

# エージェントの感情と社会的評価プロセスの可視化

佐藤 幹晃<sup>1,a)</sup> 寺田 和憲<sup>1,b)</sup>

**概要：**感情表現は、物理的環境と社会的環境の両方を反映した感覚入力を評価した結果、表情などとして現れる。感情表現の観察者は、逆評価をすることで感覚入力がどのように評価されたかを解読し、他者の意図を推論する。しかし、逆評価は、同じ感情表現が異なる状況で生成されることや、同じ状況でも個人差によって感情表現が異なることから、逆問題となる。この問題を克服するためには、個人が適切な評価モデルを持つ必要がある。その為、他者の心的状態を理解することが困難な自閉スペクトラム症者などでも、適切な評価モデルの獲得をできるように訓練できるインタフェースが求められる。そこで、本研究では感情的評価過程を可視化したインタフェースを開発した。このインタフェースによって社会的スキルが獲得できるかどうかを調査することが今後の課題である。

## 1. はじめに

社会は非ゼロ和であり、選好の違いや生産能力の違いを考慮して協力することで Win-Win な結果を望める。一方で、協力的態度の搾取は簡単である為、他者の表現から得られる情報によって、他者の心的状態を考慮し、裏切る人と協力的な人を識別する必要がある。この識別をする上で重要な情報として感情表現が挙げられる。感情表現は、物理的・社会的世界の主観的な評価の結果として現れるため、他者の精神状態を推測するために重要な情報を含んでいる [1]。

評価プロセスでは、現在の状況（または心の中で想像された状況）を信念や目標と比較する。ある状況が期待され、目標達成を促進するものであれば、その状況は肯定的に評価される。一方、ある状況が好ましくなく、行為者の目標達成を妨げるものであれば、その状況は否定的に評価される。例えば、自分の利益だけでなく相手の利益も大切にす協力的な人は、2人で利益を共有している状況の方が、利益を独占している状況よりも、喜びなどの肯定的な感情を表現することがある。肯定的な表現は、現在の状況が有利であり、継続が望まれていることを示し、相手に伝えられる。観察者の課題は、観察した現在の状況と感情表現に応じて、適切な判断をすることである。そのためには、行為者の意図を正しく認識する必要がある。意図は逆評価プロセスによって解読される [2], [3]。

逆評価は、同じ感情表現が異なる状況で生成され、同じ

状況でも個人差によって感情表現が異なるため、逆問題となる。この問題は、観察者が適切な評価モデルを用いてベイズ推論などを行うことで克服できる [3]。

心的状態の推論は、社会生活を生き抜くための基本的な計算能力の一つであるが、自閉スペクトラム症者など、他者の精神状態を理解することが困難な人もいる。そこで本研究では、評価プロセスを数学的・図式的に表現することで、表現された感情と心的状態との関係を理解することを容易にするインタフェースを作成した。

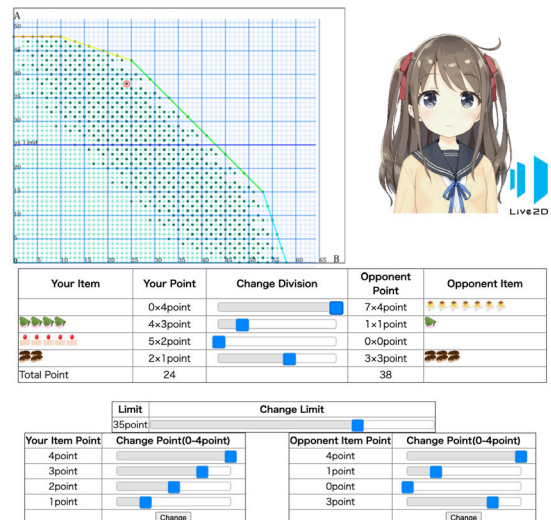


図 1 複数論点最後通牒ゲームにおいて、現在の配分がどのように評価され、どのような感情表現となるか可視化したインタフェース。アニメキャラクターの画像は Live2D Cubism, Live2D Inc. から入手したものである。デモはオンラインで公開されている。<https://yosemite.ai.info.gifu-u.ac.jp/cgi/pareto-hiyori>

<sup>1</sup> 岐阜大学  
<sup>a)</sup> motoakist@gmail.com  
<sup>b)</sup> terada@gifu-u.ac.jp

本研究では、ユーザとエージェントが解決すべき対象課題として、複数論点最後通牒ゲームを用いて、エージェントの評価プロセスを可視化するインタフェースを開発した。ユーザーは、スライダーバーを操作することで、どのような嗜好、現在の配分がどのような感情表現をもたらすかを探り、理解することができる。

## 2. 実験

図1は、複数論点最後通牒ゲームにおいて、エージェントが現在の配分をどのように評価し、どのような感情表現となるかを可視化したインタフェースである（本インタフェースはオンラインで公開されている：[https://yosemite.ai.info.gifu-u.ac.jp/cgi/pareto\\_hiyori\\_2.0](https://yosemite.ai.info.gifu-u.ac.jp/cgi/pareto_hiyori_2.0)）

提案者（ユーザ）と応答者（エージェント）の2人のプレイヤーが、一方通行かつ一度きりの交渉で複数のアイテムを配分する。標準的な最後通牒ゲームはゼロ和ゲームであるが、複数論点最後通牒ゲームは、提案者と応答者で選好に違いがあれば非ゼロ和ゲームとなる。例えば、図1では、ユーザーにとってのショートケーキのポイントは2ポイントであるが、エージェントにとってのポイントは0ポイントである。ユーザーとエージェントにとって同じ4ポイントの価値があるプリン（ゼロ和）とは異なり、ショートケーキには利害の対立がない（非ゼロ和）。ユーザーにとって価値のないエクレーアをエージェントに多く配分し、エージェントにとって価値のないショートケーキをユーザーに多く配分することで、ユーザーもエージェントもより高いポイントを得ることができる。このような好みの違いが、両者にとってより満足度の高い win-win な結果につながる。

ユーザーはスライダーバーを動かしてアイテムを割り当てることができる。エージェントは、アイテムがどのように割り当てられているかを評価し、その評価に応じて表現を変える。ユーザがスライダーバーを左端に動かすと、エージェントへのアイテムの割り当てはゼロになり、ユーザへのアイテムの割り当ては最大になる。ユーザーがスライダーバーを右端に動かすと、エージェントへのアイテムの割り当てが最大となり、ユーザへのアイテムの割り当てはゼロとなる。ユーザーは、各アイテムのポイントを設定することができる。ユーザーがあるアイテムに高いポイントを設定した場合、例えば4ポイントであれば、エージェントがそのアイテムを気に入っていることを意味する。一方で、ユーザがあるアイテムに低いポイントを設定した場合、例えば0ポイントであれば、エージェントはそのアイテムに全く興味がないことを意味する。エージェントは状況を評価し、ユーザがポイントの高いアイテム、すなわち好みのアイテムを割り当てた場合には、肯定的な表情を示す。しかし、ユーザが0点のアイテムを割り当てた場合は、いくら割り当ててもエージェントの表情は変わらない。ユーザは、スライダーバーを動かすことで、エージェントの好み

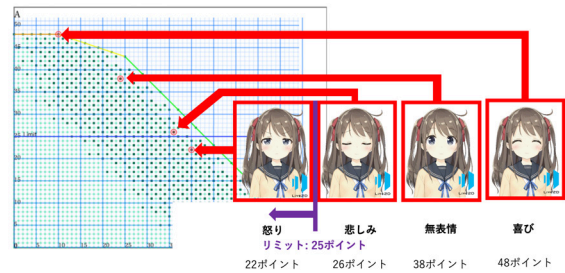


図2 ユーザーから提示されたポイント进行评估した結果の表情。提供されたポイントの量とリミットによって表情が変化する。ユーザがエージェントに多くのポイントを提供した場合、エージェントは喜びを表し、ユーザが少ないポイントを割り当てた場合、エージェントは悲しみを表す。また、ポイントの合計がリミットを下回ると、エージェントは怒りの表情を見せる。

を簡単に設計することができる。図1の左側のグラフは、ユーザとエージェントの合計ポイントを示している。エージェントは、現在のアイテムの配分をポイントとして評価する。赤い丸は、エージェントが評価したアイテムの合計ポイントを示している。ポイントの合計が許容できる限界（リミット）よりも低い場合、エージェントは怒った表情となる。ポイントの合計がリミットより高い場合、リミットに近いときは悲しげな表情、ポイントが増えると無表情、さらにポイントが増えると嬉しそうな表情をする。また、ユーザーがリミットを変更することもできる。リミットが低い場合、エージェントは怒りの表情を抑え、ポイントの配分が少なくても喜びの表情を見せる。リミットが高い場合、エージェントは怒りを示し、ほとんど笑顔を見せない。リミットを変更することで、ユーザーはエージェントを強気にしたり、寛大にしたりすることができる。

スライダーバーの動きに応じて、表情が滑らかに変化する（図2）。使用したのは、Live2D Inc. のソフトウェア「Live2D Cubism」に収録されている、大きな目をした日本のアニメ調の女性像「百瀬ひより」である。Live2D Cubismで目、眉、口のパラメータを操作することで、図2の表情を含む18種類の表情をGIFファイルとして作成した。

## 3. おわりに

現在のエージェントは個人主義、すなわち自分の利益のみに関心があり、自己利益が大きくなると肯定的な表現をするようになってきている。今後は、他者の利益を考慮する協力的なエージェントや、他者と比較して自分の利益が高くなることを重視する競争的なエージェントなど、他の社会的価値志向性のあるエージェントを実装する予定である [4]。評価プロセスを数学的・図式的に表現することで、特に他者の心を理解することが困難な人にとって、表出された感情と心的状態の関係の理解が容易になると考えている。今後の研究で、本インタフェースによって社会的スキルが向上するかどうかを調査する予定である。

## 参考文献

- [1] K. R. Scherer, "Appraisal Theory," in *Handbook of Cognition and Emotion*, John Wiley & Sons, Ltd, 2005, pp. 637–663.
- [2] C. M. de Melo, P. J. Carnevale, S. J. Read, and J. Gratch, "Reading people' s minds from emotion expressions in interdependent decision making," *J. Pers. Soc. Psychol.*, vol. 106, no. 1, pp. 73–88, 2014, doi: 10.1037/a0034251.
- [3] Y. Wu, C. L. Baker, J. B. Tenenbaum, and L. E. Schulz, "Rational Inference of Beliefs and Desires From Emotional Expressions," *Cogn. Sci.*, 2017, doi: 10.1111/cogs.12548.
- [4] S. Bogaert, C. Boone, and C. Declerck, "Social value orientation and cooperation in social dilemmas: A review and conceptual model," *Br. J. Soc. Psychol.*, vol. 47, no. 3, pp. 453–480, 2008, doi: 10.1348/014466607x244970.