4D2-OS-37d-1

# 人のインタラクションからの教師なしルール学習

Unsupervised rule learning from humans' interaction

押川　慧 *1 Satoru Oshikawa　中村　友昭 *1 Tomoaki Nakamura　長井　隆行 *1 Takayuki Nagai　岩橋　直人 *2 Naoto Iwahashi
船越　孝太郎 *3 Funakoshi Kotaro　金子　正秀 *1 Masahide Kaneko

*1電気通信大学 The University of Electro-Communications　*2岡山県立大学 Okayama Prefectural University
*3(株) ホンダ・リサーチ・インスティチュート・ジャパン Honda Research Institute Japan Co., Ltd.

There are various rules of interactions in human's society, and humans can find and learn the rules by observing such interactions. In order for robots to coexist with the humans, it is important for robots to find and learn the rules in an unsupervised way. In this paper, we propose a method for segmenting continuous motions of interactions between two people, and learning rules from segmented motions. The validity of the proposed method is examined using a simple interaction game.

## 1.　はじめに

人間社会には地域や文化によって異なるインタラクションが存在し，人間は他者とのインタラクションや他者同士のインタラクションの観察から，そのインタラクション内に存在するルールを見つけ，学習している．例えば，幼児は身近な人のふるまいを見て，それを模倣して遊ぶことができる．この時，幼児はこの行動の意味などが分かっているわけではないが，その行動を観察から模倣することができる [神谷 11]．このような学習や模倣は社会性獲得の基盤であるといえる．そのためロボットが人間と共存していくためには，ロボットも人間と同様に教師なしでルールを学習することが求められる．

ロボットのルール学習に関する先行研究として文献 [谷口 09] がある．この研究では役割反転模倣を通じて連続的な実時間インタラクションからロボットがインタラクション内の応答戦略を獲得しており，役割を推定することによってルールを学習している．しかし，この研究ではロボットの行動がシミュレーション上で事前に定義されているため，それ以外の行動やルールを学習することはできない．また，実験もシミュレーション上で行われており，実際の人のインタラクションには適用されていない．また，文献 [Soltoggio 13] では遅延報酬を考慮したゲームのルール学習が行われており，実際のロボットを用いて実験を行っているが，この研究においても行動が事前に定義されており，それ以外の行動やルールを学習することはできない．このように，ルール学習に関する研究はなされているが，事前に行動が設定されているものが多く，行動とルールの双方を学習する研究は十分になされていない．

本稿では文献 [中村 16] において提案された GP-HSMM(Gaussian process-hidden semi Markov model) を用いて実際の二人のインタラクションの連続的な時系列データを分節化することで行動を学習し，さらに行動の因果関係を抽出することによってルールを学習する手法を提案する．

## 2.　提案手法

図 1 が提案手法の概要図である．本稿では，二者間のインタラクションを kinect により測定して得られる連続的な手先座標からルールを学習することを想定している．kinect から取得する手先座標は連続的な時系列データであるため，そこから行動を獲得するためには連続的な時系列データを分節化し，行動ごとに教師なしで分類する必要がある．提案手法では，GP-HSMM により手先座標を分節化し，分節化されたデータから互いの行動の因果関係をルールとして抽出する．また，実際のインタラクションにおける行動は必ずしも同時に行われるわけではなく，時間のずれが発生する場合もあるため，そのような時間差も考慮しルールを学習する．

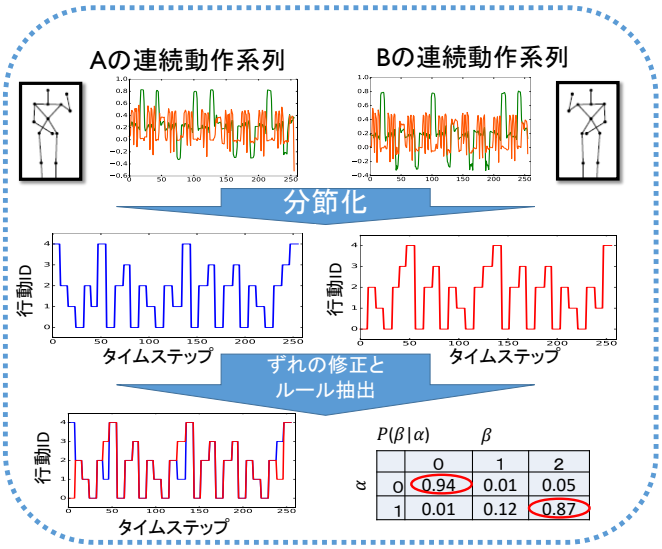


図 1: 提案手法の概要

### 2.1　GP-HSMM

図 2 に GP-HSMM のグラフィカルモデルを示す．図において $c_j$ が単位系列のクラス（動作クラス）を表している．また $X_c$ はクラス $c$ のガウス過程のパラメータであり，単位系列 $x_j$

連絡先: 押川慧，電気通信大学大学院情報理工学研究科，東京都調布市調布ヶ丘 1-5-1，s_oshikawa@radish.ee.uec.ac.jp

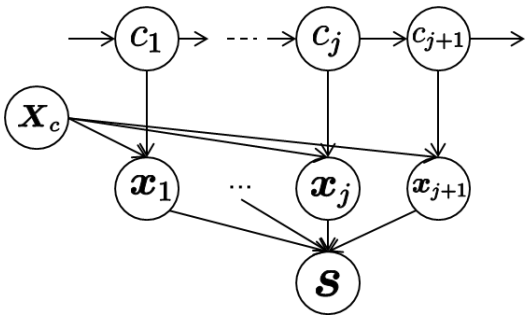


図 2: GP-HSMM のグラフィカルモデル

は $X_{c_j}$ をパラメータとするガウス過程よって生成される．

$$c_j \sim P(c|c_{j-1}) \tag{1}$$

$$\boldsymbol{x}_j \sim \mathcal{GP}(\boldsymbol{x}|\boldsymbol{X}_{c_j}) \tag{2}$$

これらの単位系列を結合することによって観測系列 $S$ が生成される．ガウス過程を用いることで取得した時系列データを離散化することなく，連続的な分布として表現することができる．ガウス過程では同一クラスに属する動作データのタイムステップ $i$ に対する出力 $x$ の組 $(\boldsymbol{i}, \boldsymbol{x})$ が得られた時，タイムステップ $t^{new}$ における出力 $x^{new}$ の予測分布は式 (6) のガウス分布となる．

$$p(x^{new}|t^{new}, \boldsymbol{x}, \boldsymbol{i}) \propto N(\boldsymbol{k}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{i}, c - \boldsymbol{k}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{k}) \tag{3}$$

得られた予測分布を用いることで，$\boldsymbol{x}_t$ がクラス $c$ に対応するガウス過程から生成される確率 $\mathcal{GP}(\boldsymbol{x}_t|\boldsymbol{X}_c)$ を以下のように求めることができ，この確率を用いることで似た単位系列を同じクラスに分類することが出来る．

$$\begin{aligned}\mathcal{GP}(\boldsymbol{x}_t|\boldsymbol{X}_c) =& p(x_{t,0}|t, \boldsymbol{X}_{0,c}, \boldsymbol{I}) \\ &\times p(x_{t,1}|t, \boldsymbol{X}_{1,c}, \boldsymbol{I}) \\ &\times p(x_{t,2}|t, \boldsymbol{X}_{2,c}, \boldsymbol{I}) \cdots\end{aligned} \tag{4}$$

#### 2.1.1 学習

GP-HSMM では，時系列データはガウス過程を出力分布とする隠れマルコフモデルによって生成されると仮定し，観測系列内の単位系列とそのクラスをサンプリングすることで推定する．モデルの学習には一つの観測系列内の単位系列とそのクラスをまとめてサンプリングする Blocked Gibbs Sampler を用いる．まず，全ての観測系列をランダムに分類し，ある観測系列をクラスから除き，各クラスのガウス過程のパラメータ $X_c$ と遷移確率 $P(c|c')$ を更新する．次に文献 [Uchiumi 15] と同様に，Forward filtering-Backward sampling を用いて単位系列の分節とそのクラスを同時に推定する．Forward filtering では観測系列のあるタイムステップ $t$ を終点とする長さが $k$ の分節化がクラス $c$ となる確率を次のように計算する．

$$\begin{aligned}\alpha[t][k][c] =& \mathcal{GP}(\boldsymbol{x}_{t-k:k}|\boldsymbol{X}_c) \\ &\times \sum_{k'=1}^{K}\sum_{c'=0}^{C} p(c|c')\alpha[t-k][k'][c']\end{aligned} \tag{5}$$

$C$ はクラス数，$K$ は単位系列の最大の長さとなる．$p(c|c')$ はクラスの遷移確率を表しており，以下の式のようになる．

$$p(c|c') = \frac{N_{c'c} + \alpha}{N_{c'} + C\alpha} \tag{6}$$

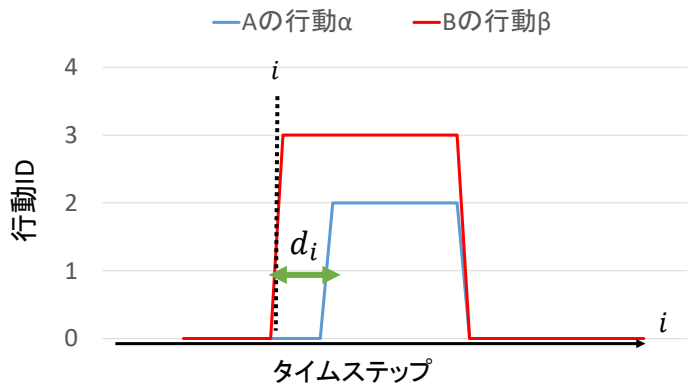


図 3: インタラクションの時間のずれ

ただし，$N_{c'}$ と $N_{c'c}$ は，それぞれクラスが $c'$ となった分節の数と，クラスが $c'$ から $c$ に遷移した回数を表している．これらの式により全ての $k$ と $c$ の組み合わせの確率を計算することが出来る．次に Forward filtering によって求めた確率に従い，Backward sampling により分節の長さとそのクラスをサンプリングすることで観測系列の全ての分節とクラスを決定することが出来る．この分節とクラスを用い，各クラスのガウス過程のパラメータ $X_c$ と遷移確率 $P(c|c')$ 更新する．以上の操作を繰り返すことで，観測系列の分節点とガウス過程のパラメータを同時に推定することが出来る．

### 2.2 ずれの修正およびルール抽出

kinect により取得した A と B 二者のそれぞれの連続的な動作を分節化し，分節化された行動の因果関係を確率で表現する．その確率が高いものがルールであるといえる．時刻 $i$ の GP-HSMM によって分節・分類された A の行動を $\alpha_i$，B の行動を $\beta_i$ とし，以下の 3 つの確率を考える．

1. A が行動 $\alpha_i$ をした時に，B が行動 $\beta_i$ をする確率 $P(\beta_i|\alpha_i)$
2. A が行動 $\alpha_i$ をした後に，A が行動 $\alpha_{i+1}$ をする確率 $P(\alpha_{i+1}|\alpha_i)$
3. A が行動 $\alpha_i$，B が行動 $\beta_i$ をした後に，A が行動 $\alpha_{i+1}$ をする確率 $P(\alpha_{i+1}|\alpha_i,\beta_i)$

ただし，ある行動を行っている間は $\alpha$ と $\beta$ は同じ値が続くため，自己遷移確率を含めた場合，自己遷移確率が極端に大きくなるため，これらの確率は自己遷移確率を除いて計算する．

さらに，現実のインタラクションにおける行動は，必ずしも同時に行われるわけではなく図 3 のように時間のずれが生じる場合がある．本稿ではこの時間差も教師なしで推定して前述の 3 つの確率を計算する．時刻 $i$ に A が行動 $\alpha_i$ を，B が行動 $\beta_i$ をしている場合，このインタラクションの時間差 $d_i$ がガウス分布に従うと仮定する．$d_i$ はギブスサンプリングを用いて，以下の式よりサンプリングし，$P(\beta|\alpha)$ を更新する．

$$d_i \quad \sim P(d_i)P(\beta_{i+d_i}|\alpha_i) \tag{7}$$

これにより時間差 $d_i$ とルールを表現する確率 $P(\beta|\alpha)$ を同時に推定することができる．以上のように，二者間のインタラクションにおける行動のずれを修正し，その因果関係を確率で表現することで，行動のずれを考慮したルール抽出を行うことが可能となる．

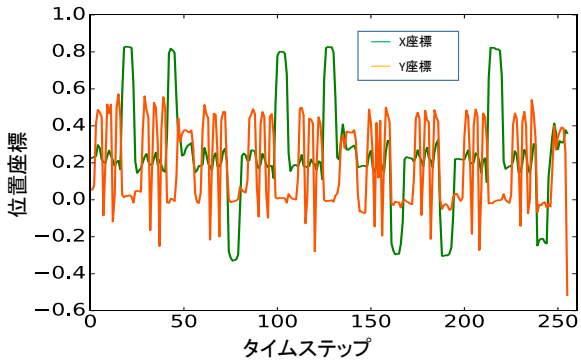


図 4: A の手先座標

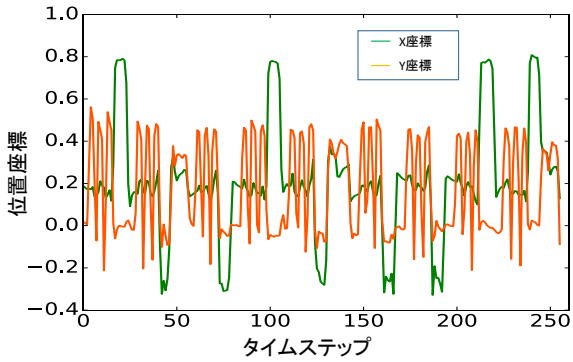


図 5: B の手先座標

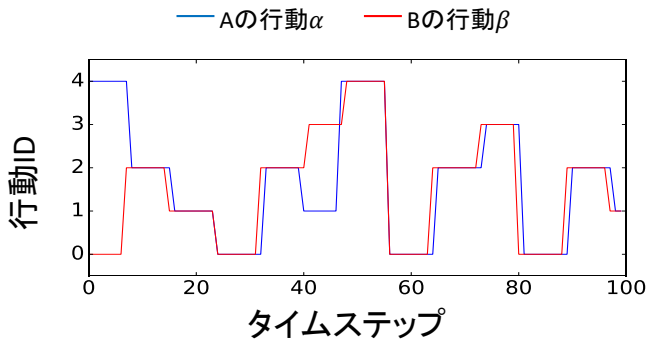


図 6: 分節化データ

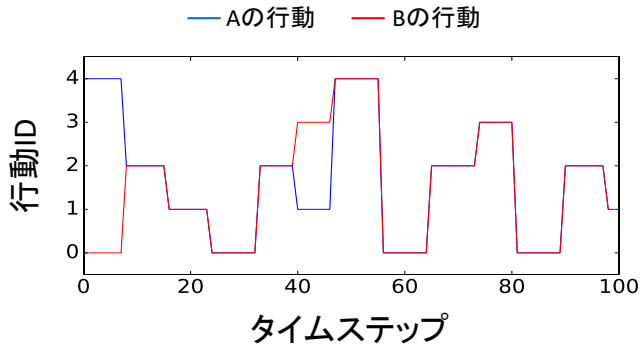


図 7: ずれの修正後

## 3. 実験

提案手法の有効性を検証するために，簡単なルールを組み込んだゲームを作成し，実際にそのゲームを用いて実験を行った．実験には kinect を用いて取得した人の右手の位置を用い，kinect から得られる 30fps の位置座標を，4fps へダウンサンプリングし使用した．この値は，計算量を削減するため，動作の特徴が失われない範囲で経験的に決定した．

### 3.1 方向合わせゲーム

実験には簡単なルールを含む人工的に作成したインタラクションゲームである方向合わせゲームを用いた．ゲームの流れは以下の通りである．

1. 互いに向かい合って右腕を正面に出して待機
2. 互いに腕を三回振る
3. 左右どちらかの方向に大きく腕を振る
4. 互いに腕を振った方向が一致したら 5，一致しなかったら 1 に戻る
5. 互いにハイタッチをして 1 に戻る

### 3.2 結果

図 4，5 に kinect より取得した A，B それぞれの右手の手先座標の時系列データを示す．それぞれ緑線が $x$ 座標を，橙線が $y$ 座標を表している．また，取得した二人の右手の位置座標を GP-HSMM を用いて分節化した結果の最初の 100 フレームが図 6 である．それぞれの行動が分節され行動 ID ごとに分類されており，次のような行動が教師なしで学習された．

0. 手を前に出す動作
1. 手を右にふる動作
2. 手を三回ふる動作
3. 手を左にふる動作
4. ハイタッチをする動作

今回設定したルールと比べると，妥当な行動が学習できたといえる．また，手先座標と分節化結果を比べると，A の行動 $\alpha$ の最初の分節が行動 4 に分類されてしまっており，正解である手を前に出す動作（行動 0）に分類することが出来ていなかったが，その他は正解に近い分節・分類が出来ている．

しかし，図 6 を見ると互いの行動にずれが生じている．これに提案手法であるずれの修正とルール学習を適用すると図 7 のようになった．この図より，ずれが修正されていることが確認できる．学習された行動の因果関係を表す確率は表 1〜3 である．ただし，表 3 は特に確率が高い行動のみを抽出して表記した．表 1 を見ると B が行動 0(手を前に出す) を行った時，A も行動 0(手を前に出す) を行う確率が 96%となっている．この行動と方向合わせゲームのルールを照らし合わせると，ゲームルールにおける互いに手を前に出すというルールが抽出できていることが確認できる．また，B が行動 0(手を前に出す) を行った時，A が行動 4（ハイタッチ）を行う確率が 18%となっているが，これは分節化における前述の誤分類によって生じた確率である．表 2 も表 1 と同様に考えることができ，A が行動 0(手を前に出す) を行った後に，行動 2（手を三回振る）を行う確率が 100%となっている．他にも，A が行動 2(手を三回振る) を行った後に行動 1（手を右に振る）と行動 3（手を左に振る）を行う確率が互いにおおよそ 50%となっており，三

表 1: $P(\beta|\alpha)$

| | | B の行動 ID$\beta$ | | | | |
|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 |
| A の行動 ID$\alpha$ | 0 | 0.96 | 0.01 | 0.01 | 0.01 | 0.01 |
| | 1 | 0.01 | 0.69 | 0.01 | 0.28 | 0.10 |
| | 2 | 0.01 | 0.01 | 0.96 | 0.01 | 0.01 |
| | 3 | 0.01 | 0.36 | 0.01 | 0.61 | 0.01 |
| | 4 | 0.18 | 0.01 | 0.04 | 0.01 | 0.76 |

表 2: $P(\alpha_{i+1}|\alpha_i)$

| | | A の行動 ID$\alpha_{i+1}$ | | | | |
|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 |
| A の行動 ID$\alpha_i$ | 0 | 0.00 | 0.00 | 1.00 | 0.00 | 0.01 |
| | 1 | 0.70 | 0.00 | 0.00 | 0.00 | 0.30 |
| | 2 | 0.00 | 0.53 | 0.00 | 0.47 | 0.00 |
| | 3 | 0.55 | 0.00 | 0.00 | 0.00 | 0.45 |
| | 4 | 0.71 | 0.00 | 0.29 | 0.00 | 0.00 |

回振った後にどちらかに腕を振るというルールを表現できていることがわかる．また，この表の確率が高い遷移として，A が行動 1（手を右に振る）を行った後に，行動 0(手を前に出す) を行う確率が 70%，行動 4(ハイタッチ) を行う確率が 30%となっており，行動 1(手を右に振る) の後には行動 0(手を前に出す) が来る可能性が高いというルールも抽出されている．さらに表 3 を見ると，A が行動 1（手を右に振る），B が行動 1(手を右に振る) をした後に，A が行動 0（手を前に出す）を行う確率が 100%となっている．また，A が行動 1（手を右に振る），B が行動 3(手を右に振る) の後に，A が行動 4（ハイタッチ）を行う確率も 100%となっている．この結果から表 2 では判断できなかったルールも抽出できており，事前に設定したルールを教師無しで正しく学習できていることが分かる．

## 4. おわりに

本稿では人同士のインタラクションを対象にし，GP-HSMM に基づく分節化を通してインタラクションに含まれているルールを学習する基礎的な手法を提案し，その有効性を検証するための実験を行った．実験結果より，提案手法を用いることで二者間のインタラクションゲームにおける動作を教師なしに分節・分類し，お互いの行動の時間のずれも考慮した上で，行動の因果関係を確率で表現することができた．この確率から，因果関係の高いものを抽出することでルールを学習することが出来た．

今後の課題として，より複雑なインタラクションを対象にした実験が挙げられる．今回の実験は簡単な動作とルールを使用し，また行動のずれも小さかったため上手くルール抽出をすることが出来ていると考えることもできる．そのため，より複雑な動作やルールにおいても適用可能かを今後検証する．また，分節化とルール学習を統合したモデルへと拡張することも挙げられる．現在の手法では分節化を行い，その後ずれの修正およびルール抽出を行っている．これらを統合することでそれぞれを相補的に学習することが可能となると考えている．

表 3: $P(\alpha_{i+1}|\alpha_i,\beta_i)$

| | | A の行動 ID$\alpha_{i+1}$ | | | | |
|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 |
| A，B の行動 ID $\alpha_i,\beta_i$ | 0，0 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 |
| | 1，1 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | 1，3 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| | 2，2 | 0.00 | 0.53 | 0.00 | 0.47 | 0.00 |
| | 3，1 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| | 3，3 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |

## 参考文献


[神谷 11] 神谷 友里，吉川はる奈，幼児の役割遊びにおける役割取得の特徴に関する研究，埼玉大学紀要. 教育学部，Vol. 60, No. 2, pp. 19-28, 2011

[谷口 09] 谷口忠大， 岩橋直人，中西弘門，西川郁子，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，人工知能学会全国大会，1F2-OS7-12，2009

[Soltoggio 13] Andrea Soltoggio, Felix Reinhart, Andre Lemme and Jochen Steil，Learning the Rules of a Game: Neural Conditioning in Human-Robot Interaction with Delayed Rewards，IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics，pp. 1-6, 2013

[中村 16] 中村友昭，Attamimi Muhammad，長井隆行，持橋大地，小林一郎，麻生英樹，金子正秀，ガウス過程の隠れセミマルコフモデルに基づく身体動作の分節化，人工知能学会全国大会，1O3-5，2016

[Uchiumi 15] Uchiumi, K., Hiroshi, T., and Mochihashi, D.: InducingWord and Part-of-Speech with Pitman-Yor Hidden Semi-Markov Models, in ACL-IJCNLP，pp. 1774-1782，2015