他者とのインタラクションを通して獲得される内部状態の変化に関するシミュレーション

Analyzing the Impact of Agent's Interaction with Others on Its Internal State Dynamics in Spatial Environments

坂本孝丈 [1*]
Takafumi Sakamoto[1]

[1] 静岡大学
[1] Shizuoka University

Abstract: エージェントが公共場面において社会的に振る舞うためには，相手の内部状態に応じて自身の内部状態を調整する必要がある．本研究では，この内部状態の変化のダイナミクスを強化学習により獲得可能なエージェントを設計する．そこから，エージェントの他者とのインタラクションの経験が内部状態の変化に及ぼす影響についてシミュレーションを通して検証する．

1 はじめに

コミュニケーションロボットが社会的に受け入れるためには，与えられた役割に従事しつつ，周囲の人々の行動や状況に合わせて行動を調整する必要がある．特に人とロボットが公共場面においてコミュニケーションを開始する場合では，今まさにここで遭遇している相手に対する適応的な行動の生成と，その場面で遭遇し得る他者の集団に対して適応的な行動を学習していく機構が必要となる．例えば，公共場面では見知らぬ他者同士が同一の空間内を行き交うことから，過度にコミュニケーションが生じないように，移動方向や移動速度，視線方向などが制御されている [1]．このようなコミュニケーションのダイナミクスをモデル化することで，人社会に受け入れられやすいロボットの振る舞いを設計できると考えられる．

ロボットが人とのコミュニケーションを開始するための接近行動や，衝突の回避や不快感を低減するための回避行動に関する研究は数多くみられる（レビューとして [2] や [3]，[4] を参照）．また，公共場面においてロボットが人の行動を分類または予測し，適切な話しかけ相手の選択と接近を行う手法が提案されている [5, 6]．一方で，エージェント同士のインタラクションをシステムダイナミクスとして記述したモデルはほとんど見られない．エージェントは目的に応じて自律的に行動する存在であるため，自身の行動を促進（または抑制）する内部状態を持つ．エージェント同士が社会的なインタラクションを成立させるためには，状況や他のエージェント（他者）に合わせて，エージェント自身の内部状態を調整する必要がある．そのため，社会的な場においてエージェントが他者の内部状態を認知するプロセスも含めたインタラクションのモデル化を行う必要がある．

これに対して，これまでの研究ではコミュニケーション開始場面における接近・回避行動を生成するためのモデルを提案し計算機シミュレーションによる検証を行ってきた [7]．また，配慮を伴う行動として相手の内部状態に合わせて自身の内部状態の値を変化させるエージェントのモデル化を行い [8]，その有用性について検証してきた [?]．しかし，エージェントの内部状態の変化の仕方はトップダウンに設計したものであり，最適な変化のダイナミクスが検討できているとはいえない．

そこで本研究では，エージェントの内部状態の変化を表す関数を強化学習により構築する手法について検証を行う．エージェントの内部状態の変化は，そのエージェントがこれまで置かれてきた環境内に存在する他者とのインタラクションにより構築されると考えられる．内部状態の変化の仕方を強化学習を通して構築することで，エージェントは自身が従事する役割やタスク，周囲の他者の特性に適応することができる．エージェントの役割やタスクは報酬関数で表現され，周囲のエージェントの行動や内部状態の変化の特性はパラメータとして表現される．このパラメータにより構築される内部状態のダイナミクスに生じる差異を検証することで，エージェントの社会性を表現する方法について検討する．

*連絡先： 静岡大学
〒 422-8037 静岡県静岡市駿河区大谷 836
E-mail:sakamoto@sapientia.inf.shizuoka.ac.jp

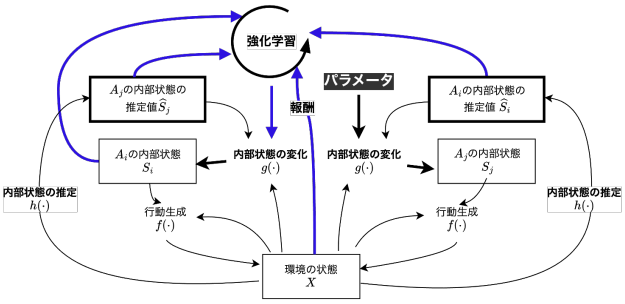

図 1: 強化学習を用いた内部状態の変化

2 内部状態の変化のダイナミクスと強化学習

他者への配慮などの社会的な行動をモデル化するためには，エージェントがインタラクションを通して互いの内部状態（欲求）を調整する過程を記述する必要がある．本研究では内部状態の変化を強化学習により計算する．内部状態の変化は，自身の現在の内部状態，相手の内部状態の推定値，相手からみた自身の内部状態の推定値により決まるものとする．内部状態が変化することで，行動が変化し，環境の状態が更新されることで報酬が得られる（図 1）．相手エージェントの内部状態の変化の仕方によって獲得される内部状態のダイナミクスは異なると考えられる．

以下では，2 体のエージェント間のコミュニケーション開始場面における接近・回避行動に基づくインタラクションのモデルについて述べる．内部状態と行動生成，内部状態の推定について，先行研究 [7, 8] と重複する部分については概要のみを示す．

2.1 内部状態と行動生成

内部状態の値に応じたエージェントの行動生成の関数については先行研究 [7] のモデルを用いる．このモデルでは 2 体のエージェント A_1，A_2 の間の身体的なインタラクションを環境 $\mathbf{x}_{12} = \{r_{12}, \theta_{12}, \theta_{21}\}$ の時間的な変化により表す．なお，r_{12} は A_1–A_2 の距離を表し，θ_{12} と θ_{21} はそれぞれ A_1 からみた相対角度の絶対値と A_2 からみた相対角度の絶対値を表す．

エージェントの内部状態は行動を促進または抑制する変数を表す．コミュニケーション開始場面における内部状態の変数として，ここでは自身から相手への関与に対する選好（Control）と，相手から自身への関与に対する選好（Acceptance）の 2 つの変数を扱う．$\mathbf{A}_1$ の $\mathbf{A}_2$ に対する内部状態を $\mathbf{S}_{1\to2} = (c_1, a_1)$，$\mathbf{A}_2$ の内部状態を $\mathbf{S}_{2\to1} = (c_2, a_2)$ で表す．

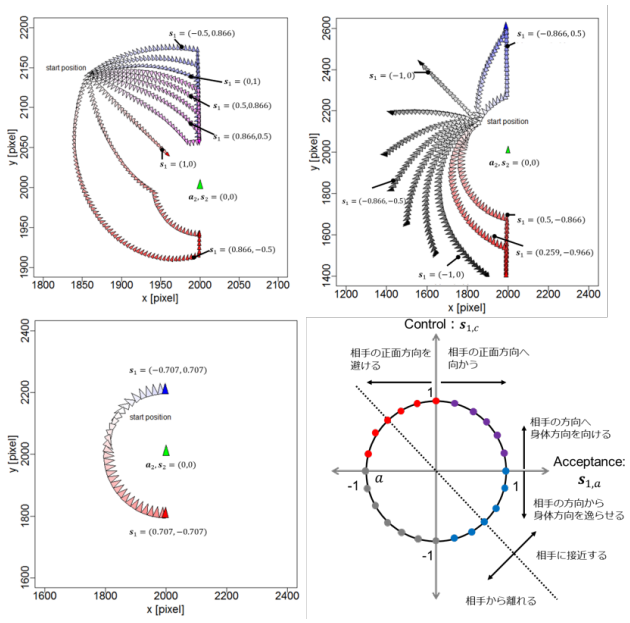

図 2: 内部状態の値に応じて生成される接近・回避行動の例 [7]

エージェントの行動それぞれは $\mathbf{x}_{12}$ の時間的な変化により表される．A_1 の行動を $\Delta_1\mathbf{x}_{12}$ は行動生成の関数 f を用いて以下の式で表される．

$$\Delta_1\mathbf{x}_{12} = f\left(\mathbf{x}_{12}, \mathbf{s}_{1\to2}; \phi_1\right) \tag{1}$$

ただし，ϕ_1 は移動速度の最大値などの行動特性を表す値とする．$\mathbf{s}_{1\to2}$ は A_1 の内部状態の値を表し，$\mathbf{s}_{1\to2} = (c_1, a_1) \in [-1, 1]^2$ とする．

内部状態と行動生成の関数 f のみを規定した場合に生成される接近・回避行動を図 2 に示す．

2.2 内部状態の推定関数

配慮を伴う行動を生成するうえで，他のエージェントの内部状態を推定するプロセスを扱う必要がある．ここでは，先行研究と同様に式 1 の行動生成の関数を用いて内部状態の推定を行う [8]．A_1 が A_2 の行動 $\Delta_2\mathbf{x}_{12}$ から内部状態を推定する場合，$\mathbf{x}_{12}$ は観測可能な変数である．また，ϕ_2 は推定しなければならないが，行動特性を表す変数であるため，行動を観察することでおおよそ推定可能であると仮定する．そのうえで，A_1 により $\mathbf{s}_{2\to1}$ の推定値 $\hat{\mathbf{s}}_{2\to1}$ は関数 h を用いて以下の式で表される．

$$\begin{aligned}\hat{\mathbf{s}}_2^{(t)} &= h\left(x_{12}, \Delta_2\mathbf{x}_{12}, \hat{\phi}_2\right) \\ &= \underset{\hat{\mathbf{s}}}{\operatorname{argmax}}\left(L\left(\Delta_2\mathbf{x}_{12}; \hat{\phi}_2\right)\right)\end{aligned} \tag{2}$$

ただし

$$L(\cdot) = 1 - l\left(f(x, \hat{s}) - \Delta_2 \mathbf{x}_{12}; \hat{\phi}_2\right) \tag{3}$$

とする．l は $0 \leq l(\cdot) \leq 1$ となる f により生成可能な行動と $\Delta_2 \mathbf{x}_{12}$ の差を規格化する関数を表す．これにより，$\Delta_2 \mathbf{x}_{12}$ と最も類似している行動を生成し得る内部状態が推定値になる．このとき，$\mathbf{s}_{2\to1}$ は 2 次元の変数であるため，グリッド探索により近似的に解を求めることができる．

2.3 内部状態の変化を表す関数

内部状態の変化は関数 g を用いて表す．行動結果に基づく内部状態の変化のみを扱う場合 A_1 の内部状態の変化 $\Delta \mathbf{s}_{1\to2}$ は以下の式で表される．

$$\Delta \mathbf{s}_{1\to2} = g(\mathbf{x}_{12}, \mathbf{s}_{1\to2}; \psi_1) \tag{4}$$

ここで，ψ_1 は A_1 の内部状態の変化の速さなどの認知的な特性を表す．先行研究と同様に，A_1 が A_2 に対して配慮する場合の内部状態を変化を表現するために，関数 g を以下の式に拡張する．

$$\Delta \mathbf{s}_{1\to2} = g(\mathbf{x}_{12}, \mathbf{s}_{1\to2}, \hat{\mathbf{s}}_{2\to1}; \psi_1) \tag{5}$$

関数 g やパラメータ ψ の値をどのように規定するかによって，そのエージェントがどの程度，相手に配慮するのかが決まる．

先行研究では配慮を表すパラメータを ψ の値で表し，ψ の値が人とエージェントのインタラクションに及ぼす影響を検証した．具体的には，A_1 の内部状態の変化を以下の式で表し，単位時間あたりに自身の内部状態を変化させる量を配慮のパラメータとしている．

$$\Delta c_1 = -\psi_1 \cdot (\hat{a}_2 - c_1) \tag{6}$$

$$\Delta a_1 = -\psi_1 \cdot (\hat{c}_2 - a_1) \tag{7}$$

これは他者の Control の値に対して自身の Acceptance の値を一致させるフィードバックであり，相補性として捉えることができる．

一方で，相補性により内部状態を調整するべきか否かはエージェントが従事する役割や周囲の他者の性質に依存すると考えられる．本研究では他者とのインタラクションを通して内部状態の変化を学習させる．

2.4 強化学習を用いた内部状態の変化

エージェントの内部状態の変化は，内部状態の変化の結果生じる相手エージェントとのインタラクションによって得られる報酬に基づいて決定される．報酬関数は想定されるエージェントの役割によって異なる．ここでは，エージェントが環境内の他者に何らかの行動を仕掛ける役割に従事していると過程し，相手エージェントと接近し，向き合うことができれば報酬が得られるものとする．また，エージェントが接近することで相手エージェントが自身を忌避するような行動を示した場合はペナルティを受け取る．この 2 つの報酬により，他者とのインタラクションの経験をもとに内部状態の変化の仕方を改善していく．

3 シミュレーション

本研究では他者とのインタラクションを通してエージェントが内部状態の変化を学習していく過程について，強化学習を用いて検証を行う．ここでは他エージェントは以下の 4 つの特徴のいずれかを持つものとする．

(a) 接近してくる相手を忌避するエージェント（$c_0, a_0 = 0$，$\psi = -0.025$）

(b) 接近してくる相手に応対するエージェント（$c_0, a_0 = 0$，$\psi = 0.025$）

(c) 相手に積極的に関わろうとするエージェント（$c_0, a_0 = 1.0$，$\psi = 0.0$）

(d) 相手に影響されず目的地に向かって移動し続けるエージェント（$c_0, a_0 = 0.0$，$\psi = 0.0$）

学習のエピソードごとに上記から相手エージェントがランダムに選択される．本研究では強化学習の手法のうち最も単純な Q 学習を用いてシミュレーションを行う．状態空間は，自身の内部状態，相手エージェントの内部状態の推定値，相手からみた自身の内部状態の推定値とする．行動空間は，1 ステップあたりの内部状態の変化値であり，Control と Acceptance のそれぞれについて $-0.25, 0, 0.25$ のいずれかの値をとる．エージェントは視野内に相手エージェントが存在しない，または，内部状態の値が閾値（0.5）未満である場合は，目的位置に向かって進む．学習段階ではエージェントの初期値と目的位置はランダムとし，1 エピソードあたり 150 ステップ移動を行う．移動の結果，違いに向かい合いった状態で接近した場合に報酬に 1 を加算し，そのエピソードを終了する．また，相手エージェントが自身を忌避する行動を示した場合は報酬から 0.1 を減算する．

3.1 シミュレーション結果と考察

上述の相手エージェントの割合が等しい場合の学習結果を図 3，4 に示す．図 4 の赤色のので示されるエー

ジェントが強化学習により内部状態の変化を学習さたエージェントであり，2000 エピソード分の学習を行った段階のそれぞれの相手に対する行動が示されている．接近する相手を忌諱するエージェント (a) に対しては，相手が忌避行動を示した時点で相手を追いかけずに静止している．それ以外の条件については，相手の正面方向に移動することで，接近に成功している．

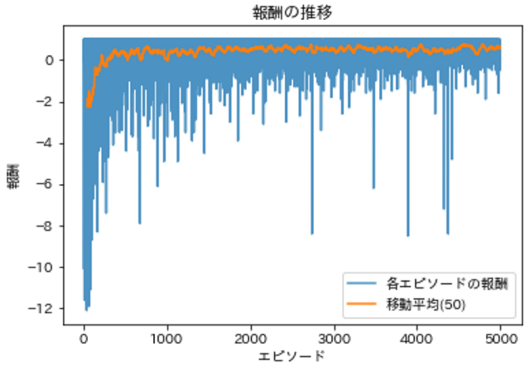

図 3: 学習結果の例（2000 エピソード）

相手エージェントの内部状態の推定値とエージェント自身の内部状態の変化の例を図 5 に示す．接近する相手を忌諱するエージェント (a) に対して，行動から内部状態を推定し，相手の内部状態に合わせて自身の内部状態を調整していることが確認された．これは先行研究において仮定していた相手の内部状態に対する相補性が最適化された内部状態のダイナミクスである可能性を示唆している．

今後は相手エージェントの特徴の割合や，報酬関数が内部状態の変化の学習結果に及ぼす影響についてさらなる検証を行う．

参考文献

[1] Erving Goffman. *Behavior in public place*. Free Press, 1963.

[2] Thibault Kruse, Amit Kumar Pandey, Rachid Alami, and Alexandra Kirsch. Human-aware robot navigation: A survey. *Robotics and Autonomous Systems*, Vol. 61, No. 12, pp. 1726–1743, 2013.

[3] Jorge Rios-Martinez, Anne Spalanzani, and Christian Laugier. From proxemics theory to socially-aware navigation: A survey. *International Journal of Social Robotics*, Vol. 7, No. 2, pp. 137–153, 2015.

[4] Yuxiang Gao and Chien-Ming Huang. Evaluation of socially-aware robot navigation. *Frontiers in Robotics and AI*, p. 420, 2021.

[5] Takayuki Kanda, Dylan F Glas, Masahiro Shiomi, and Norihiro Hagita. Abstracting people's trajectories for social robots to proactively approach customers. *IEEE Transactions on Robotics*, Vol. 25, No. 6, pp. 1382–1396, 2009.

[6] Satoru Satake, Takayuki Kanda, Dylan F Glas, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. A robot that approaches pedestrians. *IEEE Transactions on Robotics*, Vol. 29, No. 2, pp. 508–524, 2012.

[7] Takafumi Sakamoto and Yugo Takeuchi. Simulation of spatial behavior based on an agent model in human-agent initial interaction. In *Proceedings of the 6th International Conference on Human-Agent Interaction*, pp. 310–317. ACM, 2018.

[8] 坂本孝丈, 竹内勇剛. 他者への配慮を伴う接近・回避行動のモデル化. HAI シンポジウム 2021, p. G21, 2021.

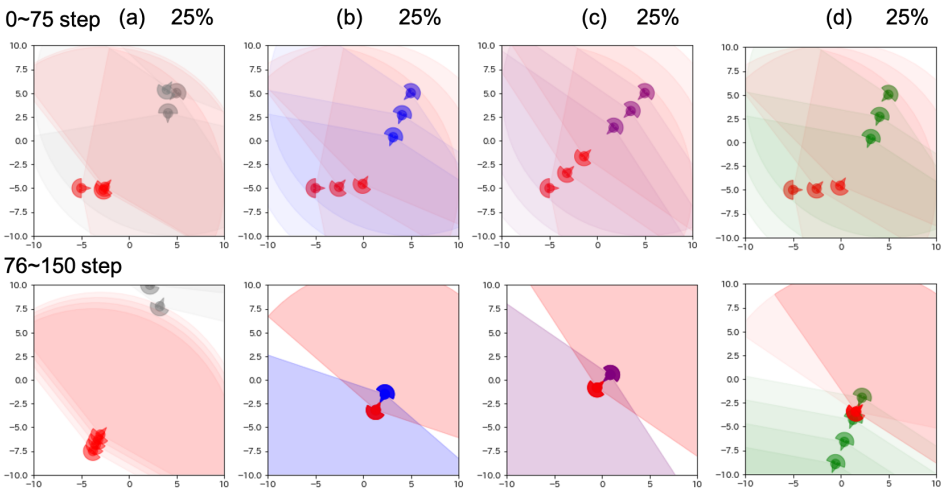

図 4: 学習結果の例（2000 エピソード）

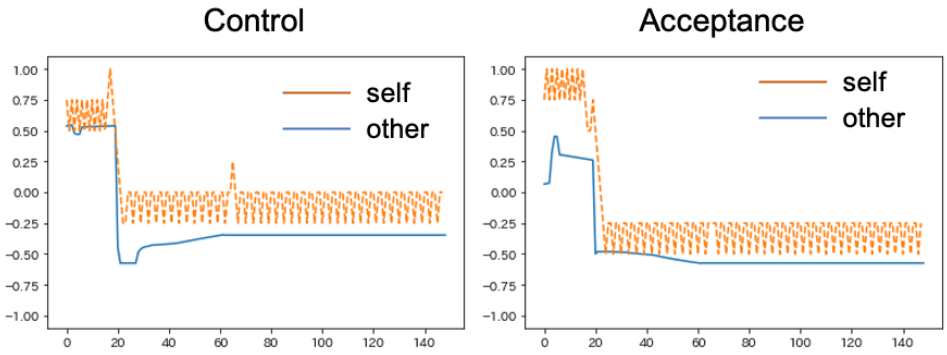

図 5: 相手エージェント (a) の内部状態の推定値と学習した内部状態の変化