

# 物体操作対話タスクにおける能動学習に基づく対話戦略最適化

○杉浦孔明, 岩橋直人, 河井恒, 中村哲 (情報通信研究機構)

## Dialogue Strategy Optimization Based on Active Learning in Object Manipulation Dialogue Tasks

\*Komei Sugiura, Naoto Iwahashi, Hisashi Kawai, and Satoshi Nakamura  
National Institute of Information and Communications Technology

**Abstract**— In an object manipulation dialogue, a robot may misunderstand an ambiguous command from a user, such as “Place the cup down (on the table),” potentially resulting in an accident. Although making confirmation questions before all motion will decrease the risk of this failure, the user will find it more convenient if confirmation questions are not made under trivial situations. This paper proposes a method for estimating ambiguity in the commands by introducing an active learning framework with Bayesian logistic regression to human-robot spoken dialogue. We conducted physical experiments in which a user and a manipulator-based robot communicated in spoken language to manipulate toys.

**Key Words:** active learning, Bayesian logistic regression, confidence, robot language acquisition

### 1. はじめに

多くのロボットの対話処理機構では、動作コマンドの伝達を目的としているにも関わらず、動作と音声は別々に処理されている。ユーザの発話の意味は動作にグラウンドされない知識に基づいて解釈されるため、動作が状況にふさわしいかどうかは音声認識時には考慮されない。そのため、ロボットが予期しない動作を行ってしまう危険性がある。

本研究では、この危険性を減少させることを目的とする。具体的なタスクとして、物体操作対話タスクを扱う。物体操作対話タスクとは、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を指す。物体操作対話タスクにおいて、ユーザの発話の意味が適切に理解されるためには、(1) 言語によるオブジェクト参照、(2) 言語による動作参照、における曖昧性を解消する必要がある。

(1) の曖昧性解消に対しては人工知能や自然言語生成の分野で多くの研究が行われてきた (例えば [4])。これに対し、(2) の曖昧性解消と関連が深い研究としては、動作の言語化を目指す試みが近年注目されてきている [9]。高野らは、運動の分節化を通じてヒューマノイドロボットが獲得した原始シンボルを用いて、運動認識・生成を行っている [8]。Ogata らは、動作系列と記号列の間の多対多対応問題を扱うリカレントニューラルネットワークに基づく手法を提案している [3]。一方、我々は (1)(2) の曖昧性を解消するアプローチとして、言語獲得フレームワーク LCore を提案している [2]。

本論文では、LCore を拡張し、能動学習の枠組みにより確信度を学習する手法 LCore-AL を提案する。今まで我々が提案してきた手法 [7] では、ユーザの発話を受けてロボットが行動し、その結果をもとに確信度の学習を行っていたため、学習フェーズで行動失敗 (負事例) を行わせて、過学習を防ぐ必要があった。しかし、学習フェーズにおける行動失敗であってもハードウェアの故障などを引き起こす可能性があるため、ユーザの予期しないような行動失敗は減少させることが望ましい。これに対し本手法では、確信度関数の学習に必要



Fig.1 物体操作対話の例

な発話をロボットに選択させ、その発話を受けてユーザが物体を操作する。発話の選択には、能動学習 [10] の一種である Expected Log Loss Reduction (ELLR) [5] を用いる。

提案手法の独自性は、以下の 2 点である。

- 対話システムへの能動学習の導入
- 確信度の学習に Bayesian Logistic Regression (BLR) [1] を用いる

### 2. タスク環境

#### 2-1 物体操作対話タスク

本研究では、ロボット/ユーザが相手の発話にしたがって、テーブル上のオブジェクトを操作するような状況を想定する。ユーザは Fig. 1 のようにロボットと向かい合って座るものとする。実験に用いるオブジェクトを Fig. 2 左図に示す。

オブジェクトの名前、動作、文法などは言語獲得手法 LCore [2] を用いて、あらかじめ学習されているものとする。LCore では、オブジェクトやその動かし方に関する非言語知識と、単語や文法 (単語や節の並び) など言語知識がモデル化されている。このモデルのパラメータは、ユーザとのインタラクションから学習サンプルを得ることで推定される。

タスクは以下の三つのフェーズからなる。

1. ロボット発話-ユーザ行動フェーズ  
ロボットがユーザに指示を与える。ユーザの行動結果により確信度関数を学習する。
2. ユーザ発話-ロボット行動フェーズ

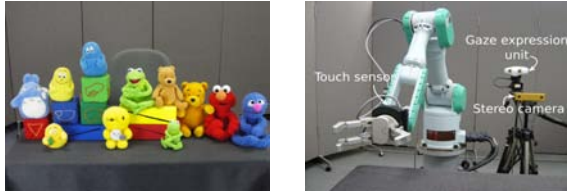


Fig.2 実験に用いたオブジェクト (左) およびロボット (右)

ユーザがロボットに指示を与える。ロボットの行動結果により確信度関数を学習する。

### 3. 実行フェーズ

ユーザがロボットに指示を与える。実験結果は、[7]を参照されたい。

## 2.2 ロボットシステム

実験に用いたロボットシステムを Fig. 2 右図に示す。ロボットシステムは、7 自由度のロボットアーム (三菱重工製 PA-10)、4 自由度のロボットハンド (Barrett Technology 製 BarrettHand)、マイクロフォン、ステレオカメラ (Point Grey Research 製 Bumblebee 2)、視線表出ユニット (Directed Perception 製 PTU-46-70 にロボットヘッドを取付) からなる。

オブジェクトの画像特徴および2次元座標 (カメラ座標系) は、固定されたステレオカメラを用いて得る。なお、オブジェクトの抽出およびトラッキングは、色およびステレオカメラから得られる距離に基づくヒューリスティックな手法により行う。カメラのフレームレートは 30[frame/sec] であり、解像度は 320×240 である。画像特徴量として、色 3 次元 ( $L*a*b^*$ )、形状 3 次元 (オブジェクト領域  $f_{area}$ 、四角らしさ  $f_{sq}$ 、縦横比  $f_{whr}$ ) を用いる。

## 3. LCore における発話理解

LCore [2] 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。本論文では、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  $\Psi$  と呼ぶ。

### 3.1 LCore における動作生成

LCore による動作生成では、[6] で提案した参照点に依存した隠れマルコフモデル (HMM) に基づく手法を用いる。この手法では、物体操作軌道がトラジェクタ (動かされるオブジェクト) と参照オブジェクトとの相対軌道としてモデル化される。参照オブジェクトとは、動作の基準となるオブジェクトのことを指し、トラジェクタそのもの、あるいはランドマーク (トラジェクタの動きの基準となるオブジェクト) から選択される。いま Fig. 1 において、ユーザがロボットに対して、オブジェクト 1 (バーバブライト) をオブジェクト 2 (赤い箱) にのせるように指示したとする。この場合、トラジェクタはオブジェクト 1、参照オブジェクトはオブジェクト 2 である。

### 3.2 共有信念モデルに基づく発話理解

ユーザの発話  $s$  は、概念構造  $z = [(\alpha_1, W_{\alpha_1}), (\alpha_2, W_{\alpha_2}), (\alpha_3, W_{\alpha_3})]$  と対応づけて解釈される。ここに、 $\alpha_i \in \{T, L, M\} (i = 1, 2, 3)$  は文節の属

性を表し、トラジェクタ ( $T$ )、ランドマーク ( $L$ )、動作 ( $M$ ) のいずれかをとることとする。よって、 $W_T, W_L, W_M$  は、それぞれトラジェクタを表す文節、ランドマークを表す文節、動作を表す文節を意味する。例えば、Fig. 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ようになる。

$$[(T, [\text{バーバブライト}]), (L, [\text{赤い, 箱}]), (M, [\text{のせて}])]$$

文節および単語の順序の規則は、統計的言語モデル  $G$  でモデル化される。 $G$  は、属性  $\alpha_i$  の系列の出現頻度と、文節内での単語インデックス列の出現頻度からなる。

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。 $O$  において可能な行動の集合  $A$  は以下により与えられる。

$$A = \{(i, i_r, C_V^{(j)}) \mid i = 1, \dots, O_N, i_r = 1, \dots, R_N, j = 1, \dots, V_N\} \\ \triangleq \{a_k \mid k = 1, 2, \dots, |A|\}, \quad (1)$$

ここに、トラジェクタのインデックスを  $i_t$ 、参照オブジェクトのインデックスを  $i_r$ 、 $O$  中のオブジェクトの数を  $O_N$ 、動作を表す単語数を  $V_N$ 、 $j$  番目の動作モデル  $C_V^{(j)}$  に対して可能な参照オブジェクトの数を  $R_N$  とする。従って、物体操作対話タスクでは、 $s$  に対し正しい行動  $a_k$  を選択することが求められる。以下、本稿では  $a_k$  のように動作インデックスと関連オブジェクトインデックスで表される離散的な表現 (記号レベル) を「行動」と称し、軌道レベルの物体操作を表すために「動作」という表現を用いる。

各信念モジュールを以下のように定義する。まず、音声信念  $B_S$  を、文法  $G$  の下での、発話  $s$  に対する  $z$  の条件付き確率の対数として表す。視覚信念  $B_I$  は、オブジェクト  $i$  の視覚特徴量  $\mathbf{x}_I^{(i)}$  に対する確率モデル (ガウス分布) の対数尤度である。動作信念  $B_M$  は、トラジェクタ  $i_t$  の位置  $\mathbf{x}_p^{(i_t)}$  が与えられたうえでの  $\hat{\gamma}_k$  に対する動作モデルの対数尤度で表される。この動作モデルには参照点に依存した HMM により表現される。ここに、 $\hat{\gamma}_k$  は  $a_k$  に対する最尤軌道を表す。動作モデル  $\lambda$  は、 $\mathbf{x}_p^{(i_t)}$  および、ランドマークの位置  $\mathbf{x}_p^{(i_r)}$ 、動作インデックス  $C_V^{(j)}$  から [6] の手法により得られる。動作-オブジェクト関係信念  $B_R$  は、オブジェクト  $(i, j)$  の視覚特徴量に対する確率モデル (ガウス分布) の対数尤度である。行動コンテキスト信念  $B_H$  は、コンテキスト  $\mathbf{q}^{(i)}$  のもとの、指示対象としてのオブジェクト  $i$  の適切さ (スコア) を表す。各信念モジュールの詳細については [2] を参照されたい。

以上より、共有信念関数  $\Psi$  を、各信念モジュールの重み付き和として定義する。

$$\Psi(s, a_k, O, \mathbf{q}^{(i)}) = \\ \max_z \left\{ \begin{aligned} & \gamma_1 \log P(s|z)P(z; G) & [B_S] \\ & + \gamma_2 \left( \log P(\mathbf{x}_I^{(i)} | W_T) + \log P(\mathbf{x}_I^{(i_r)} | W_L) \right) & [B_I] \\ & + \gamma_3 \log P(\hat{\gamma}_k | \mathbf{x}_p^{(i_t)}, \mathbf{x}_p^{(i_r)}, C_V^{(j)}) & [B_M] \\ & + \gamma_4 \log P(\mathbf{x}_I^{(i)}, \mathbf{x}_I^{(i_r)} | C_V^{(j)}) & [B_R] \\ & + \gamma_5 \left( B_H(i_t, \mathbf{q}^{(i)}) + B_H(i_r, \mathbf{q}^{(i_r)}) \right) \end{aligned} \right\}, \quad (2)$$

ここに、 $\mathbf{x}_p^{(i)}$  はオブジェクト  $i$  の位置、 $\gamma = (\gamma_1, \dots, \gamma_5)$  は、各信念に対する重み (非負の実数) を表す。  $\gamma$  の学習には、MCE 学習を用いる。  $\Psi$  により、発話  $s$  と行動  $a_k$  の対応の適切さを評価することができる。

#### 4. 発話理解確信度の能動学習

##### 4.1 統合確信度による発話理解確率のモデル化

前節の共有信念関数を用いると、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。

$$\hat{a} = \underset{k}{\operatorname{argmax}} \Psi(s, a_k, O, \mathbf{q}) \quad (3)$$

行動  $a_j$  と、最適行動  $\hat{a} (k \neq j)$  のマージンを以下の関数  $d$  により定義する。

$$d(s, \hat{a}, O, \mathbf{q}) = \Psi(s, \hat{a}, O, \mathbf{q}) - \max_{j \neq k} \Psi(s, a_j, O, \mathbf{q}) \quad (4)$$

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、統合確信度関数  $f(d)$  をロジスティックシグモイド関数を用いて以下のように定義する。

$$f(d; \mathbf{w}) = \frac{1}{1 + \exp(-(w_1 d + w_0))} \quad (5)$$

ここに、パラメータ  $\mathbf{w} = (w_0, w_1)$  である。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。

##### 4.2 統合確信度関数の学習

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  $f(d; \mathbf{w})$  のパラメータ  $\mathbf{w}$  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  $u_i$  の組として以下のように与える。

$$\mathbb{T}^{(N)} = \{(d_i, u_i) | i = 1, \dots, N\}, \quad (6)$$

ただし、 $u_i$  は 0 (不正解) または 1 (正解) の 2 値であるとする。実験では、 $u_i$  は音声 (「はい」「いいえ」など)、または触覚センサなどを通じて入力される。

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する。本手法では、BLR [1] を用いて  $\mathbf{w}$  の推定を行う。以下のように、 $w_j (j = 0, 1)$  の事前分布として、平均  $m_j$ 、分散  $\tau_j$  のガウス分布を用いる。

$$P(w_j | m_j, \tau_j) = \mathcal{N}(w_j, \tau_j) = \frac{1}{\sqrt{2\pi\tau_j}} \exp \frac{-(w_j - m_j)^2}{2\tau_j}$$

##### 4.3 ELLR に基づく発話選択

提案手法のブロック図を Fig. 3 に示す。はじめに、発話候補生成モジュールにより各  $a_k$  に対応する言語表現が生成され、マージンが計算される。入力されたマージン集合に対して、スコア計算モジュールが ELLR 基準により最適なマージンが選択される。選択されたマージンと結びつけられた発話  $s$  が音声合成され、ユーザに指示が与えられる。真偽判定モジュールはユーザが行った動作が  $s$  に対する正解行動  $\hat{a}$  と等しいか否かを判定する。この判定結果 (0 または 1) は、スコア計算モジュールにおいて、BLR の学習サンプルとして用いられる。

いま、データ集合  $\mathbb{T}^{(N)}$  を用いて訓練された確信度関数を  $\hat{f}^{(N)}(d)$  と表すとする。このとき、対数損失  $L(\mathbb{T}^{(N)})$

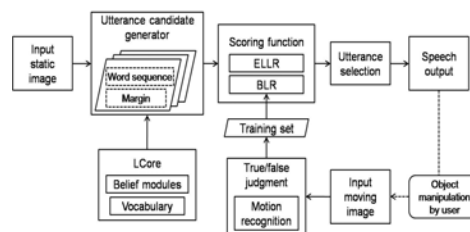


Fig.3 提案手法のブロック図

を以下で定義する。

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

シーン  $O$  で可能な発話を  $V = \{v_j | j = 1, \dots, |V|\}$ 、 $v_j$  から得られるマージンを  $e_j$  とする。  $V$  は、学習済みの単語を用いて生成できる、可能な単語列集合である。ただし、単語列の長さには制約を設ける。このとき、発話  $v_j$  を選択することに対する期待対数損失  $E(\mathbb{T}^{(N)}, e_j)$  は以下ようになる。

$$E(\mathbb{T}^{(N)}, e_j) = \hat{f}^{(N)}(e_j) L(\mathbb{T}_+^{(N+1)}) + (1 - \hat{f}^{(N)}(e_j)) L(\mathbb{T}_-^{(N+1)}),$$

$$\mathbb{T}_+^{(N+1)} \triangleq \mathbb{T}^{(N)} \cup (e_j, 1), \quad \mathbb{T}_-^{(N+1)} \triangleq \mathbb{T}^{(N)} \cup (e_j, 0)$$

つまり、ラベル付けされていない入力  $e_j$  に対して、ラベルが 0 または 1 であった場合の損失を求め、 $N$  サンプル時の確信度 (確率) で重み付ける。提案手法では、上記の  $E(\mathbb{T}^{(N)}, e_j)$  を最小化する  $v_j$  を選択する。

## 5. 実験

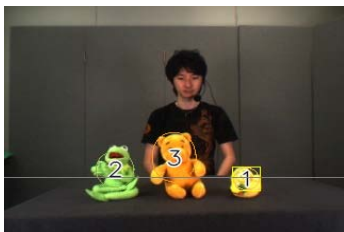
### 5.1 設定

提案手法を評価するために、(1) 確信度関数の能動学習、(2) 事前分布としての学習結果の評価、の 2 種類の実験を行った。実験 (1) の目的は、学習に必要なサンプルを調査することである。また、実験 (2) の目的は、能動学習により得られた確信度関数を、事前知識として用いることの有効性評価である。あらかじめ、[6] で提案した手法により、23 単語 (名詞 8 語、形容詞 8 語、動詞 7 語) を学習させた。

実験 (1) では、提案手法を用いてロボットに発話を行わせ、被験者にオブジェクトを操作させた。「ロボットによる発話、被験者による動作、ロボットによる正解・不正解の評価」を 1 つのエピソードとして、30 エピソードまで確信度関数の学習を行わせた。統合確信度関数のパラメータ  $\mathbf{w}$  に関するハイパーパラメータの設定においては、平均  $m_0, m_1$  が標準ロジスティックシグモイド関数の値、すなわち  $m_0 = 0, m_1 = 1$  になるように設定した。分散に関しては、 $\tau_0 = \tau_1 = 100$  と設定した。 ( $W_T, W_L, W_M$ ) の長さの最大値を (3, 3, 1) とした。

実験 (2) では、最適行動  $\hat{a}$  のマージン  $d$  と  $\hat{a}$  が正解か不正解かを表すラベル  $u$  の組  $(d, u)$  から確信度関数を学習させる。実験 (2) で用いる訓練およびテストデータは以下のように収集した。実験 (1) と同様の実験環境で、被験者にロボットにオブジェクトを操作させるための発話を行わせ、カメラ画像と音声を 60 セット収録した。得られた画像・音声セットに、ユーザが意図した行動を正解としてラベル付けした。収録データのうち半数の 30 個を訓練集合、残りの 30 個を評価集合とした。実験 (1) で得たパラメータを事前分布として訓練を行った場合を提案手法による結果とする。ペー





【状況】オブジェクト 1 が直前に操作された  
 R: プーサンカーミットとびこえさせて  
 U: (動作実行: オブジェクト 3 を持って 2 の上を飛び越えさせる)

Fig.4 学習フェーズにおける対話例.

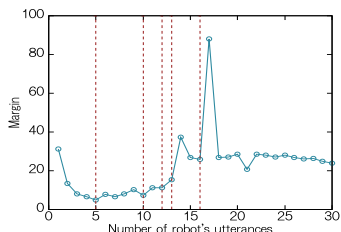


Fig.5 選択された発話の-marginの変化

スラインとしては、パラメータチューニングを行っていないような条件を想定し、初期値を実験 (1) と同じく  $(m_0, m_1, \tau_0, \tau_1) = (0, 1, 100, 100)$  と設定する。

さらに実験 (2) では、学習結果が収束条件を満たすまでに起こったロボットの行動失敗数についても評価を行う。ここで収束条件は、[7] の実験結果からテストセット尤度  $L$  が  $L < -20$  を満たすこととする。また、はじめて収束条件を満たしたエピソード数を  $i_c$  とする。

## 5.2 結果 (1): 確信度関数の能動学習

Fig. 4 に、学習フェーズにおける被験者 (U) とロボット (R) の対話例を示す。このとき、可能な動作とオブジェクトの組は合計 45 種類であった。この中で、「プーサンカーミットとびこえさせて」という単語列でオブジェクト 2 とオブジェクト 3 を表現するものが、対数損失を最小化する発話であった。このときの最適-margin は 13.4 であった。

Fig. 5 は、エピソード数 (ロボットの発話数) に対する、選択された-margin の変化を示す。図において、点線はユーザの動作が失敗したエピソードを示す。図より、ユーザの行動失敗の次のエピソードでは、より大きい-margin を持つ (曖昧性の低い) 発話が選択されていることがわかる。

## 5.3 結果 (2): 事前分布としての有効性評価

Fig. 6 左図は、各手法のテストセット尤度を比較したものである。図において、(1) は能動学習を事前知識とする提案手法を表し、(2) はベースラインを表す。図には、10 種類の異なる訓練集合と評価集合を用いた場合の平均を示した。図より、特に学習の初期において、提案手法が (2) を上回ることがわかる。

さらに、提案手法による尤度が、3 サンプル時にほぼ収束しているのに対し、(2) による尤度が収束するには 10 サンプル程度必要である。訓練集合中の負事例の割合は 18% であったので、収束までの負事例数の期待値は、提案手法が 0.54 であるのに対し、(2) では 1.8 となる。このことから、能動学習で得られた学習結果を事前分布に用いることにより、学習が収束するまでにロボットが動作を失敗する回数の期待値を減少で

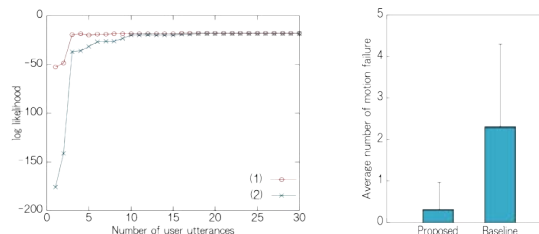


Fig.6 左: 提案手法 (1) とベースライン (2) のテストセット対数尤度. 右: 行動失敗の平均回数.

きることを示唆される。

Fig. 6 右図は、 $i_c$  エピソードまでの行動失敗数の平均値を表す。図より、 $i_c$  エピソードで学習を打ち切れれば、ベースラインに比べて行動失敗を低減できることがわかる。この結果は、提案手法 (ロボット指示による事前学習を行う) が有効であることを示している。

## 6. おわりに

一般の音声対話システムと、音声対話を行うロボットとの違いのひとつは、ロボットがユーザと類似した身体性を有していることである。つまり、既存の対話システムでは「チケットを予約する」などをシステムがユーザに指示しても意味を成さないが、物体操作対話の語彙 (「載せる」など) はユーザ・ロボット共に実行可能である。本研究ではこの点に着目し、ロボットが学習に最適な発話指示を行う手法を実現した。実験結果から、本手法により学習フェーズの行動失敗を低減できることが示された。

## 謝辞

本研究の一部は、日本学術振興会科学研究費補助金 (基盤研究 (C) 課題番号 20500186) および国立情報学研究所による研究助成を受けて実施されたものである。

## 参考文献

- [1] Genkin, A., Lewis, D. and Madigan, D.: Large-scale bayesian logistic regression for text categorization, *Technometrics*, Vol. 49, No. 3, pp. 291–304 (2007).
- [2] Iwahashi, N.: Robots That Learn Language: Developmental Approach to Human-Machine Conversations, *Human-Robot Interaction* (Sanker, N. et al., eds.), I-Tech Education and Publishing, pp. 95–118 (2007).
- [3] Ogata, T., Murase, M., Tani, J., Komatani, K. and Okuno, H. G.: Two-way translation of compound sentences and arm motions by recurrent neural networks, *Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and System*, pp. 1858–1863 (2007).
- [4] Roy, D.: Learning visually grounded words and syntax for a scene description task, *Computer Speech and Language*, Vol. 16, No. 3, pp. 353–385 (2002).
- [5] Roy, N. and McCallum, A.: Toward optimal active learning through sampling estimation of error reduction, *Proceedings of 18th International Conference on Machine Learning*, pp. 441–448 (2001).
- [6] Sugiura, K. and Iwahashi, N.: Learning object-manipulation verbs for human-robot communication, *Proceedings of the 2007 workshop on Multimodal interfaces in semantic interaction*, pp. 32–38 (2007).
- [7] Sugiura, K., Iwahashi, N., Kashioka, H. and Nakamura, S.: Bayesian Learning of Confidence Measure Function for Generation of Utterances and Motions in Object Manipulation Dialogue Task, *Proceedings of Interspeech*, pp. 2483–2486 (2009).
- [8] 高野渉, 中村仁彦: 統計的相関に基づく動作パターンのリアルタイム教師なし分節化と原始シンボルの自律的獲得, *日本ロボット学会誌*, Vol. 27, No. 9, pp. 1046–1057 (2009).
- [9] 谷口忠大: コミュニケーションするロボットは創れるか - 記号創発システムへの構成論的アプローチ, NTT 出版 (2010).
- [10] 渡辺澄夫, 萩原克幸, 赤穂昭太郎, 本村陽一, 福水健次, 岡田真人, 青柳美輝: 学習システムの理論と実現, 森北出版株式会社 (2005).