

# 表情認識によるエージェントの感情想起の精度向上

奈良女子大学生生活環境学部情報衣環境学科生活情報通信科学コース 4 回生  
新出研究室 16480323 玉森しおり

2020 年 2 月 5 日

## 概要

近年、ロボットに感情を付与する研究が進んでいる。ロボットが周囲の環境によって適切な感情を生起及び他者の感情を推測することが出来れば、人間に近い行動を選択することが可能になると考えられる。従来研究では OCC theory の 22 種類の感情を度合いや時間経過に伴う減衰なども含めて BDI モデルを使った基論理式により実装されている。その際に、度合いの決定方法の適切さについては検討されておらず、また画像から情報獲得する機構については未実装であることから、本研究では表情画像からの情報獲得を機械学習で行うとともに、それを用いて従来よりも適切な度合い決定を行った。本論文では感情の度合いの決め方や画像の収集などの機械学習について、それに伴う実験とその結果、課題について述べる。

## 1 はじめに

人工知能が著しい発展を見せる近年において、人間と意思疎通の出来るロボットに関する研究が進められている。特に、簡単な受け答えをするだけでなく、相手及び自らの感情によって対応が変わるロボットであればより人間に近いコミュニケーションを行うことが可能であると考えられる。当研究室では、より人間に近い感情表現を持つエージェントの実現を目指している。

常に周囲の環境が変化していく実世界において、問題解決のための行動を決定するロボットの実装には BDI アーキテクチャが有効とされている。BDI アーキテクチャとは、人間の行動を「信念」「願望」「意図」という 3 つの心的状態でモデル化した BDI モデルによる行動決定式を計算機上で実現したものである。また、感情の形式化には、心理学に基づいた感情の生起に関する OCC theory と呼ばれる理論が多く使用されている。これは、人間の包括的な感情を形式化し、信念、願望などの心的状態を用いて感情を分類、特徴づける理論である。この OCC theory は感情の生起条件が明確であるため、比較的理解しやすいとされ、また、これは BDI モデルが明示的に持つ信念や願望などの概念を用いて感情を特徴付けているため、BDI モデルと親和性が良く、どちらも記号論理で扱うことができる。実際、後述のように BDI アーキテクチャの既存の実装である Jason 上の信念ベースを用いて実装が可能である。

従来研究 [1][2][3] では、Adam ら [4] の論理式を基に既存の BDI アーキテクチャの実装である Jason の信念ベースを用いてこれらの実装や、複数の感情の同時生起、時間経過によって生起した感情の度合いが減衰し、削除されるエージェントの実装を行っている。そして、BDI モデルでの形式化に用いられる論理体系、BDI logic に対し複数の新たなオペレータを導入し、用いることによって、OCC theory によって扱われる 22 種類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。

他者の感情の推測についても研究されている。ただし、[3] でその度合いの決め方の改良が行われているが、決め方の適切さについては検討されていない。また、従来研究 [1][2][3] ではエージェントの感情想起を文章入力からの情報のみで行っている。画像から情報を得る機能は未実装であるため、その情報を入力する際は入力者の判断で文章化する必要があり、特に表情認識において客観性に欠けると考えられる。

そこで、本研究ではそれらの課題を解決するために、表情画像からの情報獲得を機械学習で行うとともに、それを用いて従来よりも適切な度合い決定を行った。

## 2 OCC theory

OCC theory [5] とは Ortony, Clore, Collins らが提唱した、人間の感情を信念や願望といった心的状態を用いて特徴づけた理論である。感情は 22 種類に分類されており、それぞれの特徴づけが明確であるため扱いやすく、計算機科学分野において広く用いられている。OCC theory によって分類された 22 種類の感情については表 1 に示す。なお、混合型とは複数の感情が混合して生成される感情を示す。



## 3.2 教師用画像データ収集

Python ライブラリである iCrawler を使用し、Google 画像検索から画像収集を行った。画像収集には、他に Beautiful Soup と Requests を用いる方法や Scrapy を用いる方法など存在するが、プログラム記述の単純さと収集画像の精度から iCrawler を選択した。収集の際、特にイラストに関しては、できるだけデフォルメされた画像を選び、写実的な画像（例：スケッチ、模写など）は除いた。

次に、収集した画像から顔を認識して切り出しを行った。その認識には、Haar-like 特徴を用いた、ブースティングされた分類器のカスケードを使用したプログラムを [6] を一部修正して利用した。これは、Haar-like 特徴とよばれる明暗で構成される矩形のパターンを検出に使う手法で、Haar-like 特徴と対象画像を重ね合わせ、パターンがあるかないかを判定して顔画像のみを指定サイズで切り出す。切り出した画像を教師用データとするには、目視で分類を行う必要があるため、ある程度の大きさが必要なことから、視認性を考慮し 128 × 128 サイズで切り出した。

その後、不適格な画像などを取り除き、強い喜び、弱い喜び、悲しみの 3 カテゴリーをイラストと写真それぞれで判別するため、それぞれにラベルをつけ、28 × 28 のグレースケールデータに変換し、教師用データとした。

## 3.3 エクマンの基本 6 感情

Charles Darwin が唱えた「感情を表す顔の表情が万国共通であってそれぞれの文化で別々に学習される性質のものではない」という説をもとに P.Ekman, V.Friesen ら [7] は実験を行い、「少なくともある種の顔つきは確かに万国共通であることが明らかにされたが、こうした表情がどんなときに示されるかと言うことでは、文化的な違いがみられる」と主張した。彼らの主張する基本 6 感情の分類は広く知られ、今日の表情認識分野でも指標として使用されている。

彼らは顔を、目の上の眉の周辺部分、目の周りの部分、口の周辺部分の 3 領域に分けて分析し、各領域に現れる特徴を元に表情を怒り・嫌悪・恐怖・喜び・悲しみ・驚きに分類した。ここでは特に Joy を喜び、Distress を悲しみの表情と定義し、画像分類を行った。

## 3.4 Joy

以下の定義 [8] に従い、目視での分類を行った。

- 唇の両端が耳の方向へ水平に引かれる
  - これにより、口角が多少上方に上がる、口は開いている、或いは閉じている
  - これにより、目じりにしわが寄る、また鼻唇溝が菱形にはっきり浮き出る
- 頬が持ち上げられる

ここで、本研究では Joy の度合いの決め方の適切さも検討していくために、Joy-H（強い喜び）、Joy-L（弱い喜び）を定めた。先行研究 [9] より、眼裂形状に関しては「上眼瞼が滑らかで、眼裂が狭く、下眼瞼が直線的である」こと、開口形状に関しては「面積の大小に関係なく、下唇の傾きが笑顔のアイコン形状の特徴である円弧を反映したものである」ことが共に揃っていれば Joy-H の定義とし、それらを片方しか持たない、またはそれに近い笑顔画像を Joy-L とした。

また、5.1 節と 5.2 節の検証実験で使用するために Joy-H の度合いを 0.8、Joy-L の度合いを 0.4 と数値で定義した。

### 3.5 Distress

以下の定義 [8] に従い、目視での分類を行った。

- 眉の内側が持ち上げられる
- 眉の内側が引き寄せられる
- 眉の内側の下が三角形になる
- 頬が持ち上げられる
  - これにより、目じりにしわが寄る
- 唇の両端が押し下げられる
  - これにより、下唇が押し上げられる

イラストに際しては涙がデフォルメされて描かれているものも悲しみに分類するものとした。また、P. Ekman [7] による

悲しみの強さはちょっとした「陰うつ」、または陰うつな感情から、悲嘆のときに感じられる極端な悲しみまで様々である。悲しみの強さは苦痛ほどに強くはなく、強さの質が違っている。悲しみは苦痛に比べて一掃静かで騒々しくなく、言葉で表されることは少なく、抗議調ではない。そして、悲しみは他の感情と混ざりやすく、怒りと恐怖の感情との混合が見られることが多い。特に怒りは苦痛や悲しみの表情を隠蔽するために誇張されたり、擬態されたりすることもあるわけである。

という主張を考慮し、画像単体から悲しみの度合いを目視で判断するのは難しいと判断し、悲しみを表す画像であると分類するにとどめた。

## 4 学習実験

ディープラーニングフレームワーク Caffe を用い、ネットワーク定義は LeNet-5 を用いた。次節以降で述べる 6 種の学習データと、2 種のパラメータ違い (詳細は 4.4.1 節及び 4.4.2 節で述べる) のものを含む 3 通りの solver 定義を用意し学習をそれぞれ行った。train データと test データの比率は 5:1 で作成、バリデーションデータを用いた検証に関しては、今回使用した画像総数から必要ないと判断し、行っていない。

なお、本研究では、デフォルメされたイラストや表情としてはっきり現れているもの及び映画の場面スクリーンショットといった表情が判別し辛いものの両方を判別対象にすることを目的としたため、train データ及び test データにおいて、一般人画像と表情をはっきり作れる著名人の画像を区別せずに学習を行った。

### 4.1 手順

- ネットワーク定義を行う。
  - 設定ファイル (lenet\_train\_test.prototxt) に、処理順番 (Convolution 層/Pooling 層など) とパラメータ設定を定義する。
- solver 定義ファイルを用意する。

- 重みの学習方法を設定するための solver 定義ファイルを作成する。実験ごとに solver 定義ファイルの「test\_iter」「base\_lr」「max\_iter」「solver\_mode」を変更して実験を行う。

- コマンドで学習を開始する。

以下は、solver 定義ファイル例 [10] である。

```
# The train/test net protocol buffer definition
net: "/home/tamamori/Caffe/gazou/lenet_train_test.prototxt"
# test_iter specifies how many forward passes the test should carry out.
# In the case of MNIST, we have test batch size 100 and 100 test iterations,
# covering the full 10,000 testing images.
test_iter: 11
# Carry out testing every 500 training iterations.
test_interval: 500
# The base learning rate, momentum and the weight decay of the network.
base_lr: 0.01
momentum: 0.9
weight_decay: 0.0005
# The learning rate policy
lr_policy: "inv"
gamma: 0.0001
power: 0.75
# Display every 100 iterations
display: 100
# The maximum number of iterations
max_iter: 20000
# snapshot intermediate results
snapshot: 5000
snapshot_prefix: "lenet"
# solver mode: CPU or GPU
solver_mode: CPU
```

## 4.2 実験 1

データ：人種、男女、年齢など区別なく無作為に集めたもの。

test_iter	4
base_lr	0.01
max_iter	15000,20000
solver_mode	CPU

区分	train	Test
Distress	299	40
Distress-I	37	7
Joy-H	1053	210
Joy-L	344	78
Joy-IH	104	20
Joy-IL	60	12
合計	1897	367
計	2264	

学習回数	Accuracy
15000	0.78
20000	0.778

### 4.3 実験 2

データ：人種、男女など区別なく無作為に集め、帽子や眼鏡を装着した画像をノイズとして取り除いたもの。

test_iter	4
base_lr	0.01
max_iter	15000,20000
solver_mode	CPU

区分	Train	Test
Distress	282	38
Distress-I	36	6
Joy-H	930	200
Joy-L	235	68
Joy-IH	89	20
Joy-IL	66	12
合計	1638	344
計	1982	

学習回数	Accuracy
15000	0.763
20000	0.765

実験 1 に比べてわずかに精度が落ちているため、帽子や眼鏡を装着したデータもこれ以後画像データに含めることとする。

#### 4.4 実験 3

データ：実験 2 のデータに日本人の著名人の画像追加及びイラスト数を増加したもの。

日本人の著名人 アイドル、モデル、俳優といった「表情をはっきりと作ることが出来る人間」を指す。

test_iter	8
base_lr	0.01
max_iter	15000,20000
solver_mode	CPU

区分	Train	Test
Distress	532	90
Distress-I	80	14
Joy-H	1697	420
Joy-L	1030	146
Joy-IH	400	40
Joy-IL	358	24
合計	4097	734
計	4831	

学習回数	Accuracy
15000	0.815
20000	0.785

#### 4.4.1 実験 3-1

データ：実験 3 と同一データ、学習率のみ変更したもの。

test_iter	8
base_lr	0.05
max_iter	15000
solver_mode	CPU

学習回数	Accuracy
15000	0.02

#### 4.4.2 実験 3-2

データ：実験 3 と同一データ、学習率のみ変更したもの。

test_iter	8
base_lr	0.001
max_iter	15000,20000
solver_mode	CPU

学習回数	Accuracy
15000	0.801

#### 4.4.3 実験 3-1,3-2 まとめ

実験 3 (学習率 0.01 の場合) と比べ精度が落ちているため、以後学習率は 0.01 で固定した。

#### 4.5 実験 4

ここで、佐藤氏 [11] らによるエクマン理論への反証「日本人と西洋人の表情は同じ名前がついていても異なる表現が現れる」という説にそって以下の実験を行った。

データ：実験 3 のデータから日本人以外を取り除いたもの。

test_iter	7
base_lr	0.01
max_iter	15000,20000
solver_mode	CPU

区分	Train	Test
Distress	393	57
Distress-I	57	14
Joy-H	1323	269
Joy-L	862	194
Joy-IH	335	66
Joy-IL	286	60
合計	3256	660
計	3916	

学習回数	Accuracy
15000	0.798
20000	0.805

結果、実験 3 より精度が落ちていることが確認できる。

#### 4.6 実験 5

データ：実験 3 のデータに JAFFE Database 画像などの追加及びイラスト数を増加したもの。JAFFE とは、The Japanese Female Facial Expression の略で、JAFFE Database とは 10 人の日本人女性の表情差分画像集である。

test_iter	11
base_lr	0.01
max_iter	15000,20000
solver_mode	CPU

区分	Train	Test
Distress	849	108
Distress-I	95	14
Joy-H	2341	492
Joy-L	1352	291
Joy-IH	350	70
Joy-IL	308	62
合計	5295	1037
計	2264	

学習回数	Accuracy
15000	0.860
20000	0.859

この実験において、精度が一番高いことを確認した。

#### 4.7 実験 6

データ：対照実験として画像が極端に少なかった時の検証用のもの。

test_iter	2
base_lr	0.01
max_iter	15000,20000
solver_mode	CPU

区分	Train	Test
Distress	100	20
Distress-I	100	20
Joy-H	100	20
Joy-L	100	20
Joy-IH	100	20
Joy-IL	100	20
合計	600	120
計	720	

学習回数	Accuracy
15000	0.585
20000	0.570

画像が少ない場合、非常に精度が落ちることを確認した。

#### 4.8 実験結果まとめ

結果として、実験 5 (4.6) つまり日本人、西洋人を区別しない最もデータ数の多いものが精度が高いことが分かった。仮説としては、エクマン理論の反証を踏まえた実験 4 の日本人のみのデータが一番精度が高くなるとの予想をしていたが、その予想は外れた。実験データ収集の際、表情をはっきり出すことが出来る被写体の画像を多く収集したために、「表情をはっきり出せない日本人＝一般人」の前提が崩れたことと、またデータ数の多さがより大きく働くものと考えられる。

## 5 検証実験

自身の感情生起の適切さを示すために、最も精度の高かった実験 5 の学習結果を使い、2 セットの 3 コマの画像 (図 3、図 4) を用いて、この画像を文章に起こしたシナリオによる度合い付き感情生起が画像から読み取れる度合い付き感情と一致するかどうかの検証実験を行った。実験は、生起した感情を基に行動を選択するプランを Jason 上で記述することによって行った。

シナリオ：犬は散歩に行きたかった。犬は散歩に連れて行ってもらう。犬は喜んだ。  
このシナリオを元に、登場する犬の感情を生起する実験を行う。(図 3、図 4 共通)



図 3



図 4

ここで、程度の強さを持つ Joy(喜び) の生起条件は

$$Joy_{f(d)}^i \varphi := Bel^i \varphi \wedge Des_d^i \varphi$$

である。これは「エージェント  $i$  がイベント  $\varphi$  が成立することを信じ、かつそのことがエージェント  $i$  にとって  $d$  程度望ましいことであるならば、 $i$  は  $\varphi$  の成立に対し、程度  $f(d)$  の Joy という感情を生起する」ということを意味する。

Joy の程度に関する変数  $d$  として、「エージェント  $i$  においてイベント  $\varphi$  がどの程度望ましいか ( $Des_d^i \varphi$ )」を導入している。イベントの望ましさ  $d$  が大きいほど、Joy の程度も大きくなると考えられているため、度合いを計算する関数  $f(d)$  は、[12] の実装では、 $d$  を受け取り、そのまま Joy の程度として返す関数としている。

分類した結果の Joy-H 及び Joy-L の数値を ( $Des_d^i \varphi$ ) の  $d$  の値として使い、その結果、例えば  $d = 0.8$  なら  $f(d)$  も 0.8 となり、それが Joy の度合いになる。

以後の検証実験では、この条件式によって Joy の度合いが生起するものとする。

## 5.1 図3 シナリオの検証

図3のシナリオに沿って以下のように信念を与えて実験した。結果は以下ようになった。左の数は説明のために付加した行番号である。

```
1 Connected
2 Jason Http Server running on http://127.0.1.1:3272
3 [inu] 追加
4 [inu] done check!
5 [inu] love0.7
6 [inu] done check!
7 [inu] done check!
8 addPercept(inu, perc(addbel,1,des(inu,sanpo)[degOfCert(0.8)]))
9 [inu] !b des(inu,sanpo)[degOfCert(0.4)]
10 addPercept(inu, perc(addbel,2,bel(inu,sanpo)))
11 [inu] !b bel(inu,sanpo)
12 [inu] done check!
13 [inu] joy0.4
14 [inu] done check!
15 addPercept(inu, perc(goal,3,act1(inu,sanpo)))
16 [inu] !act1(inu,sanpo)
17 [inu] joy
18 [inu] 0.18744000000000002
19 [inu] 少し喜ぶ
20 [inu] done check!
```

8・9行目で、画像から Joy-L を判断し、3.4節で定義した数値 0.4 を代入している。

13行目で joy0.4 が生起されている。

18・19行目で、散歩に行くというイベントに対し、少し喜ぶ(弱い喜び)が生起されている。これは図3の3コマ目のイラストから読み取れる度合いと一致する。

## 5.2 図4 シナリオの検証

図4のシナリオに沿って以下のように信念を与えて実験した。結果は以下ようになった。左の数は説明のために付加した行番号である。

```
1 Connected
2 Jason Http Server running on http://127.0.1.1:3272
3 [inu] 追加
4 [inu] done check!
5 [inu] love0.7
6 [inu] done check!
7 [inu] done check!
8 addPercept(inu, perc(addbel,1,des(inu,sanpo)[degOfCert(0.8)]))
9 [inu] !b des(inu,sanpo)[degOfCert(0.8)]
10 addPercept(inu, perc(addbel,2,bel(inu,sanpo)))
11 [inu] !b bel(inu,sanpo)
12 [inu] done check!
13 [inu] joy0.8
14 [inu] done check!
15 addPercept(inu, perc(goal,3,act1(inu,sanpo)))
16 [inu] !act1(inu,sanpo)
17 [inu] joy
18 [inu] 0.6716000000000001
19 [inu] 喜ぶ
20 [inu] done check!
```

8・9行目で、画像から Joy-H を判断し、3.4節で定義した数値 0.8 を代入している。

13行目で joy0.8 が生起されている。

18・19行目で、散歩に行くというイベントに対し、喜ぶ(強い喜び)が生起されている。これは図3の3コマ目のイラストから読み取れる度合いと一致する。

## 5.3 検証実験まとめ

5.1 実験と 5.2 実験の結果から、表情画像から Joy の度合いを読み取り、その情報をもとに画像からシナリオを文章化する際の違いを明確にし、その結果生起する感情の度合いが異なることを確認した。

## 6 終わりに

本研究では従来の研究をふまえ、新たに機械学習を用いた表情画像の分類によりエージェントの感情 Joy, Distress の度合いを決定する機構の作成を行った。分類ができれば、その精度と関係なく既存の感情生起機構との連携はできるが、当然、分類精度が高い方が求まる度合いも妥当なものとなるので、今後の課題としては、分類器のより一層の精度向上がある。また、視覚情報一般から感情生起の条件となるような情報の抽出を実現することも課題として考えられる。これにより、人間に近いコミュニケーションが出来るロボットの実現への寄与を期待したい。

## 7 謝辞

本研究の遂行及び本論文の作成にあたり、指導教員の新出尚之准教授から何度も繰り返し丁寧なご指導、ご助言を賜りました。心からの感謝の気持ちと御礼を申し上げたく、謝辞にかえさせていただきます。

## 参考文献

- [1] 浅井沙良. エージェントの感情生起について—新たな感情の追加—. 2017 年度卒業論文, 奈良女子大学生生活環境学部情報環境学科生活情報通信科学コース, 2018.
- [2] 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, 奈良女子大学生生活環境学部情報環境学科生活情報通信科学コース, 2018.
- [3] 吉井優佳. 自律エージェントの感情表現の度合いと他者からの情報の信頼性. 2018 年度卒業論文, 奈良女子大学生生活環境学部情報環境学科生活情報通信科学コース, 2019.
- [4] Carole Adam, Andreas Herzig, and Dominique Longin. A logical formalization of OCC theory of emotions. *Synthese*, Vol. 168, No. 2, pp. 201–248, 2009.
- [5] A. Ortony, G. L. Clore, and A. Collins. *Cognitive Structure of Emotions*. Cambridge University Press, 1988.
- [6] Tatsuro64. <https://qiita.com/Tatsuro64/items/95b0ce48b6bb155bfe29>. 美女を見分けられない機械はだだの機械だ : OpenCV, Python3 による画像からの顔抽出 (2020 年 2 月 5 日閲覧).
- [7] Ekman P. and Friesen W. V. 表情分析入門. 誠信書房, 1987.
- [8] 工藤力. しぐさと表情の心理分析. 福村出版, 1999.
- [9] 菅原徹. 笑顔の形状と表情筋活動の分析. 可視化情報, Vol. 34, No. 133, pp. 64–69, 2014.
- [10] 石橋崇司. Caffe をはじめよう. オライリージャパン, 2015.
- [11] Sato W., Hyniewska S., Minemoto K., and Yoshikawa S. Facial expressions of basic emotions in Japanese laypeople. *Frontiers in Psychology*, Vol. 10, No. 259, pp. 1–11, 2019.
- [12] 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, 奈良女子大学生生活環境学部情報環境学科生活情報通信科学コース, 2016.