個別の性格に適応できる典型他者モデルの分類
Classification of typical models of other that can adapt to individual personalities

長原 令旺 †，大澤 正彦 †
Reo Nagahara, Masahiko Osawa
† 日本大学
Nihon University
chre23013@g.nihon-u.ac.jp

概要

他者モデルは，他者の心的状態や行動を予測するためのモデルである．他者モデルは「平均他者モデル」「典型他者モデル」「個別他者モデル」の 3 種類に段階別に分類できる．平均他者モデルは一般的な行動パターンを捉える基礎であり，典型他者モデルは性格ごとに分類されて高い予測精度を持つ．個別他者モデルは特定の個人に基づき，最も精度が高い．本論文では典型他者モデルを 4 つに細分化し，それらの特性について議論した．

キーワード：他者モデル，認知アーキテクチャ，性格推定

1. はじめに

人が日常生活で他者と円滑にコミュニケーションをとるためには，他者の心的状態や行動を予測する必要がある（日永田他, 2018）．しかし，現在のコミュニケーションロボットにおいて他者の意図や感情を正確に理解することは難しく，それらを誤って解釈することは誤解や対立を引き起こす要因となりうる．この課題を解決するために，著者らは他者モデルという他者の心的状態や行動を予測・解釈するためのフレームワークの研究を行っている (大澤他 2020).

Dennett の提唱する意図スタンスは，他者モデルの理解において重要な概念である（Dennett et al. 1989）．意図スタンスは，他者の行動をその意図に基づいて解釈し，予測するための枠組みである．Bratman の意図の理論（Bratman et al.1987）や Rao と Georgeff の BDI モデル（Belief-Desire-Intention モデル）も同様に，他者の行動選択をその内部状態に基づいて説明している．HAI の分野で他者モデルに関する研究が多く行われてきた (横山他 2009,2011. 長田他 2006.)．また，他者モデルは「平均他者モデル」「典型他者モデル」「個別他者モデル」の 3 つに分類できることを説明してきた（阿部他 2023）．それぞれの他者モデルは異なる特性を持ち，他者とインタラクションする状況や精度において使い道が異なる．

平均他者モデルは，これまで関わってきた人を平均化した他者モデルであり，一般的な行動パターンを捉えるための基礎となる．そのため，個別の他者に対して他者の行動を予測する精度が低く，情報が限られているインタラクションの初期段階で主に利用される．

典型他者モデルは，これまでかかわってきた人を特定の性格に基づいて分類し，それぞれの性格ごとに構築された他者モデルである．ここで性格とは心的状態や行動の起こりやすさとする．これは，平均他者モデルと比べて高い予測精度を持つ．数回のインタラクションで構築済みのモデルの中から適切なものを選択するため，比較的高速で高精度に予測することが可能である．また，著者らは比較的高速で他者の性格予測精度が良い典型他者モデルに注目して研究してきた（長原他 2022.）．

個別他者モデルは，特定の個人に対して構築されたモデルであり，過去のインタラクションのデータに基づいて構築される．そのため，個別の他者モデルを作成する際には長期的なインタラクションが必要なため，瞬時に適応することができないが，最も精度の高い予測が可能となる．

インタラクション開始前から開始直後，その後の長期的なインタラクションまでの様々なフェーズがある．各フェーズにおいて，より高速かつ高精度な他者への適応を実現するためには，平均・典型・個別の他者モデルをそれぞれ組み合わせることが有効と考えられる．つまり，開始前には平均他者モデルを活用し，開始後数回のインタラクションの中で典型他者モデルを選びながら活用して，長期的なインタラクションが続く場合には個別他者モデルを構築するといった運用が想定される．

本研究では，さらに典型他者モデルを詳細に分類し，それぞれの特性について議論する．

2. 典型他者モデルの分類

2.1 典型他者モデルの分類方法

本研究では，典型他者モデルの分類方法として 2 種類考える．1 つ目の分類方法は，典型他者モデルの獲得時の性質による分類であり，2 つに分類する．1 つはインタラクションを通じて蓄積される経験から構築されるものである．もう 1 つは外部から得られる情報や定義に基づいて構築するもので，獲得過程において自ら繰り返しのインタラクションを必ずしも体験する必要がない．前者を獲得時 N ショット，後者を獲得時 0 ショットと表記することとする．

2 つ目の分類方法は，典型他者モデルの選択時の性質による分類であり，典型他者モデルがどのような情報に基づいて選択されるかによって 2 種類に分類する．1 つは，繰り返しインタラクションを行うことを通して推測・選択される性格に基づく分類方法である．もう 1 つは，見た目や肩書き，評判など観測可能な情報を活用して間接的に性格を推測する方法であり，場合によってはインタラクション開始前に典型他者モデルの選択を実現することができる．前者を選択時 N ショット，後者を選択時 0 ショットと表記することとする．

2.2 典型他者モデルの分類

典型他者モデルは，特定の性格に基づいて構築される他者モデルであり，平均他者モデルよりも高い予測精度を持つ．本研究では典型他者モデルを前節で説明した 2 つの分類方法によって（2 x 2 =）4 つのカテゴリに詳細に分類する．ここで，4 つのカテゴリをそれぞれ（獲得時: N or 0）x （選択時: N or 0）ショットとそれぞれ表記する．

2.2.1 0x0 ショット

0x0 ショットは，獲得も選択も瞬時に行える典型他者モデルである．このモデルは，絶対に判断を間違ってはいけない（判断を間違うと危険な）場面で有効である．例えば，「刃物を持った人物は危険なので避ける」や「知らない人にはついていかない」といった，自らの体験に必ずしも基づかない知識による，即座に判断が求められる状況に適している．しかし，インタラクション中に他者に適応しないため，インタラクション開始時に予測エラーがあったときに他者に適応ができない．例えば，「刃物を持っているが実は危険じゃない人だと気が付く」「知らない人だったが仲良くなった」といった関係性や他者への理解の促進は 0x0 ショットの他者モデルのみをもつエージェントでは実現が困難である．なぜなら，そもそもインタラクションをしない判断を瞬時にするため，他者に対する適応が進まない．また，偏見や差別の元凶になってしまう危険性もある．

2.2.2 0xN ショット

0xN ショットは，獲得は瞬時に行えるが，選択には繰り返しのインタラクションが必要な典型他者モデルである．すなわち，あらかじめ構築されたモデルをインタラクション中で選択することで他者に適応する．例えば，「親切な人なら助けてくれるはず」という知識を持ちながら，実際のインタラクションを通じてその人が親切であることを確認するケースが該当する．このモデルは，事前に用意された情報を基にしつつ，実際の経験を通じて他者を評価するため，行動推定精度が 0x0 ショットよりも高い．

著者らはこれまでに囚人のジレンマタスクを用いてこの他者モデルに関する研究を行ってきた（長原他, 2022）．実験では，繰り返し囚人のジレンマゲームにおいて，プレイヤー A は 3 つの典型他者モデル，「協力的な性格を予測する典型他者モデル」，「非協力的な性格を予測する典型他者モデル」，「合理的な性格を予測する典型他者モデル」をあらかじめ構築し，ルールに基づいて行動するプレイヤー B に対して適切な典型他者モデルを選択する予測精度を評価した．結果として，プレイヤー A が既知の性格には適応できる 1 方で，確率的な行動をする相手や予測不可能な性格には対応が難しいことも判明した．この研究結果からインタラクション開始時の予測にエラーがあった時でも他者に適応できる．

2.2.3 Nx0 ショット

Nx0 ショットは，獲得には個人の主観的な体験が必要だが，選択は瞬時に行える典型他者モデルである．例えば，「○○に所属している人たちとは気が合う」というような，ラベル情報に基づいて判断する個人的な感覚が該当する．このモデルは，主観的な経験を基に構築されているため，選択は迅速に行える．見た目や肩書，評判など観測可能な情報を基に即座に性格を推測し，適応することが可能である．

著者らは，見た目で判断するバイアスに関する研究を行っており，インタラクション開始前に典型他者モデルを選定する妥当性について検討してきた（Nahagara et al. 2023）．この研究では，見た目の特徴や文化的背景に基づいて他者を迅速に分類し，自分と似た個体同士で集団を形成することを示した．

2.2.4 NxN ショット

NxN ショットは，獲得と選択の両方に個人の主観的体験と繰り返しインタラクションが必要な典型他者モデルである．このモデルは，長期的な関係性の構築に

適しているが，予測精度が個別他者モデルも高いわけではない．むしろ，過去の経験に基づいて形成された感覚や直感に依存する部分が大きいと考える．

例えば，「なんとなくこういう感覚で関わっている時は，相手が警戒しているだろう」といった主観的な感覚が該当する．このような感覚は，具体的なデータや明確なパターンに基づくものではなく，個人の経験と直感に依存している．

3. おわりに

本研究では，典型他者モデルをさらに 4 つに分類し，それらが予想される性質について議論した．今後は 4 つの典型他者モデルを実装しそれぞれ評価する他，それらを組み合わせることで柔軟に他者に適応できる他者モデルの実現を目指す．

文 献

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.

阿部 将樹・田足井 昇太・長原 令旺・大森 隆司・大澤 正彦 (2023). 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, HAI シンポジウム

Bratman, M.: Intention, plans, and practical reason (1987).

Carlson, S. M., Koenig, M. A., and Harms, M. B. (2013).Theory of mind. Wiley Interdisciplinary Reviews: Cognitive Science, 4 (4), 391–402. https://doi.org/10.1002/wcs.1232

Dennett, D. C.(1989): The intentional stance, MIT press .

日永田 智絵・長井 隆行（2018.）. サービスロボットに他者モデルは必要か？日本認知科学会第 35 回大会論文集, OS12-7, pp. 378–383. 日本認知科学会,

長田悠吾・石川悟・大森隆司・森川幸治（2006）他者意図の推定に基づく協調行動の計算モデル化. 人工知能学会全国大会論文集, No. 0, pp. 28–28,

長原 令旺・田足井 昇太・佐々木 康輔・大森 隆司・大澤 正彦（2022）“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.

Nagahara et al.（2023） Model Simulation of Physical Distance and Cultural Constraints in Maintaining Genetic Diversity.ISIS

大澤正彦・奥岡耕平・坂本孝丈・市川淳・今井倫太 (2020).：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

Rao, A. S. and Georgeff, M. P.(1997): Modeling rational agents within a BDI-architecture, Readings in agents, pp. 317–328 .

横山絢美・大森隆司 (2009） 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,

横山絢美・大森隆司・阿部香澄・長井隆行（2011）他者の状態推定に基づく対人インタラクションロボットの行動戦略. 2011 年度日本認知科学会第 28 回大会,