ヒューマン・ロボットインタラクションを通した役割反転模倣に基づく実時間応答戦略獲得

Acquisition of interaction strategy based on role-reversal imitation through human-robot interaction

谷口 忠大[*1*2] Tadahiro Taniguchi　岩橋 直人[*2] Naoto Iwahashi　中西 弘門[*1] Hiroto Nakanishi　西川 郁子[*1] Ikuko Nishikawa

[*1]立命館大学 Ritsumeikan University　[*2]情報通信研究機構 National Institute of Information and Communications Technology

In this paper, we describe a novel imitation learning method which enables an autonomous robot to acquire interaction strategy through human-robot real-time interaction. The robot becomes able to respond to human user's social action correctly. We constructed the learning method based on role reversal imitation which is found in human infants in developmental psychological researches. A probabilistic model is proposed which assumes that delayed reactions are stochastically generated by initiative actions. In an experiment, we show a robot hand became able to exhibit correct reaction and estimate whether another's action is an initiative action or a reaction.

1. はじめに

人間社会は物理的なダイナミクスの多様性のみならず，地域や文化に依存した社会的・記号論的多様性を有している．たとえ同一の地域であっても，家庭毎に，集団毎に言葉を始めジェスチャや間に持たせる意味は変容する．そのような記号論的多様性の中に生きるロボットは事前に記号の意味や価値，表出の仕方を知る事は出来ず，設計後に学習し獲得していく事が求められる．

1.1 応答戦略の獲得

社会で人間を相手に活動するロボットは，ユーザがどのような行動を行った時に，自らがどのようなリアクションを返すべきかという応答戦略を実時間的に獲得すべきである．また，そのようにユーザとの相互作用を通して，行動の規範自体を変容させる事はペットロボットで求められるようにユーザ・ロボット間の関係性構築の視点からも重要である．応答戦略の獲得には強化学習や進化的方法を用いる事も考えられるが，このような評価関数に基づく手法では多くの試行錯誤を必要とし，ユーザが期待する時間許容度とはかけ離れる可能性が大きい．ユーザが相互作用を通じてロボットに何かを覚えさせようとする際，せめて分スケールでの学習の達成と同時に，インタラクションが明示的に区切られる事無く実時間的な連続的な相互作用の中で達成されることが期待される．本稿では人間の幼児にみられる役割反転模倣 (role-reversal imitation) に着目し，実時間的にロボットが応答戦略を獲得する為の枠組みを提案する．

1.2 役割反転模倣

Tomasello は子供がことば（記号）を覚える社会的基盤の説明の一部するために役割反転模倣 (Role reversal imitation) という言葉を用いている [3]．役割反転模倣とは，共同行為を達成するにあたり対象への役割を反転し模倣することである．二者間のインタラクションであっても，バイバイに対してバイバイを返すなどといった事はこの二つのアクションの組により初めて達成される共同行為であると捉える事が出来る．このような共同行為を学習する際にお互いの役割を反転する事で，学習者はその行為の社会的な利用方法を学んでいく事が出来ると考えられる．また，この際には，自らの行為がその共同作業を構成する役割の内のどちらであるかという役割認識と役割の動的な変更を行う必要がある．

栗山らは，乳児の親子間のやりとりにおける随伴性に注目して，移動エントロピーを利用することで応答戦略を適応的に獲得する手法を提案した [1]．谷口らは時間遅れを伴うリアクションの発生回数を数え上げる事で時間遅れを伴う同時確率分布を推定することで応答戦略を獲得する手法を提案した [4]．しかしながら，これらの学習則は基本的に発見的手法に基づいている[*1]．本稿では，EM アルゴリズムにより役割反転模倣を達成する学習則を提案する．

2. EM アルゴリズムによる応答戦略獲得

2.1 役割の定義

二者の共同行為として見たインタラクションは，自発的な動作とそれに対する他者の応答としての動作により成り立っていると捉えられる．本研究では前者をイニシアティブアクションと呼び，後者をリアクションと呼ぶ事にする[*2]．以下本稿ではイニシアティブアクションを IA，リアクションを RA と略記する[*3]．また，前者を行う側をイニシエーター，後者を行う側をリアクターと呼ぶ．一つの共同行為を実現するためには，二者がこれらの役割を適切に分担しながら動作を出力する必要がある．本研究ではイニシエーターとリアクターの役割を反転する事を役割反転とし，動作を出力する規則を応答戦略と呼ぶ (図 1)．

これらの学習は，明示的にどの RA がどの IA に対応するかという関係づけが明示的な系では容易であるが，本稿で扱うように，連続時間であり，かつ役割が明示的に与えられないような系での学習は決して自明ではない．本稿では IA に対する RA の発生を確率的な生成モデルに基づいてモデル化し，また，

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 − 1，077-561-5829, taniguchi@ci.ritsumei.ac.jp

*1 谷口らは不完全データ尤度最大化基準から更新式を導出しているとしているが，実際に用いられている更新式は不完全データ尤度最大化基準から導出出来るものではなく，結果的に発見的手法となっている [4].

*2 本研究でアクションと言うときには，この二つを包含する概念を指す.

*3 自発的な動作の中にも他者の応答を求めるものと，そうでないものが存在するとも考えられるが，本稿では何らかの動作に対する RA で無い動作全てを IA として扱う.

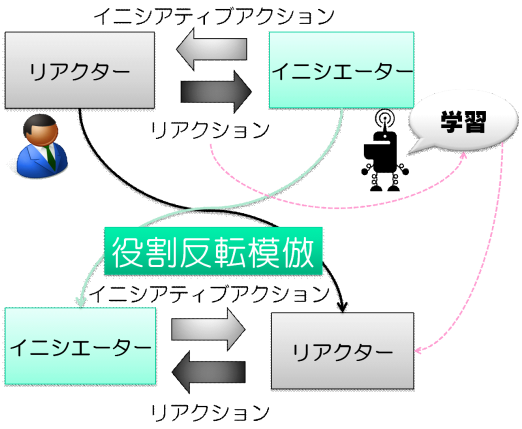

図 1: 役割反転模倣の概念図

潜在変数としての役割[*4]を明示的に推定をする事で，ロボットの学習フェーズと行動フェーズを分離することなく人間との実時間相互作用から応答戦略獲得を行う学習則を構築する．

2.2 リアクション発生確率

図 2 に添って RA 生成モデルを説明する．まず，ロボットの取り得る単位動作の集合を $A = \{a_1, a_2, \cdots, a_m\}$ とし，人間の取り得る単位動作の集合を $B = \{b_1, b_2, \cdots, b_m\}$ する．さらにこれらに「何もしない動作」a_ϕ, b_ϕ を追加した集合を A', B' とする．ここで，人間の単位動作とロボットの単位動作は身体的な対応関係を表わす全単射写像 c が存在し $c(a_i) = b_i$ を通して一対一対応の関係が存在すると仮定する[*5]．本研究では RA は IA を原因として，時間遅れを伴い確率的に発生すると仮定する．

時刻 s において IAa_i を原因とした何らかの RA が発生する確率 $W^a_{a_i,s}$ を**リアクション発生確率**と呼び，時間と共に指数的に減衰する確率関数で定義する．これを以下の様に毎ステップ逐次的に更新することで計算する．ここで s は連続時間を Δt 間隔でサンプリングした時の離散時間ステップを指す．

$$W^a_{a_i,s+1} \leftarrow \begin{cases} W^a_{a_i,s} + \frac{\Delta t}{\tau} & (IA\ a_i\ is\ observed) \\ \gamma W^a_{a_i,s} & (IA\ a_i\ is\ not\ observed) \\ \lambda & (i = \phi) \end{cases} \quad (1)$$

$$\gamma = \exp(-\frac{\Delta t}{\tau}) \quad (2)$$

ここで τ は IA が発生してから RA が発生するまでの時間間隔に比例した時定数である．この確率 W^a に基づいて IA から RA が生成されるとする．また,「何もしない動作」a_ϕ, b_ϕ を原因として発生するアクションを IA と見なす事で，IA を RA の一種として一体的に取り扱う．IA は毎ステップ λ の確率で発生するとする．このリアクション発生確率には λ, τ の二つのパラメータがあるが，これらは後に示すように相互情報量最大化基準により求める．しかしながら，W^a を求める為には各時刻の a_i が IA であるか否かを知る必要があるが，これは一般的には観測する事が出来ない潜在変数である．

*4 イニシエーターであるかリアクターであるかということ
*5 この意味で本研究では模倣者と被模倣者の身体の対応関係を問題にする correspondence problem は取り扱わない [2].

2.3 EM アルゴリズムに基づくリアクション選択確率推定

リアクション出力確率に従い RA の発生が決定した後に，原因となった IA に依存し出力される具体的な RA の種類が確率的に選択される．IA a_i に対して RA b_j が出力される確率 $Z(b_j, a_i)$ を**リアクション選択確率**と呼ぶ．ロボットは図 1 のように人間の RA の観測に基づき，EM アルゴリズムに従って学習する．

一定時間のインタラクションを通してロボットと人間の動作ログ $A^O = \{a(s)|a(s) \in A, s \in S_a\}$ および $B^O = \{b(s)|b(s) \in B, s \in S_b\}$ が得られたとする．ここで S_a, S_b はそれぞれロボットと人間がアクションを起こした時刻の集合である．ここでロボットが人間の応答戦略を学習するとすると対数尤度 L は以下の様になる．

$$L = \sum_{s \in S_b} \log(\sum_{a_i \in A'} W^a_{a_i,s} Z(b(s), a_i)) \quad (3)$$

ここで $b(s)$ は時刻 s での人間の選択したアクションである．これを最大化する事でロボットは人間のリアクション選択確率を推定し，それを反転する事で人間の応答戦略を共有する必要がある．この時，B^O の内のどの動作が A^O のどの動作を原因として起きたかが分からないという部分観測性が問題となる．このような潜在変数を持つ尤度最大化問題は EM アルゴリズムを用いる事で解くことが出来る．Q 関数を以下のように定義する．

$$Q(Z|Z^{(t)}) = \sum_{s \in S_b} \sum_{a_i \in A'} P(a_i|b(s), s) \log(Z(b(s), a_i))$$

$$P_b(a_i|b(s), s) = \frac{W^a_{a_i,s} Z^{(t)}(b(s), a_i)}{\sum_{a_j \in A'} W^a_{a_j,s} Z^{(t)}(b(s), a_j)} \quad (4)$$

ここで $Z^{(t)}$ は t 反復でのパラメータ推定値，$P_b(a_i|b(s), s)$ を時刻 s で $b(s)$ が生じた際にその原因が a_i である確率とする (E ステップに相当)．これより，更新式

$$M^{(t+1)}(b_j, a_i) = \sum_{s \in S_{b_j}} P(a_i|s, b_j) \quad (5)$$

$$Z^{(t+1)}(b_j, a_i) = \frac{M^{(t+1)}(b_j, a_i)}{\sum_j M^{(t+1)}(b_j, a_i)} \quad (6)$$

$$J^{t+1}(b_j, a_i) = \frac{M^{(t+1)}(b_j, a_i)}{\sum_{i,j} M^{(t+1)}(b_j, a_i)} \quad (7)$$

ここで S_{b_j} は S_b の内，出力動作が b_j であるものの集合である．$M^{(t+1)}$ は IA-RA 関係の頻度の数え上げに相当し，$J^{(t+1)}$ は同時確率分布の推定値に相当する．

2.4 役割推定とリアクション出力確率の推定

W^a が決定すれば前節のように EM アルゴリズムで漸近的に推定出来るが，どの動作が IA であるかはターンテイクが動的に生じるインタラクション中では明確ではなく，結果的にリアクション出力確率 W^a も確定的には決定できない．よって，EM アルゴリズムの E ステップに於いて，同時にどのアクションが IA であるかを推定する必要がある．前節の P_b の逆に，ロボットの s ステップ目の動作 $a(s)$ が人間の動作 b_i へのリアクションである確率を $P_a(b_i|a(s), s)$ とする．このとき，

$$P_a(b_i|a(s), s) = \frac{W^b_{s,b_i} Z(c(a(s)), c^{-1}(b_i))}{\sum_{b_j \in B'} W^b s, b_j Z(c(a(s)), c^{-1}(b_i))} \quad (8)$$

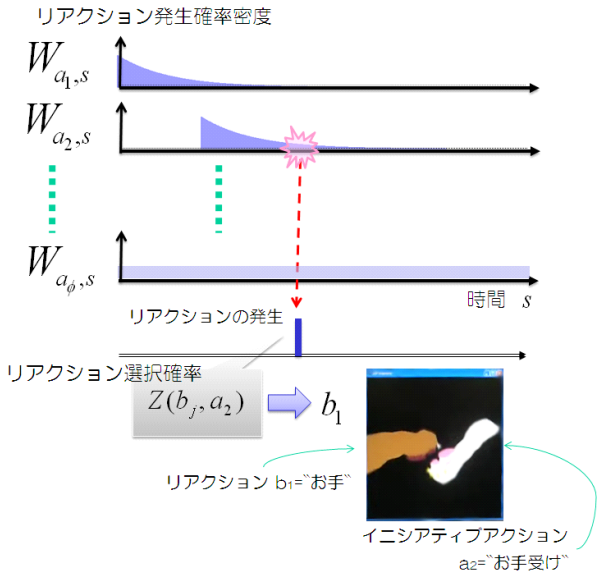

図 2: イニシアティブアクションからリアクションが確率的に発生する生成確率モデルの概要

これらを用いて

$$W^a_{a_i,s+1} \leftarrow \begin{cases} W^a_{a_i,s} + P_a(b_\phi|a(s),s)\frac{\Delta t}{\tau} & (a_i \ is \ observed) \\ \gamma W^a_{a_i,s} & (a_i \ is \ not \ observed) \\ \lambda & (i=\phi) \end{cases}$$

$$W^b_{b_i,s+1} \leftarrow \begin{cases} W^b_{b_i,s} + P_b(a_\phi|b(s),s)\frac{\Delta t}{\tau} & (b_i \ is \ observed) \\ \gamma W^b_{b_i,s} & (b_i \ is \ not \ observed) \\ \lambda & (i=\phi) \end{cases}$$

として各時刻のリアクション出力確率を逐次的に推定し，EM アルゴリズムにおける更新に利用する．また各アクションに対して，$P_a(b_i|a(s),s)$, $P_b(a_i|b(s),s)$ を見ることで，それぞれのアクションが IA なのか RA なのかの推定が可能である．つまり，$\phi = \mathrm{argmax}_i\, P_b(a_i|b(s),s)$ であれば，$b(s)$ は IA と推定される*6.

2.5 役割反転による行動選択

前記の手法に基づきロボットは行動を選択する際に人間のリアクション行動選択則 $Z(b,a)$ を $\hat{Z} \leftarrow Z^{(T)}$ として推定する．T は反復回数である．この推定した人間のリアクション行動選択則と身体動作の対応関係を表わす関数 c を用いて，イニシアティブアクション b_j に対するリアクション a_i を以下の確率で選択する．

$$P(a_i|b_j) = \hat{Z}(c(a_i)|c^{-1}(b_j)) \quad (9)$$

最終的に時間ステップ s に於いて行動 a_i を選択する確率は

$$P(a_i|s) = \sum_{b_j \in B'} W^b_{b_j,s} \hat{Z}(c(a_i)|c^{-1}(b_j)) \quad (10)$$

となる．

*6 ここでの $W_{a_j,s}$ の推定は s ステップまでの情報までを考慮したものであり，厳密性には欠けるが本稿ではこの推定値を採用する．

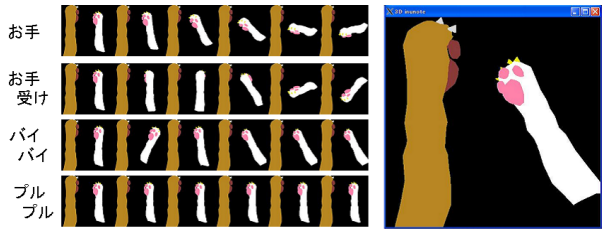

図 3: シミュレーション画面と用意した単位動作

3. インタラクション実験

本研究で提案した手法の有効性を確認するために，計算機上でインタラクション可能なロボットを実装し実験を行った．その内容を以下に示す．

3.1 実験条件

実験では図 3 のように計算機上に二つの犬の手を表わし，これをユーザとロボット（学習プログラム）が動かす事で相互作用を行う空間を準備した．手前がユーザの操作する手であり，奥がロボットが操作する手となっている．ユーザはキーボードを押すことで予め設定された単位動作を出力することが出来る．アクションには a_1 = "お手"，a_2 = "お手受け"，a_3 = "バイバイ"，a_4 = "プルプル" と名付けた四つの単位動作を設定した．本実験ではユーザが操作する側の手もロボットが操作する側の手も同型であるので b_i は a_i と同様に与えた．キーボードの 1 キーから 4 キーまでがそれぞれが a_1〜a_4 に対応付けられており，ユーザがキーを押すと遅滞なく選択された行動が開始され，ロボットは時間遅れも認識誤り無く瞬時にその行動を認識できるものとする．フレームレートは 10[Hz] とした．ユーザにはロボットに

1. お手受けに対してお手をさせる行動
2. バイバイにバイバイを返す行動

を学習させる事を求めた．ユーザはロボットがお手受けを出したときにお手をし，バイバイをしたときにバイバイを返し，また，他の時には無規則に行動を表出した．また，今回は応答戦略のパラメータ推定と役割推定に焦点を当てる為，ロボットの行動出力については今回は式 10 の生成モデルを直接用いる事はせず，簡単の為ロボットには約 2[s] に一回リアクション出力を行わせそのリアクション選択は獲得したリアクション選択確率に基づくという生成手法をとった．

3.2 実験結果

インタラクションは約 300 秒行われ，その間ユーザは基本的に実験条件に示したような応答を示すよう心がけた．実験は $\tau = 0.2, \lambda = 0.001$ とし，EM アルゴリズムは 5 回繰り返した．得られたリアクション選択確率 $Z(b_j,a_i)$ を図 4 に示す．特に，お手受け，バイバイに対し，それぞれお手とバイバイを返す応答戦略が適切に獲得されている事がわかる．

また，λ, τ の値を相互情報量を最大化させるように探索した，相互情報量 I は同時確率分布 $J(b,a)$ から計算される．

$$I(B,A) = \sum_{b\in B}\sum_{a\in A'} J(b,a)\log\frac{J(b,a)}{J(b)J(a)} \quad (11)$$

図 5 に示すように，相互情報量は $\lambda = 0.01, \tau = 0.5$ 周辺で最大化された．これは与えられたインタラクション履歴 A^O, B^O

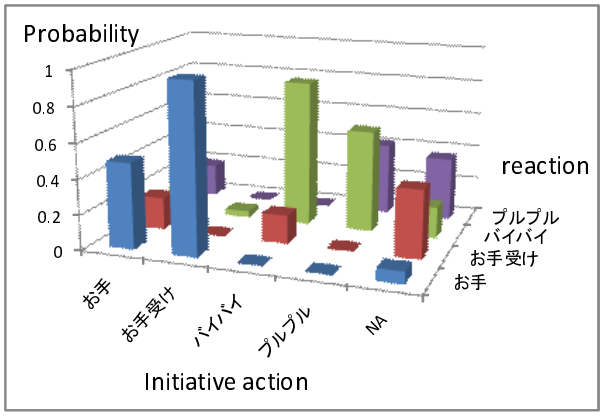

図 4: 学習後のリアクション選択確率 $\hat{Z}$

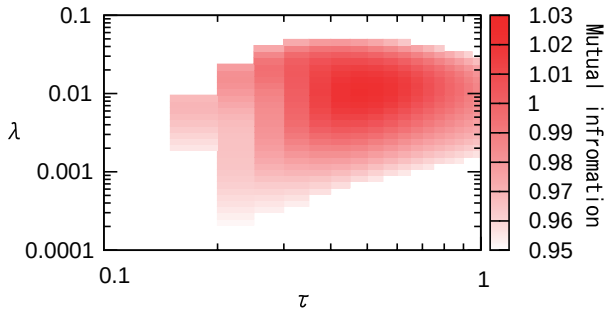

図 5: λ, τ による相互情報量の変化

が連続的な時間の中で最も高い随伴性で生じていると解釈するためには，IA に対して RA がどのくらいの遅れで生じるとするべきかを同定している．

また，役割推定も十分な精度で行われていた．インタラクションに於ける真の役割を決定することが困難なため定量的な評価は出来ていないが，役割推定の典型的な結果の一部を表 3.2 に示す．(1)613〜630 ステップと (2)1326〜1331 ステップの推定結果について示している．何も動作が生じなかった時間帯については表示を省略している．(1) でははじめユーザが”お手受け”を出した後にロボットが”お手”をした際にユーザのお手受けが IA，ロボットの”お手”がユーザの”お手受け”に対するリアクションと推定されている．その後，0.6 秒後にロボットが”バイバイ”を出しているが，これはユーザの”お手受け”に対するリアクションとは解釈されずに IA と推定されている．その後のユーザの”バイバイ”はロボットの”バイバイ”に対するリアクションと推定されている．これより，実時間的にターンテイクが行われ，かつ，それの役割の交代がロボットに正しく認識されている事がわかる．(2) ではロボットが”お手受け”をした後に”お手”を出しているが，その直後 0.2 秒後にユーザから出された”お手”は，その時点でロボットが出している”お手”ではなく，その直前に出した”お手受け”に対するものであるとロボットが解釈出来ている事を示している．

しかしながら，役割推定の全履歴を見ると，ロボットの動作の直後に人間が行った動作（またはその逆）は，その動作のリアクションと解釈されやすい傾向が見られた．これは，リアクション発生確率 W を時間方向の指数分布として定義している為，IA 発生直後は，極端にそれに対する RA が最も発生しやすくなる為である．しかし，IA に対する実際の RA はそのように分布せず，IA から一定の遅れを持った時刻を中心にピークをもって分布すると考えられる．この点から W の分布モデルを変更することも考えられるべきである．

step	動作主体	アクション	推定されたアクションの原因
613	ユーザ	お手受け	IA
618	ロボット	お手	お手受け
624	ロボット	バイバイ	IA
630	ユーザ	バイバイ	バイバイ
1326	ロボット	お手受け	IA
1329	ロボット	お手	IA
1331	ユーザ	お手	お手受け

表 1: 役割推定結果の一部

4. まとめ

本研究では有限個の単位動作を元に連続的な実時間インタラクションから役割反転模倣を通じ自律ロボットが応答戦略を獲得する手法を提案した．手法は主に動作の役割と，表出された動作が，どの IA に対する RA であるかという関係を隠れ変数と見なし，EM アルゴリズムを用いて構築した．また，その有効性を計算機上に構築した仮想空間上のインタラクションを通じて検証した．また，実時間的なインタラクションを通じて動的に移り変わる役割を推定する手法についても提案した．人間同士のインタラクションでは動的にイニシアティブが切り替わるが，このような役割推定手法を用いることで，そのような自然なインタラクションをヒューマン・ロボットインタラクションの中で実現する事が可能となる．今後の課題としては本手法の有効性を実ロボットで検証することや，当手法を時間のみならず動作についても連続的なインタラクションへと展開することなどが考えられる．

謝辞

本研究は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」及び科学研究費補助金 若手研究（スタートアップ）20800060「非分節な人間機械相互作用を通じた自己組織化型模倣学習機構の構築」，科学研究費補助金 学術創成「記号過程を内包した動的適応システムの設計論」19GS0208 の一部支援を受けた．

参考文献

[1] T. Kuriyama and Y. Kuniyoshi. Acquisition of Human-Robot Interaction Rules via Imitation and Response Observation. In *Proceedings of the 10th international conference on Simulation of Adaptive Behavior: From Animals to Animats*, pp. 467–476. Springer, 2008.

[2] C.L. Nehaniv and K. Dautenhahn. The correspondence problem. In *Imitation in Animals and Artifacts*, pp. 41–61. MIT Press, 2002.

[3] マイケル・トマセロ. 心とことばの起源を探る. 勁草書房, 2006.

[4] 谷口忠大, 岩橋直人, 中西弘門, 西川郁子. 役割反転模倣に基づく実時間相互作用からの応答戦略獲得. 第 36 回知能システムシンポジウム, pp. pp.127–132, 2009.