

世界の部分的な心的操作モデルの獲得 による他者の背景世界の推定

The imagination of Umwelt of other based on a partial manipulable world model

高橋英之¹ 橋川莉乃¹ 堀部和也¹ 岡田浩之²
Hideyuki Takahashi, Rino Hashikawa, Kazuya Horibe, Hiroyuki Okada

1. 大阪大学大学院 基礎工学研究科・基礎工学部 2. 玉川大学 工学部
Osaka university, Tamagawa university
takahashi@irl.sys.es.osaka-u.ac.jp

概要

人間はどのように自分と他者、そして世界との関係を学んでいるのであろうか？本発表では、「自分と他者の関係」と「世界の心的操作可能なモデル」を反復的にもう一方の学習に際してプロジェクションさせることで、他者との関係性の理解と世界そのものの理解が相補的に展開していくモデルを提案する。そしてこのようなモデルにもとづいてコミュニケーションを捉えることで、これまでとは異なるコミュニケーションの価値についての視座が得られるのではないか、という議論を行いたい。

キーワード：心の理論、世界の心的操作可能なモデル

1. はじめに

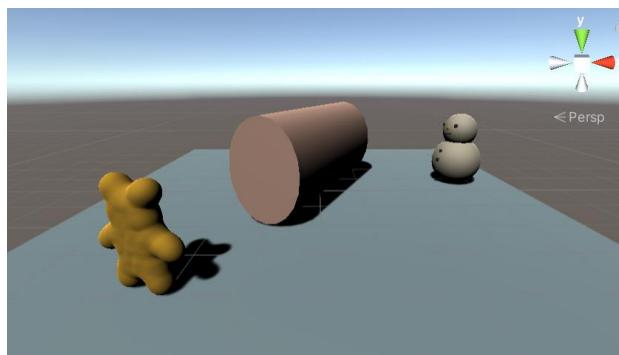


図1. シミュレーションで想定する風景の例

他者の視点に立って、他者の心を推定する能力（心の理論）の研究は、これまでに心理学、神経科学、哲学などで幅広く研究がなされてきた。しかしそれらの研究の多くは、自分と他者の同一性や、自分と他者を取りまく世界が既知なものであることが暗黙の前提として置かれていることが多かった。一方、同じ種である人間であっても、それぞれの視点や価値観は千差万別であり、また自分たちが置かれている世界についての知識も十分であるとは言えない。従って、他者を完全に理解することが可能である、とい

うことを前提とするモデルは現実的ではない、と考えている。

そこで我々はコミュニケーションにおいて他者の心を推定するということは、「自分と他者の視点や価値観の違いをまず認識し、その上で他者がみている世界（背景世界）を想像することにある」、という前提を置く、その上で、どのようにこのようなプロセスが理論化可能か、そのプロセスをモデル化し、思考実験にもとづくシミュレーションにより、このプロセスのダイナミクスをとらえることを目指す。

2. シミュレーションの概要

具体的には、二種類のエージェントそれぞれについて、三次元空間上での立ち位置と視線方向を設定し（それぞれのエージェントの空間上の位置は、相手からは観測することはできない）、それぞれの視線方向が交わる位置に立体図形のオブジェクトを置く場面を想定したシミュレーションを行う（図1）。この立体図形のオブジェクトは、世界全体を抽象化した存在であり、それぞれのエージェントはこのオブジェクトの像を二次元画像としてのみ取得することができる。このシミュレーションの肝として、エージェント単体では、オブジェクトの立体像を想像することができず、他のエージェントからの情報を統合することで、初めてオブジェクトの立体像をとらえることが可能になる。

このシミュレーションが難しい点として、二つのエージェントの相対的位置関係は未知であり、また同時に推定すべきオブジェクトの立体像の形も未知である点である。従ってこのシミュレーションにおいて、エージェントが正確にオブジェクトの立体像を推定するためには、自分と相手のエージェントの相対的な位置関係と、オブジェクトの立体構造の推定を同時に

行う必要がある。

本シミュレーションにおいて、エージェント間の通信は、それぞれのエージェントが観測した二次元画像をオートエンコーダによって低次元のバイナリ行列にした人工言語を使用する。従って、エージェントは自ら観測したオブジェクトの二次元画像、相手から伝達された圧縮された不完全なバイナリ行列の情報、そして自分と相手エージェントの相対的な位置関係の推定値にもとづき、観測しているオブジェクトの立体像を推定する。このシミュレーションにおいて、エージェントは自分と相手エージェントの位置関係の想定を固定してオブジェクトの立体像を予測するフェーズと、オブジェクトの立体像の想定を固定して相手エージェントとの相対的位置関係を予測するフェーズを交互に繰り返すことで、相手エージェントの背景世界の想像がどのように変化していくのかについて検討する。

今回の発表では、エージェントのパラメータをどのように設定することで、エージェントが自分と相手の相対的位置関係、さらに観測しているオブジェクトの立体像の安定した推定を行うことができるのかを検討する。また本研究では、このモデルが一方向（自分エージェントが他者エージェントの背景世界を想像する）である場合と、双方向（二つのエージェントがそれぞれ相手の背景世界を想像する）である場合で、どのようにモデルのダイナミクスが変化するのかについても検討する（図2）。

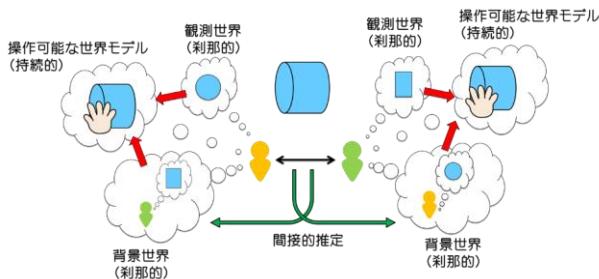


図2. 双方向性を仮定したモデルの例

このモデルの面白いところは、自分エージェントと他者エージェントの視点の相違性の推定をすることが、結果的に（部分的ではあるが）世界の心的操縦モデルを獲得することにつながり、それが結果的には自分と他者の相対的な関係性の推定を可能にする、という点である。

3. 今後の展開

このように「自分と他者の関係」と「世界の心的操縦可能なモデル」を反復的にもう一方の学習に際してプロジェクションさせることで、他者との関係性の理解と世界そのものの理解が相補的に展開していく点は、コミュニケーションにおける新たな価値関数に対する視座を与えてくれる。すなわち他者の背景世界を推定しながらコミュニケーションを行うことは、世界の操作モデルを獲得することにつながり、結果として自らの内的世界の拡張を実現する。このような自らの内的世界の拡張を報酬系として組み込んだコミュニケーションのモデルを提案することは、従来の互恵性をベースとしたコミュニケーションのモデルとは全く異なる価値関数の提案につながるのではないかと期待している。今回の発表では、その点についても、心理学や哲学などの関連領域の知見も交えながら議論をしたい。

文献

- [1] Yang, D. Y. J., Rosenblau, G., Keifer, C., & Pelpfrey, K. A. (2015). An integrative neural model of social perception, action observation, and theory of mind. *Neuroscience & Biobehavioral Reviews*, 51, 263-275.
- [2] Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience*, 20(8), 495-505.