# 繰り返し囚人のジレンマを題材にした典型他者モデルの獲得

# Acquisition of Typical Other's Mind Models in the Repeated Prisoner's Dilemma

阿部 将樹[1*]　田足井 昇太[1]　長原 令旺[1]　大森 隆司[1,2]　大澤 正彦[1]
Masaki Abe[1],Shota Tatarai[1],Reo Nagahara[1],Takashi Omori[1,2],Masahiko Osawa[1]

[1] 日本大学
[1] Nihon University
[2] 玉川大学
[2] Tamagawa University

**Abstract:** これまでの他者モデル研究では，多くの他者に共通して適用するものや，特定の個体に適用するものが主に検討されてきた. しかし前者は個別の個体への適応ができず，後者は学習に時間がかかる. そこで筆者らは大まかな性格ごとの他者モデルを典型他者モデルと呼び，その有効性を検証してきた. 本研究では，最適な戦略が 2 通りある繰り返し囚人のジレンマタスクの中で典型他者モデルを獲得する方法を設計し，計算機シミュレーションによって有効性について検証した.

## 1　はじめに

近年 AI コンシェルジュやサービスロボットの実用化が進み，人とロボットのインタラクションが増加している．そのような場面で他者の心的状態や行動の予測を行うことで円滑なコミュニケーションを行うことが可能になる．そこで他者の心的状態や行動を予測することを目的とした他者モデルが提案されている [1, 2].

他者モデルとはインタラクションする相手の心的状態を推定し，その行動を予測するためのモデルである. 個々人の他者に対しては，その他者とインタラクションを重ねることで個別の他者モデルを形成することができ，円滑なコミュニケーションや関係性の構築が可能になる．しかし個別に他者モデルを形成するためには，個々の相手に対してそれぞれ大量のインタラクションや行動データが必要になるため，それを獲得しながら運用することは現実には困難である.

一方で他者モデル研究では，個々人に対してではなくインタラクションをした全ての他者に対するための平均的な他者モデルを用いている場合も多い．平均的な他者モデルを一つ作ることができて幅広い他者に対応できるとすれば有用性は高いが，個々の他者の性格の違いなどを扱うことは難しく，個別の他者に対して高い精度で心的状態や行動の予測をすることは難しい.

*連絡先：日本大学文理学部
〒 156-8550 東京都世田谷区桜上水 3-25-40
E-mail: chma22006@nihon-u.ac.jp

そこで著者らは大まかな性格ごとの他者モデルとして典型他者モデルを提案した [3]．典型他者モデルは全ての他者の平均的な予測の用いる他者モデルと特定他者に対して形成した他者モデルの中間的な位置付けであり，平均的な他者モデルより個人に対する適応能力が高く，個別に形成した他者モデルよりも少ないインタラクションで獲得することが期待できる.

著者らは，これら 3 種類の異なる他者モデルを，平均他者モデル · 典型他者モデル · 個別他者モデルと呼ぶ．平均他者モデルは全他者の平均的なモデルのことを指し個人適応はできず，他者の行動予測の精度も高くはない．典型他者モデルは他者をいくつかの典型的な性格ごとに分類し，その性格ごとに調整された他者モデルとしている．インタラクションの相手の行動特性に適した典型他者モデルを選んで当てはめるため，部分的に個人適応が可能になると同時に学習も早いことが期待される．個別他者モデルは，個々の他者の性格に対して個別に形成するため，インタラクションを重ねるたびに変化していき高い精度の個人適応が可能になる.

典型他者モデルに関しては，長原らが事前に形成した典型他者モデルを用いて他者の性格に合わせた行動を素早く取れるようになることを示している．しかし典型他者モデルの獲得に関しては述べていない [3].

ここまでの経緯を受けて本研究では，二人のゲームプレイヤーが両者にとって最適な戦略が 2 通りある繰

り返し囚人のジレンマタスクを行う場面を題材に，一方のプレイヤーエージェントが相手の典型他者モデルを獲得することで他者の戦略に対応した行動を素早くとれるようになることを，計算機シミュレーションによって検証したので，その結果を報告する．なお本論文では性格とは他者モデル内で扱う心的状態/行動の起こりやすさと定義し，戦略と性格は同義とする．

# 2 背景

## 2.1 他者モデル

他者モデルとは相手の心的状態や行動を予測する手法を計算モデル化したものであり, これまでさまざまな研究が行われてきた．横山らは，2 体の獲物を 2 体のエージェントが 1 体ずつ捕獲するハンタータスクにおいて，それぞれのエージェントに他者の意図を推定してその行動を予測する他者モデルを持たせることで，タスクの解決が容易になることを示した [4]．しかし形成されたのは個別他者モデルであり，他の他者に対して適応するためには新たに学習を行う必要があるため学習に時間がかかってしまうという問題があった．

また阿部らは，子供と遊ぶロボットに他者モデルを持たせた [5]．そして実際に子供と遊ぶ過程で相手の心的状態を推定し，ロボットの行動を変化させる実験を行い，その有用性を示唆した．しかしこの他者モデルは様々な子供とのインタラクションを学習に用いた平均他者モデルであり，適応できる他者と適応できない他者が存在するという課題が残った．

## 2.2 典型他者モデル

著者らは学習にかかる時間を減らしつつすばやく個人に適応する方策として，他者の大まかな性格ごとに他者モデルを用意する方式を考え，それを典型他者モデルと呼んでいる．長原らは 3 種類の性格ごとに行動の起こりやすさが異なるエージェントと各性格の典型他者モデルを持ったエージェントで繰り返し囚人のジレンマゲームを行った [3]．その事例では，典型他者モデルの切り替えを行いながら他者の性格を推定し，その推定に基づいた自己の行動決定を行った．その結果，事前に用意されている性格に対しては適応が可能であることを示した．しかし典型他者モデルはあらかじめ作り込んだものを使用しており，典型他者モデルの獲得に関しては未解決のままとなっている．それを受けて本研究では，繰り返し囚人のジレンマゲームを題材として典型他者モデルの獲得を行う．そして，その典型他者モデルを用いて新たにゲームを行う他者の行動を推定し，自身の行動決定を行うことが可能であることを示す．

表 1: 利得票

| A\B | 協力 | 裏切り |
|---|---|---|
| 協力 | 5 \5 | 0 \10 |
| 裏切り | 10 \0 | 1 \1 |

## 2.3 競合学習

競合学習 (Competetive Learning) とは，教師なし学習で多く使われる学習方式で，入力データに対して最も反応したニューロンを入力データに近づけるように更新する学習アルゴリズムである．入力に対して最も反応するという点が競合となっているため，この名前がついている．本研究では典型他者モデルを獲得するための手法として競合学習を用いた．

## 2.4 繰り返し囚人のジレンマゲーム

囚人のジレンマゲームは 2 体のプレイヤーを対象としたゲーム理論のモデルである．人間の行動や経済的な意思決定のモデル研究において，計算機シミュレーションにより頻繁に使用されている．本研究では，2 体のプレイヤーは「協力」·「裏切り」のどちらかを選択し，利得表 1 のような報酬が与えられる．両者が協力を選択した場合は互いに 5 点を得られる．片方が協力でもう一方が裏切りを選択した場合は協力を選択したプレイヤーは 0 点，裏切りを選択したプレイヤーは 10 点が利得として与えられる．両者が裏切りを選択した場合は利得は互いに 1 点となる．

このゲームでは，ナッシュ均衡の状態がお互いに裏切りをする場合であるため，1 回で終わるゲーム条件では 2 体とも裏切りを選択してしまうが囚人のジレンマゲームを繰り返し行うことで他者の心的状態を考慮した上で行動を選択し得点を伸ばすことが可能となる．

# 3 競合学習を用いた典型他者モデルの獲得

本研究における典型他者モデルとは戦略ごとの行動の起こりやすさを表す．そのため他者の行動確率 $MM_p$ を学習によって獲得をする．学習方法は競合学習の手法を用いており複数の典型他者モデルの行動確率と実

際の行動確率の類似性が高い方の値を以下の式 1 で更新する．類似性は 4.2 章で述べる手法を用いて測る．

$$MM_p = \alpha P_i + (1-\alpha) MM_p \tag{1}$$

# 4 実験

本実験は繰り返し囚人のジレンマゲームで戦略ごとの典型他者モデルを獲得し，その典型他者モデルを用いて新たにゲームをする他者の戦略を推定しながら適応した行動を取ることが可能であるか検証することを目的としている．

## 4.1 実験設定

本実験では典型他者は協力的エージェントと交互エージェントの 2 種類を想定する．協力的エージェントは 0.9 の確率で協力を選択し，交互エージェントは協力と裏切りを交互に繰り返すエージェントである．協力的なエージェントは互いに協力を出し合うことで継続的に得点を得ることを狙っており，交互に繰り返すエージェントは互いに協力と裏切りを出し合うことで交互に得点を得ることを狙っている．このような設定では，典型他者モデルとして獲得されるのは他者エージェントの協力行動の行動確率 P になる．

実験は学習フェーズと検証フェーズの 2 フェーズで行った．学習フェーズでは協力と交互のエージェントを 10 体ずつ用意し，検証フェーズではそれぞれ 5 体ずつ用意した．3 章の手法を用いて学習係数 $\alpha = 0.01$ で学習を行った．自己は学習フェーズ，検証フェーズで各エージェントと 25 回ずつ繰り返し囚人のジレンマゲームを行った．検証フェーズでは獲得した典型他者モデルと実際の行動確率を元に，相手がどちらの典型他者モデルに近いエージェントであるか推定する．そして推定した典型他者モデルに基づき自己の行動決定を行う．相手エージェントのモデルの推定方法は 4.2 章の手法を用いて類似度を求め，より近い，すなわち類似度の高いモデルを相手の典型他者モデルとする．本実験では，それぞれの典型他者にどの行動が最適かは事前にわかっているものとした．すなわち，他者が協力確率が 0.9 に近い典型他者であると推定した場合には，その典型他者モデルの確率の通りに自己の行動決定をおこない，協力確率が 0.5 に近い典型他者であると推定した場合には，他者エージェントの次の行動を予測して相手とは逆の手を出すようにあらかじめ設計をした．

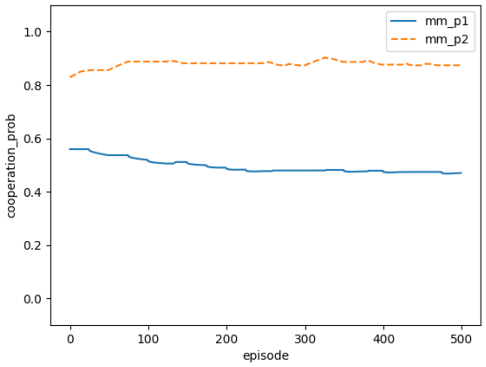


図 1: 獲得フェーズ：典型他者モデルの確率遷移

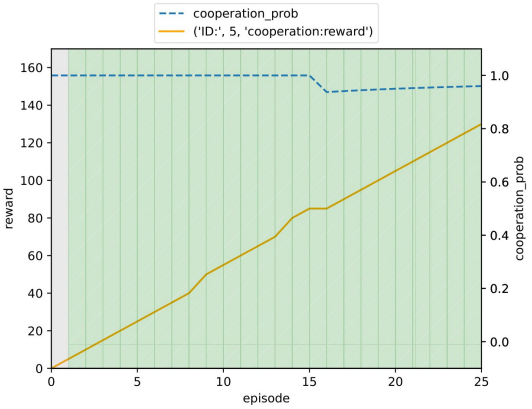


図 2: 協力的エージェントとのゲームの結果

## 4.2 モデルの類似性

検証フェーズでは，2 つの典型他者モデルと実際の他者エージェントの行動確率を KL ダイバージェンス (Kullback-Leibler divergence) を用いて比較する．KL ダイバージェンスとは 2 つの確率分布の類似性を求める尺度であり，以下の式で求められる． ここで $P$ は実際の行動確率分布，$Q$ は典型他者モデルの確率分布，$a_1$ と $a_2$ は協力と裏切りをそれぞれ表す．

$$KL(P|Q) = \sum_{X=a_1,a_2} P(X) \log \frac{P(X)}{Q(X)}$$

## 4.3 実験結果

提案手法による典型他者モデルの獲得の結果を図 1 に示す．破線は協力の行動確率が 0.87 に収束した典型他者モデルの遷移を表している．協力エージェントの

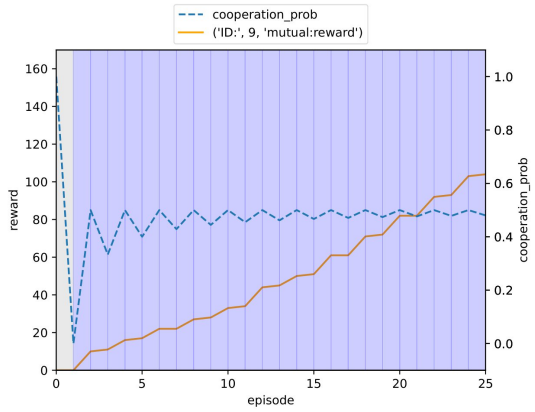


図 3: 交互エージェントとのゲームの結果

協力行動確率が 0.9 であるため典型他者モデルが獲得できたと言える．実線は協力の行動確率が 0.47 に収束した．交互エージェントは協力行動確率が 0.5 であるため典型他者モデルが獲得されている．

次に，獲得した典型他者モデルにより他者の戦略を推定しつつ行った繰り返し囚人のジレンマゲームの結果を図 2, 図 3 に示す．背景色がどちらの典型他者モデルと推定したかを表している．1 ゲーム目はランダムに行動を決定しているためグレーとなっており，2 ゲーム目以降は緑色が協力確率が 0.87 の典型他者モデルとの推定を，青色が協力確率が 0.47 の典型他者モデルとの推定を表している．凡例にエージェント ID とエージェントの戦略を表記した．この図から早い段階で他者の正しい戦略が推定できていることがわかる．ただし協力的エージェントの場合，裏切りの確率が上振れて交互エージェントと推定されることもあった．破線は相手のエージェントの実際の行動確率の遷移を表しており実線は自己の獲得点数の遷移を表している．

## 5　考察

協力的エージェントに対しては継続的に 5 点を獲得できているため，相手の戦略を正確に予測し，その戦略に適した行動を選択することができている．交互エージェントに対しては，確率のゆらぎの影響で行動選択が相手と合わない場合もあるが，おおよそ狙った通りの戦略での行動が実現できた．

今回は競合学習により典型他者モデルとして協力行動確率のみを獲得した．そのため，相手が出した手によって行動の確率が変わってくるような戦略に対してはそのままでは適応できないことが予想される．今後，典型他者モデルの獲得方式はタスクなどに応じて考える必要がある．

また実験では獲得する典型他者モデルは 2 個になるように事前に設計していた．いずれは，他者行動の分布に従って典型他者モデルの数を調整できる手法が必要になると考える．また，典型他者モデルの獲得に競合学習を用いたことで，ゲームを行いつつ学習を進めることができた．そのため，学習エージェントの戦略に偏りがあった場合でも理想とする典型他者モデルを獲得できる．

## 6　おわりに

本研究では競合学習を用いて典型他者モデルの獲得を行い，繰り返し囚人のジレンマゲームを題材に典型他者モデルの有効性を検証した．その結果，確率で表現できる典型他者についてはそのモデルを獲得して新規の他者の戦略が推定できた．今後は典型他者モデルの数の自動調整や，より複雑な確率分布に従う他者についても汎化できる典型他者モデルの形式を検討したい．

## 参考文献

[1] 佐々木 康輔, 田足井 昇太, 森口 昌和, 野田 尚志, 大森 隆司, 宮田 章裕, 大澤 正彦 “自己・他者モデルの枠組みによる代理存在の実現にむけたキャンパスガイドシミュレーションによる検討” HAI シンポジウム, 2022.

[2] 大澤 正彦, 奥岡 耕平, 坂本 孝丈, 市川 淳, 今井 倫太, “認知的インタラクションフレームワークに基づいた他者モデルの提案” HAI シンポジウム, 2020.

[3] 長原 令旺, 田足井 昇太, 佐々木 康輔, 大森 隆司, 大澤 正彦 “繰り返し囚人のジレンマゲームを題材とした典型他者モデルの切り替えによる個人適応” HAI シンポジウム 2022.

[4] 横山 絢美, 大森 隆司 “協調課題における意図推定に基づく行動決定過程のモデル的解析” 電子情報通信学会論文誌 A Vol. 92, No.11, pp.734–742, 2009.

[5] 阿部 香澄, 岩崎 安希子, 中村 友昭, 長井 隆行, 横山 絢美, 下斗米 貴之, 岡田 浩之, 大森 隆司 “子供と遊ぶロボット: 心的状態の推定に基づいた行動決定モデルの適用” 日本ロボット学会誌 Vol. 31, No.3, pp.263–274, 2013.