

# 人間とロボットの安全・安心なインタラクションのための動作の言語化技術の研究

## —Verbalization of Motion for Secure Human-Robot Interaction—

1081004 (登録番号)

研究者代表 (独)情報通信研究機構知識・創成コミュニケーション研究センター  
音声コミュニケーショングループ 専攻研究員 岩橋 直人  
共同研究者 (独)情報通信研究機構知識・創成コミュニケーション研究センター  
音声コミュニケーショングループ 専攻研究員 杉浦 孔明

### [研究の目的]

介護や福祉などの生活支援ロボットにおいて、ユーザが自然な発話によって安心して動作を指示できる機能の実現が望まれている。しかし、ロボットが発話の意味を実世界状況に依存して適切に理解できずにユーザが予期しない動作を行ってしまう危険性があることが問題であった。本研究では、この問題を解決するため、実際に動作を実行する前に対話により指示動作を明確化するための発話生成技術の開発を目的とする。

### [研究の内容、成果]

#### 1. タスク設定

具体的なタスクとしては、ユーザが発話によりロボットにオブジェクトを操作させるタスク(物体操作指示)を対象とする。物体操作指示において、ユーザの発話の意味が適切に理解されるためには、(1) 言語による動作参照、(2) 言語によるオブジェクト参照、における曖昧性を解消する必要がある。(1)(2)の曖昧性をともに解消する研究は今までにない。一方我々は、実世界にグラウンドした動作のイメージをユーザとロボットが共有する手法 LCore を提案している。よって LCore に確信度に基づく確認発話生成を導入することで、上記の危険性を解消できる可能性がある。本稿では、ユーザの発話の曖昧性を定量化し、タスク達成の効用を最大化する応答(動作あるいは確認発話)を生成する手法 LCore-DEC を提案する。提案手法の独自性は、(1) 音声・視覚・動作などを統合したユーザモデルに対

する確信度を、統計的学習手法に基づいて学習すること、(2) 確信度を用いて過不足ない自然な確認発話を生成すること、である。さらに、ユーザの動作指示発話を受けてロボット(図1参照)が行動する場合には、動作の確率モデルをもとに、状況に応じて最も適切な動作軌道を計画する。この動作はユーザの教示によりロボットが学習したものであるため、ユーザにとってロボットの動作がイメージしやすい。

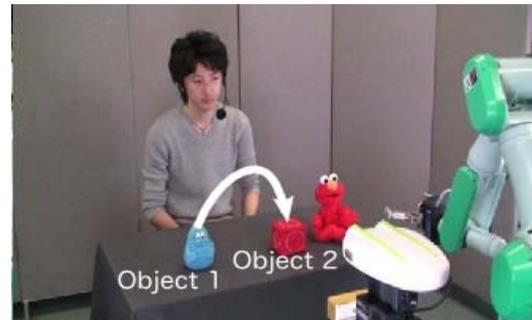


図1 物体操作対話の例

#### 2. 共有信念モデルに基づく発話理解

提案手法では、マルチモーダル入力から学習されたユーザモデルを用いてユーザの発話を理解する。さらにこのユーザモデルを用いて、適切な行動や発話を生成する。本稿では、音声・画像・動作などの各モダリティに対応するユーザモデルを信念モジュールと呼ぶ。また、(1) 音声、(2) 動作、(3) 視覚、(4) 動作-オブジェクト関係、(5) 行動コンテキスト、の5つの信念モジュールを統合したユーザモデルを共有信念関数  $\Psi$  で表す。信念モジュールと共有信念関数  $\Psi$  は、教師あり学習の枠組みにより学習される。教師データは、ユーザとロボットの

実世界インタラクションを通じて収集される。

ユーザの発話  $s$  は、トラジェクタ（動かされるオブジェクト）を表す文節  $W_T$ ，ランドマーク（動作の規準となるオブジェクト）を表す文節  $W_L$ ，動作を表す文節  $W_M$  からなる概念構造  $z = (W_T, W_L, W_M)$  と対応づけて解釈される。例えば，図 1 に示すシーンにおいて，ユーザが「バーバブライト，赤い箱のせて」と発話したとする。このとき，正しく分割された文節は以下のようなになる。

$W_T$ : バーバブライト,  $W_L$ : 赤い箱,  $W_M$ : のせて

また，ランドマークを必要としない動作概念では， $z = (W_T, W_M)$  である。いま，シーン  $O$  において発話  $s$  が与えられたとしよう。 $O$  は，カメラ画像中の全オブジェクトの画像特徴量および位置を表す。物体操作対話タスクでは， $s$  に対しシーン  $O$  において正しい行動  $a$  を選択することが求められる。

### 3. 発話理解確信度の推定に基づく発話と動作の生成

#### 3.1. 全体確信度関数による発話理解確率のモデル化

提案手法は，行動を指示する発話の曖昧性を定量化し，あらかじめ定義された効用を最大化する応答（動作あるいは確認発話）を生成するものである。ここでは，まず共有信念関数  $\Psi$  に基づく曖昧性の尺度について説明する。

共有信念関数  $\Psi$  を用いると，コンテキスト  $q$ ，シーン  $O$ ，発話  $s$  が与えられたときの最適行動  $\hat{a}$  は以下で得られる。

$$\hat{a} = \arg \max_a \Psi(s, a, O, q)$$

最適行動  $\hat{a}$  のマージンを以下の関数  $d$  により定義する。

$$d(s, \hat{a}, O, q) = \Psi(s, \hat{a}, O, q) - \max_{A \neq \hat{a}} \Psi(s, A, O, q)$$

ここで，マージンを用いて  $\hat{a}$  に対する確信度を得ることを考える。提案手法では，全体確

信度関数  $f(d)$  をシグモイド関数を用いて以下のように定義する。

$$f(d) = \frac{1}{1 + e^{-(w_1 d + w_0)}}$$

この  $f(d)$  により， $d$  のもとで発話が正しく理解される確率をモデル化する。 $f(d)$  の形から  $0 < f(d) < 1$  であり，マージン  $d$  が大きいほど  $f(d)$  が 1 に近づくことがわかる。

#### 3.2. 全体確信度関数の学習

マージンと正解ラベルを学習サンプルとして，ロジスティック回帰により  $f(d)$  のパラメータ  $\mathbf{w} = \{w_0, w_1\}$  を推定することを考える。学習サンプル集合を入力  $d_i$  と教師信号  $u_i$  の組  $\{(d_i, u_i) | i = 1, \dots, N\}$  として与える。ここで， $u_i$  は最適行動  $\hat{a}_i$  が正解なら 1，不正解なら 0 の 2 値をとる。いま，入力  $d_i$  を与えたときの出力  $f(d_i)$  を，入力  $d_i$  のもとで教師信号  $u_i$  が 1 である確率の推定値であると考え。このとき，学習サンプル集合に対する全体確信度関数の対数尤度は，以下で与えられる。

$$L = \sum_{i=1}^N \{u_i \log f(d_i) + (1 - u_i) \log(1 - f(d_i))\}$$

全体確信度関数  $\Psi$  の学習には，尤度最大化規準により行われる。

#### 3.3. 期待効用最大化に基づく応答の決定

ユーザの発話  $s$  に対してロボットが行った動作応答が，ユーザがロボットに行わせたい行動と異なることは，安全性の観点から望ましくない。これに対し，全体確信度を用いれば，このような危険を回避できる可能性がある。例えば，発話  $s$  に対する最適行動  $\hat{a}$  の全体確信度が小さければ，ユーザに  $\hat{a}$  を行うか否かを確認する発話をすればよい。本節では，応答に対する効用を導入し，これを最大化する応答（最適応答）に

関する意志決定を行わせる手法について述べる。いま応答として，動作応答  $b_1$  と確認発話応答  $b_2$  があ

るとする。統合確信度  $f(d)$  は，マージン  $d$  のもとで発話が正しく理解される確率をモデル化

するものであったから、応答 $b_i (i = 1, 2)$ に対する期待効用 $\mathbf{E}[R_i]$  および最適応答 $\hat{b}$ を以下のように推定することができる。

$$E[R_i] = r_{i1}f(d) + r_{i2}(1 - f(d))$$

$$\hat{b} = \arg \max_b E[R_i]$$

ただし、 $r_{i1}$ ,  $r_{i2}$  はそれぞれ、 $\hat{a}$  が正解、不正解であったときの応答 $b_i$ に対する効用である。図2右図に効用をまとめた。ここで、 $r_{12} < r_{22} < r_{21} < r_{11}$  であるとする。 $\mathbf{E}[R_i]$  は $f(d)$ の線形関数であるので、このとき等式 $\mathbf{E}[R_1] = \mathbf{E}[R_2]$ は $0 < \theta_0 < 1$ なる解 $\theta_0$ を持つ(図2参照)。つまり、 $\theta_0$ を閾値として最適応答が選択できる。

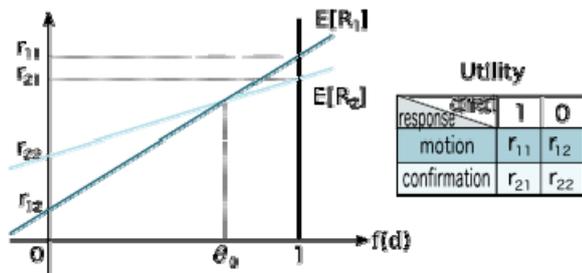


図2 全体確信度と期待効用の関係

次に、確認発話において、共有信念として学習されたユーザモデルを言語表現の生成に用いることを考える。例えば食器が複数ある状況では、「四角くて白い皿」のように最も曖昧性が減少し、かつ冗長でない表現でオブジェクトを表現できることが望ましい。提案手法では、ユーザの発話に対しマージンを最大化する単語を加えることで曖昧性を減少させる。ただし、加える単語数は、 $f(d) \geq \theta_0$ を満たす最小の単語数とする。

#### 4. 実験に用いたハードウェア

実験に用いたロボットシステムを図1に示す。ロボットシステムは、7自由度のロボットアーム、4自由度のロボットハンド、マイクロフォン、ステレオカメラ、視線表出ユニットからなる。

オブジェクトに関する画像特徴や座標は、ステ

レオカメラから得られた画像から抽出される。カメラのフレームレートを30[frame/sec]とし、解像度を320×240とした。画像特徴量として、色3次元(L\*a\*b\*)、形状3次元を用いる。

#### 5. 実験結果

全体確信度関数の学習に関する定性的結果を図3示す。図におけるそれぞれの曲線は、訓練サンプル数を変えたときの回帰結果である。図より、サンプル数20までに収束していることがわかる。

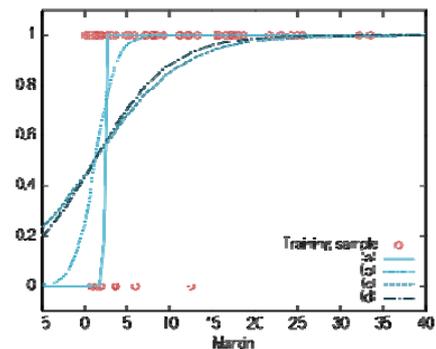
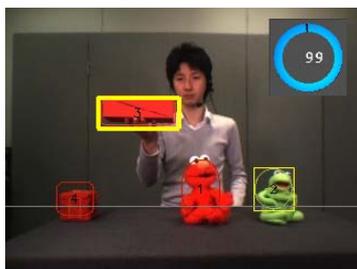


図3 全体確信度関数の学習結果。○は訓練サンプルを表す。(a)～(d)は、それぞれ訓練サンプル数10, 15, 20, 25のときの結果に対応する。

次に、定性的な結果について述べる。図4, 5にユーザ(U)とロボット(R)の対話例を示す。これらはともに $\theta_0 = 0.7$ と設定したときに得られたものである。各図において右上の数値は全体確信度を表す。図4では、ユーザはトラジェクタおよびランドマークについて発話しなかったものの、ロボットの動作は正しいもの(「オブジェクト2をオブジェクト3に載せる」)であった。この理由は以下のように考えられる。図4に示すシーンでは、正解動作の軌道に対する尤度(動作信念モジュールから得られる尤度)は比較的小さい。具体的には、正解動作軌道の尤度は、動作候補60通りのうち24位であった。しかしながら、動作-オブジェクト関係の信念モジュールと、コンテキスト信念モジュールから得られるスコアを加えることで、正解動作に対するスコアが動作候補のなかで

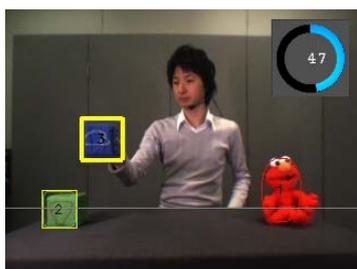
最大になった. さらに, 確信度  $f(d) = 0.998 > \theta_0$  であったので, 動作を実行する効用が確認発話をする効用を上回り, 動作を実行する意志決定が行われた. 図5では, 最適行動の確信度は  $f(d) = 0.47813 < \theta_0$  であった. よって確認発話が最適応答であり, 「アオイハコ」という言語表現が生成された. この言語表現は, オブジェクト2と3の視覚的特徴のなかで最も異なる属性について述べており, ユーザにとって理解しやすい.

実験結果より, 曖昧性が大きい発話に対しては, ユーザにとって自然な確認発話を生成することで, 不適切な動作を実行前に中止させて実行失敗率を減少させることが可能になったことが示された.



【状況】オブジェクト2が直前に操作された  
U: のせて  
R: (動作実行: オブジェクト2を3に載せる)

図4 対話例(1) 確認発話を行わなかったケース



【状況】オブジェクト2が直前に操作された  
U: ハコ エルモ ちかづけて.  
R: ミドリハコをちかづけて?  
U: いいえ.  
R: アオイハコをちかづけて?  
U: はい.  
R: (動作実行: オブジェクト3を1に近づける)

図5 対話例(2) 動作実行前に確認発話を行ったケース.

スクとしたが, より複雑で実用的な部屋の空間などでの物体操作や行動支援をタスクとして, 提案手法の有効性の検証を行ってゆく予定である. その場合, より高度な画像処理技術の開発が必要になる. さらに, ユーザの行動も複雑になるので, ユーザモデルや全体確信度関数を, いかにか効率的に学習するかが課題となる.

[成果の発表、論文等]

1. Sugiura, K., Iwahashi, N.: Motion Recognition and Generation by Combining Reference-Point-Dependent Probabilistic Models, *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 852-857, (2008)
2. 中村慎也, 岩橋直人, 長井隆行: 実世界における人とロボットの共有信念に基づいた相互適応的な発話生成: 電子情報通信学会音声研究会, IEICE-SP-338, pp.91-96, (2008)
3. 杉浦孔明, 岩橋直人: 物体操作タスクにおける発話理解確信度の推定に基づく発話と動作の生成, 計測自動制御学会第36回知能システムシンポジウム予稿集, pp.29-34, (2009)
4. 中村慎也, 岩橋直人, 長井隆行: 人とロボットの共有信念に基づいた発話の生成, 電子情報通信学会総合大会, p.109, (2009)
5. 岩橋直人, 中村慎也, 長井隆行: 人間とロボットの共有信念の推定に基づいた相互適応的な発話生成手法, 人工知能学会全国大会予稿集, (2009発表予定)
6. 杉浦孔明, 岩橋直人: 対話戦力最適化としての発話理解確信度の能動学習, 人工知能学会全国大会予稿集, (2009発表予定)
7. 岩橋直人, 中村慎也, 長井隆行: 人とロボットの共有信念に基づいた相互適応的な発話生成, ファジーシステムシンポジウム, (2009発表予定)

[今後の研究の方向、課題]

本研究では, 机の上の空間での物体操作をタ

