OCC theory に基づくエージェントの感情表現と時間経過に関する論理モデル

今井 那緒　浅井 沙良　塚本 麻衣　新出 尚之

近年、エージェントが人間とコミュニケーションを取ることが重要視されており、その一環としてエージェントに感情を持たせることが求められている。OCC theory と呼ばれる、心理学的見地に基づいた人間の感情を 22 種類にモデル化した理論と、この理論で扱われている感情を論理式として形式化し、さらにこれを基に感情の強さを取り入れた研究がある。しかし、これらの研究では時間経過による感情の強さの減衰を扱っていないこと、また、OCC theory で分類された 22 種の感情のうち love と hate の 2 種類は扱えなかったことから、人間らしい感情表現としては不十分であった。そこで、本論文では、従来研究を基に論理式として形式化されていなかった 2 種の感情の形式化と、時間経過による感情の強さの減衰を表現できる論理体系を提案し、この体系で表現される感情の性質などについて述べる。

1　はじめに

近年、人間らしい感情を持つロボットの開発が進んでいる。現実世界において、ロボットが人間とコミュニケーションをとるためには、ロボットが感情を持っていることが望ましい。ロボットが人間らしい感情を持つことで、自ら人間に近い行動を決定することが求められる。

周囲の環境が常に変化する現実世界において、問題解決のために行動するロボットの実現として BDI アーキテクチャーが有効である [5]。BDI アーキテクチャーとは、人間行動を信念、願望、意図の 3 つの心的状態でモデル化した BDI モデル [4] による行為決定方式を計算機上で実現したものである。それにより、人間の合理的思考に基づいた行動をとる BDI エージェントを構築することが可能である。また、この BDI モデルは、BDI logic という論理モデルを持つ。

一方、OCC theory [3] と呼ばれる、人間の感情を分類する理論がある。この OCC theory では、信念や願望などの心的状態を用いて人間の感情を分類、特徴付けており、論理モデルによって表現が可能で、BDI logic を持つ BDI モデルとの親和性が良い。

Adam らの研究 [1] では、OCC theory の理論を BDI モデルに取り込むことで、感情を論理式でモデル化している。この研究で、BDI logic に「不確実だが起こると期待されている事象」等の新たなオペレータを複数導入することで、OCC theory で扱われている 22 種類のうち 20 種の感情を論理式として形式化し、BDI モデルに取り込んでいる。そこで、我々は、Adam らの形式化を踏まえ、BDI アーキテクチャに基づく記述言語 AgentSpeak の処理系である Jason [2] で実現、ライブラリを実装した [6]。この [6] では、1 つの行動につき 1 つの感情しか生起できず、また、複合的な感情については扱っていなかった。それに加えて、生起する感情には度合いがなかった。そのため、[8] では、度合い付きの感情が生起できるような論理体系の提案をし、[7] で、その論理体系に基づいた実装を行った。さらに、[9] [10] では、その度合いが時間経過によって減衰し、度合いが 0 になった感情の削除を実装した。しかしこれらの研究 [6] [7] [8] [9] [10] では、時間経過による感情の強さに対する論理体系が提案されておらず、また、人間の持つ主要な感情である好意や嫌悪についての感情 (love, hate) が表現できず、人間らしい感情表現としては不十分である。

A Logical Model of Emotional Expressions and Time Course based on OCC theory

Nao Imai, 奈良女子大学人間文化研究科情報科学専攻, Nara Womens's University.

そこで本論文では、Adam らの形式化を踏まえつつ、OCC theory で定義されている感情のうち形式化されていなかった 2 種の感情の形式化と、時間経過による感情の強さの減衰を実現する論理体系を提案する。

2 構文

本章では、我々が提案する論理体系での論理式の定義を与える。我々の論理体系は、LTL(線形時相論理)に感情表現の定義に用いるさまざまな様相オペレータを導入し、さらに述語論理に拡張したものである。

まず、以下のものを定めておく。

- 命題記号の無限集合を $ATM = \{P_1, P_2, \ldots\}$ とする。また、命題記号は論理式である。
- エージェントを表す記号の無限集合を $AGT = \{i_1, i_2, \ldots\}$ とする。
- 述語記号の無限集合を $PRD = \{R_1, R_2, \ldots\}$ とする。
- 定数記号の無限集合を $CST = \{c_1, c_2, \ldots\}$ とする。
- 変数記号の無限集合を $VRB = \{x_1, x_2, \ldots\}$ とする。
- 関数記号の無限集合を $FUC = \{f_1, f_2, \ldots\}$ とする。

以上を定めておいたうえで、項および論理式を以下のように定義する。

- $c \in CST$ ならば、c は項である。
- $x \in VRB$ ならば、x は項である。
- $f \in FUC$、$t_1, t_2, ..., t_n (n \geq 0)$ が項ならば、$f(t_1, t_2, ..., t_n)$ は項である。
- $r \in PRD$、$t_1, t_2, ..., t_n (n \geq 0)$ が項ならば、$r(t_1, t_2, ..., t_n)$ は論理式である。これを原始論理式と呼ぶ。
- $x \in VRB$、$\varphi(x)$ が論理式ならば、$\forall x \varphi(x)$、$\exists x \psi(x)$ は論理式である。
- φ_1, φ_2 が論理式ならば、$\neg\varphi_1, \varphi_1 \vee \varphi_2$ も論理式であるとする。また、$\wedge, \rightarrow, \leftrightarrow$ の省略形を以下のように定義する。
  - $\varphi_1 \wedge \varphi_2 := \neg(\neg\varphi_1 \vee \neg\varphi_2)$
  - $\varphi_1 \rightarrow \varphi_2 := \neg\varphi_1 \vee \varphi_2$
  - $\varphi_1 \leftrightarrow \varphi_2 := (\varphi_1 \rightarrow \varphi_2) \wedge (\varphi_2 \rightarrow \varphi_1)$
- φ_1, φ_2 が論理式ならば、$\neg\varphi_1, \varphi_1 \vee \varphi_2, \varphi_1 U \varphi_2$ も論理式である。
- φ が論理式、$i \in AGT$、d(desirability) が $0 \leq d \leq 1$ を満たす実数の時、$Des^i_d \varphi$ は論理式である。
  直感的には、$Des^i_d \varphi$ は「エージェント i にとって φ は度合い d で望ましい」ことを表す。
- φ が論理式、$i \in AGT$、l(likelihood) が $0 \leq l \leq 1$ を満たす実数の時、$Prob^i_l \varphi$ は論理式である。
  省略形として、以下を定義する。
  $$Bel^i \varphi := Prob^i_{1.0} \varphi$$
  直感的には、$Prob^i_l \varphi$ は「エージェント i は φ が見込み l で真になる可能性があると信じている」、$Bel^i \varphi$ は「エージェント i は φ が真であると信じている」ことを表す。
- φ が論理式ならば、$G\varphi$ は論理式である。
  省略形として、以下を定義する。
  $$F\varphi := \neg G \neg \varphi$$
  直感的には、$G\varphi$ は「φ は現在時刻以降ずっと成り立つ」、$F\varphi$ は「φ は現在時刻以降のある時点で成り立つ」ことを表す。
- φ が論理式ならば、$H\varphi$ は論理式である。
  省略形として、以下を定義する。
  $$P\varphi := \neg H \neg \varphi$$
  直感的には、$H\varphi$ は「φ は現在時刻までずっと成り立つ」、$P\varphi$ は「φ は現在時刻までのある時点で成り立つ」ことを表す。
- φ が論理式ならば、$X\varphi$ は論理式である。
  直感的には、$X\varphi$ は「φ は現在時刻より 1 時刻後で成り立つ」ことを表す。
- φ が論理式ならば、$Y\varphi$ は論理式である。
  省略形として、以下を定義する。
  - $$Bel\_arise^i \varphi := Bel^i \varphi \wedge \neg Y Bel^i \varphi \quad (1)$$
  - $$Prob\_arise^i_l \varphi := Prob^i_l \varphi \wedge Y Prob^i_0 \varphi \quad (2)$$
  直感的には、$Y\varphi$ は「φ は現在時刻より 1 時刻前で成り立つ」ことを表す。
- φ が論理式、$i \in AGT$、e が $0 \leq e \leq 1$ を満たす実数の時、$Effort^i_e \varphi$ は論理式である。
  直感的には、$Effort^i_e \varphi$ は「エージェント i は φ を真にするため、e 程度努力した」ことを表す。
- φ が論理式、$i \in AGT$、v(value) が $0 \leq v \leq 1$ を満たす実数の時、$Deserve^i_v \varphi$ は論理式である。
  直感的には、$Deserve^i_v \varphi$ は「エージェント i にとっ

て φ が程度 v で相応しい」ことを表す。

- φ が論理式、$i \in AGT$、p(praiseworthiness) が $0 \le p \le 1$ を満たす実数の時、$Praise^i_p\varphi$ は論理式である。

  直感的には、$Praise^i_d\varphi$ は「エージェント i が φ を真にすることは、程度 p で評価される」ことを表す。

- φ が論理式、$i \in AGT$、a が $0 \le a \le 1$ を満たす実数の時、$Appealing^i_a\varphi$ は論理式である。

  直感的には、$Appealing^i_a\varphi$ は「エージェント i は φ に a 程度魅力を感じる」ことを表す。

- φ が論理式、$i \in AGT$、u が $0 \le u \le 1$ を満たす実数の時、$Unappealing^i_u\varphi$ は論理式である。

  直感的には、$Unappealing^i_u\varphi$ は「エージェント i は φ に u 程度魅力を感じない」ことを表す。

- φ が論理式、$i \in AGT$、f が $0 \le f \le 1$ を満たす実数の時、$Familiar^i_f\varphi$ は論理式である。

  直感的には、$Familiar^i_f\varphi$ は「エージェント i は φ に f 程度親しみを感じる」ことを表す。

この他、$\wedge, \vee, \neg$ などのオペレータ間に一般的な優先順位を導入する。

述語記号、定数記号、変数記号、関数記号を組にしたものを一階言語と呼ぶことにする。一階言語を定めると、一階言語の論理式の集合が決まるのでこれを AF と書く。

3 意味論

3.1 構造

領域の無限集合を $U = \{u_1, u_2, \ldots\}$ とし、以後、U の要素は全て CST に属するものとする。また、以下、$\mathbb{B} = \{\top, \bot\}$ とする。

以下のものを定めておく。

- 可能世界の集合 $W(\neq \emptyset)$
- 各世界 w ごとに、領域 U のもとでの一階述語論理の解釈を１つ決めると、w での原始論理式の真偽が決まる。

  これを関数

  $$V : W \times AF \to \mathbb{B}$$

  で表す。

- AGT の各要素 i に対し、$W \times W$ から [0,1] への関数

  $$\mathscr{B}^i : W \times W \to [0,1]$$

  ただし任意の $i \in AGT$、$w \in W$ に対し、$\sum_{w' \in W} \mathscr{B}^i(w, w') = 1$ であること。また、エージェントが自身の心的状態に関して完全な信念を持つという、通常の BDI logic と同じ性質を成り立たせるため、$\mathscr{B}^i$ は以下を満たすものとする。

  $$\mathscr{B}^i(w, w') > 0 \text{ ならば、} \mathscr{B}^i(w, w'') = \mathscr{B}^i(w', w'')$$

  $\mathscr{B}^i$ は、エージェント i にとっての度合い付きの信念を表す関数である。

- AGT の各要素 i に対し、w から [0,1] への関数

  $$\mathscr{D}^i : w \to [0,1]$$

  エージェント i にとっての各世界の望ましさを定める関数である。

- 可能世界の集合 W から W への全単射

  $$\mathscr{X} : W \to W$$

  この $\mathscr{X}$ は可能世界 w の未来の可能世界 $\mathscr{X}(w)$ を表す関数である。同様に、$\mathscr{X}^{-1}$ も可能世界 w の過去の可能世界 $\mathscr{X}^{-1}(w)$ を表す関数である。

- AGT の各要素 i に対し、$w' \in \mathscr{G}(w)$ を満たす w と w' の組から、[0,1] への関数

  $$\mathscr{E}^i : \{(w, w') | w' \in \mathscr{G}(w)\} \to [0,1]$$

  エージェント i の努力を表す関数である。

- AGT の各要素 i に対し、W から [0,1] への関数

  $$\mathscr{V}^i : W \to [0,1]$$

  エージェント i にとっての各世界の相応しさを定める関数である。

- AGT の各要素 i に対し、W から [0,1] への関数

  $$\mathscr{P}^i : W \to [0,1]$$

  エージェント i にとっての各世界の称賛度を定める関数である。

以上を組にしたものを構造 M と呼ぶ。

W から、2^W への関数 $\mathscr{G}, \mathscr{H}$ を以下のように定める

- $\mathscr{G} = \{w' \,|$ ある非負整数 n が存在して $w_0 = w,$
  $w_1 = \mathscr{X}(w_0), w_2 = \mathscr{X}(w_1), \ldots,$
  $w_n = \mathscr{X}(w_{n-1}), w' = w_n\}$
- $\mathscr{H} = \{w' \,|$ ある非負整数 n が存在して $w_0 = w,$
  $w_1 = \mathscr{X}^{-1}(w_0), w_2 = \mathscr{X}^{-1}(w_1), \ldots,$
  $w_n = \mathscr{X}^{-1}(w_{n-1}), w' = w_n\}$

3.2 解釈

論理式 φ と構造 M、および M の世界 w に対し、φ の M, w での解釈を $[\![\varphi]\!]_{\langle M,w\rangle}$ を以下のように定義する。

- $P \in ATM$ ならば、$[\![P]\!]_{\langle M,w\rangle} = \top$ **iff** $V(w,P) = \top$
- $[\![\neg\varphi]\!]_{\langle M,w\rangle} = \top$ **iff not** $[\![\varphi]\!]_{\langle M,w\rangle} = \top$
- $[\![\varphi \vee \psi]\!]_{\langle M,w\rangle} = \top$ **iff** $[\![\varphi]\!]_{\langle M,w\rangle} = \top$ **or** $[\![\psi]\!]_{\langle M,w\rangle} = \top$
- $[\![\forall x\, \varphi(x)]\!]_{\langle M,w\rangle} = \top$ **iff** $u \in U$ を満たす全ての u に対し、$[\![\varphi(u)]\!]_{\langle M,w\rangle} = \top$
- $[\![\exists x\, \varphi(x)]\!]_{\langle M,w\rangle} = \top$ **iff** $u \in U$ を満たす、ある u に対し、$[\![\varphi(u)]\!]_{\langle M,w\rangle} = \top$
- $[\![Prob^i_l \varphi]\!]_{\langle M,w\rangle} = \top$ **iff** $\sum_{\substack{w' \in W \text{ かつ} \\ w' \text{で} \varphi \text{ が真}}} \mathscr{B}^i(w,w') = l$
- $[\![Effort^i_e \varphi]\!]_{\langle M,w\rangle} = \top$ **iff** $w \in \mathscr{G}(w')$ を満たす w' が存在して、$[\![\varphi]\!]_{\langle M,w'\rangle} = \top$ **and** $\mathscr{E}^i(w,w') = e$
- $[\![Des^i_d \varphi]\!]_{\langle M,w\rangle} = \top$ **iff** $\dfrac{\sum_{\substack{w' \in W \text{ かつ} \\ w' \text{で} \varphi \text{ が真}}} (\mathscr{D}^i(w') \times \mathscr{B}^i(w,w'))}{\sum_{\substack{w' \in W \text{ かつ} \\ w' \text{で} \varphi \text{ が真}}} \mathscr{B}^i(w,w')} = d$
- $[\![Deserve^i_v \varphi]\!]_{\langle M,w\rangle} = \top$ **iff** $\dfrac{\sum_{\substack{w' \in W \text{ かつ} \\ w' \text{で} \varphi \text{ が真}}} (\mathscr{V}^i(w') \times \mathscr{B}^i(w,w'))}{\sum_{\substack{w' \in W \text{ かつ} \\ w' \text{で} \varphi \text{ が真}}} \mathscr{B}^i(w,w')} = v$
- $[\![Praise^i_p \varphi]\!]_{\langle M,w\rangle} = \top$ **iff** $\dfrac{\sum_{\substack{w' \in W \text{ かつ} \\ w' \text{で} \varphi \text{ が真}}} (\mathscr{P}^i(w') \times \mathscr{B}^i(w,w'))}{\sum_{\substack{w' \in W \text{ かつ} \\ w' \text{で} \varphi \text{ が真}}} \mathscr{B}^i(w,w')} = p$
- $[\![G\phi]\!]_{\langle M,w\rangle} = \top$ **iff** 任意の $w' \in \mathscr{G}(w)$ に対して $[\![\phi]\!]_{\langle M,w'\rangle} = \top$
- $[\![H\phi]\!]_{\langle M,w\rangle} = \top$ **iff** 任意の $w' \in \mathscr{H}(w)$ に対して $[\![\phi]\!]_{\langle M,w'\rangle} = \top$
- $[\![X\phi]\!]_{\langle M,w\rangle} = \top$ **iff** $[\![\phi]\!]_{\langle M,\mathscr{X}(w)\rangle} = \top$
- $[\![Y\phi]\!]_{\langle M,w\rangle} = \top$ **iff** $[\![\phi]\!]_{\langle M,\mathscr{X}^{-1}(w)\rangle} = \top$

任意の構造 M の任意の世界 w に対し、$[\![\varphi]\!]_{\langle M,w\rangle} = \top$ が成り立つ時、φ が恒真であるといい、$\vdash \varphi$ と書く。

4 感情

本章では、OCC theory で定義されている感情を、Adam らの形式化を踏まえて、2 章で定義したオペレータを用いて定義する。我々の従来研究では、ある感情の生起を状況に関する、ある信念の存在を基に定義していた。しかし、これではその信念が存続する限り、同様に感情が生起することになり、感情の時間経過による減衰を表現することは困難である。そこで、本論文では 2 章で略記として導入した、信念が生起したことを表すオペレータ *Bel_arise* を用いて、ある感情が生起したことを表す論理式を定義し、それを用いて、感情が時間経過により減衰することを表す論理式を定義する。

また、OCC theory で定義されている 22 個の感情は、以下の 7 グループに分類され、同じグループに属する感情は、生起条件が同様の形で定義されている。

① Well-being emotions：Joy，Distress

② Fortunes-of-others emotions：
Happy-For，Sorry-For，Resentment，Gloating

③ Prospect emotions：Hope，Fear

④ Confirmation emotions：
Satisfaction，Fear-Confirmed，Relief，Disappointment

⑤ Attribution emotions：
(Self-focued)：Pride，Shame
(Other-focued)：Admiration，Reproach

⑥ Well-being/Attribution compounds emotions：
(Self-focued)：Gratification，Remorse
(Other-focued)：Gratitude，Anger

⑦ Attraction emotions：Love，Hate

4.1 に示すように、同じグループに属する感情は、同じ形の論理式で記述できる。よって、4.2 以降では 1 グループあたり 1 つの感情についてのみ、論理式での記述を示す。

4.1 Well-being emotions

Joy

Joy の生起条件は、従来研究[8] では OCC theory に沿って「ある事柄 φ の生起を信じ、かつそれが望ましい事柄であるとき、その事柄に関する Joy の生起が生起する」としていた。我々もこれを踏襲するが、我々は、感情の時間経過による減衰を考慮するため、以下のような定義を行う。他のグループの感情につい

ても同様である。

$$Joy\_arise^i_{f_{J'}(d)}\ \varphi := Des^i_d\ \varphi \wedge Bel\_arise^i\ \varphi$$

$Joy^i_{f_J(d,t)}\varphi$ を新たなオペレータとし、$Joy^i_{f_J(d,t)}\varphi$ の解釈を以下のように定義する。

$$[\![Joy^i_{f_J(d,t)}\varphi]\!]_{\langle M,w\rangle} = \top$$
$$\textbf{iff}\quad [\![Joy\_arise^i_d\varphi]\!]_{\langle M,\mathscr{X}^{-t}(w)\rangle} = \top$$

ただし、関数 $f_{J'}$ は、d に関する増加関数であり、関数 f_J は、t に関する減少関数で、$f_J(d,0)=d$ を満たすものとする。関数 f_J、関数 $f_{J'}$ のどちらも、値域は $[0,1]$ とする。

また、$\mathscr{X}^{-t}(w)$ は、w に関数 $\mathscr{X}^{-1}$ を t 回適用して得られる世界で、w の t 時刻前の世界である。

Distress

Distress の生起条件を以下のように定義する。

$$Distress\_arise^i_{f_{D'}(d)}\ \varphi := Des^i_d\ \varphi \wedge Bel\_arise^i\ \varphi$$

$Distress^i_{f_D(d,t)}\varphi$ を新たなオペレータとし、$Distress^i_{f_D(d,t)}\varphi$ の解釈を以下のように定義する。

$$[\![Distress^i_{f_D(d,t)}\varphi]\!]_{\langle M,w\rangle} = \top$$
$$\textbf{iff}\quad [\![Distress\_arise^i_d\varphi]\!]_{\langle M,\mathscr{X}^{-t}(w)\rangle} = \top$$

ただし、関数 $f_{D'}$ は、d に関する減少関数であり、関数 f_D は、t に関する減少関数で、$f_D(d,0)=d$ を満たすものとする。関数 f_D、関数 $f_{D'}$ のどちらも、値域は $[0,1]$ とする。

4.2 Fortune-of-others emotions

HappyFor

HappyFor の生起条件を以下のように定義する。

$$\begin{aligned} HappyFor\_arise^i_{f_{HF'}(d_i,d_j,v)}\ \varphi \quad := \quad & Bel\_arise^i\ \varphi \\ & \wedge Des^i_{d_i}Bel^j\ \varphi \\ & \wedge Bel^i Des^j_{d_j}\ \varphi \\ & \wedge Bel^i Deserve^j_v\ \varphi \end{aligned}$$

$HappyFor^i_{f_{HF}(d,t)}\varphi$ を新たなオペレータとし、$HappyFor^i_{f_{HF}(d,t)}\varphi$ の解釈を以下のように定義する。

$$[\![HappyFor^i_{f_{HF}(d,t)}\varphi]\!]_{\langle M,w\rangle} = \top$$
$$\textbf{iff}\quad [\![HappyFor\_arise^i_d\varphi]\!]_{\langle M,\mathscr{X}^{-t}\rangle} = \top$$

ただし、関数 $f_{HF'}$ は、d_i,d_j,v に関する増加関数、関数 f_{HF} は、t に関する減少関数で、$f_{HF}(d,0)=d$ を満たすものとする。関数 f_{HF}、関数 $f_{HF'}$ のどちらも、値域は $[0,1]$ とする。

4.3 Prospect emotions

Hope

Hope の生起条件を以下のように定義する。

$$\begin{aligned} Hope\_arise^i_{f_{H'}(d,l)}\ \varphi \quad := \quad & Des^i_d\ \varphi \\ & \wedge Prob\_arise^i_l\ \varphi \\ & \wedge l \neq 1 \end{aligned}$$

$Hope^i_{f_H(d,t)}\varphi$ を新たなオペレータとし、$Hope^i_{f_H(d,t)}\varphi$ の解釈を以下のように定義する。

$$[\![Hope^i_{f_H(d,t)}\varphi]\!]_{\langle M,w\rangle} = \top$$
$$\textbf{iff}\quad [\![Hope\_arise^i_d\varphi]\!]_{\langle M,\mathscr{X}^{-t}\rangle} = \top$$

ただし、関数 $f_{H'}$ は、d,l に関する増加関数、関数 f_H は、t に関する減少関数で、$f_H(d,0)=d$ を満たすものとする。関数 f_H、関数 $f_{H'}$ のどちらも、値域は $[0,1]$ とする。

4.4 Comfirmation emotions

Satisfaction

Satisfaction の生起条件を以下のように定義する。

$$\begin{aligned} Satisfaction\_arise^i_{f_{St'}(x,e,l)}\ \varphi \quad := \quad & Bel^i PHope^i_x\ \varphi \\ & \wedge Bel^i PEffort^i_e \varphi \\ & \wedge Prob\_arise^i_l\ \varphi \end{aligned}$$

$Satisfaction^i_{f_{St}(d,t)}\varphi$ を新たなオペレータとし、$Satisfaction^i_{f_{St}(d,t)}\varphi$ の解釈を以下のように定義する。

$$[\![Satisfaction^i_{f_{St}(d,t)}\varphi]\!]_{\langle M,w\rangle} = \top$$
$$\textbf{iff}\quad [\![Satisfaction\_arise^i_d\varphi]\!]_{\langle M,\mathscr{X}^{-t}\rangle} = \top$$

ただし、関数 $f_{St'}$ は、x,e,l に関する増加関数、関数 f_{St} は、t に関する減少関数で、$f_{St}(d,0)=d$ を満たすものとする。関数 f_{St}、関数 $f_{St'}$ のどちらも、値域は $[0,1]$ とする。

4.5 Attribution emotions

Pride

Pride の生起条件を以下のように定義する。

$$\begin{aligned} Pride\_arise^i_{f_{Pr'}(p,l)}\ \varphi \quad := \quad & Praise^i_p\ \varphi \\ & \wedge Bel^i PProb^i_l \varphi \\ & \wedge Bel\_arise^i\ \varphi \\ & \wedge l \neq 1 \end{aligned}$$

$Pride^i_{f_{Pr}(d,t)}\varphi$ を新たなオペレータとし、$Pride^i_{f_{Pr}(d,t)}\varphi$

の解釈を以下のように定義する。

$$\begin{aligned}&[\![Pride^i_{f_{Pr}(d,t)}\varphi]\!]_{\langle M,w\rangle} = \top\\ \textbf{iff}\quad &[\![Pride\_arise^i_d\varphi]\!]_{\langle M,\mathscr{X}^{-t}\rangle} = \top\end{aligned}$$

ただし、関数 $f_{Pr'}$ は、p に関する増加関数、l に関する減少関数であり、関数 f_{Pr} は、t に関する減少関数で、$f_{Pr}(d,0)=d$ を満たすものとする。関数 f_{Pr}、関数 $f_{Pr'}$ のどちらも、値域は $[0,1]$ とする。

4.6 Well-being/Attribution compounds emotions

Gratification

Gratification の生起条件を以下のように定義する。

$$Gratification\_arise^i_{f_{Grtf'}(x,y)}\varphi \;:=\; Joy\_arise^i_x\varphi \wedge Pride\_arise^i_y\varphi$$

$Gratification^i_{f_{Grtf}(d,t)}\varphi$ を新たなオペレータとし、$Gratification^i_{f_{Grtf}(d,t)}\varphi$ の解釈を以下のように定義する。

$$\begin{aligned}&[\![Gratification^i_{f_{Grtf}(d,t)}\varphi]\!]_{\langle M,w\rangle} = \top\\ \textbf{iff}\quad &[\![Gratification\_arise^i_d\varphi]\!]_{\langle M,\mathscr{X}^{-t}\rangle} = \top\end{aligned}$$

ただし、関数 $f_{Grtf'}$ は、x,y に関する増加関数、関数 f_{Grtf} は、t に関する減少関数で、$f_{Grtf}(d,0)=d$ を満たすものとする。関数 f_{Grtf}、関数 $f_{Grtf'}$ のどちらも、値域は $[0,1]$ とする。

4.7 Attraction emotions

この感情グループには、時間経過による感情の強さの減衰は考慮しない。というのも、現実世界において好意や嫌悪は時間経過より、イベントに影響を受けると考えられるからである。

Love

OCC theory において、Love の生起条件は、「liking an appealing object」と定義されており、魅力的なオブジェクトに対する好意の感情タイプであり、また、強さに影響する変数については、以下の変数が定義されている。

- オブジェクトがどの程度魅力的か
- オブジェクトとどの程度親しみがあるか

この変数を、魅力の程度を変数に持つオペレータ *Appealing* と、親しみの程度を変数に持つオペレータ *Familiar* で表現し、オペレータ *Love* を以下のように定義する。

$$Love^i_{f_L(a,f)}\varphi := Appealing^i_a\varphi \wedge Familiar^i_f\varphi$$

この関数 f_L は、a に関する増加関数、f に関する増加関数とする。

Hate

OCC theory において、Hate の生起条件は、「disliking an unappealing object」と定義されており、非魅力的なオブジェクトに対する嫌悪の感情タイプである。また、強さに影響する変数については、以下の変数が定義されている。

- オブジェクトがどの程度魅力的ではないか
- オブジェクトとどの程度親しみがあるか

この変数を、魅力的ではない程度を変数に持つオペレータ *Unapealing* と、親しみの程度を変数に持つオペレータ *Familiar* で表現し、オペレータ *Hate* を以下のように定義する。

$$Hate^i_{f_{Ha}(u,f)}\varphi := Unappealing^i_u\varphi \wedge Familiar^i_f\varphi$$

この関数 f_{Ha} は、u に関する増加関数、f に関する増加関数とする。

4.8 感情の強さの減衰の表現

ここでは、感情の強さの減衰が本体系でどのように表現されるかを、直感的な議論で例示する。

例えば、ある事象 φ がエージェント i にとって 0.8 ほど望ましく ($Des^i_{0.8}\varphi$)、かつ、ある時刻に φ が起きたことを i が信じたとする。この時、φ が成り立っているという信念はしばらく続くのが自然である。

従来研究では、φ に関する感情 Joy が起きる場合、その強さはその時点で信念 φ があることと、その時の φ の望ましさで決まっていた。望ましさは時刻によって変化しないという仮定を置いているため、信念 φ がある限り、φ に関する Joy の強さは変化しないことになり、感情の強さの減衰は表現できなかった。

これに対して、本論文での体系では、信念 φ が生じた時刻 t に $Bel\_arise^i\varphi$ が成り立ち、この時刻に φ に関する *Joy_arise* がある強さで生起する。以後の時刻において信念 φ が保持されていたとしても、$Bel\_arise^i\varphi$ は成立せず、成り立った時刻 t の *Joy_arise* から次第に減少していく強さで、その時刻での Joy の強さが決まる。

また、このように、信念の保持ではなく生起によって感情の強さが決まるため、信念の生起をトリガリングイベントとしてエージェントの内部状態や行動の変化を引き起こす、実際のエージェントシステムの実装方法とも整合する。

5　感情に関する性質

本論文では、我々の従来研究に比べ、感情の減衰を表現可能となった点が異なる一方で、論理を用いて感情を形式化しているため、従来研究で成り立っていた「感情に関する妥当と思われる性質が、恒真な論理式として示せる」という利点も保持しており、それは本研究で提案した形式化の妥当さを示すものとも言える。その一例としてここでは、対立関係にある感情が両立しないという性質が、論理式として示せることを述べる。

ここで示すのは、同じグループに属する、相反する感情は同時に生起されないという性質である。ここでは、well-being emotions についてのみ説明するが、他の感情グループでも同様に示せるので、他の感情グループについての対立関係については省略する。

- $t = 0$ かつ、*Joy* の $f_{J'}$ が、$f_{J'}(d) = 1.0$ **iff** $d = 1.0$ を満たし、かつ、*Distress* の $f_{D'}$ が、$f_{D'}(d) = 1.0$ **iff** $d = 0$ を満たすとき、
  $\vdash \neg(Joy^{i}_{1.0}\varphi \wedge Distress^{i}_{1.0}\varphi)$

証明

仮定と Joy と Distress の定義より、$f_{J'}(d) = 1.0$ であるなら、$f_J(f_{J'}(d), t) = 1.0$ で、かつ、$d = 1.0$、$f_{D'}(d) = 1.0$ であるなら、$f_D(f_{D'}(d), t) = 1.0$ で、かつ、$d = 0$ である。どちらの感情も d が表すのは $Des^{i}\varphi$ の強さであるため、矛盾する。

よって、

$$\vdash \neg(Joy^{i}_{1.0}\varphi \wedge Distress^{i}_{1.0}\varphi)$$

は成り立つ。

6　考察

本章では、本論文で提案した論理体系の妥当性と、今後の課題について述べる。

Adam らは、感情に関する望ましい性質を恒真な論理式として示すことで、論理体系の妥当性を示している。我々の論理体系は、Adam らの論理体系に準拠し、一階述語論理への拡張したものに新たなオペレータを導入をしているため、Adam らの論理体系で成り立つ感情に関する自然な性質のほとんどは、我々の形式化でもほぼ成り立つ。例えば、相反する感情は同時に生起されないという性質 (ここでは、Joy と Distress)、$\vdash \neg(Joy^{i}_{1.0}\varphi \wedge Distress^{i}_{1.0}\varphi)$ は、我々の論理体系でも成り立つ。この証明については、5 章で述べている。

また、本論文で提案した論理体系によって、時間経過による感情の強さの減衰が表現出きるようになった。それによって、従来では、生起した感情は強さが変わらず、時間が経過してもずっと残ったままになるが、我々の論理体系では、生起した感情の強さが時間の経過によって減衰することが可能となった。これによって、人間の感情のように、時間が経過するにつれて和らいでいく感情を表現することができるようになった。

加えて、信念の生起を考えるようになったこと (p.2 の (1) や (2)) で、より実態および実装を反映出来るようになった。我々の実装では、信念の生起によって感情が生起するため、論理体系と実装の整合性がとれるようになった。

今後の課題については、本論文で提案した論理体系の現実世界における妥当性の検討である。新たに形式化した 2 種の感情も含め、全ての感情について、どの程度人間に近く、もっともであるかの検討が不十分である。また、感情の強さを決定する関数や減衰する関数については、適当な値を返す関数としか決めていない。今回は、減衰については単調に減少していく関数と考えているが、人間の感情は複雑なもので、単調な式で表現可能かという問題もあり、もっともらしさも考慮していかなければならない。

7　終わりに

本論文では、OCC theory で定義されている感情を論理式として形式化した Adam らの論理体系を基に、Adam らの形式化では形式化されていなかった 2 種の感情と、時間経過による感情の強さの減衰を表現可能とするため、新たなオペレータの導入と一階述語

論理への拡張することで、従来では表現できなかった2種の感情と時間経過による感情の強さの減衰を表現できる論理体系を提案した。

今後は、6章で述べたような問題を検討していくことが課題となる。

参考文献

[1] Carole Adam, Andreas Herzig, and Dominique Longin. A logical formalization of OCC theory of emotions. Synthese, Vol. 168, No. 2, pp. 201-248, 2009.

[2] Rafael H. Bordini, Jomi Fred H́ubner, and Michael-Wooldridge. Programming MultiAgent Systems in AgentSpeak using Jason. John Wiley & Sons, 2007.

[3] A. Ortony, G. L. Clore, and A. Collins. The Cognitive Structure of Emotions. Cambridge University Press, 1988.

[4] Anand S. Rao, Munindar P. Singh, and Michael P. Georgeff. Formal Methods in DAI:Logic-Based Representation and Reasoning. Massachusetts Institute of Technology, 1999.

[5] 藤田恵, 片山寛子, 新出尚之, 高田司郎. 実世界の多様性に適応したBDIロボットについて. 情報処理学会論文誌数理モデル化と応用, Vol. 5, No. 1, pp. 50-64, 2012.

[6] 山根瑞樹. OCC theoryによる感情表現を持つエージェントの実現. 2013年度卒業論文, 奈良女子大学理学部情報科学科, 2014.

[7] 石川葉子. 感情に程度の強さを取り入れたエージェントの実現. 2014年度卒業論文. 奈良女子大学理学部情報科学科, 2015.

[8] 池之内彰子. OCC theoryに基づくエージェントの感情表現の論理モデル. 2014年度卒業論文, 奈良女子大学人間文化研究科情報科学専攻, 2015.

[9] 向井香里. 時間経過による感情の程度の減衰を取り入れたエージェントの実現. 2015年度卒業論文, 奈良女子大学理学部情報科学科, 2016.

[10] 今井那緒. 複数の感情の同時生起および削除を取り入れたエージェントの実装. 2015年度卒業論文, 奈良女子大学理学部情報科学科, 2016.