

# 表情認識における顔パーツの着目度に関する 人と深層学習モデルとの比較

伊藤 彩世† 飯山 将晃†

滋賀大学 データサイエンス学部†

## 1. はじめに

人間の心理状態や感情を理解するうえで、表情は大きな役割を持つ。コミュニケーションの中では表情の種類を読み取ることに加え、表情の大小も読み取り、相手がどれだけ強くその感情を抱いているかを感じ取ることも重要である。

近年、深層学習モデルによる表情認識性能が向上しているが、本当に人間によるそれと同等の捉え方をできているのかは未知である。モデルによる推論は、そのメカニズムがわかりにくい。そのため、モデルの認識精度は向上しているものの、注目している部分や誤った判断をする画像の特徴が人間とは異なる可能性がある。仮に精度が高かったとしても誤り傾向が人間のそれと異なると、人の表情認識機能の代替としては不十分である。そこで本研究では深層学習モデルによる表情認識を人によるそれと近づけることを最終目標とし、その予備的調査として人と機械の2つの表情認識機能の比較を行う。

## 2. 分析方法

深層学習モデルの認識能力と人の認識能力との差異を明らかにするために、それぞれが顔のどこに注目して表情を認識しているのかを調査する。そのために本研究では通常の顔画像に加え、口、目、鼻、頬の4か所を隠したマスク画像を用いて表情認識タスクの性能を比較する。まず、これらの画像を用いて、エクマンの基本6表情[1]に無表情を加えた7表情分類を行うモデルを構築する。次に、人間にも同様の画像を提示し、どの表情に最も近いと感じるかを回答させる。その結果とモデルの性能とを比較する。また、同様の調査を表情表出の大小を図る表情

Comparison of human and deep learning models on the degree of focus on facial parts in facial expression recognition.

†Sayo Ito †Masaaki Iiyama  
Faculty of Data Science, Shiga University

表出レベル推定[2]でも行う。なお、表情表出レベル推定に関して、少なくとも人にとっては表出レベルを数値化することは困難であり、数値化の基準も個人差が大きい。そのため、2つの画像のうちどちらの顔画像の方がより強く感じるかを回答させることで相対的な表出レベルの大小により認識性能を評価する。顔画像データセットにはBU-3DEF[3]の二次元の顔画像を用い、マスク画像は顔の特徴点であるランドマーク推定に基づいてマスク領域を決定する。

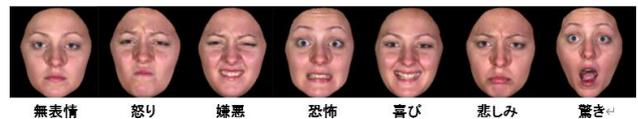


図1. 7表情の分類の例

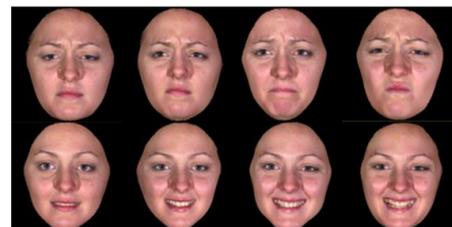


図2. 表情表出レベルの例。左から右に向かって強い表出である。

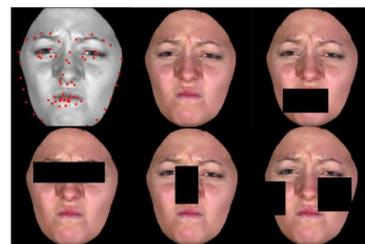


図3. ランドマーク推定と各マスク画像の例

## 3. 比較結果

被験者12人から得られたアンケート結果のうちマスク無し画像に対する7表情分類タスクにおけるモデルと人間による結果の比較を表1に、表情表出レベル推定タスクにおけるモデルと人間による結果の比較を表2に示す。

表1. 表情認識結果の深層学習(M)と人間(H)との比較 (マスク無し画像) 14 画像×12 人

M/H	NE	AN	DI	FE	HA	SA	SU
NE	11	0	0	1	0	0	0
AN	0	7	16	0	0	13	0
DI	1	5	9	1	1	7	0
FE	0	0	0	0	0	0	0
HA	1	1	0	1	33	0	0
SA	9	1	6	0	0	8	0
SU	4	7	8	4	2	3	8

無表情(NE)と喜び(HA)の表情に対する2つの認識結果は近いことがわかった。それに対し、異なる認識をする傾向にあったのは怒り(AN)であった。モデルが怒りと認識する表情を人間は嫌悪(DI)や悲しみ(SA)と捉え、人間が怒りだと感じる画像をモデルは驚き(SU)だと感じる人が多い。また、モデルが嫌悪や悲しみ、驚きと判断する表情に対する人間の表情の捉え方には大きく個人差がある結果となった。

表2. 表情表出レベル推定の比較 (マスク無し画像) 12 問×12 人

M/H	○	×
○	71	37
△	10	26
×	0	0

共に2つが誤った認識をした画像は存在しなかった。全体の結果と比べ、モデルで正解、人間で不正解だった数の割合と人間で正解、モデルで同じレベルが出力された数の割合が大きく、マスク画像よりもマスクなしの場合の方が、2つの認識傾向の違いが大きくなる結果だった。

同様の手順で口、目、鼻、頬にマスクを与えた画像でも表情認識を行い、2つの結果を比較した。

7 表情分類においては、(1)両者ともに口と目にマスクを与えた場合に認識精度が顕著に低下する。(2)人間はモデルが怒りや嫌悪だと判断する表情を無表情や悲しみだと感じる傾向がある。(3)反対に人間が無表情や恐怖だと捉える画像をモデルは悲しみや怒りだと出力する傾向にある。このように誤認識する画像の傾向は一致していない。

一方、表情表出レベル推定に関しては、(1)モ

デルの認識精度はマスク無しと鼻マスク画像で下がり、反対に口と目のマスク画像では高い。これはモデルが口や目から得られる情報によって認識が阻害されていることを示唆する。これは表情認識とは異なる傾向である。(2)人間はマスク無しと口マスク画像で誤る傾向が強く、モデルとは異なり、口を頼りに判断していることを示唆する。(3)人間はマスク無しでも誤りが多いことから、顔の各パーツの特徴に影響され、正しい認識が困難になっていることが考えられる。

このように、表情表出レベル推定ではそれぞれ隠す顔のパーツによってモデルと人間とで異なる傾向となった。

#### 4. まとめ

本研究では、深層学習モデルと人間の表情認識機能の差異を明らかにするため、マスク画像を用いて2つの性能比較を行った。7 表情分類において2つの認識機能を近づけるためには、人間が無表情や恐怖、怒り、嫌悪だと感じる顔画像を再学習させる必要がある結果だった。一方、表情表出レベル推定において2つを近づけるためには、モデルが特に口を注視して正しく表情の強さを感じ取れるように学習させる必要がある。さらに、両者が認識を誤る顔画像にはそれぞれ特徴があったため、人間が得意とし、かつモデルが苦手な画像をよりモデルに学習させることも重要である。しかし、今回の顔画像データセットはレベル付けのために作成されたものであり、人間にとってナチュラルさに欠ける表情である可能性があることから、学習させるデータセットについて改善する必要がある。

#### 参考文献

- [1] Paul Ekman. "Universal facial expressions of emotion.", Culture and Personality: Contemporary Readings/Chicago, 1974
- [2] O. Rudovic, V. Pavlovic, and M. Pantic, "Multi-output Laplacian Dynamic Ordinal Regression for Facial Expression Recognition and Intensity Estimation," Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 2634-2641, 2012.
- [3] "https://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE\_Analysis.html"