相手の表情を用いた自律エージェントの感情生起

奈良女子大学 生活環境学部 情報衣環境学科情報通信科学コース 4 回生
新出研究室 16483060 一色亜梨沙

2020 年 2 月 5 日

概要

近年、感情が付与され、人間と意思疎通できるロボットの開発が進んでいる。ロボットが動的環境下において適切な感情を生起することで、人間に近い行動選択が可能になると考えられている。本研究では、視覚情報による感情生起を行うことで、より人間らしい感情を持つ自律エージェントの実現を目指した。先行研究では、感情に関する理論 OCC theory の 22 種類の感情を、度合いの大きさや時間経過に伴う変化も含めて実装しているが、感情の度合いを決めるパラメータは人間がストーリーから主観的に推測して入力しているため、客観性に欠ける。本研究では、機械学習によって相手の表情から相手に対する好感度を判定し、Love と Hate の感情を生起する度合いに影響させ、さらにそれを他者に関する感情の度合い決定に利用したときの結果検証を行った。また、表情を考慮した相手の信頼度の設定と結果検証も行った。

1 はじめに

近年、人工知能の発達により、人間と意思疎通できるロボットの開発が進んでいる。ロボットに感情を付与することで、より人間に近い行動選択をすることが期待できる。

周囲の環境が動的に変化する実世界において、そのようなロボットの実現には、BDI アーキテクチャが有効である [13] 。BDI アーキテクチャとは、人間の合理的行動を信念 (Belief)、願望 (Desire)、意図 (Intention) の 3 つの心的状態を用いてモデル化した BDI モデル [5] による行動決定方式を、計算機上で実現したものである。これにより、合理的思考に基づいた行動を選択するロボットの構築が可能となる。

一方で、OCC theory[4] と呼ばれる理論は、人間の包括的な感情を 22 種類に分類し、特徴付けした理論である。この OCC theory は、論理モデルによって表現することが可能で、BDI logic をもつ BDI モデルとの親和性が高く、既存の BDI アーキテクチャの実装である Jason[2] 上の信念ベースを用いて実現することが可能である。

先行研究 [9][12][7] では、OCC theory によって分類された 22 種類の感情すべてが、度合いの大きさや時間経過に伴う減衰も含めて表現されている。また、情報を伝えた他者の信頼性が感情生起に及ぼす影響についても考慮されている。しかし、先行研究では、感情の度合いを決めるパラメータは人間がストーリーから主観的に推測して入力しているため、客観性に欠ける。また、初対面の他者 (対象物) と接する場合などに感情の度合いを適切に推定する方法も示されていない。

本研究では、笑顔、真顔、怒り顔の 3 つに分類した他者の表情から好感度 (魅力) を抽出し、OCC theory 22 種類の感情の中の Love, Hate の感情の生起時における感情の度合い決定に影響させた。なお、視覚情報から抽出した好感度 (魅力) を Love, Hate の感情生起時の度合いに影響させた理由は、先行研究で Love, Hate の実装において、度合いを決定するパラメータのひとつとして魅力が用いられているためである。

その後、その生起させた Love, Hate の感情を、これらを生起条件とする他エージェントの感情生起において、その度合い生成に利用したときの結果検証を行い、視覚情報による感情生起の度合いが妥当であるか確認した。また、他者の信頼性についても、他者の表情を考慮した信頼度の設定を行った。

2 先行研究

先行研究 [9][12] では、BDI モデルと OCC theory を元に、感情を論理式で表現している。本節では、OCC theory 22 種類の感情分類と Adam らによる形式化、さらに先行研究の課題を述べる。

2.1 OCC theory

OCC theory[4] とは、Ortony, Clore, Collins らが提唱した、心理学的見地を基にした 22 種類の感情タイプをモデル化した理論である。感情の特徴付けが明確であるため、計算機科学分野において広く用いられている。以下に分類を示す。

1. 結果の望ましさに関して
   (a) 自分にとっての望ましさに関して
      i. イベントに関して
         - Joy(喜び), Distress(悲しみ)
      ii. 予想に関して
         A. 単なる予想に関して
            - Hope(望み), Fear(恐怖)
         B. 予想していた事象が起こったことに関して
            - Satisfaction(満足), FearConfirmed(恐れていたことが確定)
         C. 予想していた事象が起きなかったことに関して
            - Relief(安堵), Disappointment(落胆)
   (b) 他者にとっての望ましさに関して
      i. 他者が良い結果を得たことに関して
         - HappyFor(共に喜ぶ), Resentment(憤る)
      ii. 他者が悪い結果を得たことに関して
         - SorryFor(共に残念に思う), Gloating(ほくそ笑む)
2. 行動に対する賞賛度に関して
   (a) 行動に関して
      i. 自分の行動に関して
         - Pride(誇り), Shame(羞恥心)
      ii. 他者の行動に関して
         - Admiration(称賛), Reproach(非難)
   (b) 行動とイベントに関して
      i. 自分の望ましさに対する自分の行動に関して (Joy, Distoress との混合型)
         - Gratification(満足), Remorse(後悔)
      ii. 自分の望ましさに対する他者の行動に関して (HappyFor, Resentment との混合型)
         - Gratitude(謝意), Anger(怒り)
3. 対象物の好き嫌いに関して
   - Love(好き), Hate(嫌い)

2.2 Adam らによる形式化

Adam らの形式化 [1] では、感情を BDI logic の論理式で表現している。OCC theory 22 種類の分類を以下の 6 グループに分類することで、グループ毎に同じ形式をもつ感情生起条件が定義されている。

① Well-being emotions : Joy, Distress
② Prospect emotions : Hope, Fear
③ Confirmation emotions : Satisfaction, FearComfirmed, Relief, Disappointment
④ Fortunes-of-others : HappyFor, Resentment, SorryFor, Gloating
⑤ Attribution emotions : Pride, Shame, Admiration, Reproach, Gratification, Remorse, Gratitude, Anger
⑥ Attraction emotions : Love, Hate

Adam らは Love, Hate を除くの感情のみ形式化を行ったが、先行研究では 22 種類の感情すべてが形式化されており、感情の度合いが時間経過に伴う減衰を含めて実装されている。

2.3 先行研究の課題

先行研究 [9][12][7] では、感情の度合いが定められており、それらは周囲の状況などに関するエージェントの信念の度合いや、他者が関わる感情の場合はその他者の好感度などから決められているが、それらは実行時にはストーリーなどから人間が推測して入力している。そのため、客観性を示すことが困難である。また、初対面の他者 (対象物) と接する場合には、好感度を推定する方法がなく、感情の度合いを適切に定めることができない。

3 視覚情報による感情生起

本研究では、課題解決のため、感情生起に用いられる他者の好感度の情報を、相手に関する視覚情報から得る機構を開発した。機械学習によって相手の顔画像を笑顔、真顔、怒り顔の 3 通りに分類し、いずれであるかに応じて相手への好感度を自動に設定し、これを感情の度合いの決定に用いるようにした。

相手に関する視覚情報から得る機構を開発するにあたり、オープンソースの深層学習ライブラリである Caffe を用いて学習を行った。

3.1 Caffe

Caffe[8] は、処理の高速性とモジュール性を考慮した深層学習 (Deep Learning) のオープンソースフレームワークであり、主に画像処理分野に利用されている。Yangqing Jia がカリフォルニア大学のバークレー校の博士課程に在籍していた頃に開発され始めたフレームワークで、現在は、コミュニティのコミッター達によって GitHub 上で開発されている。Caffe には、Python と MATLAB のインターフェースが用意されており、本研究では Python インターフェースを使用した。

3.1.1 深層学習

深層学習とは複数の隠れ層を持つような多層ニューラルネットワークを用いた機械学習の総称である。多層ニューラルネットワークは、隠れ層をいくつも持つようなニューラルネットワークで、隠れ層が多ければ多いほど、そのニューラルネットワークは深いということになる。

3.1.2 LeNet-5

ネットワークを 0 から設計するのは多大な労力が掛かるため、本研究では、既存のネットワークアーキテクチャを使用した。使用したネットワークアーキテクチャは LeNet-5[3] で、1989 年に Yann LeCun らが手書き文字認識等を行うために提案した、入力層を除いて全 7 層で構成されているアーキテクチャである。

先述したとおり LeNet-5 は手書き文字認識用のネットワークであるが、既存研究 [11] で、画家の絵画画像を用いて特徴抽出を行い流派を分別するという研究もあったことから、顔画像の表情の分類程度ならば十分な精度が、AlexNet などと比べて少ない計算資源で得られると考えられたため、採用した。

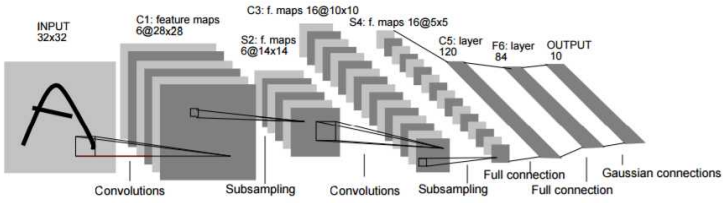

図 1　LeNet-5 ([3] より)

3.2 顔画像分類

表情がはっきりと分かる顔写真を用意し、笑顔、真顔、怒り顔に手動分類して、これを教師として機械学習を行った。学習結果としては、表 1 のような値が得られた。

表 1　学習結果

学習データ数	テストデータ数	学習回数	accuracy	loss
540	135	1 万	0.56609	1.6021
540	135	1 万 (fine-tuning)	0.570451	1.73555
1080	270	1 万	0.889323	0.587153
1620	405	1 万	0.938346	0.281864

学習結果は、画像数が少ない場合は、既存のモデルの一部を再利用し、新しいモデルを構築するという fine-tuning をしても正確性は上がらなかったが、画像数を増やすことで、かなり高い正確性を得ることができた。学習データ 540 枚、テストデータ 135 枚で 1 万回学習させたところ、accuracy = 0.56609, loss = 1.6021 であったのに対し、 学習データ 1620 枚、テストデータ 405 枚で 1 万回学習させたところ、accuracy = 0.938346, loss = 0.281864 となった。loss とは、ネットワークの出力が教師データの正解とどの程度異なるかを、二乗和誤差や交差エントロピー誤差などで数値化した値であり、学習が進んでいるかの判断や学習パ

ラメータの変更などの判断材料に用いられるが、本研究ではパラメータの変更は行っておらず、loss の値は Caffe の出力をそのまま記載したものである。

図 2 に、最も正確性が高かった、学習データ 1620 枚、テストデータ 405 枚で実行した学習の進行による accuracy と loss の変化の様子を示す。

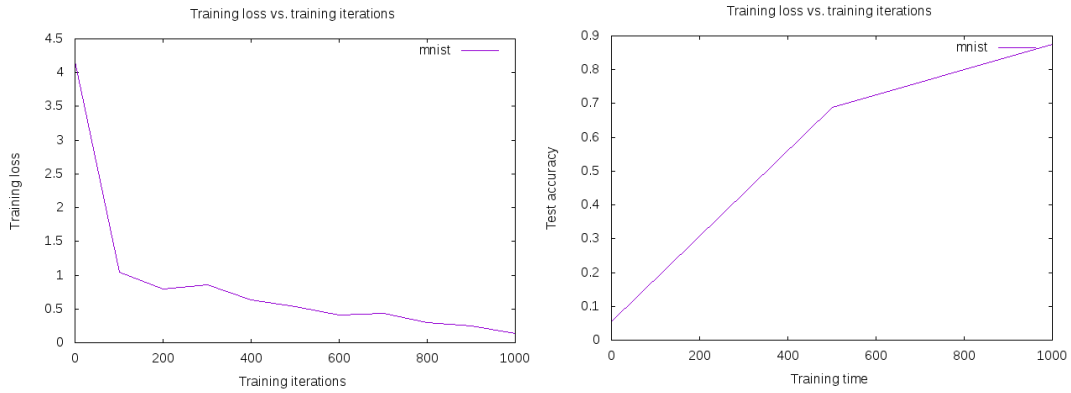

図 2　学習の進行状況

学習回数が増えるにつれ、loss が減り、accuracy が増えているのが分かる。なお、本研究では、バリデーションデータを用いた検証は行っていない。

4　Love, Hate の感情生起

顔画像を笑顔、真顔、怒り顔に分類した後、分類から好感度 (魅力) を定め、感情 Love, Hate の生起の際の度合いの設定に用いた。

4.1　好感度の設定

Love, Hate の感情生起に用いるため、Love, Hate の感情の度合い設定に用いられる好感度の設定を行った。好感度は、既存研究 [6] に基づいて設定した。同研究では、190 人の被験者が笑顔、真顔、しかめ顔から、親しみやすさについてどのような感情を抱いたかが記されている。親しみやすさについては、「思わない」1 点「あまり思わない」2 点「どちらともいえない」3 点「やや思う」4 点「思う」5 点の 5 段階で評価されている。この評価を被験者全体で平均し、表 2 のような値が得られている。

表 2　好感度の評価結果

顔	値
笑顔	4.25
真顔	2.1
しかめ顔	1.51

この結果に基づき、本研究では、5 段階評価であることから 5 で割って、笑顔の場合の好感度を 0.85、真顔の場合を 0.4、しかめ顔の場合を 0.3 とした。その後、好感度 0.5 を 0 程度魅力かつ 0 程度魅力でない「中

間値」とし、笑顔を 0.35 程度魅力である、真顔を 0.1 程度魅力でない、怒り顔を 0.2 程度魅力でないと設定した。

4.2 結果検証

4.1 節で定めた魅力度を、先行研究 [9] による感情の定義の論理式を用いて、Love, Hate の度合いを定めた。Love, Hate の度合いを定めるもうひとつの要因である親密度は、視覚情報からは影響を受けない値なので、初対面の他者と接する場合の親密度は 0 と設定する。*Love* の生起条件を定める論理式は

$$Love^i_{f_L(a,b)}(x) := Appealing^i_a(x) \wedge Familiar^i_b(x)$$

である。これは、「エージェント i がある人物 x のことを a 程度魅力に感じており、かつ、b 程度親密に思っていることが成立した時、i は x に対して $f_L(a,b)$ 程度 *Love* という感情が生起する」ということを意味する。ここで f_L は、引数 a, b に関する増加関数と定められており、先行研究 [9] では $f_L(a,b)$ が、程度 a, b の値を受け取り、$a+b \leq 1$ ならば $a+b$ を、さもなければ 1 を返す関数として実装されている。また、*Hate* の生起条件を定める論理式は

$$Hate^i_{f_H(c,d)}(x) := Unappealing^i_c(x) \wedge Familiar^i_d(x)$$

であり、これは、*Love* の条件の「a 程度魅力に感じている」の部分が「c 程度魅力的に感じていない」という意味になる。関数 $f_H(c,d)$ についても *Love* と同様に、程度 c, d の値を受け取り、$c+d \leq 1$ ならば $c+d$ を、さもなければ 1 を返す関数として実装されている。

我々の実装において、実際の入力と、実験によって生起した感情の出力結果を下記に示す。
初対面の笑顔の人物には、

入力：person appealing(boy)[degOfCert(0.35)]　←人は男の子を 0.35 程度魅力と思う
入力：person familiar(boy)[degOfCert(0)]　←人は男の子を 0 程度親密に思う
出力結果：[person] love0.35　←人は男の子のことが 0.35 程度好き

計算式： $a = 0.35 \quad b = 0 \quad f_L(a,b) = a + b = 0.35$

Love 0.35 が生起し、初対面の怒り顔の人物には、

入力：person unappealing(boy)[degOfCert(0.2)]　←人は男の子を 0.2 程度魅力でないと思う
入力：person familiar(boy)[degOfCert(0)]　←人は男の子を 0 程度親密に思う
出力結果：[person] hate0.2　←人は男の子のことが 0.2 程度嫌い

計算式： $c = 0.2 \quad d = 0 \quad f_H(c,d) = c + d = 0.2$

Hate 0.2 が生起するという結果になり、妥当性を確認した。

5 HappyFor, Resentment の感情生起

他者に関する感情である HappyFor, Resentment, SorryFor, Gloating は、生起条件の論理式に Love, Hate を含むため、画像から定めた魅力度を Love, Hate の度合いに反映させることによって、これらの 4 つの感情にも反映させることができる。本研究では、先行研究 [9] で論理式が実装されている HappyFor, Resentment について示す。 Love, Hate 同様、先行研究による感情の定義の論理式を用いて、HappyFor, Resentment の度合いを定めた。*HappyFor* の生起条件を定める論理式は

$$HappyFor^{ij}_{f_H(a,b)}\phi := Love^i_a(j) \wedge Guess\_Joy^{ij}_b\phi$$

である。これは、「エージェント i がエージェント j のことを a 程度好きで、かつ、i が j がイベントに対して b 程度喜んでいることを推測したら、i は j に対して $f_H(a,b)$ 程度の *HappyFor* という感情を生起する」ということを意味する。ここで f_H は、引数 a, b に関する増加関数と定められており、先行研究 [9] では、$f_H(a,b)$ が、程度 a, b の値を受け取り、$a+b \leq 1$ ならば $a+b$ を、さもなければ 1 を返す関数として実装されている。また、*Resentment* の生起条件を定める論理式は

$$Resentment^{ij}_{f_R(c,d)}\phi := Hate^i_c(j) \wedge Guess\_Joy^{ij}_d\phi$$

であり、これは、*HappyFor* の条件の「a 程度好き」の部分が「c 程度嫌い」という意味になる。関数 $f_R(c,d)$ についても *HappyFor* と同様に、程度 c, d の値を受け取り、$c+d \leq 1$ ならば $c+d$ を、さもなければ 1 を返す関数として実装されている。

我々の実装において、実際の入力と、実験によって生起した感情の出力結果を下記に示す。

エージェントが、初対面の笑顔の人物が「宝くじが当たる」という、0.5 程度喜んでいると推測できる事象に遭遇したときには

初期信念：[person] love0.35　←人は男の子のことが 0.35 程度好き
入力：[person] guess joy(person,boy,takarakuji)0.5　←人は男の子が 0.5 程度喜んでいると推測する
出力結果：[person] happyfor(person,boy,takarakuji)0.85　←人は男の子と 0.85 程度共に喜ぶ

計算式： $a = 0.35 \quad b = 0.5 \quad f_H(a,b) = a + b = 0.85$

HappyFor 0.85 が生起し、初対面の怒り顔の人物が「宝くじが当たる」という、0.5 程度喜んでいると推測できる事象に遭遇したときには、

初期信念：[person] hate0.2　←人は男の子のことが 0.2 程度嫌い
入力：[person] guess joy(person,boy,takarakuji)0.5　←人は男の子が 0.5 程度喜んでいると推測する
出力結果：[person] resentment(person,boy,takarakuji)0.7　←人は男の子に 0.7 程度憤る

計算式： $c = 0.2 \quad d = 0.5 \quad f_R(c,d) = c + d = 0.7$

Resentment 0.7 が生起するという結果になった。

6 表情を考慮した相手の信頼度

先行研究 [7] により、信頼度の定義は、他者から伝えられた情報がどの程度信頼できるかという度合いを 0 から 1 の間で定めたもので、従来の感情生起条件である信念の一部としてエージェントの初期信念に追加されている。信頼度の論理式は、

$$Trust^{i}_{y}j$$

であり、これは「エージェント i がエージェント j に対して y 程度信頼していることを表す」ということを意味する。先行研究では、他者への信頼度の度合いは、不変なものとして設定されており、信頼度を変える場合はソフトウェアを実行し直す必要がある。

6.1 信頼度の設定

本研究では、表情を考慮した相手の信頼度の設定を行った。これにより、信頼度を固定でないものとした。3.2 節で行った笑顔、真顔、怒り顔の画像分類を、既存研究 [10][14] に基づいて、信頼度の設定にも影響させた。

既存研究 [10] では、笑顔と真顔では信頼度に差がなく、怒り顔のみ信頼度が落ちるとされている。また、既存研究 [14] では、24 人の被験者が、怒り顔、真顔、喜び顔から、信用度についてどのような感情を抱いたかが、「全く信用できない」1 点から「非常に信用できる」9 点までで評価されている。この評価を被験者全体で平均し、表 3 のような値が得られている。

表 3　信用度の評価結果

顔	値
喜び顔	4.31
真顔	4.61
しかめ顔	3.94

そこで [10] に基づき、相手が喜び顔の場合と真顔の場合とでは信頼度に差をつけず、怒り顔の場合のみ信頼度を下げることとした。[14] で、しかめ顔の信用度の度合いが、喜び顔と真顔の信用度の度合いの平均値の 0.88 倍であったことから、本研究では、怒り顔のみ、$Trust^{i}_{y}j$ で与えられる信頼度 y の値を 0.88 倍した。

6.2 結果検証

我々の実装において、実際の入力と、実験によって生起した感情の出力結果を下記に示す。

先行研究

入力：trust(boy)[degOfCert(0.7)]　出力結果：prob:0.6 trust:0.7 expect:0.42

先行研究では、0.7 程度信頼している人物から、0.6 程度起こると考えられる事象が起こると伝えられたとき、常に「0.42 程度事象が起きることを予期する」のに対し、

本研究

笑顔、真顔の場合
入力：trust(boy)[degOfCert(0.7)]　出力結果：prob:0.6 trust:0.7 expect:0.42
怒り顔の場合
入力：trust(boy)[degOfCert(0.616)]　出力結果：prob:0.6 trust:0.616 expect:0.3696

本研究では、笑顔や真顔のときは、先行研究と同じように 0.42 程度事象が起きることを予期し、怒り顔のときは、信頼度が 0.88 倍の 0.616 になるため、「0.3696 程度事象が起きることを予期する」という結果になった。

7　おわりに

本研究では、ストーリーからの推測だけでなく、第一印象から抱く好感度の判定に使われる「表情 (顔画像)」から Love, Hate の度合いを生起させ、初対面の相手と接した時の感情表現を生むことで、より人間に近い感情表現の出来るロボットの実現を目指した。また、表情を考慮した信頼度を設定することで、信頼度を固定ではないものにした。

今後の課題としては、表情から生起させる好感度の正確性の向上や、表情に限らない一般的な視覚情報から感情生起に影響を与える情報を得ることなどが挙げられる。また、現時点では学習による分類結果の出力が感情の度合い設定へ自動的に利用されるようにはなっていないので、そこを自動化する必要もある。さらに、好感度の設定や信頼度の設定を参照できる過去の研究が少なく、根拠が十分とは言えないため、度合いを評価する際の評価方法についても検討していきたい。

8　謝辞

本研究の遂行及び本論文の執筆にあたり、指導教員の新出尚之准教授には、丁寧なご指導、ご助言を賜りました。心からの感謝の気持ちとお礼を申し上げたく、謝辞にかえさせて頂きます。

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 Hubner, and Michael Wooldridge. *Programming MultiAgent Systems in AgentSpeak using Jason.* John Wiley and Sons, 2007.

[3] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278–2324, 1998.

[4] A. Ortony, G. L. Clore, and A. Collins. *The Cognitive Structure of Emotions.* Cambridge University

Press, 1988.

[5] Anand S. Rao, Munindar P. Singh, and Michael P. Georgeff. Formal methods in DAI: Logic-based representation and reasoning. In *Multiagent Systems*, pp. 331–376. The MIT Press, 1999.

[6] 井上清子. 表情が初対面の相手に与える印象. 生活科学研究 (文教大学生活科学研究所紀要), Vol. 36, No. 1, pp. 183–194, 2014.

[7] 吉井優佳. 自律エージェントの感情生起における感情の度合いと他者からの情報の信頼性. 2018 年度卒業論文, 奈良女子大学生活環境学部情報衣環境学科生活情報通信科学コース, 2019.

[8] 石橋崇司. Caffe をはじめよう. 株式会社オライリージャパン, 2016.

[9] 浅井沙良. エージェントの感情生起について. 2017 年度卒業論文, 奈良女子大学生活環境学部情報衣環境学科生活情報通信科学コース, 2018.

[10] 大薗博記ほか. 表情と言語的情報が他者の信頼性判断に及ぼす影響. 社会心理学研究, Vol. 26, No. 1, pp. 65–72, 2010.

[11] 長島秀明, 清水郁子. H-032 ディープニューラルネットワークによる画風の特徴抽出. 情報科学技術フォーラム講演論文集, Vol. 13, No. 3, pp. 133–138, 2014.

[12] 塚本麻衣. 感情表現を持つ自律エージェント. 2017 年度卒業論文, 奈良女子大学生活環境学部情報衣環境学科生活情報通信科学コース, 2018.

[13] 藤田恵, 片山寛子, 新出尚之ほか. 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.

[14] 布井雅人, 吉川左紀子. 受け手の体勢が表情画像の印象に及ぼす影響—対人援助場面を想定した検討—. 日本心理学会大会発表論文集, Vol. 82, No. 1, p. 514, 2018.