

脳波指標と心拍変動指標の簡便な計測機器と特徴量選択による感情推定モデルの構築と精度検証

鈴木圭¹ ラオハカンワンウィットティポーン¹ 菅谷みどり¹

概要：機械学習を応用した脳波と心拍変動指標による感情推定において、2つの課題が存在する。1つ目は、生理指標の計測機器が高価であり、装着が簡便でない点である。2つ目は、精度向上が期待できる、感情推定に不要な指標の削除が行われていない点である。そこで本研究では、これら2つの課題を解決しつつ、精度が良い感情推定モデルの構築を目的とする。安価かつ装着が簡便な単一電極の脳波計と光電式脈波計を用いた。また、多角的な複数手法により不要な指標の削除を行った。男性18名、女性7名の計25名の実験参加者から機械学習に用いるデータを収集した。そして、深層学習により感情推定モデルを構築し、そのモデルの精度検証には、推定を試みる対象人物の生理指標のデータを学習データに含める方法と学習データに含めない方法の2種類を用いた。その結果、前者では4感情分類において99%の精度を得た。一方、後者では前者と比べ著しく低い23%の精度を得た。これは学習をせず無作為な推定をするモデルの精度25%を下回る結果である。この結果により、安価な機材でも感情推定をする対象人物の生理指標のデータを学習データに含めることができれば精度よく感情推定でき、目的が達成できることが示唆された。しかし、含めることができれば低い感情推定精度となり、目的が達成できないことが示唆された。

キーワード：感情推定，機械学習，特徴量抽出，特徴量選択，脳波，心拍変動指標

Construction and Accuracy Verification of Emotion Estimation Model Using a Simple Measurement Device and Feature Selection for Electroencephalography and Heart Rate Variability Indices

KEI SUZUKI^{†1} TIPPORN LAOHAKANGVALVIT^{†1} MIDORI SUGAYA^{†1}

1. はじめに

近年、工学的な分野において、人の感情を推定する研究が数多くなされており、教育やマーケティング、医療などの複数分野で応用が期待されている[1], [2]

このように複数分野での応用が期待されている感情推定技術の手法に、生理指標を用いる手法が存在する。生理指標は人の顔の表情や声よりも恣意的に変更することは困難であることから、客観性が高い[2]。そのため、近年では生理指標を用いることによる感情推定に注目が集まっている[3]。

こうした生理指標を用いる感情推定において、心理学分野におけるラッセルの円環モデルやこれと類似した Arousal Valence Space モデルが用いられる(図 1)。これらは人の基本感情を覚醒度(Arousal)と快適度(Valence)の2軸で表現するモデルである[4], [5]。これらのモデルは、生理指標を用いる感情推定において利用実績のある一般的なモデルであり[1]、多くの感情を2軸のみで表現できるうえに定量的に感情を評価できるという利点がある[5]。

これらのモデルが生理指標とともに用いられる例として池田らの研究が存在する。池田らは、ラッセルの円環モデルにおける Arousal 軸に覚醒度指標として脳波指標を対

応づけ、Valence 軸に快適度指標として心拍変動指標を対応づけた[6]。脳波は、集中状態の測定に用いられることや主観評価された覚醒度と負の相関があることが報告されている[1], [7]。そのため、脳波指標を Arousal に対応づけることは一定の妥当性がある。一方、心拍変動指標は不快感情と関係のあるストレスの検出において、信頼性の高い指標として用いられてきた[8]。また、心拍変動指標が低い場合はリラックスした状態を示す[1]。そのため、心拍変動指標を Valence に対応づけることは一定の妥当性がある。

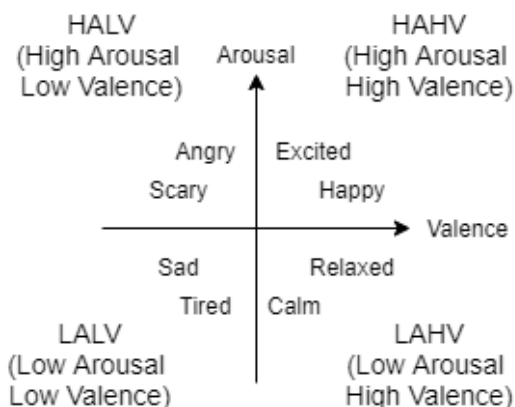


図 1 Arousal-Valence Space Model

¹ 芝浦工業大学
Shibaura Institute of Technology

しかし、池田らの方法は、生体指標で発生する個人差を考慮していないという課題が存在する。彼らは、心拍変動指標の1つである pNN50 を Valence に対応付け、0.3 以上ならば high valence(快適)、0.3 未満ならば low valence(不快)とした。このように、0.3 という固定された閾値を設けることで、生体指標で発生する個人差が考慮されていないという課題が存在する。

この課題を解決した研究に浦部らの研究が存在する。浦部らは、これらの生理指標と、本人の主観により得られた感情情報をもとに、機械学習を行う方法を提案した[8]。浦部らは、脳波指標や心拍変動指標を説明変数とした深層学習により、個人ごとに感情推定モデルを構築することで、個人差が考慮されていないという課題を解決した。このモデルの感情推定精度を検証した結果、図 1 の Arousal-Valence Space モデルの各象限である HAHV, HALV, LALV, LAHV の 4 分類において平均 0.80、最大 1.00 の分類精度が得られることが報告された。

しかし、医療などの分野への応用を考えた場合、平均精度が 80%では、まだ不十分である可能性がある。一般的に、機械学習による推定モデルの構築にあたり、特徴量の算出・抽出や選択により、推定に有用な特徴量の発見や不要な特徴量の除去を行うことで、精度を向上させることができるとされている[9]。浦部らの研究で用いた特徴量の数は、脳波指標が 5 つ、心拍変動指標が 1 つの計 6 個であり、特徴量の抽出や選択が十分になされていないことが、精度が十分に達していない要因の 1 つであると考えられる。

脳波と心拍変動指標を説明変数として用いた他の研究に Stamos ら研究が存在する[5]。Stamos らは 14 個の電極を持つ脳波計から 42 個の特徴量を抽出し、心電図から心拍変動指標を含めて 71 個の特徴量を抽出した。これらすべての特徴量を用いて感情推定をした結果、Low Arousal、と High Arousal、Low Valence と High Valence の 2 値分類において、それぞれ 62%程度の精度が得られたことを報告した。

Stamos らは、浦部らや Tong らよりも抽出した特徴量の数が多いが、特徴量選択が行われていない。また、生理指標の計測機器に課題が存在する。脳波計は、電極の装着にあたり頭皮に触れる必要があり、髪の毛の合間を縫って電極を装着する必要がある。また、装着にあたり電気抵抗を減らすために食塩水や専用のジェルが必要な脳波計も存在する。ユーザの快適性や装着の簡便性を高めるために電極が少ないものが推奨される[3]が、Stamos らは 14 個もの電極を用いている。また、心拍変動指標を算出するにあたり心電図を用いているが、光電式脈波計はより安価に心拍変動指標を計測できる[1]。

本研究では、これまでに述べた従来の研究と比べてより高い精度の感情推定モデルを構築すること、および、安価で装着が簡便な生理指標の計測機器を用いつつ感情推定することを目的とし、脳波・心拍変動指標の特徴量を抽出・

選択する。そして、深層学習により感情推定モデルを構築し、そのモデルの精度評価を行う。この際、ユーザの快適性や装着の簡便性を高めるために、安価で単一電極である簡易脳波計を用いる。さらに光電式脈波計を用いることで安価に心拍変動指標を算出する。これにより、感情推定技術をより簡単に利用できるようにする。また、特徴量を選択する際、実績がある複数の特徴量選択手法を用いることで多角的に行う。

本論文の構成は以下の通りである。2 節にて、機械学習に用いるために抽出した脳波・心拍変動指標について述べる。3 節にて、機械学習に用いるデータの収集方法について述べる。4 節にて、特徴量選択とその結果について述べる。5 節にて機械学習による感情推定モデルの構築とその精度検証について述べる。6 節にて、本論文のまとめを述べる。

2. 脳波・心拍変動指標の説明

2.1 脳波指標

脳波とは、脳内の電気信号を電極などを用いて記録したものである。脳波は周波数帯により種類を分類でき、周波数帯により反映する心理状態や脳の活動状態が異なる[10]。一般的に、 α 、 β 、 γ といった比較的広い周波数帯ごとに指標として用いられるが、本研究では、周波数帯を細分化することで、Low α 、High α といった、詳細に人の状態を表現できる指標も用いた。

以上の脳波指標の他に、区間を 15 とした脳波指標の移動平均を用いた。脳波指標には、激しい変動が存在する。この激しい変動が、脳波指標を用いた機械学習アルゴリズムの内部における、目的関数の最小化や閾値算出の障害になると考える。移動平均を算出することで、これらの障害を減らすことができ、より精度高く感情推定できる可能性があると考えたため、検討した。本研究での検討の対象とする脳波指標と周波数帯などを表 1 に示す。

表 1 脳波指標の周波数帯と解釈

指標	周波数帯	解釈
δ	1-3Hz	夢を見ない深い睡眠、無意識
θ	4-7hz	直感的、創造的、夢
Low α	8-9hz	リラックス、平穏、意識的
High α	10-12hz	リラックスしているが集中している
Low β	13-17hz	思考、事故及び環境の認識
High β	18-30hz	警戒、動揺

Low γ	31-40hz	記憶, 高次精神活動
Mid γ	41-50hz	視覚情報処理
α^{*1}	8-12hz	リラックス
β^{*2}	13-30hz	ストレス
γ^{*3}	31-50hz	認知
MA15 x^{*4} ($x=\theta, \delta,$ Low α , High α , Low β , High β , Low γ , Mid γ , α, β, γ)	-	-

*1... α = Low α + High α として算出

*2... β = Low β + High β として算出

*3... γ = Low γ + Mid γ として算出

*4...区間を 15 とした指標 x の移動平均

脳波計には、侵襲性の低い単一電極の簡易脳波計である NeuroSky 社製の Mindwave Mobile2 を用いた。この脳波計のサンプリングレートは 512hz である。測定部位は国際 10-20 法における AF3 である。Mindwave Mobile2 からは α , β , γ の脳波指標を直接取得できないが、 α = Low α + High α , β = Low β + High β , γ = Low γ + Mid γ として算出した。

2.2 心拍変動指標

心拍変動指標とは、心拍間隔である IBI から算出される指標であり、隣接する心拍の時間間隔の変動である。IBI の標準偏差や変動係数を算出することにより、快適度や感情により影響が及ぼされる交感神経や副交感神経の指標となる。本研究で検討の対象とする心拍変動指標と、その算出方法および、その値が反映する状態などを表 2 に示す。

表 2 心拍変動指標の算出方法

指標	算出方法	解釈
IBI	心拍間隔	緊張・落ち着き
Heart Rate(HR)	1 分間の拍動回数	緊張・落ち着き
pNNx ^{*1} ($x=10, 20, 30, 40, 50$)	隣接する IBI の絶対値が x ms を超えた割合	副交感神経
SDNN ^{*1}	IBI の標準偏差	交感神経と副交感神経
RMSSD ^{*1}	IBI の差の	副交感神経

	自乗平均平方根	
SDNN / RMSSD ^{*1}	SDNN/RMSSD	交感神経
CVNN ^{*1}	IBI の変動係数	交感神経と副交感神経
LF ^{*2}	IBI を周波数解析した 0. 04-0. 15hz のパワー値	交感神経と副交感神経
HF ^{*2}	IBI を周波数解析した 0. 15-0. 40hz のパワー値	副交感神経
LF/HF ^{*2}	LF/HF	交感神経

*1...IBI 値が取得されるごとに、区間を 30 として値を算出する

*2...IBI 値が取得されるごとに、200 秒間の IBI データを用いて値を算出する。

心拍変動指標である LF, HF は周波数解析により算出した指標である。周波数解析の手順は次に述べる 1)~6)である。1)200 秒間の IBI データを 3 次スプライン補完する。2)1hz にリサンプリングする。3)リサンプリングしたデータにハニング窓をかける。4)高速フーリエ変換する。5)LF, HF それぞれの周波数帯のパワースペクトルを積分する。6)以上の手順 1)~5)を IBI データが取得されるごとに行う。

心拍変動指標は心拍間隔を表す IBI から算出されるが、その IBI の算出には Switch Science 社の光電式脈波センサと Arduino の組み合わせたものを用いた。なお、この脈波計のサンプリングレートは 500hz である。

3. データ収集

機械学習による感情推定モデルの構築にあたり、感情を伴っている状態の脳波・心拍変動指標のデータや目的変数に用いるどの感情を抱いているのかを判別するデータが不可欠である。また、感情を抱いている状態を作り出すための感情喚起の方法は重要である。この節では、これらデータの取得するにあたり行った実験方法について述べる。また、収集したデータをもとに構築する、機械学習に用いるデータセットの構築方法について述べる。なお実験参加者の人数は、男性18名、女性7名の計25名である。

3.1 感情喚起の方法

本研究では、感情喚起を行うために、ユヴァスキュラ大学の研究者らが作成した音楽データベース上の音楽を用いた[11]。このデータベース上の音楽は、Arousal-Valence Spaceモデルにおける各象限に属する感情のスコアが評価されている。このスコアを参考に、Valence-Arousal Spaceモデルにおける各象限に対応する感情を喚起するための音楽を任意に選択した。

3.2 感情データの計測方法

本研究では、どの感情を伴っているのかのデータを計測する方法として主観評価を行った。その際、言語などに左右されず、できるだけ客観的に主観評価を行うために、

Self-Assessment Manikin(SAM)を用いた(図 2)[12]. 実験参加者は自分の感情に最も近いマネキンを SAM 上から選択することで感情を評価する. 各マネキンには 1~9 の値が割り振られており, マネキンを選択することによって定量的な Arousal, Valence の主観評価ができる.

SAM による主観評価の結果から, 感情が Arousal-Valence space モデルにおけるどの感情に属するのかを判断した. 具体的には, $Arousal > 5$ かつ $Valence > 5$, または $Arousal = Valence = 5$ では第 1 象限の HAHV, $Arousal \leq 5$ かつ $Valence > 5$ では第 2 象限の HALV, $Arousal < 5$ かつ $Valence \leq 5$ では第 3 象限の LALV, $Arousal > 5$ かつ $Valence < 5$ では第 4 象限の LAHV に属すると判断した.

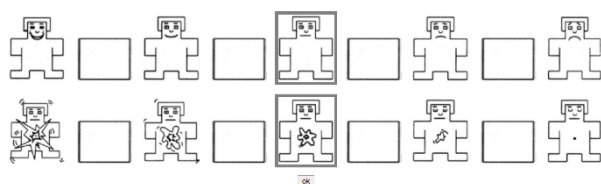


図 2 SAM

3.3 実験手順・環境

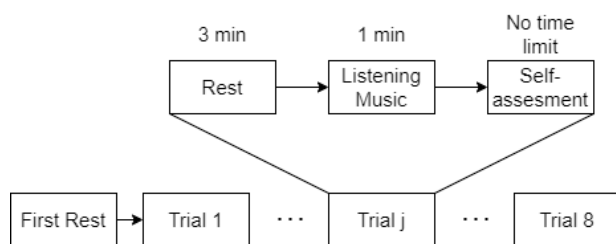


図 3 実験手順

実験参加者が実施した実験手順は次に述べる 1)~4)になる(図 3).

- 1) 椅子に座るとともに, 脳波計・脈波計・イヤホンの装着をする.
- 2) 実験の練習として, 手順 3)~4)以降の手順を簡略化したものを実施する.
- 3) 10 分間, 安静の状態待機する.
- 4) 2 分間の安静, 1 分間の音楽視聴, 時間制限無しの, 音楽により喚起された感情の主観評価を 1 施行として, 8 施行行う.

安静時・音楽視聴時には, 実験参加者の前に配置したディスプレイ上に, 背景を灰色とした中心に黒い十字がある画像を表示した. 実験参加者には, できるかぎり真ん中の十字に視点を置くように指示をした. 実験場所は静かで目がくらまない程度の明るい部屋で実施し, 部屋はできるだけ静かにした. そのうえ, 実験協力者には実験中においてイヤホンを装着するように指示をした.

3.4 データセットの構築

実験手順 3)~4)において, 脳波・心拍変動指標の記録を続けた. データを用い, まず脳波の移動平均を算出した. その後, 記録されたデータの中から実験協力者に対して音楽刺激が加えられた時間帯のデータを取り出し, これを機械学習に用いるデータセットとした.

脳波計として用いる Mindwave Mobile からは 1 秒に 1 回算出された脳波指標を取得される. 一方, 心拍変動指標の算出に用いられる IBI は 0.5-1.5 秒ほどの周期で取得される. これらの同期しないデータをもとにデータセットを構築するにあたり, 1 秒ごとに記録された脳波指標と心拍変動指標の最新のデータを取り出すことでデータセットを構築した. こうすることで, 1 秒につき 1 個データが生成される. 音楽刺激は一施行につき 60 秒程度, 実験参加者に加えられる. そのため, 一施行につき 60 個程度のデータが生成される. この 60 個程度のデータに対して, 音楽刺激直後の主観評価結果を割り当てる. 以上のデータセットの構築方法により Arousal-Valence Space モデルのそれぞれの象限において, 3558, 2175, 2704, 3312 個得られた.

4. 特徴量選択

特徴量選択にあたり, それぞれの指標について, 感情推定に寄与すると予測される度合を複数の手法により定量化し, これを統合した. 複数手法の結果を統合し, 推定に有用な特徴量を多角的に得ることで, より推定精度が高い特徴量を選択ができることは Haq らによって検証され, 一定の成果を収めている[13]. そのため, 本研究にて行った複数手法の結果を統合することによる特徴量選択は一定の妥当性があると考えられる.

特徴量選択の手法には, 相関比, 相互情報量, ランダムフォレストの重要度, SVM L1 正則化の重み係数を用いた. これらを用いた理由としては, 既の実績があること, 重要度が算出されることで, 特徴量選択を行いやすくすることができるという点を考慮した[14][15][16].

本節では, まず, それぞれの特徴選択手法について説明した後, それらの手法を統合する手法と統合結果, および, その考察について述べる.

4.1 相関比

相関比とは, 質的データと量的データとの関連性が定量化された値である. 本研究では, Arousal-Valence Space モデルのどの各象限に属するのかを識別するためのカテゴリデータを設けている. 質的データであるこのカテゴリデータと量的データである脳波・心拍変動指標の関連性を観察するために相関比を用いた.

算出方法は, 式(1)により表される. 式中的変数の意味を説明すると次のようになる. η^2 : 相関比. a : 質的データの種類の数. n_i : i 番目の質的データに属する特徴量 x のデータ数. $\bar{x}_i$: i 番目の質的データに属する特徴量 x の平均値. $\bar{x}$: 特徴

量 x の平均値. $x_{i,j}$: i 番目の質的データに属する j 番目の特徴量 x の値である.

$$\eta^2 = \frac{\sum_{i=1}^a n_i (\bar{x}_i - \bar{x})^2}{\sum_{i=1}^a \sum_{j=1}^{n_i} (x_{i,j} - \bar{x})^2} \quad (1)$$

4.2 相互情報量

相互情報量とは、2 変数間の関係性が定量化された値である. 本研究においては、喜怒哀楽を表すカテゴリデータと脳波・心拍変動指標との関係性を定量化した.

質的データと量的データとの相互情報量の算出式は、式(2)のようになる[15]. 式の中の変数の意味を説明すると次のようになる. $I(X; Y)$: X と Y の相互情報量. $p(x)$: x の確率. $p(y)$: y の確率. $p(x, y)$: x と y の条件付き確率である.

$$I(X; Y) = \sum_{x,y} P(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) \quad (2)$$

4.3 ランダムフォレストの重要度

ランダムフォレストの重要度とは、ランダムフォレストによる機械学習において算出される、各特徴量の推定における寄与の度合いが定量化された値である. ランダムフォレストは複数の決定木を作成し、それぞれの決定木におけるノードにおいてデータが分類されてゆく. それらの決定木による分類結果の投票で、最終的な推定をするアルゴリズムである.

算出方法は、式(3)、式(4)により表される. 式の中の変数の意味を説明すると、次になる. I_x : 特徴量 x の重要度. N : 特徴量 x により分岐されたノード数. $\Delta I_{x,n}$: 特徴量 x により分岐された n 番目のノードにおける不純度の減少量. G_{parent} : n 番目のノードの親ノードにおける不純度. $G_{Child Left}$: n 番目のノードにおける左側の子ノードにおける不純度. $G_{Child Right}$: n 番目のノードにおける右側の子ノードにおける不純度. m : n 番目のノードのデータ数. m_{Left} : n 番目のノードにおける左側の子ノードのデータ数. m_{Right} : n 番目のノードにおける右側の子ノードのデータ数である.

$$I_x = \sum_{n=1}^N \Delta I_{x,n} \quad (3)$$

$$\Delta I_{x,n} = G_{parent} - \frac{m_{Left}}{m} * G_{Child Left} - \frac{m_{Right}}{m} * G_{Child Right} \quad (4)$$

複数の決定木における各ノードにおいて、式(3)のような正解データの分類による不純度の減少量が算出される. 不純度が減少したということは、純度が増加したと解釈できる. 純度が増加したということは分類・推定に寄与したと解釈できる. そのため、不純度の減少量を式(4)に示すように、総和をとることによって、推定における寄与の度合いが定量化される.

4.4 SVM L1 正則化の重み係数

SVM L1 正則化の重み係数とは、SVM の目的関数に L1 ノルム正則化項を導入した際の、特徴量ごとの重み係数である. 正則化項を導入することで推定に有用でない特徴量

の重み係数が 0 に近づき、推定に有用でない特徴量からの影響が低減することで、推定精度が向上する[17]. この際、重み係数が 0 に近く推定に有用でないと示された特徴量を取り除くことで、特徴量選択が行われる. SVM の目的関数に L1 ノルム正則化項が導入された式は(5)になる[15]. 式の中の変数の意味を説明すると次になる. $\|w\|_1$: L1 ノルム項. w : 重み係数. C : L1 ノルム項による影響の程度を制御する変数である.

$$\min_{w_0, w} \sum_{i=1}^n [1 - y_i (w_0 + \sum_{j=1}^q w_j x_{i,j})] + C \|w\|_1 \quad (5)$$

4.5 特徴量選択と考察

結果の統合にあたり、それぞれの手法ごとに算出された値を、最大値 1、最小値 0 になるように正規化した. その後、指標ごとに、正規化された値の平均値を算出した.

図 4 に、特徴量選択の結果、HAHV, HALV, LALV, LAHV の 4 感情推定において有用と予測された指標の上位 10 個を示す.

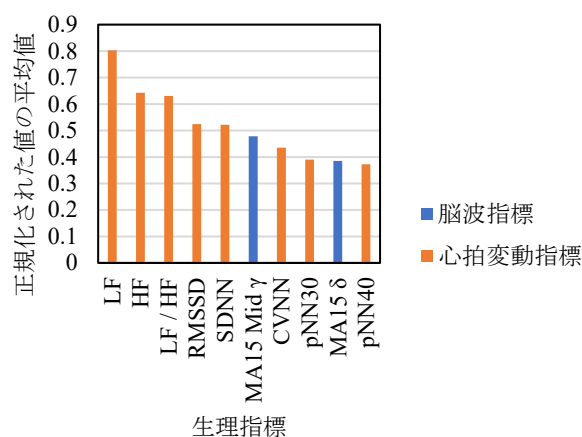


図 4 HAHV, HALV, LALV, LAHV の 4 感情推定における特徴量選択の結果

図 4 において、脳波指標は移動平均した γ 値の高周波数帯の脳波が、感情推定に寄与すると示唆された. これは複数電極を用いた Wang らの報告[18]や γ 周波数帯が感情の変化に最も敏感であるという報告[2]と一致した. しかし、前頭葉 (AF4) のみからの脳波を取得した本研究においては、低周波数帯の δ も比較的、感情推定に寄与することが示唆された.

心拍変動指標においては、周波数解析区間が長い LF, HF, LF/HF が感情推定に寄与することが示唆された. 周波数解析区間が長いがために、音楽による感情喚起以外の安静時の状態を反映している可能性が高く、感情推定に寄与しづらいと予測できるが、この予想と反する結果であった. 心拍変動指標は交感神経や副交感神経の指標であるが、刺激を受けてから、心拍変動指標がその刺激の影響を反映する

までに時間差が存在する．こうした点から，十分に刺激による影響が反映された，信頼できる心拍変動指標を算出するために，解析区間を 24 時間や 5 分とすることが多い[19]．本研究において，LF, HF, LF/HF は比較的長く解析区間が設定されたことにより指標の信頼性が高まり，結果として感情推定に寄与すると示唆されたと考えられる．次いで，感情推定への寄与の度合いが高い RMSSD は，短い解析区間でも 5 分間の解析区間と同等の信頼性があることが示唆されている指標である[19]．本研究により，感情推定においても短い解析区間にも関わらず信頼性が高い指標であることが示唆された．

5. 機械学習による感情推定モデルの構築とその精度検証

機械学習により感情推定モデルを構築するにあたり，本研究の特微量選択手法の有効性を示すために複数の特微量の組合せで機械学習を行い，感情推定モデルを特微量の組合せごとに構築した．本節では，これら複数の特微量の組合せやそれぞれの特微量の組合せから学習し構築された感情推定モデルの精度検証方法，および，その結果について述べる．

5.1 特微量の組合せ

特微量の組合せを表 3 に示す．表 3 に示す特微量の組合せ名は次のようになる．EEG は，本研究で用いた EEG から取得した指標である．MA15 EEG は，EEG の指標を区間 15 として移動平均した指標である．Time Domain HRV(TD HRV)は，IBI の周波数解析により算出した指標である．TD HRV + FD HRV は，TD HRV と FD HRV の指標を組み合わせた指標である．ALL は，EEG と MA15 EEG と TD HRV と FD HRV の指標を組み合わせた指標である．HAHV, HALV, LALV, LAHV は，図 4 における，HAHV, HALV, LALV, LAHV の 4 感情分類における感情推定への寄与の度合いが高い上位 10 個の指標である．

表 3 特微量の組合せ

特微量の組合せ名	特微量
EEG	θ , δ , Low α , High α , Low β , High β , Low γ , Mid γ , α , β , γ
MA15 EEG	MA15 θ , MA15 δ , MA15 Low α , MA15 High α , MA15 Low β , MA15 High β , MA15 Low γ , MA15 Mid γ , MA15 α , MA15 β , MA15 γ
TD HRV	IBI, HR, CVNN, SDNN, RMSSD, SDNN/RMSSD, pNN10, pNN20, pNN30,

	pNN40, pNN50
FD HRV	LF, HF, LF/HF
TD HRV + FD HRV	IBI, HR, CVNN, SDNN, RMSSD, SDNN/RMSSD, pNN10, pNN20, pNN30, pNN40, pNN50, LF, HF, LF/HF
ALL	θ , δ , Low α , High α , Low β , High β , Low γ , Mid γ , α , β , γ , MA15 θ , MA15 δ , MA15 Low α , MA15 High α , MA15 Low β , MA15 High β , MA15 Low γ , MA15 Mid γ , MA15 α , MA15 β , MA15 γ , IBI, HR, CVNN, SDNN, RMSSD, SDNN/RMSSD, pNN10, pNN20, pNN30, pNN40, pNN50, LF, HF, LF/HF
HAHV, HALV, LALV, LAHV	LF, HF, LF/HF, RMSSD, SDNN, MA15 Mid γ , CVNN, pNN30, MA15 δ , pNN40

5.2 交差検証

機械学習により構築したモデルの精度を評価するにあたり交差検証を用いた．交差検証には，Stratified K-fold cross validation(SKF) と Leave One Subject Out cross validation(LOSO)を用いた．SKF とは，訓練データ・テストデータを K 個に分割する際，目的変数のそれぞれの種類のデータ数の比率を同率にする交差検証である[20]．訓練データとテストデータにおいて，正解データのそれぞれの種類のデータ数の比率に偏りが生じた場合，訓練データではクラス 1 のデータ数が多くなる一方，テストデータではクラス 1 のデータ数が少なくなるといった，データ数が不公平な精度検証となる可能性がある．こうした問題を低減するために，SKF を用いた．なお，本研究では $k=10$ とした．この交差検証により，学習データに含まれている人に対して感情推定する際の精度を算出する．

LOSO とは，実験参加者 1 名のデータをテストデータとし，その他の実験参加者のデータを訓練データとして機械学習を行う操作を 1 試行とし，テストデータを入れ替えて，実験参加者の人数分だけ試行を実施する交差検証方法である．この交差検証により，学習データに含まれていない人に対して感情推定する際の精度を算出する．

5.3 精度評価指標

精度評価指標には，Macro F1 を用いた．Macro F1 とは，2 値分類において用いられる精度評価指標である F1-score をマルチラベル分類に拡張したものである．F1-score の算出方法は，以下の式(6)，(7)，(8)により表される．

$$precision = \frac{TP}{TP + FP} \quad (6)$$

$$recall = \frac{TP}{TP + FN} \quad (7)$$

$$F1\text{-score} = \frac{2 * recall * precision}{recall + precision}$$

(8)

正例と負例の2値分類において、式中の変数の意味を説明すると次のようになる。TP:予測値が正例で、その予測が正しいデータ数。TN:予測値が負例で、その予測値が正しいデータ数。FP:予測値が正例で、その予測が誤りであるデータ数。FN:予測値を負例として、その予測が誤りであるデータ数。*precision*は、正例の誤検知を少なくしたい場合に重視される精度指標であり、*recall*は、正例の見逃しを避けたい場合に重視される精度指標である。*F1-score*はこれらの精度指標の調和平均をとることにより、バランスをとった指標である。*Macro F1*とは、以上の*F1-score*を目的変数の種類ごとに算出し、それらの平均値を算出した精度指標である[20]。

深層学習モデルの構造は、浦部らに習い、中間層:256次元の3層、中間層の活性化関数:ReLU、出力層の活性化関数:Softmax、最適化アルゴリズム: Stochastic Gradient Descent(SGD), Dropout: 0.0とした。

5.4 精度検証結果と考察

以上の交差検証、精度指標、深層学習の構造を用いて精度検証をした結果を表4に示す。なお、精度比較のベースラインとして学習せずにランダムな予測を返す分類器の精度(baseline)を記載した。

表4 HAHV, HALV, LALV, LAHV 4感情分類の精度

特徴量の組合せ名(表3)	SKF	LOSO
EEG	38%	22%
MA15 EEG	94%	22%
TD HRV	93%	20%
FD HRV	57%	21%
TD HRV + FD HRV	99%	18%
ALL	98%	20%
HAHV, HALV, LALV, LAHV	99%	23%
baseline	25%	24%

表4において、SKFではEEGの38%やFD HRVの57%、HAHV, HALV, LALV, LAHVの99%などとbaselineの25%よりも精度が高かった。一方、LOSOではEEGの22%やTD HRV + FD HRVの18%、HAHV, HALV, LALV, LAHVの23%などとbaselineの24%よりも低い結果となった。SKFは感情推定をする対象人物のデータが含まれているが、LOSOでは感情推定をする対象人物のデータが含まれていない交差検証方法である。これにより、対象人物の生理データが含まれたデータから学習された感情推定モデルでは最大99%と高い精度の感情推定ができるが、対象人物の生理データが含まれていないデータから学習された感情推定

モデルでは精度よく感情推定ができないことが示唆された。この理由の1つを生理指標の個人差によるものだと考える。生理指標には、個人差が存在する。感情推定する対象人物の生理データを学習データに含めることができれば個人差を学習でき、baselineの精度を超えるような感情推定精度になると考える。一方、学習データに含まれず個人差が機械学習により学習されなければ、LOSOのようにbaselineを下回る低い感情推定精度になると考える。LOSOにおいても感情推定を可能とするため方法の1つとして、個人差による影響が少ないといわれているエントロピーやカオス性を表現する非線形特徴量の利用を考える[21]。個人差による影響が少ないがために、学習データに含めることで機械学習に個人差を学習させずとも感情推定が可能になると考える。実際に、脳波から抽出した非線形特徴量により感情推定ができることは実証されている[22]。今後は、これら非線形特徴量を加えたうえで、脳波・心拍変動指標により感情推定をすることで、精度向上を目的とした研究を進める。

SKFにおいて、特徴量選択の結果を反映した特徴量の組合せであるHAHV, HALV, LALV, LAHVの精度は99%であり、すべての特徴量を用いた精度は98%であり、両者の精度はほとんど同率であった。すべての特徴量を用いずとも、99%の精度がなされたため本研究で用いた特徴量選択の手法の有効性が示されたと考ええる。特徴量の数が多いほど学習に費やす時間が多くなるため、この特徴量選択により、精度を維持しつつ特徴量の数を減らすことで、学習に費やす時間を減少させることが可能であることを示した。

SKFのHAHV, HALV, LALV, LAHVにおいて、EEGとMA15 EEGの精度はそれぞれ38%と94%であった。移動平均により2倍以上もの精度向上がなされたことから、移動平均の有効性が示唆された。しかし、LOSOにおいては精度が同程度であり、学習データに含まれていない人に対する感情推定においては有効でないことが示唆された。

6. まとめ

安価で簡便な脳波計と光電式脈波計を用いたうえで、感情推定精度の向上を目的として脳波指標と心拍変動指標の特徴量抽出や特徴量選択を行った。特徴量選択の有効性を示すために、複数の特徴量の組合せで複数の感情推定モデルを深層学習により構築し、その精度を検証・比較をした。

SKFによる交差検証により、すべての特徴量を用いた場合と本研究で用いた特徴量選択を適応した場合の感情推定精度を比較した結果、精度は同程度であった。しかし、特徴量を減らすことで学習に費やす時間の低減がなされることから、本研究で行った特徴量選択手法の有効性が示唆された。

感情推定をする対象人物の生理データを学習データに含める交差検証方法であるSKFによる感情推定では、最大精

度 99%となり、安価で簡便な脳波計と光電脈波計により精度が高い感情推定が可能であることが示唆された。

一方、感情推定をする対象人物の生理データを学習データに含めない LOSO による精度検証では、ほとんどの特徴量の組合せで精度が 24%未満と低い感情推定精度にとどまることが示唆された。この要因の 1 つを生体情報の個人差による影響であると考え、この課題を解決するために、個人差による影響が少ないといわれている非線形特徴量の利用を考え、今後も研究を進める。

参考文献

- [1] A. Dzedzickis, A. Kaklauskas, and V. Bucinskas. Human Emotion Recognition: Review of Sensors and Methods. *Sensors*. 2020, vol. 20, no. 3.
- [2] J. Zhang, Z. Yin, P. Chen, and S. Nichele. Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. *Inf. Fusion*. 2020, vol. 59, p. 103–126.
- [3] S. M. Alarcão and M. J. Fonseca. Emotions Recognition Using EEG Signals: A Survey. *IEEE Transactions on Affective Computing*. 2019, vol. 10, no. 3, pp. 374–393.
- [4] J. A. Russell. A circumplex model of affect. *J. Pers. Soc. Psychol.* 1980, vol. 39, no. 6, p. 1161–1178.
- [5] S. Katsigiannis and N. Ramzan. DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals From Wireless Low-cost Off-the-Shelf Devices. *IEEE Journal of Biomedical and Health Informatics*. 2018, vol. 22, no. 1, p. 98–107.
- [6] Y. Ikeda, R. Horie, and M. Sugaya. Estimating Emotion with Biological Information for Robot Interaction. *Procedia Comput. Sci.* 2017. vol. 112, pp. 1589–1600.
- [7] G. U. Navalyal and R. D. Gavas. A dynamic attention assessment and enhancement tool using computer graphics. *Human-centric Computing and Information Sciences*. 2014. vol. 4, no. 1, p. 11.
- [8] U. Naoto and S. Midori. An Emotion Classification Method for Individuals Using EEG and Heart Rate Data and Deep Learning. *Proceedings of the Annual Conference of JSAI*. 2020, vol. JSAI2020.
- [9] S. Khalid, T. Khalil, and S. Nasreen. A survey of feature selection and feature extraction techniques in machine learning. in *2014 Science and Information Conference*. 2014, p. 372–378.
- [10] C. A. Lim, Wai Chong Chia, and Siew Wen Chin. A mobile driver safety system: Analysis of single-channel EEG on drowsiness detection. in *2014 International Conference on Computational Science and Technology (ICCST)*. 2014, p. 1–5.
- [11] T. Eerola and J. K. Vuoskoski. A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*. 2011, vol. 39, no. 1, p. 18–49.
- [12] M. M. Bradley and P. J. Lang. Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. *J. Behav. Ther. Exp. Psychiatry*. 1994. vol. 25, no. 1, p. 49–59.
- [13] A. U. Haq, D. Zhang, H. Peng, and S. U. Rahman. Combining Multiple Feature-Ranking Techniques and Clustering of Variables for Feature Selection. *IEEE Access*. 2019, vol. 7.
- [14] J. Wagner, Jonghwa Kim, and E. Andre. From Physiological Signals to Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification. in *2005 IEEE International Conference on Multimedia and Expo*. 2005, p. 940–943.
- [15] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, and B. Hu. Exploring EEG Features in Cross-Subject Emotion Recognition. *Front. Neurosci.* 2018, vol. 12, p. 162.
- [16] C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*. 2007, vol. 8, no. 1, p. 25.
- [17] A. Y. Ng. Feature selection, L_1 vs. L_2 regularization, and rotational invariance. in *Proceedings of the twenty-first international conference on Machine learning*, Banff, Alberta, Canada. 2004, p. 78.
- [18] X.-W. Wang, D. Nie, and B.-L. Lu. Emotional state classification from EEG data using machine learning approach. *Neurocomputing*. 2014, vol. 129, p. 94–106.
- [19] H. J. Back, C.-H. Cho, J. Cho, and J.-M. Woo. Reliability of Ultra-Short-Term Analysis as a Surrogate of Standard 5-Min Analysis of Heart Rate Variability. *Telemedicine and e-Health*. 2015, vol. 21, no. 5.
- [20] 門脇 大輔, 阪田 隆司, 保坂 桂佑, 平松 雄司, *Kaggle で勝つデータ分析の技術*. 技術評論社, 2019.
- [21] S. Byun *et al.*. Entropy analysis of heart rate variability and its application to recognize major depressive disorder: A pilot study. *Technol. Health Care*. 2019, vol. 27, no. S1, p. 407–424.
- [22] B. García-Martínez, A. Martínez-Rodrigo, R. Alcaraz, and A. Fernández-Caballero. A Review on Nonlinear Methods Using Electroencephalographic Recordings for Emotion Recognition. *IEEE Transactions on Affective Computing*. 2018.