

## 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

### Personal Adaptation by Switching Typical Other 's Mind Model Using The Repeated Prisoner's Dilemma Game

長原 令旺<sup>1\*</sup> 田足井 昇太<sup>2</sup> 佐々木 康輔<sup>1</sup> 大森 隆司<sup>1</sup> 大澤 正彦<sup>1,2</sup>  
Reo Nagahara<sup>1</sup>, Shota Tatarai<sup>1,2</sup>, Kosuke Sasaki<sup>1</sup>, Takasi Omori<sup>1</sup>, Masahiko Osawa<sup>1,2</sup>

<sup>1</sup> 日本大学

<sup>1</sup> Nihon University

<sup>2</sup> 専修大学

<sup>2</sup> Senshu University

**Abstract:** 従来の研究では、全他者の平均的な他者モデルや他者の性格に対して個別に形成する他者モデルを構築する研究が行われてきた。しかし、前者は個人適応できず、後者は個別の他者とのインタラクションや学習データ数が多く必要である。本研究では、典型的な性格グループごとに他者モデルを構築し、繰り返し囚人のジレンマゲームにおいて、他者の性格に類似したモデルの切り替えを行うことで、比較的高速かつ高精度に個人適応する方法を提案する。

## 1 はじめに

近年ではサービスロボットの技術開発や実用化が進み、人とロボットのインタラクションが増加している。一方で、ロボットカウンセラーのようなコミュニケーションロボットも注目を集めている。そのようなロボットの多くは表情や声の抑揚から人の心的状態や行動を推定している。しかし、顔で笑って心で泣くというように、表情からは必ずしも人の心的状態を推定できず、結果として正確な行動の推定もできないことがある。本研究の目的は、現状の技術レベルで社会実装可能な人の心的状態や行動の推定方法を確立することである。

心的状態や行動を推定する方法として、他者モデルの研究 [1, 2] が行われている。他者モデルとは、コミュニケーション相手の心的モデルであり、環境からの情報により相手の心的状態を推定し、行動を予測するためのものである。インタラクションする他者一人ひとりに対して個別に他者モデルを形成することができれば円滑なコミュニケーションや人との関係構築することが可能である。

しかし、これらの研究で扱われた他者モデルは一人ひとりのモデルではなく、人の平均的なモデルである。人の平均的なモデルでは、他者の性格の違いを扱うことができず、一人ひとりに対して高い精度で心的状態

を予測することは難しい。一方で一人ひとりに対して個別に他者モデルを構築するためには、一人ずつに対して大量のデータやインタラクションの経験が必要であるため、個人の性格に応じた他者モデルの形成を社会実装することは困難である。したがって性格が異なる複数の他者に対して、一定以上の予測の精度と、高速な適応を担保できる他者モデルの枠組みが必要である。なお本論文において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する。

本研究では、全ての他者の平均的な他者モデルと、一人ひとり個別の他者モデルの中間的位置づけとなる、典型的な性格グループごとの他者モデルである典型他者モデルを提案する。典型他者モデルは、平均的な他者モデルよりも個人への適応性能が高く、個別の他者モデルよりも必要とする学習データが少なく高速に獲得することが期待できる。実験では、典型他者モデルの枠組みを活用して繰り返し囚人のジレンマゲームを用いた計算機シミュレーションを行った。このゲームの中で典型他者モデルの枠組みで他者に適応できる可能性の評価や、フレームワークの性能の評価を行う。

\*連絡先：日本大学文理学部  
〒156-8550 東京都世田谷区榎上水 3-25-40  
E-mail: chre19060@g.nihon-u.ac.jp

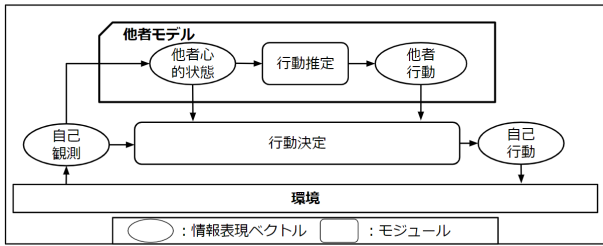


図 1: 他者モデルアーキテクチャ

## 2 背景

### 2.1 他者モデル

既存の他者モデルの工学研究 [1] では事前に他者モデルを作り込み、心的状態や行動を予測できても、人の性格に応じて他者モデル内のパラメータは変化しないため他者に適応することは困難であった。

大澤ら [3] は図 1 に示す他者モデルアーキテクチャを提案している。エージェントは環境からの情報をもとに他者の心的状態を推定し行動を予測する。そして、他者モデルに基づき、自身の行動を決定する。すなわち、自身の行動は他者の行動の予測結果に基づき計画する。

### 2.2 Accumulator Based Arbitration Model (ABAM)

人は行動を変化させるときに過去の経験から行動の選択を行っている。そして、人の脳に存在する神経細胞はアキュムレーターとしてモデル化できる可能性が示唆されている [4, 5]。アキュムレーターモデルとは、ある行動を引き起こすための証拠を蓄積していくモデルであり、蓄積された証拠が設定した閾値を超えたとき、そのモジュールを採用する。

Accumulator Based Arbitration Model (ABAM) を図 2 に示す。

## 3 典型他者モデルの設計

### 3.1 典型他者モデルの概要

典型他者モデルとは、特定の性格ごとに形成された複数の他者モデルである。エージェントはインタラクションの途中で他者の性格に応じて、あらかじめ形成していた典型他者モデルのうちの 1 つを選び出して利用する。この性質によって、平均的な唯一の他者モデルを形成する方法では全くできなかった、他者の性格に対する適応が、部分的に可能となる。一方、学習時には似た性格の複数の他者のデータに基づいてあらか

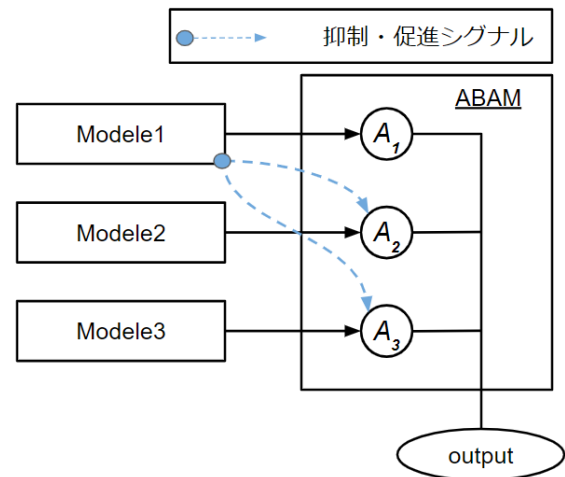


図 2: ABAM

じめ学習ができるため、一人ひとりに対する個別の他者モデルを形成するよりも現実的であるうえ、初めてインタラクションする他者に対しても高速に適応できる可能性がある。

### 3.2 ABAM を利用したモデル切り替え

典型他者モデルに必要な特有の要素技術として、他者の性格を分類し、形成したどの典型他者モデルを割り当てるかを判断する機構が挙げられる。そこで当該機構の実現方法として、2.2 節で説明した ABAM を応用する。

本提案において行動を引き起こすための証拠として、式 (1) 内の各典型他者モデルの予測誤差を集約した  $y_n^t$  を用いる。累積値  $A_n^t$  が、式 (2) 中の閾値  $th_n$  を超えたとき、あらたに典型他者モデル  $n$  を採用する。あらたな典型他者モデルが採用された際、直前まで選択されていた典型他者モデルの予測誤差が大きくなり、切り替わったときにその典型他者モデルに対応する閾値  $th_n$  を上昇させる。これによって一度適用したが予測エラーが大きかった典型他者モデルが再度選ばれにくくなるように調整する。

$$A_n^t = rA_n^{t-1} + y_n^t \quad (1)$$

$$A_n^t > th_n \quad (2)$$

ここで、 $r$  が割引率である。 $A_n^t$  は  $A_n^t$  に対応するモジュールの出力であり、 $y_n^t$  は  $A_n^t$  に対応するモジュール以外の予測誤差の合計値とする。本実験では、モジュールの出力はエージェントが典型他者モデルを選択したということである。また安定化の工夫として、 $A_n^t$  が閾値を超えた場合、 $A_n^t$  を 0 にリセットすることで同じモジュールが連続して出力されないようにする。

表 1: 利得票

A\B	協力	裏切り
協力	5 \ 5	0 \ 10
裏切り	10 \ 0	2 \ 2

## 4 評価実験

### 4.1 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り、人間の行動や経済の動きをシミュレーションするための手法として使用されている。2体のプレイヤーA,Bは「協力」または「裏切り」のどちらかを選択し、その利得として表1のような報酬が与えられる。両者が協力を行ったときは、お互いの利得点は5点となる。一方で、両者が裏切りの行動を行ったときは、お互いの利得点は2点となる。どちらかのプレイヤーが協力を選択し、もう一方のプレイヤーが裏切りを選択した場合は、裏切りを選択した方に10点となる。しかし、囚人のジレンマゲームでは、ナッシュ均衡がお互い裏切りのため、他者の心的状態を考慮せずにゲームが行われると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次の行動を選択することが可能となる。

### 4.2 実験目的

本実験の目的は、提案した枠組みによって他者の行動から他者の性格に合わせた典型他者モデルを選定できるか、また選定の際に提案手法がどのような挙動をみせるかを確認することである。

### 4.3 実験手順

繰り返し囚人のジレンマゲームは以下の手順で行われる。プレイヤーAは他者の心的状態に合わせて「協力」または「裏切り」を選択し、プレイヤーBは決められたルールで「協力」または「裏切り」を選択する。1ゲームにつき100試行する。1試行における勝負の結果、利得票にしたがってエージェントは得点を得る。

1ゲームごとにプレイヤーAが獲得した得点と選択した典型他者モデルを色に分けてグラフにした。

### 4.4 エージェントの設定

プレイヤーAを自己エージェントとする。プレイヤーAは3つの他者モデルを所有している。1つ目は協力的な性格に基づく典型他者モデル (Cooperative), 2

つ目は非協力的な性格に基づく典型他者モデル (Non-Cooperative), 3つ目はしっぺ返し戦略を行う性格の典型他者モデル (Tit for tat) である。しっぺ返し戦略は1手目に協力をを行い2手目以降は相手の前回の手を真似る戦略である。プレイヤーAはゲーム開始時、相手がどのような他者なのかを見極める必要があるためランダムに行動する。そして、相手の性格に合わせて典型他者モデルの切り替えを行い、典型他者モデルによる予測に基づく行動を行う。

プレイヤーBを他者エージェントとする。プレイヤーBは後述する6つのルールに従って行動する。

ここで、プレイヤーBがプレイヤーAにとって学習済みのエージェント (i.e. 3つの典型他者モデルのいずれかで行動の予測が可能なエージェント) だった場合の予想される振る舞いについて整理する。もし、プレイヤーAが、プレイヤーBにしっぺ返しの典型他者モデルを当てはめているときに、お互いが協力的行動を続けることで安定することが想定される。この際、採用している他者モデルの予測誤差は0となる。しかし一方で、相手に協力的な性格の典型他者を割り当てた場合にもこの状況で発生する予測誤差は0となる。さらにプレイヤーBが実際に協力的な性格のエージェントだった場合、プレイヤーAは裏切りをした方がスコアが高くなる。同様に、互いに裏切り行動をとり続けて安定している場合にも、プレイヤーBは非協力的な性格と、しっぺ返し戦略の両方の可能性が考えられ、場合によってはより獲得スコアが高い協力的行動をお互いに取り合うことができる可能性がある。このような状態を提案手法が適切に判断し、より高いスコアを獲得できるかが1つの論点となる。

### 4.5 実験条件

本研究では、1ゲームごとにプレイヤーBの性格を変更し以下6つの条件で繰り返し囚人のジレンマゲームを行う。

条件1: プレイヤーBが協力的な行動をする。

条件2: プレイヤーBが非協力的な行動をする。

条件3: プレイヤーBがしっぺ返し戦略をする。

条件4: プレイヤーBが協力と裏切りを交互に選択し行動する。

条件5: プレイヤーBが3割の確率で裏切りの行動をする。

条件6: プレイヤーBが序盤に協力的な行動をし最後に搾取行動をする。

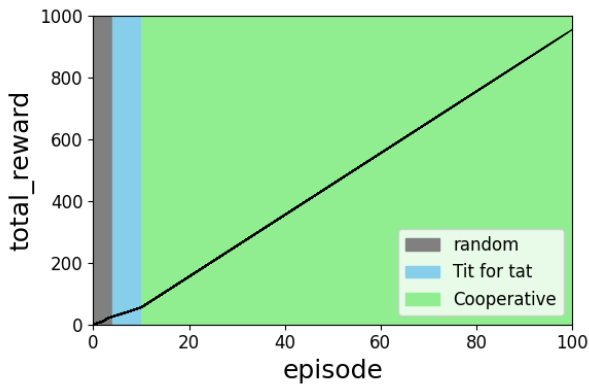


図 3: 条件 1: 協力行動をするプレイヤー B に対してプレイヤー A の予測が正判定

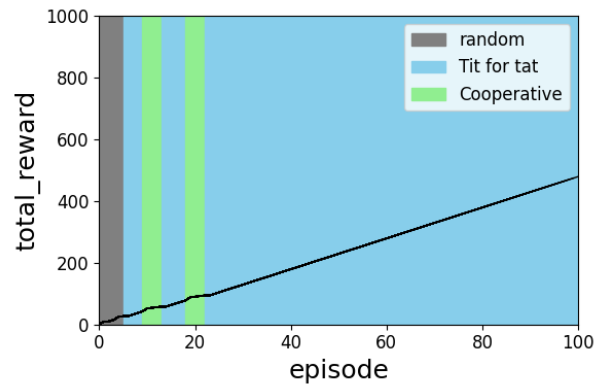


図 5: 条件 3: しっぺ返し戦略をするプレイヤー B に対してプレイヤー A の予測が正判定

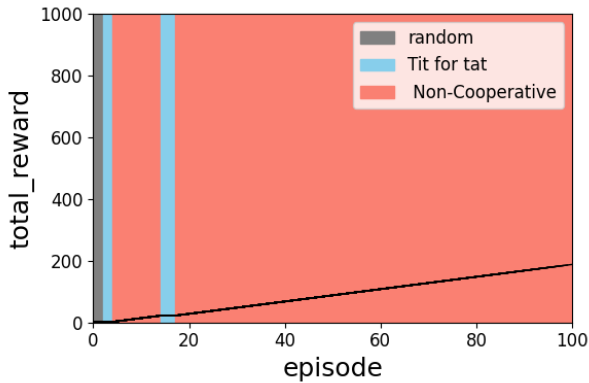


図 4: 条件 2: 非協力的な行動をするプレイヤー B に対してプレイヤー A の予測が正判定

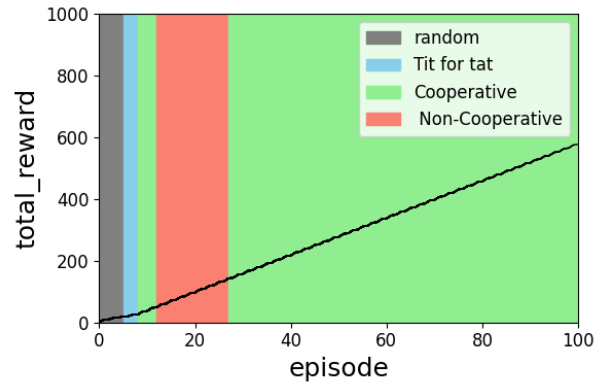


図 6: 条件 4: 協力と裏切りを交互に行動するプレイヤー B に対してプレイヤー A は協力的だと予測

## 4.6 実験結果

1 ゲームごとにプレイヤー A が獲得した得点と選択した典型他者モデルを色に分けてグラフにした。

### 4.6.1 条件 1: プレイヤー B が協力的な行動をする。

結果を図 3 に示す。図 3 を見ると協力的だと推定するモデルに収束したことがわかる。プレイヤー B は協力的な行動をしているのでプレイヤー A が選択したモデルは正しいということがわかる。

### 4.6.2 条件 2: プレイヤー B が非協力的な行動をする。

結果を図 4 に示す。図 4 を見ると非協力的だと推定するモデルに収束したことがわかる。プレイヤー B は非協力的な行動をしているので、プレイヤー A が試行錯誤しながらモデルを切り替えた結果、正しい選択をしていることがわかる。

### 4.6.3 条件 3: プレイヤー B がしっぺ返し戦略をする。

結果を図 5 に示す。図 5 を見ると、ランダムな行動から他者を協力的だと推定するモデルだと判断して行動していることがわかる。そのときの得点が伸びていないため、推定したモデルが正しくないと判断し、別のモデルに切り替えている。その結果、適切に他者を見極めることが可能となっている。

### 4.6.4 条件 4: プレイヤー B が協力と裏切りを交互に選択し行動する。

結果を図 6 に示す。図 6 を見ると協力的だと推定するモデルに収束したことがわかる。自身が想定していない他者モデルが存在しているとき、少なからず典型他者モデルを一つ当てはめなければならない。相手を協力的エージェントだと判断しているとき、協力的だと推定するモデル以外の典型他者モデルの予測が同じになってしまった。そのため、協力的だと推定するモ

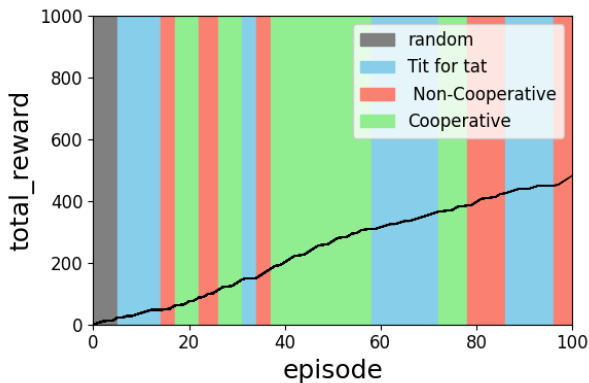


図 7: 条件 5: 3 割で裏切る行動をするプレイヤー B に対してプレイヤー A は判定不能

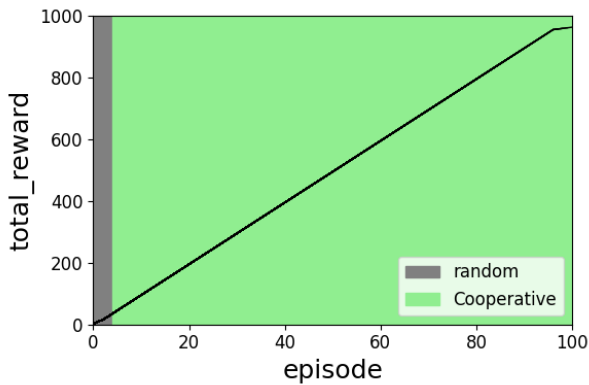


図 8: 条件 6: 最後搾取行動するプレイヤー B に対してプレイヤー A は協力的だと予測

デル以外の予測誤差が集まってしまったため協力的だと推定するモデルに収束したと考えられる。

#### 4.6.5 条件 5: プレイヤー B が 3 割の確率で裏切りの行動をする。

結果を図 7 に示す。図 7 を見ると複数の典型他者モデルを試行錯誤しているが中々定まらない結果となった。相手が確率で動くとき、乱数に依存して行動するため正確なモデルを選択できていないことがわかる。この点に関しては今後提案手法を拡張して対応できるように発展させていく必要がある。

#### 4.6.6 条件 6: プレイヤー B が序盤に協力的な行動をし最後に搾取行動をする。

結果を図 8 に示す。図 8 を見ると最後に搾取をされているが協力的だと推定してしまっていることがわか

る。1 ゲームだけでは他者の心的状態に合わせたモデルを選択しただけで終わってしまうが、他者のモデルを獲得することで、2 ゲーム目からは同じような搾取は受けないと考えられる。すなわち、1 ゲームのみのインタラクションのときは、他者を性格別に分類することが典型他者モデルの切り替えの他者への適応が限界であるといえる。

## 5 考察

ABAM を用いてシミュレーションを行った結果、プレイヤー A が認知している性格ならば正しいモデルに切り替えることが分かった。しかし、不明な性格には適応できなかったことがわかった。人と人とのインタラクションでは、自分が想定していない性格の相手であることが少なからず考えられる。そのため、インタラクション中に自分が想定していない性格のモデルを分類したいときは、新たな性格のモデルの構築ができるようなアルゴリズムが必要になる。

また、本実験では 3 割で裏切るエージェントに対して、典型他者モデルをもつエージェントがモデルの選択を試行錯誤しているようなグラフが示された。他者が確率で行動するとき、どの程度の割合でこの行動をするのか過去のデータを取りながらインタラクションする必要があると考える。したがって、確率で行動する他者の性格を分類するためには、大量のインタラクション数が必要になることが予想される。

一方で他者の性格に合わせて適応できるのかシミュレーションを行ったが、協力的な行動をするエージェントが最後に搾取をしても協力的だと判断してしまうことが示された。今後、個別に合わせて他者モデルを形成していくときには、「この他者は協力的な典型他者モデルに分類できるが、最後に裏切る性格だな」というような学習が必要になってくる。そのため、シミュレーションを通してその場面に適した他者の性格のパラメータの追加や調整が必要であると考えられる。

## 6 おわりに

本研究では、あらかじめ性格ごとに典型他者モデルを構築し、繰り返し囚人のジレンマゲームにおいてモデルの切り替えを行うことで他者に適応できる可能性及び、フレームワークの性能の評価を行った。その結果、典型他者モデルでは性格ごとには適応できたが、個別に適応することができないということが明らかとなった。今後典型他者モデルからどのように個別に適応させていくか検討する。

## 謝辞

本研究は JSPS 科研費 JP20K23351 の助成を受けたものです。

## 参考文献

- [1] 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
- [2] 港 隆史, トーマス デール, 吉川 雄一郎, 石黒 治, “予測性を評価した選好性に基づく模倣発達メカニズムの提案” 日本ロボット学会誌, Vol.28 No.8, pp.1047-1057, 2010.
- [3] 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.
- [4] Mazurek, M.E., Roitman, J.D., Ditterich, J., Shadlen, M.N.: A role for neural integrators in perceptual decision making. *Cerebral cortex* vol. 13, no. 11, pp. 1257–1269 (2003)
- [5] Hanks, T.D., Kopec, C.D., Brunton, B.W., Duan, C.A., Erlich, J.C., Brody, C.D.: Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature* vol. 520, no. 7546, pp. 220–223 (2015)