

連続的相互作用に埋め込まれた 特徴的動作系列の発見的模倣学習手法

Imitation learning method for characteristic motion
embedded in unsegmented human-robot interaction

谷口忠大*1*3
Tadahiro Taniguchi

岩橋直人*2*3
Naoto Iwahashi

*1 立命館大学
Ritsumeikan University

*2 ATR
ATR

*3 NiCT 情報通信研究機構
National Institute of Information and Communications Technology

This paper presents a novel method of a robot learning through imitation to acquire a user's key motions automatically. The learning architecture mainly consists of three learning modules: the switching autoregressive model (SARM), a keyword extractor without a dictionary, and a keyword selection filter that refers to the tutor's reactions. In most previous research on methods of imitation learning by autonomous robots, target motions that were given to robots, were segmented into several meaningful parts by the experimenters in advance. However, to imitate certain behaviors through the continuous interaction with a person, the robot needs to find segments that should be learned. In our experiment, an experimenter displayed several unsegmented motions to a robot, and the experimenter reacted to the motions the robot had acquired. The results revealed that the framework enabled the robot to obtain several meaningful motions that the experimenter expected the robot would acquire.

1. はじめに

将来的に家庭やオフィスといった人間環境下で活動する自律ロボットを実現するためには、各自律ロボットが物理的、社会的活動環境における多様性に適応することが求められる。その一方で、その適応プロセスがユーザの手による意識的な教示過程を経るものであれば、ユーザにロボットへの教示という新たなタスクを背負わせる事になりユーザの負荷をより増すことになる。これに対し、人間の子供は親の日常的な動作を観察し、自らも親に働きかける中で「勝手に」模倣する対象動作を発見的に見出し、多様なジェスチャーや動作パターンを学習することができるという高度な適応性を有する。このような学習過程を計算論的に実現することは、自律的に多様な動作パターンを学習できるロボットを作るといった目的のみならず、人間の持つ模倣学習能力を理解する点においても重要であろう。

一方で、現在までのロボットの模倣学習についての多くの研究では、学習対象となる行動は事前に実験者により分節化された意味ある単位として与えられていた。分節化されていない連続的相互作用から模倣学習を実現するためには、ロボットが人間との自然な相互作用を通じて、自ら学習すべき意味ある単位を抽出する必要があり、これは困難な問題であると考えられている。

Inamura [Inamura 04] や Sugiura [Sugiura 07] らは連続隠れマルコフモデル (continuous hidden Markov model: HMM) を用いる事でロボットに動作学習を行わせている。彼らの研究ではロボットが学ぶべき対象の時系列は事前に分割され実験者により意味づけられ与えられており*1、ロボットがいかにして自らが学ぶことを連続的な時系列の中から見いだすのかという問題は扱っていない。

これに対して、Ito らはヒューマノイドロボット QRIO にユーザの連続的な動作からいくつかの動作を recurrent neural network with parametric bias (RNNPB) を用いる事で獲得させている [Ito 06]。RNNPB は一つの RNN でありながら RNN のパラメータ空間を拡張するパラメトリックバイアスを持つことで複数の動作を一つの学習器に分散的に学習することができるようになっており、ダイナミクスの変化に基づいて、複数の動作が異なるバイアス値に対応するように獲得される。しかし、これらの研究では複数の動作間の区別には動作のダイ

ナミクス情報のみが用いられる為、軌道が似ていながら異なる動作を区別して獲得させる事は難しい。本研究では事前に分節化しない連続的な相互作用を通して自律ロボットが発見的に模倣学習を行う手法を提案する。提案手法では模倣する動作系列の分節化に RNNPB のようにダイナミクス情報を用いる以外に、全時系列の中における模倣対象動作の分布情報も用いる。この学習手法ではキーモーションの抽出に直接的な教師値も使わず、事前に設計された強化信号のような評価値も用いないために、全体としての学習過程は自己組織化的な学習過程ととらえる事ができる。

2. 発見的模倣学習手法

2.1 手法の概要

Fig. 1 に提案する学習器の概観を示す。提案学習器は学習器として主に二つの構成要素を含む。一つは Murphy によって提案された SARM [Murphy 98] であり、もう一つが梅村によって提案された辞書を用いないキーワード抽出手法である。

全体の流れを説明する。まず、ユーザは当学習器を有するロボットとインタラクションを行う。このインタラクションは動作毎に手動で分節化される事は無く、一つのセッションを通じて連続的な相互作用の履歴としてロボットに記憶される*2。SARM はこの連続的な時系列が複数の AR モデルの切り替えによって生成されているとの仮定の下、各時刻でそれらの各々のモデルが利用されたと考えられる事後確率を計算する。この事後確率が最大のものでその時刻に用いられた AR モデルと考えることで時系列を分割し、結果として時系列は SARM の隠れ状態を意味する文字の連なりに変換する。連続時系列が文書へと変換された後に、この文書を日本語などの自然言語で書かれた文書と同様の存在であると見立て、キーワードが梅村らによるキーワード抽出手法を用いることで、この文書中のキーワードを抽出する [Umemura 00]。このキーワードがセッションに含まれるユーザの提示したキーモーションに相当すると考える。多くのキーワード抽出手法では事前に準備された辞書を用いるが、梅村はもしキーワードが全文書中である種の偏った分布を持っている場合は、辞書を用いないで日本語の文章からキーワードが抽出できる手法を提案した。これと同様の特性を人間ロボット相互作用を記録した時系列情報を SARM によ

連絡先: 谷口忠大, 立命館大学情報理工学部, 滋賀県草津市野路東 1-1-1, 077-561-5839, tadahiro@tanichu.com

*1 例えば “walk”, “squat”, “kick” など。

*2 一つのセッションは例えばロボットの居る部屋にユーザが入ってきたところからスタートし、ユーザが部屋から出て行った段階で終わるといったものである。

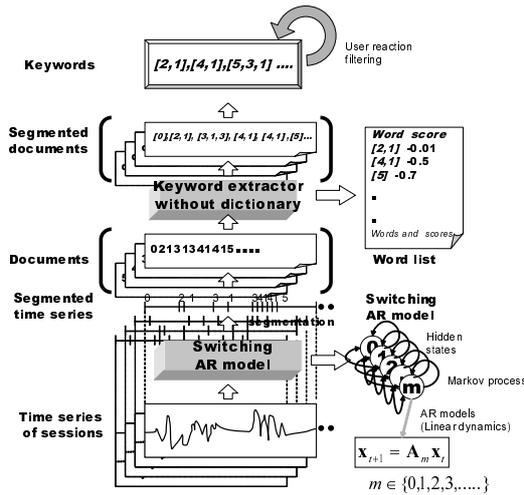


図 1: 提案学習器の概観

り変換することで生成された文書群も持っている事を仮定し当手法を用いる。しかし、本抽出手法で得られた結果は必ずしもユーザの持つ「何が意味のある動作であるか?」の基準を満たすとは限らない。そこで、本手法ではロボットが意味ある動作を示せば、それに対してユーザが意味ある動作を返す可能性が高いと想定し、ユーザの応答情報を用いて、ロボットに重要なキーワードは残し、重要でないキーワードを除去させる事にした。ロボットの視点に立てば、ユーザにできるだけ自分に対するリアクションをとらせたいと考え、ユーザからの有意なリアクションの量を最大化するように自らの動作パターンの候補を絞っていく過程となる。

2.2 Switching autoregression model (SARM)

SARM は多次元の時系列を複数の自己回帰 (AR) モデル (autoregressive model) でモデル化する手法である。各 AR モデルを表す隠れ状態はマルコフ過程に基づいて遷移すると仮定する。SARM は M 個の隠れ状態を持つ。我々は時刻 t における隠れ状態を s_t とする。隠れ状態 j は以下の AR モデルをもつ。

$$x_t = A_j x_{t-1} + v_t, \quad (1)$$

ここで x_t は時刻 t における時系列情報の状態量^{*3}、 $v_t \sim N(0, Q_t)$ はノイズを表す。まず、ロボットが時系列情報 $x_{1:T}$ ($t = 1$ から $t = T$ までの時系列情報) を観測した際に、どの AR モデルから生成されたものとして認識されるかについての計算機構について説明する。この為には事後確率 $\Pr(s_t = j | x_{1:T})$ が計算される必要がある。隠れ状態遷移がマルコフ過程に従うとすると、

$$\Pr(s_t = j | x_t, x_{1:t-1}) \quad (2)$$

$$= \frac{1}{c} \Pr(x_t | s_t = j, x_{1:t-1}) \Pr(s_t = j | x_{1:t-1}) \quad (3)$$

$$= \frac{1}{c} \Pr(x_t | s_t = j, x_{1:t-1}) \sum_i \Pr(s_t = j | s_{t-1} = i, x_{1:t-1}) \times \Pr(s_{t-1} = i | x_{1:t-1}) \quad (4)$$

$$= \frac{1}{c} L_t(j) \sum_i Z(i, j) \Pr(s_{t-1} = i | x_{1:t-1}), \quad (5)$$

ここで c は正規化定数であり、 Z は隠れ状態の遷移確率行列である。 $Z(i, j)$ が隠れ状態 i から j の遷移確率を表す。

$$L_t(j) = N(x_t; A_j x_{t-1}, Q_j) \quad (6)$$

*3 つまり、ロボットが獲得したユーザの関節角などの情報。

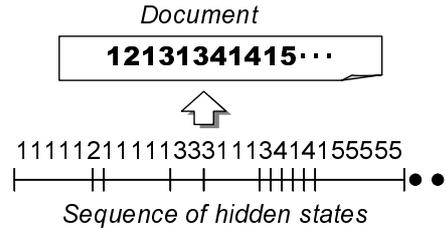


図 2: 隠れ状態系列から文書への変換

は j 番目の AR モデルによる時刻 t における予測誤差に基づく尤度を表わす。ここで N は中心を $A_j x_{t-1}$ 、そして分散共分散行列を Q_j とした多次元正規分布を表す。事後確率は backward の計算により以下のように算出される。

$$\Pr(s_t = j | x_{1:T}) = \sum_k \frac{\Pr(s_t = j | x_{1:t}) \Pr(s_{t+1} = k | s_t = j)}{\Pr(s_{t+1} = k | x_{1:t})} \times \Pr(s_{t+1} = k | x_{1:T}). \quad (7)$$

パラメータ A_j, Q_j も以下の EM アルゴリズムを用いる事で推定することができる [Murphy 98].

$$A_j = \left(\sum_l \sum_{t=2}^T W_t^j P_{t,t-1} \right) \left(\sum_l \sum_{t=2}^T W_t^j P_{t,t-1} \right)^{-1} \quad (8)$$

$$Q_j = \left(\frac{1}{\sum_l \sum_{t=2}^T W_t^j} \right) \times \left(\sum_l \sum_{t=2}^T W_t^j P_t - A_j \sum_l \sum_{t=2}^T W_t^j P'_{t,t-1} \right) \quad (9)$$

ここで $W_t^j \equiv \Pr(s_t = j | x_{1:T})$ 、 $P_t \equiv x_t x_t'$ 、かつ $P_t \equiv x_t x_{t-1}'$ 。元の論文 [Murphy 98] では初期の分布 $P(s_1 = j) = \pi_j$ と $Z(i, j)$ も推定されているが、本稿では簡単の為、この推定は行わなかった。また、も固定とし推定しなかった。

次に、計算された事後確率から最大のものを選ぶことで時系列情報を文書に変換する。

$$s_t^* = \operatorname{argmax}_j \Pr(s_t = j | x_{1:T}) = \operatorname{argmax}_j W_t^j, \quad (10)$$

ここで s_t^* は最も尤もらしい隠れ状態を指す。隣接する同じ隠れ状態を無視することで、隣接する文字が必ず異なる文書へと圧縮する (Fig. 2 参照)。これにより、セッション数と同じだけの文書を獲得する事ができる。この後に、ロボットはこの一見意味のない文字列からキーワードを抽出する。一般的に、隠れ状態の列により形成される n-gram (たとえば、[1, 3, 4, 5], や [2, 1, 2, 1] など) はある種の動作系列を表象するが、そのほとんどの動作はユーザにとって意味がない。例えば 4-gram により動作モーションが形成されると、キーモーションの候補は組み合わせから約 7000 に上る。この非常に多い候補をユーザのリアクションを通じた教示のみにより絞っていくことは困難である。そこで、その中から全セッション通じてのインタラクションで得られた統計的な情報を用いてキーワードを抽出することが重要となる。

2.3 辞書を用いないキーワード抽出手法

梅村は辞書を用いない有用性の高いキーワード抽出手法を提案した [Umamura 00]。辞書を使わずに未知の文書から意味のある分節を切り出すことは日本語のような空白で単語が区切られていない言語では難しい。多くのキーワード抽出手法では形態素解析を用いたり辞書を用いるアプローチがなされて

いた．これに対して梅村はキーワードが文書中に偏って現れる性質に着目しキーワード抽出手法を提案した．人間によって書かれた文書群では，もし，文書のトピックがキーワードに関係するならば，キーワードは文書中に複数回あらわれやすい傾向がある事が知られている．この興味深い定量的な傾向性は Church により自然言語でかかれた文書において示されている [Church 00]．一方でキーワードとなる語は他の文書中では殆ど出てこない．この仮定に基づき，梅村は n -gram がキーワードかどうかをスコアリングする非常に簡単な手法を提案した^{*4}．スコアが高いほどその n -gram はキーワードである可能性が高いと見なされる．

統計量に基づいたそのスコアは Church により以下のように定義された “positive adaptation” [Church 00] と呼ばれる値を元に算出する．

$$Pr(+adapt) = Pr(k \geq 2 | k \geq 1) \approx DF_2 / DF_1, \quad (11)$$

ここで DF_k (Document Frequency k) は 指定の n -gram を k 個かそれ以上含む文書の数を示す． DF_k は文書検索などの領域で有名な document frequency DF を一般化したものである． DF は 指定の n -gram を含む文書の数である．

しかし，positive adaptation だけではキーワードかどうかを特定するのに不十分である．なぜならば，部分文字列も元の文字列とほぼ同じ positive adaptation を持つからである．例えば，“パイナップル”と“パイナップ”は殆ど同じ頻度で同じように現れる．この問題を解決するために，梅村はこのスコアの和が最大になるように文書を分節化することで， n -gram の境界を求めた．

n -gram のスコアは以下で表される．

$$score = \log_2(\min(UB, Pr(+adapt))) \quad (12)$$

ここで UB (Upper Bound) は設計者により決定されるメタパラメータである．一般的に， UB が 1.0 に近いと，対象の文書は細かく分節化され， UB が 0.0 に近いと，対象の文書は粗く分節化されることになる，もし $DF_2 > MA$ ならば n -gram は “wordlist” にそのスコアと共に登録される事にした (Fig. 1 参照)．The MA (Minimum Appearance) は wordlist の大きさを決めるメタパラメータである．梅村の研究では，これらのパラメータ， UB や MA は発見的に定められていた．しかし，実験結果は明白にこれらの値に依存する．

もし，対象の n -gram がランダムに文書中に分布しているならば $Pr(+adapt) = DF_1 / DN$ ，となる．ここで DN (Document Number) は文書の総数を示す．しかしながら，語が二度目に出てくる確率は一度目に出てくる確率よりも高いことはよく知られている．故に， $Pr(+adapt)$ にスケールアップのための \log をかけたものを単純にキーワードかどうかのスコアとした．

文書の最適な分節を特定するために Viterbi search を用いた．これは計算量を減らすためである．wordlist に登録されていない語のスコアは -10000 とした．

分節化を行った後に，キーワード抽出器が分節化された文書からキーワードを抽出する．本稿の実験では以下の条件を満たすものをキーワードとして抽出した^{*5}．

1. $DF_2 / DN < F_{max}$
2. $score > score_{min}$
3. keyword は 2 文字以上で構成される

*4 n -gram は文書中で連続する n 個の文字を意味する．

*5 このキーワード抽出手法は最終的に 4 つのメタパラメータを含む UB , MA , F_{max} , と $score_{min}$ である．現在は，これらは発見的に決定されているが，将来的にはその決定手法についても検討する必要がある．

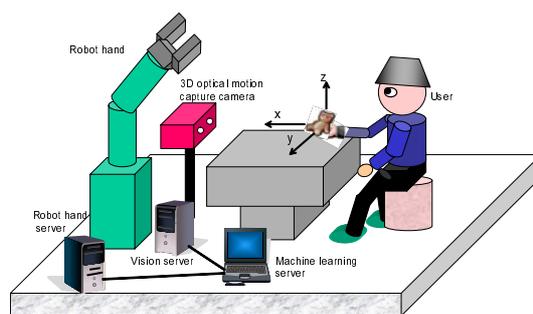


図 3: 実験環境の概観

2.4 キーワード抽出に基づくユーザの応答認識

キーワードが抽出された後に，元の時系列から，そのキーワードが用いられていたときに，それぞれの隠れ状態にどれだけの時間滞在したかの平均が算出される．また，キーワードに相当する動作が開始される初期位置についても同様に求められる．そしてこれらの情報を用いる事によって，ロボットは文字で書かれたキーワード情報を実世界の連続値の動作系列へと変換することが出来る．

ロボットがインタラクションにおいて意味のある動作をした場合，例えば「手を振る」「パイパイ」「お辞儀をする」などをした場合は，その前にいるユーザも何かしらの意味のある動作を返す可能性が高まるものと仮定する．そうすると，その応答に意味のある動作が含まれているかどうかは，獲得した SARM と keyword list を用いることで認識することができる．ゆえに，ロボットがキーモーションであると考えたものを提示した後に，ロボットがユーザの動作系列を観測し，その中にキーモーションが含まれていることが観測されれば，そのロボットの動作はユーザにとって意味あるものであったとロボットは考えることができる (正のフィードバック)．逆に，検知することができなければその意味のない行動をロボットは徐々に取り去る事ができる (負のフィードバック)．

3. 実験

提案手法の有用性を検証するために，人間・ロボット相互作用の為の実験環境を構築し，実験を行った．

3.1 Conditions

実験では長い腕を持ったロボットハンドをロボットの身体として，また人間の動作を観測するために 3D モーションキャプチャカメラを用いた (Fig. 3)．本実験では実験者の右手先の x -軸， y -軸 と z -軸の 3 D モーションキャプチャカメラによって観測される．それぞれの時刻 t における値は x_t^o, y_t^o, z_t^o とした．本稿では簡単のため手先の模倣だけを論じる．ロボットは実験者の手先座標を観測できる．状態変数は $x_t = (x_t^o, y_t^o, z_t^o, c)$ と定義した (ここで $c = 100$ は定数)．また，各次元の単位は $[mm]$ である．さらに，ロボットは教示者の手先座標を自らの手先座標に直接的に写像することができるものとする．

本実験では実験者はロボットに 21 セッション分の動作を提示した．1 セッションは約 20 秒ほどの時間であった^{*6}．サンプリングレートは $25 [Hz]$ であった．それぞれのセッションではそれぞれのセッションではそれぞれのセッションに特徴的な動作が中心的に示された．我々は Fig. 4 に示す二種類のキーモーションを準備した．

sessions 1-7 では実験者は主に up-down motion を複数回提示した．またセッション 8-14 では主に square motion を複数回提示した．他のセッションではこれらの動作や他の動作がランダムに複数回提示された．これらの時系列を実験者は明

*6 どれだけの長さのインタラクションを行うかは実験者にゆだねられた．

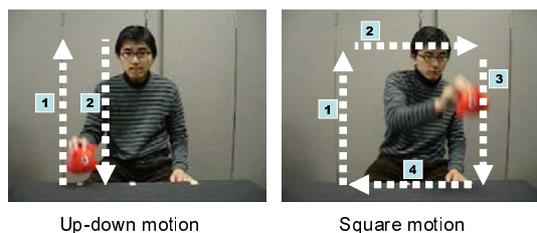


図 4: 連続的相互作用に埋め込まれた二つの動作

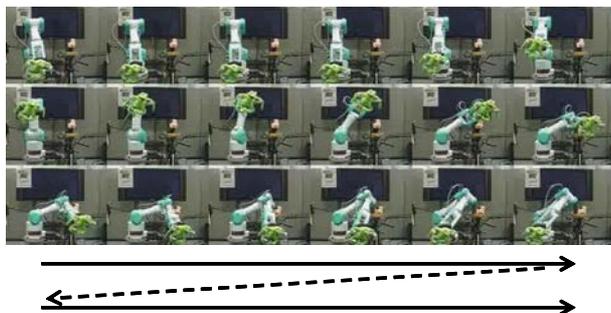


図 5: 獲得された square motion に相当する動作例

示的には分節化していない。ゆえに、ロボットはどこからどこまでが up-down motion なのか、どこからどこまでが square motion かは自ら発見しなければならない。さらにロボットは何通りの動作をユーザが教えようとしているのかについても教えられていない。Z の直行成分のを 0.964, その他の成分は 0.004 とした。隠れ状態の数は発見的に 10 と定めた。AR モデルの初期値は乱数を用いて決定した。EM アルゴリズムは 5 回繰り返して適用して AR モデルのパラメータを決定した。これらのパラメータを用いて, SARM は事後確率 $\Pr(s_t = j | x_{1:T})$ をそれぞれのセッションに対して求めた (Eq. 7 参照)。最も最もらしい隠れ状態の系列 $\{s_t^*\}$ がその事後確率から決定された (Eq. 10 参照)。この系列を圧縮することでロボットは 21 の文書を得た。この後にキーワード抽出手法が用いられ最終的にいくつかのキーワードと word list が獲得された。この実験ではパラメータは $\{UB = 0.9, MA = 2, F_{max} = 0.4, score_{min} = -0.5\}$ とした。

3.2 結果

学習結果として SARM のパラメータが推定され, word list が計算された後,

キーワード抽出が行われ, 7 つのキーワードが抽出された。それらは $[7, 9, 0, 5, 3, 0, 7]$, $[9, 0, 5, 3, 0, 7]$, $[9, 0, 5, 3]$, $[7, 9, 4, 3]$, $[0, 7, 9, 0, 5]$, $[3, 0, 7]$, と $[9, 4, 3]$ *7 であった。これらの意味を理解するために, ロボットにそれぞれに従って動作を生成させたところ $[7, 9, 4, 3]$ と $[9, 4, 3]$ が up-down motion に相当し, $[7, 9, 0, 5, 3, 0, 7]$ が square motion に相当することが分かった。 $[7, 9, 0, 5, 3, 0, 7]$ から生成された動作例を Fig. 5 に示す*8。しかしながら, キーモーションとして実験者に捉えられる以外の動作も本手法では抽出される。本手法が自己組織化型学習であり, 教師あり学習では無いため, このような想定外の学習結果を含むことは大いにあり得るが, これらが実験者にとってのキーモーションでないことにロボットはインタラクションを通じて気づく必要がある。

各キーモーション獲得の後に, ロボットに 5 回づつ各動作を提示させた。実験者はロボットがキーモーションと思われる動作をした際には, 返答として up-down motion を返し, それ

以外の動作に対しては実験者はロボットに何も返さなかった。もしロボットに返答として up-down motion が返って来たとしたらロボットはその動作をキーモーションの候補 $[7, 9, 4, 3]$ もしくは $[7, 4, 3]$ として十分な確率で認識することができる。この実験者のリアクションを用いることで最終的にロボットは実験者が事前にキーモーションと考えた, 二つのキーモーションを獲得することが出来た。

4. 結言

我々は連続的な人間ロボット間相互作用から発見的に模倣学習を行う事のできる役割反転模倣の計算論的なモデルを提案した。これにより, 外部から明示的な教示や分節を与えること無しに, 自律ロボットがユーザの行動を観察し, 相互作用する中で複数の動作を獲得することが出来るようになると思われる。しかし, 現段階では, 当手法は学習のために複数の手動で設定すべきパラメータを持っており, これらの設計法, 自動調節法が求められる。また, 学習器全体としても 3 つのプロセスが直列的に結合しており, これらを一体的な学習器として定式化することがモデル選択の適用や, 理論の精緻化, ロバスト化の視点からも求められる。

5. 謝辞

本研究は文科省 21 世紀 COE プログラム「動的機能機械システムの数理モデルと設計論」, 国立情報学研究所共同研究助成「能動的ハンドインタラクションによる実世界言語コミュニケーションの学習に関する研究」, 日本学術振興会特別研究員研究奨励費 (17-1685 及び 19-3467) の一部支援を受けた。また, ロボット環境の構築について杉浦孔明氏, 今木理英氏にただならぬ支援の頂いた。ここに感謝の念を示す。

参考文献

- [Church 00] Church, K.: Empirical estimates of adaptation: the chance of two noriegas is closer to $p/2$ than $p/2$, *Proceedings of the 17th conference on Computational linguistics-Volume 1*, pp. 180-186 (2000)
- [Inamura 04] Inamura, T., Toshima, I., Tanie, H., and Nakamura, Y.: Embodied symbol emergence based on mimesis theory., *International Journal of Robotics Research*, Vol. 23, No. 4, pp. 363-377 (2004)
- [Ito 06] Ito, M., Noda, K., Hoshino, Y., and Tani, J.: Dynamic and interactive generation of object handling behaviors by a small humanoid robot using a dynamic neural network model, *Neural Networks*, Vol. 19, No. 3, pp. 323-337 (2006)
- [Murphy 98] Murphy, K.: Switching Kalman filters, *Dept. of Computer Science, University of California, Berkeley, Tech. Rep* (1998)
- [Sugiura 07] Sugiura, K. and Iwahashi, N.: Learning Object-Manipulation Verbs for Human-Robot Communication, in *Workshop on Multimodal Interfaces in Semantic Interaction at the International Conference on Multimodal Interfaces* (2007)
- [Umemura 00] Umemura, K.: Related Word-pairs Extraction without Dictionaries, Technical report, IPA Exploratory Software Project development result report, <http://www.ipa.go.jp/archive/NBP/12nendo/12mito/mdata/10-36h/10-36h.pdf> (in Japanese) (2000)

*7 Each numeral represents a hidden state of SARM.

*8 他の獲得動作については紙面の都合上, 詳細を略す。