

人物画像からの姿勢情報を用いた感情指定

永瀬 智之¹ 岸田 嵩平² 堀田 裕弘³

富山大学工学部¹ 富山大学大学院理工学教育部² 富山大学学術研究部(都市デザイン学系)³

1. はじめに

人の感情状態と身体動作には深い関係があるとされている。しかし、どのような動作がどのような感情と関連性があるかは、個人差があり明確な統一的な見解はない。そこで、本研究では、既公開のデータベース[1]を用いて、身体動作と感情との関係を、OpenPose[2]を用いて骨格化した身体情報を用いてその関係性を検討する。

2. 実験方法

まず、既公開のデータベースより図1のような人物画像を準備する。これらはそれぞれ、怒り、恐れ、喜び、悲しみの感情を表している。



怒り 恐れ 喜び 悲しみ

図1 人物画像

図1の怒りの画像にOpenPoseを用いて骨格化すると、図2の左側のような人物の骨格データが得られる。OpenPoseは深層学習を用いて25個の関節点を検出しており、それぞれの座標と、その座標がどれだけ信頼できるかを表す信頼度を出力することができる。

OpenPoseで座標を出力したのちに正規化をしたものが図2の右側である。これは首の座標を原点とし、首から骨盤までの距離を1(基準化)として、その拡大縮小率に合わせて他の座標の位置を再計算している。

これらの座標データをニューラルネットワーク(NN)で学習し、姿勢から感情を推定する。今

Emotion estimation using posture information from human image

¹ Tomoyuki Nagase, School of Engineering, University of Toyama

² Kohei Kishida, Graduate School of Science and Engineering, University of Toyama

³ Yuukou Horita, Faculty of Sustainable Design, University of Toyama, horita@sus.u-toyama.ac.jp

回の実験ではそれぞれの感情に約60枚ずつ、計約240枚の画像を使う。

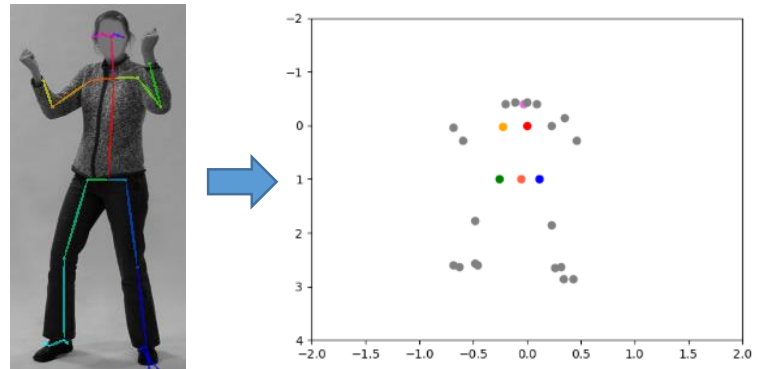


図2 正規化

ニューラルネットワークで学習させる際、どの部位が感情推定により役立つのか調べるため、図3に示すような範囲をもとにデータセットを作成する。具体的には、「基軸+第一範囲」、「基軸+第一範囲+第二範囲」、「すべての範囲」の3種類とし、これらのデータから信頼度を除いたデータセット3種類を含め計6種類である。

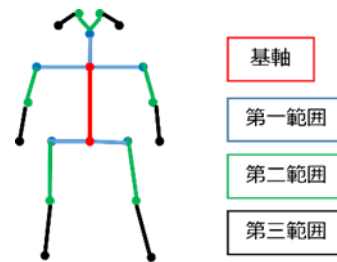


図3 特徴点の制限範囲

なお、データセット240枚からランダムに8割を選んで学習用データにし、2割を選んでテスト用データにする。それぞれの感情が8割と2割ずつになるようにする。これを10回ずつ、計60回ニューラルネットワークで学習させ、推定結果を出す[3]。

3. 実験結果

基軸+第一範囲をA群、基軸+第一範囲+第二範囲をB群、すべての範囲をC群として、10回学

習させたときの正答率の平均を表1に示す。また、各群の正答率の平均と標準偏差を図4に示す。

表1 推定結果

	正答率 (10回)
A群	0.44
A群 (信頼度なし)	0.43
B群	0.57
B群 (信頼度なし)	0.52
C群	0.73
C群 (信頼度なし)	0.68

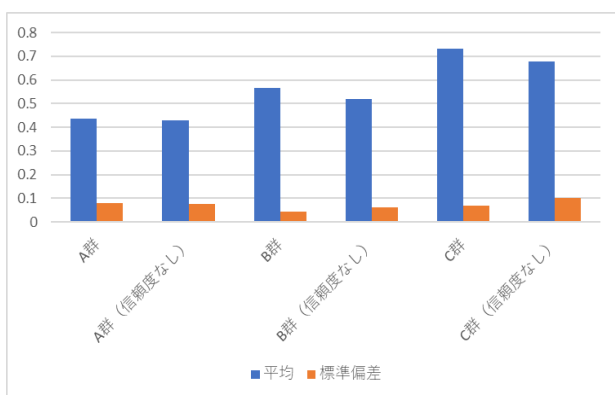


図4 平均と標準偏差

表1を見ると、学習に用いる骨格の範囲が大きいほど正答率が上がっている。手足の関節が感情推定に大きく役立っていることが見て取れる。図4において、信頼度を用いた場合では平均正答率は若干の向上が見られる。

また、用いた関節の範囲の違い、信頼度の有無の違いについて一元配置ANOVAを行った。その結果を表2と表3に示す。

表2 ANOVA(用いた関節の範囲)

信頼度の有無	p-value	
有	1.69E-09	***
無し	2.3E-06	***

(p-value,***p<0.01,**p<0.05,*p<0.1)

表3 ANOVA(信頼度の有無)

関節範囲	p-value	
第一範囲	0.877699	
第二範囲	0.072206	*
第三範囲	0.209001	

(p-value,***p<0.01,**p<0.05,*p<0.1)

表2のp値が非常に小さいことより、信頼度の有無に関わらず、用いた関節の範囲を増やすことで正答率が向上している傾向(表1)は、統計的にも有意であることが分かった。一方、表3のp値が比較的大きな値であることより、どの関節範囲においても信頼度の有無は正答率には影響がないことが示された。

次に、もっとも正答率の高かったC群における混同行列(Confusion Matrix)を表4に示す。表4を見ると、それぞれの感情の正答率が一番高くなっている。特に、悲しみが高い。また、怒りと恐れを互いに間違えているのが目立つことから、動作が似ているとポジティブな感情をネガティブな感情と推定している。

表4 C群における Confusion Matrix

解 \ 予測結果	怒り	恐れ	喜び	悲しみ
怒り	0.57	0.23	0.16	0.04
恐れ	0.31	0.57	0.11	0.00
喜び	0.04	0.08	0.85	0.03
悲しみ	0.05	0.00	0.00	0.95

4. おわりに

OpenPose とニューラルネットワークを用いて人物の姿勢から感情推定を行った。結果、正答率は関節点の座標が多いほうが精度は上がった。

今回は10回だけ学習させたときの正答率を比べたが、学習の数をもっと増やせば正答率は変わるのか、今後の課題として検討の余地がありそうである。

参考文献

- [1] Beatrice de Gelder, Jan Van den Stock “The bodily expressive action stimulus test (BEAST). Construction and validation of a stimulus basis for measuring perception of whole body expression of emotions” (2011)
- [2] Zhe Cao, et al, ”Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields”, IEEE Conference on Computer Vision and Pattern Recognition, pp.1302-1310 (2017).
- [3] 岸田嵩平, 堀田裕弘, ”画角の異なる多視点映像からの人物同定法の検討”, 信学技報, vol. 119, no. 131, IMQ2019-1, pp. 1-5, 2019年7月