

日本文化の感情表出における表情の特徴抽出に関する検討 Study on facial feature extraction in emotional expression in Japanese culture.

平野 善識[†] 奥村 健太[†]
Yoshinori Hirano Kenta Okumura

1. はじめに

近年、ビジネスや私生活に関わらず、WEB 会議システムを利用して、オンラインで意思疎通を行う機会が増えてきている。しかし、こういった WEB 会議では、対面と比較して情報の取得が困難であるために齟齬などの様々な問題が生じている。とりわけ日本においては、直接的な感情表現を避ける文化[1]も手伝い、そういった問題が深刻化する可能性が高い。そこで、表情や音声信号といった意思疎通において重要となる非言語情報を複合し「場の雰囲気」を工学的に捕捉することで、意思疎通の客観視に役立てたい。

そのために主に観測すべきは人間の感情表出であるが、非言語情報の感情表出は文化依存の差異が非常に大きい。例えば表情における既存研究[2]では、マスク着用時の表情認知判断の実験により、アメリカ人と日本人との間で口元から受ける情報取得量に大きな差があることが報告されている。このことから、欧米文化圏と比較して、日本文化圏では口元以外の表情で感情を判断していると推測できる。そこで、顔の部位や特定の特徴量に着目して表情の認識精度を評価できる仕組みを確立できれば、こういった推測について更に踏み込んだ検証をすることが可能となる。

しかし、本研究の着手にあたっては、日本文化圏のみで完結する検証を行う場合、文化圏の狭さからデータ面など環境の整備に不足が目立つ。そのため、検証環境の構築を優先し、変化の分かり易さやデータの入手性を考慮して、欧米文化圏を対象に部位ごとの影響を調査するのが妥当だと考える。欧米文化圏において口の影響が重要であることを工学的に示せれば、後の工程で日本文化圏のデータを少量でも用意できた際に、日本文化圏の表情認識に関する通説を裏付けることにもなる。

以上を踏まえ、本稿では研究の諸段階として、表情に関して、欧米文化圏の大規模データセットを利用し、部位ごとなどで特徴選択を行った場合の影響について調査する。

2. 提案手法

本稿では、日本文化圏での表情認識を用いたシステムの実用化に向けて、特定の表情特徴が文化圏の感情表出に与える影響の調査を目的とする。そのため、提案手法には特徴量を主体的に選択できること、人間の解釈可能性が高く、計算コストが膨大でないことが求められる。これらの考慮は、今後の研究となる日本文化圏にとって有用な新規特徴量を探索する際にも必要であり、継続的な検証環境の構築という観点でも重要である。

以上の理由から、本稿では表情分析において一般的な深層学習ではなく、狭義の機械学習である決定木学習を用いて実験を行う。機械学習は深層学習と比較して分類精度は低い、特徴量の影響を調査するための解釈可能性は高い。

特に、今回採用する決定木学習はデータを分類するためのルールが明確な構造を持ち、各決定点での条件やその条件に至った理由が直観的に理解しやすいという利点がある。機械学習を行う際に利用する特徴量については、既存のコンピュータビジョンライブラリを用いて、画像から顔のランドマークを検出し、それらの座標から計算的に抽出する。

3. 実験

実験は機械学習を利用したことによる表情認識精度の確認と、それに対し特定部位の表情特徴を除外した場合の影響調査の二段階に分けて行う。なお、認識対象とした表情クラスは Neutral, Happy, Sad, Surprise, Fear, Disgust, Anger, Contempt の 8 種類である。

細かな実験条件については、表 1 の通りである。データセットは、手動でアノテーションされた約 42 万枚の顔画像の内、ノイズの少ないものを選別した約 28 万枚の画像を利用することとした。また、ランドマーク検出には、468 点もの特徴点を検出でき、特徴量の創出性が高いことから MediaPipe を採用した。

利用する特徴量については、権利的に問題の無い画像[3]を用いて図 1 で示した。今回は差分のない顔画像からの取得であるため、各特徴点同士の直線距離を算出して特徴量として利用した。 F_2, F_5, F_6 以外は、左右を平均した値を算出している。また、 F_1 については眉のランドマークが上下に分かれて並行しているため、その平均値を新たな特徴点として定義した。

表 1 実験条件

データセット	AffectNet [4]
比較用手法	EmoNet [5]
提案手法	決定木学習
ランドマーク検出	MediaPipe Face Mesh [6]

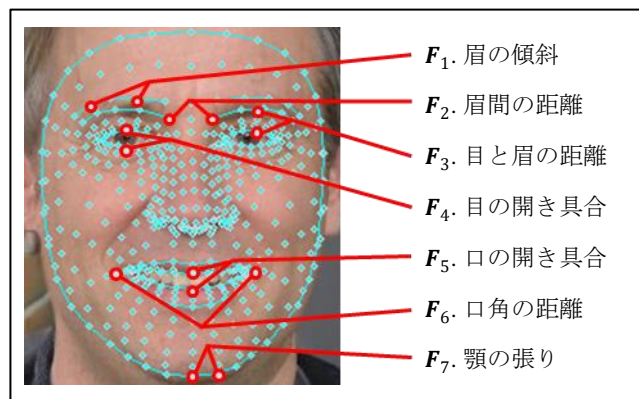


図 1 使用した表情特徴

[†]名古屋国際工科専門職大学 International Professional University of Technology in Nagoya

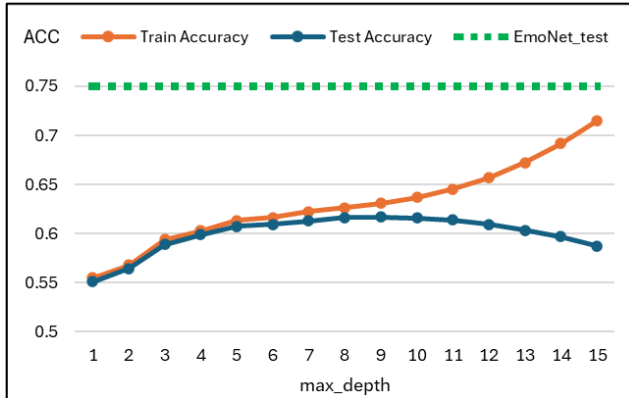


図 2 3.1 実験結果

3.1 機械学習手法の精度確認

提案手法の訓練データとテストデータの正解率を図 2 に示す。比較手法は表 1 にある EmoNet であり、本実験と同様のデータセット、表情クラスを利用している深層学習モデルである。

結果を見ると、深層学習と機械学習の性能差としては妥当な数値が出ていると言える。また、決定木の深さについては、9 あたりを境に性能低下と過学習が認められる。このことから、以降は max_depth を 9 として行えば提案手法としては最適な性能を基に実験を進めることができる。

3.2 部位ごとの特徴量検証

実験を行う上で、全特徴量 S を顔上部 $S1$ と顔下部 $S2$ に分割し、部分特徴量として定義する。それらの部分特徴量を用いて構築した各決定木を探索し、条件に従って抽出したノードについて S に対する gini 値の差分を求めた。その結果を表 2 に示す。抽出条件としては、殆どが例外処理的な働きをしていた gini 値 0 を除き、最小値のものを採用している。またサンプル数の問題から Disgust, Contempt については各決定木でそれらを有効に分類できなかった。

結果から、欧米文化圏の感情表出においては、顔の部位に依存して特徴が表れやすい表情がいくつかあることが示唆される。具体的には、Happy や Fear といった表情クラスでは顔下部の影響が強く、逆に顔上部は殆ど影響をしていないことが分かる。また、その他の表情クラスについても、 S に対する $S1$ と $S2$ で明確な差が見られない場合が多い。これを踏まえると、必ずしも顔全体の特徴を用いて分類を行う必要がないと言える。実際に、図 3 で示している部位

表 2 部位ごとの影響 (ルートノードの gini 値 = 0.695)

	Neutral	Happy	Sad	Surprise	Fear	Anger
全特徴量 $S(F_1, \dots, F_7)$	0.375	0.02	0.133	0.238	0.320	0.133
顔上部の特徴量 $S1(F_1, \dots, F_4)$	0.198	0.087	0.32	0.153	0.625	0.278
	-0.177	+0.067	+0.187	-0.085	+0.305	+0.145
顔下部の特徴量 $S2(F_5, \dots, F_7)$	0.278	0.023	0.375	0.124	0.320	0.444
	-0.097	+0.003	+0.242	-0.114	± 0	+0.311

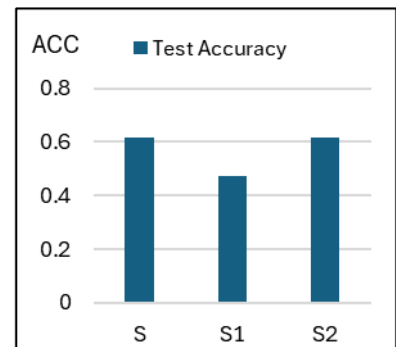


図 3 部位ごとの正解率

ごとの正解率を見ると、 $S2$ のみでテストデータを分類したときと S との正解率がほぼ変わらないと分かる。

以上から、欧米文化圏では顔上部よりも顔下部の方が感情表出に寄与していると判断できる。

4. おわりに

今回の研究により、特定の文化圏において各特徴量の影響を評価可能な仕組みが構築でき、特に顔上部と顔下部では、明確な性能差が認められた。これは、第 1 節で推測した欧米文化圏にとって口回りの表情特徴が重要であることを定量的に裏付ける結果となった。

とはいえ、今回の提案手法は単純な部位の違いを検証するに留まっており、特徴量について、算出方法を直線距離だけでなく面積にすることや扱っていない部位の取得など、更なる探求の必要性がある。それに応じて、部位特徴量の定義も見直すことも求められる。また、適用対象となるデータセットに関して、サンプルの少ない表情クラスや文化圏のデータを適切に分析するためには、少量でも今回と同様の実験結果を得られる必要がある。そのために、データベースを文化圏ごとに切り分けて用いるなど、実験条件のフィルタリングを通して、小規模な文化圏への対応を進めていかなければならない。

これらのステップを交互に繰り返していくことで、当面の目標とする日本文化圏への適用が可能となるまで検証環境を最適化し、確かな再現性をもって日本文化圏の感情表出を把握することを目指していく。これにより、未整備な環境でも日本文化圏に対応できるシステムの実用化が促進されれば、WEB 会議の客観視など、様々なコミュニケーション支援技術への応用が期待できる。

参考文献

- [1] P・エクマン, W・V・フリーゼン, 工藤 力(訳), “表情分析入門”, 誠信書房, p.33-34, (1987).
- [2] 齊藤 俊樹, 元木 康介, 高野 裕治, “マスクをした顔に対する表情認知の文化差”, 日本認知心理学会発表論文集, (2022).
- [3] thispersonnotexist.org, <https://thispersonnotexist.org>, (2024-06-14).
- [4] Ali Mollahosseini, Behzad Hasani, and Mohammad H. Mahoor, “AffectNet: A New Database for Facial Expression, Valence, and Arousal Computation in the Wild”, IEEE Transactions on Affective Computing, (2017).
- [5] Antoine Toisoul, Jean Kossai, Adrian Bulat, Georgios Tzimiropoulos, Maja Pantic, “Estimation of continuous valence and arousal levels from faces in naturalistic conditions”, Nature Machine Intelligence, Vol.3, No.1 (2021).
- [6] Google, <https://ai.google.dev/edge/mediapipe/solutions/guide>, (2024-06-14).