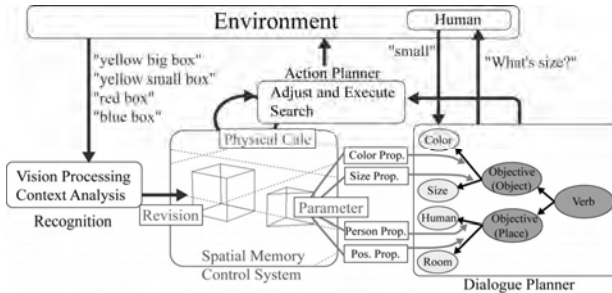


図4. 環境記憶、発話制御、行動決定モジュールの統合

4. 確率的記憶モデルに基づく協調的タスク遂行実験

以上で述べた記憶システムと行動計画システムを利用し、ヒューマノイドロボットがユーザとのインタラクションを介してタスクを実行する実験について述べる。

4.1 記憶による文脈の把握と行動制御実験

この実験では、人間がロボットにある物体を運んでくるタスクを指示する状況を取り扱う。ロボットは部屋の中を動き回るので、部屋の中のどの位置のどの物体が、ユーザの支持した物体であるのか、ということ判断する必要が生じる。

仮想環境における実験の流れを図5および図6に示す。図6左側にある、数字付きの大きな映像は確率的空間記憶モデルの状態を示している。右側の4分割されている映像は、実際の世界の状況、およびロボットが取得しているカメラ映像をそれぞれ意味している。

実験の流れは以下の通りである。まずロボットがシーン1においてサイドテーブルの上に視線を向け、黄色い物体一つを発見して記憶モデルに登録した。(実際には二つの物体があるが、ロボットの視野から外れてしまい、一つのみが認識された) 次にシーン2でユーザの目の前まで移動したロボットは、「黄色い箱を持って来て」という依頼を受けた。ロボットは空間記憶モデルを参照し、通過した箇所に戻った(シーン3)。しかしここで当初は発見していなかった二つ目の「黄色い箱」の存在に気がつき、記憶モデルに登録すると共に、タスク遂行のためのユーザへの質問生成を行い「大きい方ですか?」という質問を通じて、適切な物体を選択し運ぶ行動を実現した。

図5右部の円柱で表されるものは、確率的記憶の推移

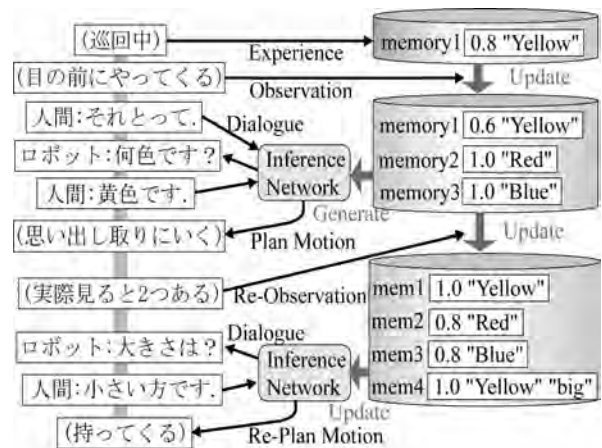


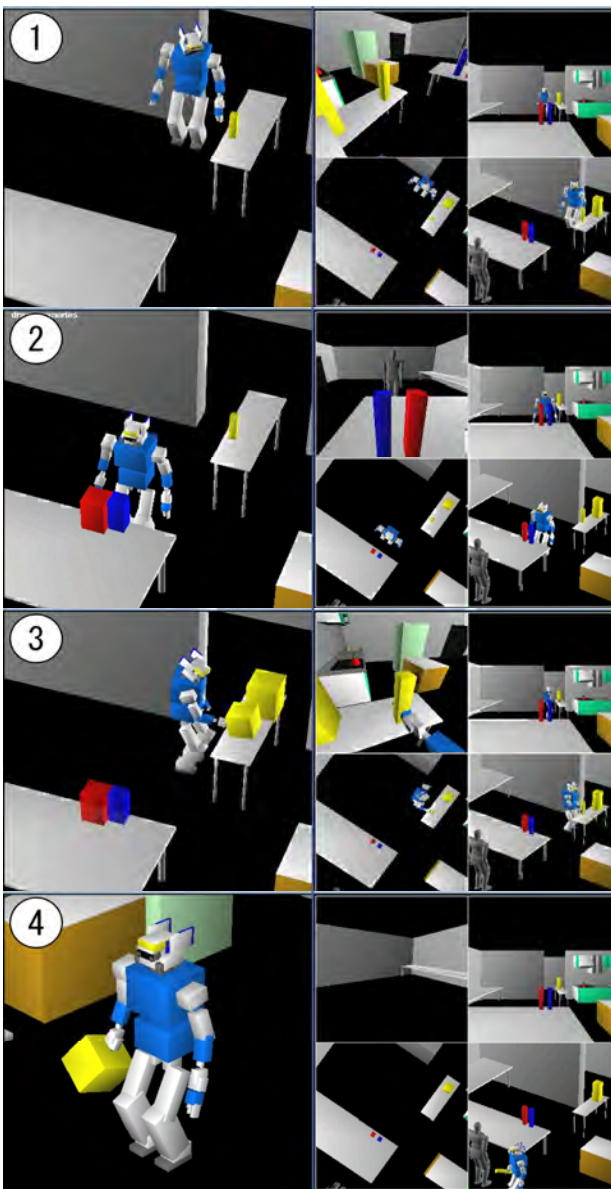
図5. 確率的記憶モデルを用いた対話制御の流れ

の状況である。時間と共にロボットが発見した物体に関する情報は記憶に蓄積され、基本的には時間の経過とともに物体の存在確率が減少していく。ロボットの視野に物体が入り、記憶と同じ場所に存在することが確認されれば存在確率は1.0に更新される。また、視覚情報として観測されなかったとしても、その物体の事を参照するような発話をユーザが行った場合、その物体の存在確率は上昇する。

4.2 日常生活の定常的な状況の表現と利用

次に、単に言語的な表現から対象物を特定するだけでなく、日常生活における物体の役割や意味を記憶させる実験例について述べる。ここでは缶とペットボトルを分別して専用のゴミ箱に捨てるタスクについて考え、目の前にある二つの物体(ゴミ箱)について、缶用のゴミ箱とペットボトル用のゴミ箱を判断するために、図Xのようなネットワークを用いる。ロボットは画像情報から得られた物体の色のRGB値を離散化した上で、ネットワークのRed, Blue, Greenノードに入力し、さらに物体の位置および大きさの情報もPosition, Sizeノードに入力する。色情報は色のシンボルに変換され、最終的にTrash Box Typeノードを対象とした推論を行うことで、缶用のゴミ箱かペットボトル用のゴミ箱かという判断を下す。

図7は、ペットボトルと缶のゴミを分別して廃棄するという状況における実験結果を示したものである。ここでは赤い缶を捨てる行動を取ろうとしており、缶用のゴミ箱(黄色)をさがすことになる。始め、ロボットは確率的空間モデルを参照し、缶用のゴミ箱がロボットの左手にある情報を入手した。シーン(1)において、ロボットは左手に視野を向け、一つの物体を発見する。次に、図8のベイジアンネットワークを用いてこのゴミ箱が缶用ではなく、ペットボトル用のゴミ箱であると判断した。そこで、缶用のゴミ箱を探すため、確率的空間モデルを参照し、



Spatial Memory Model	Vision Image	Real World(1)
	Real World(2)	Real World(3)

図6. 確率的記憶モデルを用いた対話行動実験

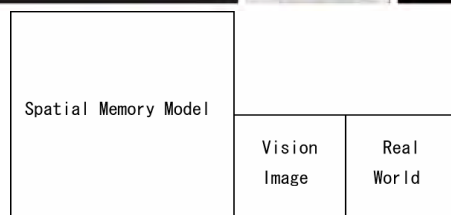
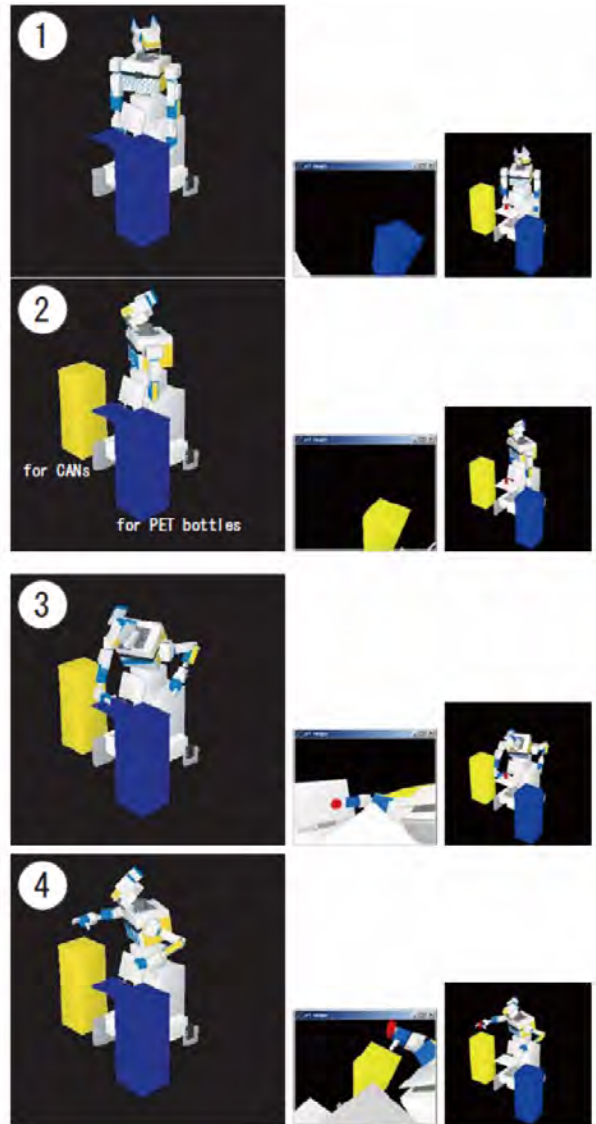


図7. 確率的記憶モデルを用いた対話行動実験

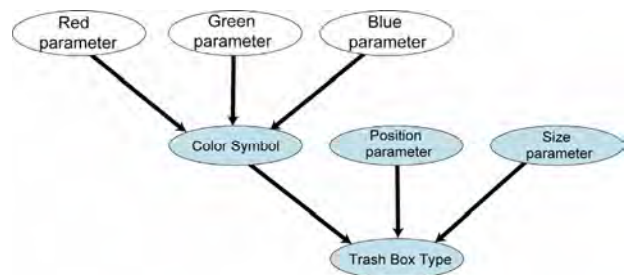


図8. ゴミ箱の種別認識用ネットワーク

まだ物体の存在確率が低い領域が右手方向であることを利用し、視野を右手に移した。そこで新たな物体を一つ発見し、空間記憶モデルに登録しつつ、ペットボトル用のゴミ箱であると認識した。最終的にこの黄色いゴミ箱に缶を捨てる行動を実現した。

5. おわりに

本稿では常に変わり続ける環境の中でロボットが状況と文脈を考慮して対話に基づいて人間の指示を理解することを目的とし、ロボットが持つべき内部表現の仕組みとこれらを確率的に管理する手法を提案した。この仕組みを用いて、ロボットが常に変化する環境の中で物体の存在を確率的にとらえることで時間的な軸に沿った情報の扱いを可能にし、自身の経験とその場の状況に基づいた推論が実現可能になった。

この枠組みは、多岐に渡った状況に応用可能である。例えば、今回の物体とロボットの関係だけでなく、人間を観察するロボットの認識モジュールとしても活用可能である。特に、指さし動作をしながら物体を指示し、その物体を判別して把持を行うような状況にはこの記憶モデルが有効となる。例えば第4章の実験で、ゴミ箱の位置を探索する例について述べたが、ユーザがそばにいる場合、指さし動作によって捨てるべきゴミの物体を指し示すジェスチャーを理解すると効率的に状況判断できる。しかし、指さしの方向をビジョンで精度良く認識するのは難しく、不確実性を考慮して3次元空間内の方向や対象物体の探索を行う必要がある。この際に確率的な表現方法を用いる手法¹³⁾も考案しており、容易に本手法と統合を行うことが可能である。

このように、空間記憶モデルは、人間とロボットが日常生活空間において共存し、フレンドリなユーザインタフェースを提供するために必要不可欠な要素技術となる。今後は発話戦略の決定や行動決定モジュールに対して、より効果的な機能を提供するための確率的表現・推論の手法について応用研究を進めて行く予定である。

謝辞 本研究は、東京大学大学院情報理工学系研究科稲葉雅幸教授および岡田慧講師のご協力のもとに行った。

引用文献

- 1) Y. K. Hwang, P. C. Chen, and P. A. Watterberg. Interactive Task Planning through Natural Language. In Proceedings of the 1996 IEEE International Conference on Robotics & Automation, pages 24--29, Apr 1996.
- 2) R. Bischoff and V. Graefe. Hermes - an intelligent humanoid robot, designed and tested for dependability. Experimental Robotics VIII, B. Siciliano and P. Dario (eds), springer tracts in advanced robotics 5, Springer, pages 64--74, 2003.
- 3) Z. Mohd Hanafiah, C. Yamazaki, A. Nakamura, and Y. Kuno. "Human-robot speech interface understanding inexplicit utterances using vision". In CHI '04 extended abstracts on Human factors in computing systems, pages 1321--1324, 2004.
- 4) A. Knoll, B. Hildebrandt, and J. Zhang. Instructing Cooperating Assembly Robots through Situated Dialogues in Natural Language. In Proceedings of the 1997 IEEE Int'l Conf. on Robotics & Automation, pages 888--894, 1997.
- 5) J. Ido, Y. Matsumoto, T. Ogasawara, and R. Nisimura. "Humanoid with interaction ability using vision and speech information". In IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems, 2006.
- 6) S. Fujie, K. Fukushima, and T. Kobayashi. "A conversation robot with back-channel feedback function based on linguistic and nonlinguistic information". In Proc. Intl. Conf. on Autonomous Robots and Agents, pp. 379--384, 2004.
- 7) Terry Winograd and Fernando Flores. Understanding Computers and Cognition : A New Foundation for Design. Addison-Wesley, 1987.
- 8) T. Inamura, M. Inaba, and H. Inoue. Dialogue control for task achievement based on evaluation of situational vagueness and stochastic representation of experiences. In In the Proc. of Int'l Conf. on Intelligent Robots and Systems, pages 2861--2866, 2004.
- 9) T. Inamura, M. Inaba, and H. Inoue. "Pexis : Probabilistic experience representation based adaptive interaction system for personal robots," Systems and Computers in Japan, 35(6):98--109, 2004.
- 10) Y. Wakamatsu, T. Kondo, and K. Ito. "A proposal of communication design for continual human-agent interaction using natural utterances," In SICE Annual Conference, pages 2315--2319, 2003.
- 11) Y. Shinyama, T. Tokunaga, and H. Tanaka. "kairai" - software robots understanding natural language. In Proc. of 3rd International Workshop on Human Computer Conversation, pages 158--163, 2000.
- 12) E. Stopp, K. Peter Gapp, G. Herzog, T. Laengle, and T. C. Lueth. "Utilizaing Spatial Relations for Natural Language Access to an Autonomous Mobile Robot," In The 18th German Annual Conference on Artificial Intelligence, pp. 39--50, 1994.
- 13) 大久保, 岡田, 稲邑, 稲葉. ロボットの日常生活支援行動のオンサイト教示のための指差しジェスチャー認識. In 電子情報通信学会技術研究報告 信学技報, volume 105, pp. 11--16, 2005. PRMU2005-65.