

EEG を用いた感情・認知の精神課題における 機械学習モデルの構築方法の比較

金井健太郎¹ 鈴木圭¹ 菅谷みどり¹

近年、人の感情の推定や、認知機能といった精神状態を客観的に推測する手法が重要となってきた。中でも EEG の出力データを用いた機械学習においては、人の情動認知に関わる多様な精神状態の推定のためのモデル構築が可能である。モデル構築にあたっては、脳波の個人差を校正するための正規化手法が重要であるが、どのような正規化が、どの精神状態の推定モデルにとって重要であるか、十分な議論がない。そこで本研究では、「感情推定」、「認知機能推定」という異なる精神課題について、異なる正規化手法を適用し、比較検討を行うものとした。結果、「感情推定」では、個人ごとに正規化をした場合のモデルの精度が高くなり、一方、「認知機能推定」では、全体の正規化もしくは、正規化なしの場合に、僅かであるが、モデルの精度が高い結果となった。

EEG in Emotional and Cognitive Mental Tasks Comparison of machine learning model building methods

KENTAROU KANAI^{†1} KEI SUZUKI^{†1} MIDORI SUGAYA^{†1}

1. はじめに

近年、人の感情の推定や、認知機能といった精神状態を客観的に推測する手法が重要となってきた。これらの手法は、医療診断の補助、ヘルスケア、ロボット分野やマーケティング分野等にも適応されており、その重要性はますます高くなっている。

精神状態を客観的に推測する手法の中でも、Electroencephalogram(EEG)は、簡易に装着することができ、かつ人の感情の変化や認知機能の状態の把握など人の精神状態を、直接的にかつリアルタイムに把握することができるとされる[1]。こうした性質から、応用性も高いことが特徴である[2]。さらに、近年では、脳波を用いた機械学習モデルにより、「感情推定」[2]や「認知機能推定」[3]などを、精度高く行う手法への関心も高まっている[4]。例えば、Yilong らは、「感情推定」を高精度に行う手法として、EEG を用いた機械学習手法を提案した。機械学習モデルの構築にあたっては、感情分類の結果を目的変数として、EEG の出力データを説明変数とすることを示した[5]。榊澤らは、人の「認知機能推定」にて、人の認知機能のレベルを目的変数とし、その際に取得した EEG の出力データを説明変数としてモデル構築を行った[3]。このように、EEG を用いた機械学習においては、高精度に推定を行うためのモデル構築が可能であることから、今後より重要性が増していくと考えられる。

EEG の出力データを用いた機械学習においては、人の情動認知に関わる多様な精神状態の推定のためのモデル構築、すなわち、精神課題を実現できる。例えば、「感情推定」と「認知機能推定」では、正解ラベルが異なるだけであるため、同一の機械学習手法を適用することで、効率よく人の精神課題の機械学習モデルを構築することが可能である

と考えられる。しかし、精神課題ごとに目的変数が異なる場合に、モデル構築のために用いる EEG の値にどのような差があるのかについては十分な議論はなされていない。例えば、脳波から「感情推定」を行う場合の EEG のデータセットの構築方法と、「認知機能推定」をする場合の EEG のデータセットの作成方法は、同一で良いのか、区別すべきか比較検討はなされていない。

脳波は一般的に大きく個人差が発生する分野であることから、EEG のデータセット作成においては、個人差の校正手順として正規化の方法が重要となる。この正規化については、次の疑問がある。(1)正規化手法の差により影響があるのか。(2)正規化の対象が「個人」ごとのデータに対する正規化が良いのか、実験協力者「全員」分のデータを対象とした正規化が良いのか、明らかではない。(1)については min-max, Z-score などの手法があるが、十分に比較されていない。また、(2)については、「個人」の特徴を活かしたモデル構築を行った方が良いのか、データ「全体」の特徴を残してモデル構築を行った方が良いのか明らかになっていない。

そこで、本研究においては、精神課題ごとの最適なモデルの構築手法の検討を目的とし、EEG の出力データにおける説明変数のデータセットの作成方法の中でも正規化に着目し、目的変数の異なる2つの精神課題である「感情推定」、「認知機能推定」について比較するものとした。

課題(1)(2)に対応するために、「感情推定」、「認知機能推定」について、同一の脳波計で値を取得するために実験を実施し、データを収集した。その上で、異なる正規化手法を適用し、比較を行う。具体的には、(1)代表的な2つの min-max, Z-score の差を比較する。(2)「個人」と「全体」の異なる対象に対する正規化の差を比較した。

結果、(1)の min-max, Z-score の手法による差はほとんど

¹ 金井健太郎, 鈴木圭, 菅谷みどり, 芝浦工業大学

^{†1} KENTAROU KANAI, KEI SUZUKI, MIDORI SUGAYA, Shibaura Institute of Technology

見られない結果となった。これに対して、(2)では差が見られた。具体的には、「感情推定」の機械学習モデルにおいては、「個人」ごとに正規化をした場合のモデルの精度が高い結果となった一方、「認知機能推定」においては、「全体」の正規化もしくは、正規化なしの場合に、僅かであるが、モデルの精度が高い結果となった。これは、感情は、認知機能と比較し、より内面的な差が出やすく、より「個人」の差が発生するためにこのような結果となったと考えられる。これに対して、認知は人全体に共通した傾向があることが窺える結果となった。

本論文の構成は以下のとおりである。まず、2節で関連研究とその課題、3節で研究の目的と提案4節で機械学習モデルの処理方法、5節で感情、認知機能の評価実験について、6節で比較結果、7節でまとめとする。

2. 関連研究

2.1 精神課題を用いた研究

近年、「感情推定」や「認知機能推定」など人の精神状態を推定する研究が多く行われてきている[2][3]。人の精神状態と脳には多くの関連が示されており[2]、人の精神状態のモデル化が行われる際に、多くの研究では脳から取れる指標が用いられている。

この節では脳と、「感情推定」と「認知機能推定」の2つの人の精神状態に関わる研究とその研究の課題について述べる。

2.2 感情推定に関する研究

「感情推定」を行う研究として、池田らは脳波指標と、心拍変動指標を用いて「感情推定」を行う方式を提案した[6]。この方式では、図1に示してあるラッセルの円環モデル[7]を感情の判定をするために使用した。使用方法としては、ラッセルの円環モデルのX軸である快-不快に関しては心拍変動指標であるpNN50を、Y軸である覚醒-眠気に関しては脳波指標を当てはめて「感情推定」を行った。この研究の課題として、脳波指標と心拍指標を組み合わせるため脳波のみで感情に対する議論が行われていない課題が存在する。



図1 ラッセルの円環モデル

機械学習を用いた感情分類モデルにおいては、Arousal(覚醒度)と Valence(快適度)による2感情分類が主流である[5]。しかし、近年では、脳波による感情分類の応用

分野として、味覚反応や[8]、空間的印象評価[1]など、対象が広がっており、これらの分野においては、人の感情についてより詳細な評価が求められることから、Arousal/Valenceの2感情分類では十分でないと考えられるケースが見られる。例えば、Narumonらは、自動運転時の快適覚醒の研究で、快適と覚醒を組み合わせた感情の評価手法を提案した[1]。この中で、快適覚醒は、Arousal(覚醒度)と Valence(快適度)が共に高い感情を「快適覚醒」として、定義されている[1]。しかし、過去の多チャンネル脳波計を用いた研究では、先に述べたように、Arousalと Valenceのそれぞれの2感情分類が主流となっており、こうした議論に適した4感情分類の場合、成果は得られていない。

これに対して鈴木らは、4感情分類の機械学習モデルを構築し、評価を行った[9]。鈴木らは、単チャンネル脳波計と心拍計による実験を実施し、これらから得られた生理指標から特徴量の選択を行った。特徴量選択により、重要度の高い指標の組み合わせを得た後、深層学習をはじめとした主要な機械学習によりモデルを構築し、4感情分類を行った。分類精度の評価では、学習データ内に含まれる対象人物の感情推定精度を検証した層化10分割交差検証では約98%の感情推定精度を達成した。しかし課題として、鈴木らが用いた NeuroSky社の MindWave[10]は測定箇所が前頭葉のみの単チャンネルであり、前頭葉の働きに依存した感情モデルの議論しかできないといった課題が存在する。

脳と感情には関連が示されており、脳の機能局在性から感情は脳の前頭葉や扁桃体との関連が、音楽を聴取する実験の際に用いられる聴覚には側頭葉との関連が示されている[4]。しかし、先行研究ではこれらの脳の特徴を活かした議論やモデルの構築は十分なされていない。

2.3 脳の認知機能に関する研究

脳の認知機能に関する研究として、竹原らは脳波を用いて認知機能評価のためのシステムを開発した[11]。竹原らは認知機能を検査するための手法として、トレイルメイキングテスト(TMT)を使用した。TMTとは、決められた順番に数字や文字をなぞっていき、なぞり終わるまでの時間を計測するテストである。脳波指標としては、頭頂部付近の8チャンネルから脳波データの取得を行い、そこから算出できる指標である事象関連電位を使用した。これらを用いて実験を行った結果、脳波を用いて行う課題の成績とTMT-Bにおいて有意な負の相関があることが判明した。一方で、TMT-Aにおいては反応時間と相関がある結果となった。この研究の課題として、脳波指標が頭頂部付近のみを取得していることから、脳の機能局在性を議論できないという点が存在する。

この課題を解決するための提案として Zakzanisらは、脳を画像として捉えるfMRIとTMTの関連について調査を行った[12]。その結果、TMTを行った際の脳活動の活発であった領域として、背外側前頭前野や、運動制御に関わる領

域である前頭前野回や帯状回、内側前頭前野などの左側前頭部が活性化していることがわかった。

しかし課題として、fMRI ではこのような結果になったが、脳波でも同じような結果になることがわかっていない。その理由として、脳波は頭部に電極を装着してデータの取得を行うため、脳波の電流が頭蓋骨に反射して本来装着している部位以外のデータも一部混在してしまうといった点が存在するためである。

3. 感情・認知機能に関する課題とデータセット

本節では、EEG から出力される値を用いて精神課題における機械学習モデルを構築する場合の方法と、その課題について詳細を述べる。

3.1 EEG を用いた感情推定・認知機能推定の機械学習のモデルの構築

まず、一般的な EEG をデータとして説明変数に用いる場合の機械学習手法について説明する。

一般的には、[3][5]にあるように、精神課題ごとに、目的変数として、感情の主観評価や認知テストの結果を用い、EEG は、説明変数として用いる。EEG からの出力データを説明変数として用いる場合には、複数の手順を実施する。具体的には、まず、実験にて取得した EEG の生データに対してノイズ処理を行い、余分な成分の除去を行い、ノイズ処理をした値から周波数解析などを行い、指標として時間領域ごとに α 、 β などの各周波数帯のバンドを作成する。

本研究で課題としてしているように、「感情推定」や、「認知機能推定」などの精神状態の推定においては、目的変数が異なる。この場合、同一の脳波指標を用いてモデルを構築する場合でも、脳波指標の有用な脳波指標（バンド帯）が、精神課題の目的ごとに異なることが予想される。そのため、精神課題の目的ごとに、有用な脳波指標の選択方法や、次元圧縮や正規化などのデータセットに対して行う処理方法が異なることが予測される。

3.2 EEG を用いた感情推定・認知機能推定の機械学習のモデルの構築の課題

同一の脳波指標を用いる場合でも、機械学習モデルの構築手法が同一で良いとは限らない。例えば、感情推定と認知機能推定課題では、脳の機能局在性から感情では前頭葉や扁桃核、認知機能には前頭葉が使用されるとされている[2]。そのため、「感情推定」と「認知機能推定」では異なる脳の特性をを用いることが予想される。さらに、脳波の出現の仕方について、個人ごとに振幅の大きさや刺激への反応の大きさなどの出方が異なる。そのため、個人ごとの脳波の出方がある程度統一して個人差を減らすために、正規化などを行い EEG データの大きさをある程度統一する必要がある。また、正規化のみを見た場合でも、min-max や Z-score といったさまざまな手法があるため、それぞれの精神

課題のモデルを構築するのに適したデータセットの作成手法を検討する必要がある。

このことから、同一の脳波指標を用いた場合でも、正規化の方法などデータセットに対する処理方法が異なることが予測される。しかし、目的変数がモデルの精度にどの程度影響を与えているか、また、正規化などのデータの前処理手法の違いが機械学習モデルの精度へ及ぼす影響は明らかになっておらず、どこを変更すべきか統一的で汎用的な議論は行われていない。

3.3 目的と提案

本研究においては、精神課題ごとの最適なモデルの構築手法の検討を目的とし、EEG の出力データにおける説明変数のデータセットの作成方法の中でも正規化に着目し、2つの課題を明らかにするために比較を行うものとした。具体的には、目的変数の異なる2つの精神課題である「感情推定」、「認知機能推定」について、(1)同一脳波計で取得した値を最小0と最大1で正規化する min-max と、平均0、分散1にスケーリングする Z-score での正規化手法の二つの異なる手法を比較する。そのため、それぞれのデータセットを作成した。具体的には、個人ごとのデータに対して正規化を行った後、データを統合してデータセットとする「個人」を対象とした方法と、データ全体を統合した後に正規化を行った「全体」を対象とした方法の2種類の比較を行った。「個人」の特徴として、個人ごとの反応の特徴を残しつつデータのスケールをある程度統一することが可能であり、「全体」の特徴として、データ全体に現れた特徴のスケールを小さくして表すことが可能である。これらの手法にて1つの精神課題にて5つのデータセットを作成し、機械学習のモデル構築、比較するものとした。

4. 機械学習モデル

4.1 機械学習モデル構築の流れ

本節では比較に用いるための機械学習モデルの構築手順やその詳細、作成するデータセットの詳細について述べる。

最初に、精神課題を用いた機械学習モデル構築の流れを図2に示した。

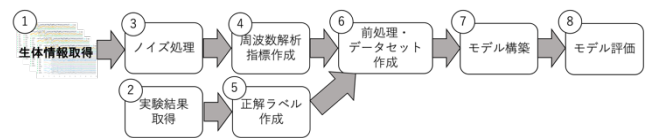


図2 機械学習モデル構築の流れ

まず、実験にて(1)生体情報の取得と(2)実験結果の取得を行う。次に生体情報については、(3)ノイズ処理を行い瞬きなどによって発生したノイズの除去・再配置を行った。ノイズ処理の終了後、(4)周波数解析を行い、機械学習に使用する指標の作成し説明変数の作成を行った。また、実験結果に関しては(5)クラス分けに従い正解ラベルの作成を行っ

た。指標の作成後、(6)これまで個人ごとに処理を行っていたデータを全員合わせたり、正規化などの前処理を行ったりしてデータセットの作成を行った。最後に(7)Deep Learningを用いてモデルの構築を行い、(8)層化10分割交差検証にて構築したモデルの精度評価を行った。

4.2 ノイズ処理・指標作成

ノイズ処理は、MATLABで使用できるソフトウェアであるEEGLABを用いて行った。行ったノイズ処理の種類としては、(1)ハイパスフィルタ、(2)ローパスフィルタ、(3)バンドストップフィルタ、(4)電極データの除去、(5)Artifact Subspace Reconstruction(ASR)による再構成、(6)再参照、(7)独立成分分析、(8)除去された脳波データの補完である。

まず、(1)ハイパスフィルタとは、汗によるノイズや独立成分分析への悪影響を低減させることを目的として、1Hz以下を低減するフィルタである。次に、(2)ローパスフィルタとは、100Hz以上を低減するフィルタであり、頭蓋骨により妨げられてノイズである可能性が高い高周波数帯の脳波の悪影響を低減させることを目的としている。(3)バンドストップフィルタについては、交流電源によりもたらされるハミングノイズと言われる50Hzと60Hz付近を低減するために使用した。(4)電極データの除去では、ノイズが多くノイズ除去が困難な電極から得られたEEGや、電極と頭皮が接触していないことで計測ができていない電極から得られたEEGの除去を行なった。(5)ASRによる再構成については、ASRアルゴリズムによりノイズが混入したと推論できる時間帯のEEGの再構成を行なった。また、再構成をすることで、除去によるデータ量の減少を防ぐことができる。続いて、(6)再参照ではまず平均値を算出し、平均値を全てのチャンネルのEEGに対して減算を行った。(7)独立成分分析は、瞬きや筋電位、心拍等の脳以外の生体活動によるノイズを低減させるために行なった。手法としては、ノイズを含む脳波データをいくつかの成分に分離させる。その成分の中でノイズである可能性が高い成分の除去を行い、ノイズの低減を行った。最後に、(8)チャンネルごとに脳波データの除去が行われた場合、その除去されたチャンネルの脳波データをEEGLABのプラグインにより補完を行った。

指標の作成には、周波数解析を用いており、周波数解析により算出したパワーバンドを指標として用いた。周波数解析はWelch法を使用した。作成した指標は5.2節の表1に示している。

4.3 データセット作成

本研究では、作成した脳波指標を説明変数、精神課題の結果を目的変数としてデータセットの作成を行った。さらに、複数の処理方法にて作成したデータセットの比較を行うために一つの精神課題につき、前処理を何も行ってないパターンと正規化2種類に対して個人ごとに行なった後にデータを合わせた「個人」のパターンと、データを全て合わせた後に正規化を行う「全体」パターンの計5種類のデ

ータセットを作成した。本研究では感情に関する精神課題、認知に関する精神課題2種類の計3種類の精神課題を行い、15個のデータセットを作成した。正規化を選んだ理由としては、生体情報は個人ごとにスケールの異なるため、そのスケールの縮小・統一をすることで個人差の影響を考慮できると考えたためである。正規化手法としては、min-max正規化とZ-score正規化の2種類を用いた。

min-max正規化とは、全てのデータのスケールを0から1にする手法であり、求めるための式は下記の通りである。

$$x'_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (i = 1 \cdot \cdot \cdot n)$$

今回min-max正規化を選んだ理由は、データのスケールを統一できるため、個人差を考慮したデータセットの構築ができると考えたためである。

続いてZ-score正規化とは、データ全体の平均を0、標準偏差を1にする手法であり、データ全体の平均を \bar{x} 、標準偏差を σ としたときの式は下記の通りである。

$$x'_i = \frac{x_i - \bar{x}}{\sigma} \quad (i = 1 \cdot \cdot \cdot n)$$

今回Z-score正規化を選んだ理由は、データの外れ値を考慮したデータセットを作成できるため、より正確なモデルが構築できると考えたためである。

4.4 モデル構築

作成したデータセットから、Deep Learningを用いてモデル構築を行った。Deep Learningのサブパラメータについて、まず中間層は[256,256,256]を使用した。続いて活性化関数について、中間層ではReLU関数を使用し、出力層ではSoftmax関数を使用した。Dropoutの値は0に設定し、最適化アルゴリズムにはSGDを用いた。

4.5 モデル評価

構築したモデルの評価として、層化10分割交差検証にて交差検証を行い、macro f1を用いて精度評価を行った。

層化10分割交差検証とは、データ全体の正解ラベルの比率が同じになるように10分割を行い、テストデータと訓練データを入れ替えながら行う交差検証です。データの比率を同じにする理由は、データのむらによる精度評価のばらつきを減少させるためである。この検証方法の特徴としては、学習データ内に含まれる人物への検証をすることができることである。

macro f1は2値以上のマルチクラス分類をする際に用いられる指標であり、それぞれのクラスごとにf1を算出した後、その値を平均することで算出される精度指標である。f1とは、混合行列(図3)を用いて算出する値である。その求め方として、まず、precision(適合率)とrecall(再現率)を算出し、下記の式を用いてf1を作成する。

$$f1 = (2 * precision * recall) / (recall + precision)$$

precisionの求め方は、TP/(TP+FP)であり、正例と予測したもののうち真の値も正例である割合を表す。recallの求め

方は、 $TP/(TP + FN)$ であり、真の値が正例のもののおどれほど正例であるのかの割合を表す。

		predict Positive	predict Negative
Positive class		True Positive (TP)	False Negative (FN)
Negative class		False Positive (FP)	True Negative (TN)

図3 混合行列

5. 感情評価及び認知機能評価の実験

5.1 データ取得のための機器

本研究では、機械学習モデルの構築のためのデータ取得を目的として二つの実験を行った。一つ目は、感情推定モデルを構築するための実験、二つ目は認知機能推定モデルを構築するための実験である。本研究では、これらの実験で EEG を取得するために多チャンネル脳波計として Emotiv 社の EPOC X を使用した[13]。この脳波計は、前頭部 6 チャンネル、側頭部 2 チャンネル、頭頂部 4 チャンネル、後頭部 2 チャンネルの計 14 チャンネル計測することができるため、脳の広範囲をカバーすることが可能である。EPOC X で計測することができるチャンネルを図 4 に示す。

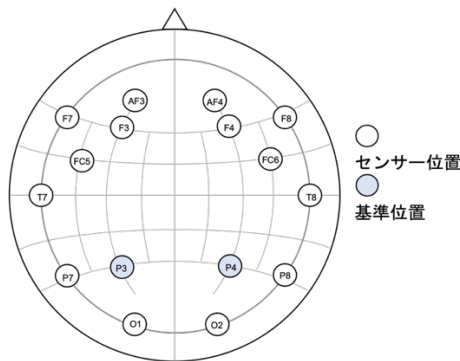


図4 EPOC X にて計測可能なチャンネル

多チャンネル脳波計を使用する利点として、脳の広範囲をカバーできることで脳の機能局在を考慮しながら議論をすることができるという点が存在する。

5.2 脳波指標

使用する脳波指標としては、各実験にて取得した脳波の生データにまずノイズ除去を行う。本研究で行ったノイズ除去としては、ダウンサンプリング、1Hz のハイパスフィルター、100Hz のローパスフィルター、50Hz、60Hz のラインノイズフィルター、電極データの再参照などである。ノイズ処理を行った脳波データに対して welch 法を用いて周波数解析を行なった。周波数解析を行なった後の指標とその帯域を表 1 に示す。

表1 周波数解析後の指標とその帯域

指標	帯域(Hz)
----	--------

δ	1-3
θ	4-7
α	8-12
β	13-30
γ	30-100
Low- α	8-9
High- α	10-12
Low- β	13-17
High- β	18-30
Low- γ	31-40
Mid- γ	41-50
High- γ	51-100
Total Power(1-30)	1-30
Total Power(1-40)	1-40
Total Power(1-100)	1-100

本研究では、これらの 15 の指標を 14 チャンネル分、計 210 個の指標を用いて、後の分析や機械学習モデルの構築を行なっていく。

5.3 データセットの作成

また、データの正規化を行う対象については、1 つ目として、全員分をまとめたデータに対して正規化を行った「全体」のデータセットと、2 つ目として、個人ごとのデータに対して正規化を行った後に全てのデータをまとめた「個人」のデータセットの 2 パターンの対象を用意した。本研究では、感情と 2 つの認知機能の精神課題にて、精神課題ごとに min-max(個人)、min-max(全体)、Z-score(個人)、Z-score (全体)、正規化なし、の計 5 つのデータセットを作成した。作成したデータセットの比較については、正規化の有無、正規化手法、正規化の対象に焦点を当てて比較を行なっていく。

5.4 感情推定モデル構築のためのデータ取得

本実験は、感情推定モデル構築のためのデータ取得を目的として、音楽による感情喚起を 20 代の男女 4 名を対象として実験を行った。実験において、感情推定モデル構築の際に説明変数として用いるために EEG を、目的変数として用いるために音楽を聴取した際の感情の主観評価を取得した。実験手順は、鈴木らの手法に従った[9]。この目的を実現するために行うことは以下の通りである。1 つ目として脳波指標の取得、2 つ目として音楽による感情の喚起、3 つ目として喚起された感情の主観評価である。

5.4.1 EEG の取得方法

EEG を取得する方法として、Emotiv 社が提供している Lab Streaming Layer を利用して EEG の取得を行った。これは EPOC X と連携しているアプリケーションである EmotivPRO がまず、PC 側から発信されたマーカーと脳波計からデータの受信を行う。その後、受信したマーカーから PC 側と連携を行い、Lab Streaming Layer が必要となる EEG を PC 側に送信するという形でデータの取得を行った。

5.4.2 音楽による感情喚起

本実験では、ユヴァキュラス大学が作成した音楽データ

ベース[14]上の音楽を用いて感情の喚起を行った。このデータベースは Anger, Fear, Sad, Happy, Tender の 5 つの個別感情の中程度と高程度の刺激が存在する。この音楽刺激をラッセルの円環モデルの各象限に均等に当てはまるように音楽を任意に選択した。

5.4.3 喚起された感情の主観評価

感情の主観評価手法として Self-Assessment Manikin(SAM)を用いて、自分の感情を言語に頼らずに客観的な評価をできるようにした[15]。(図 5)SAM は自分の感情に最も近いマネキンを選ぶことで評価を行う手法であり、本実験ではこれらの組み合わせで 4 感情の分類を行った。各マネキンには Valence(快適度), Arousal(覚醒度)それぞれに 5 つのマネキンとそれぞれの間接の計 9 つの指標が存在し、1~9 の値が割り振られている。Arousal, Valence の各マネキンを選択することによって定量的な主観評価を行う。主観評価の結果をラッセルの円環モデルにあてはめて音楽を聴取した際の感情がどの象限に属しているかを判断した。

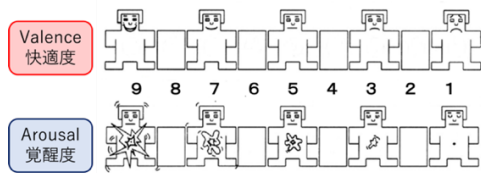


図 5 Self-Assessment Manikin(SAM)

SAM をラッセルの円環モデルに当てはめる具体的な条件として、 $Arousal > 5$ かつ $Valence \geq 5$ または $Arousal = Valence = 5$ のときには第 1 象限、 $Arousal \leq 5$ かつ $Valence > 5$ のとき第 2 象限、 $Arousal < 5$ かつ $Valence \leq 5$ のとき第 3 象限、 $Arousal \geq 5$ かつ $Valence < 5$ のとき第 4 象限に属していると判断した。

5.4.4 実験手順

実験手順としてはまず、椅子に腰かけた状態で脳波計とイヤホンの装着を行う。次に実験前安静として 2 分間安静状態をとってもらふ。その後、1 分 30 秒間の安静時間、2 分の音楽聴取、音楽聴取時の感情の主観評価を 1 試行として計 8 試行を行う。音楽聴取時に聴取してもらふ音楽はラッセルの円環モデルの各象限に属する音楽を 2 種類ずつ聴取してもらふ。聴取の順番はランダムに決めた。安静時、音楽聴取時には外からの刺激を軽減させるためにディスプレイ上に灰色の背景の中心に配置した黒い十字がある画面を見てもらふようにした。実験の様子を図 4 に示す。

5.5 認知機能推定モデル構築のためのデータ取得

本実験は、TMT による認知機能の計測、実験時の脳波指標、認知負荷を与えた前後での脳波の比較をすることを目的として 20 代の男女 5 名を対象として実験を行った。実験において、認知機能推定モデル構築の際に説明変数として用いるために EEG を、目的変数として用いるために TMT による認知機能のレベルの取得を行った。

5.5.1 EEG の取得方法

使用した機材としては感情推定モデル構築のための実験と同様のものを使用しており、脳波計として Emotiv 社の EPOC X を使用し、Lab Streaming Layer を用いて EEG の取得を行った。

5.5.2 TMT による認知レベルの計測

本実験では、トレイルメイキングテスト(TMT : Trail Maiking Test)を用いて認知レベルの 3 段階の評価を行った。今回は、千葉テストセンター 心理検査専門所が出版している「TMT-J Trail Maiking Test 日本版」を用いた[16]。TMT は、決められた順番に数字や文字をなぞっていき終了までの時間を計測するテストであり、幅広い注意、ワーキングメモリ、空間的探索、処理速度、保続、衝動性などを総合的に測定することができる。TMT の活用方法は、自動車運転の適性に関する神経心理評価法の一つとして用いられる場合や、外傷性脳損傷による高次脳機能障害、軽度認知障害と比較的軽度の認知症、前頭前野損傷に代表される比較的純粋な遂行機能障害の者等の評価法として用いられるなど、幅広い範囲で活用されている。TMT には、Part A と Part B の 2 種類が存在しており、A4 縦紙に Part A ではまるで囲まれた「1」から「25」までの数字が、Part B ではまるで囲まれた「1」から「13」までの数字と「あ」から「し」までの平仮名が散りばめられた検査用紙である。課題としては、Part A では、「1—2—3—…—25」というふうに 1 から順に 25 まで鉛筆を紙面から離すことなくできるだけ早く結び、Part B では、1 からはじめて数字と五十音を交互に「1—あ—2—い—3—う—…—13」の順番で 13 まで鉛筆を離すことなくできるだけ早く結び、終了までの所要時間と鉛筆反応の回数を記録する。本テストでの誤反応とは、鉛筆を紙から離れた回数や、誤った順番で文字をなぞることである。誤反応が起きた際は監視員が直接注意を行い、誤反応が起きた点からテストを再開する。TMT の例を図 6 に示す。

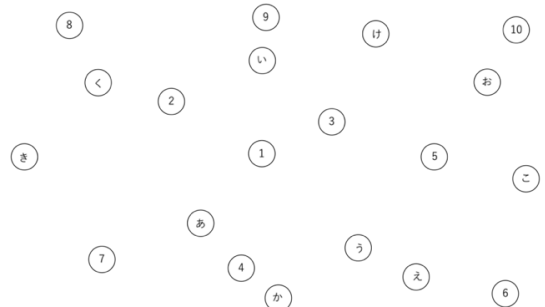


図 6 TMT の例

TMT の評価方法は、誤反応の回数と所要時間から決定を行う。所要時間は 10 歳刻みで年代が上がるほどに延長される。それぞれの年代の所要時間を表 2 に示す。

表 2 Trail Maiking Test の年代別所要時間判定

年代	Part A		Part B	
	+1SD 以内	+2SD 以内	+1SD 以内	+2SD 以内

20代	37	45	54	65
30代	37	45	55	65
40代	38	45	61	73
50代	40	46	74	88
60代	42	49	75	88
70代	57	68	113	138
80代	60	69	128	152

今回は20代のみを対象として実験を行ったため、評価指標としては20代の評価指標を用いた。また、所要時間と誤反応の回数を合わせた評価に関しては、TMT PartA については表3に、TMT PartB については表4に示す。

表3 TMT Part A の評価

		所要時間		
		+1SD 以内	+2SD 以内	延長
誤反応	2回以下	クラス1	クラス2	クラス3
	3回以上	クラス2	クラス3	クラス3

表4 TMT Part B の評価

		所要時間		
		+1SD 以内	+2SD 以内	延長
誤反応	4回以下	クラス1	クラス2	クラス3
	5回以上	クラス2	クラス3	クラス3

5.5.3 ストループ課題による認知負荷

本実験では認知負荷を与えることを目的としてストループ課題を行った。ストループ課題とは、アメリカの心理学者 Stroop などによって1935年に報告された課題である。本実験では、「あか」、「あお」、「きいろ」、「みどり」のどれかの文字が、「赤色」、「青色」、「黄色」、「緑色」のどれかの文字色で書かれているためその文字の色を答える課題である。今回は、Web 上でストループ課題を行ったため図6のように上部に大きく4つの色で書かれた文字が書いてあり、下部に「あか」、「あお」、「きいろ」、「みどり」と書かれたボタンがあり文字の色を選択する。例えば図7のように「赤色」で「あお」と書かれていた場合、「あか」と答える。

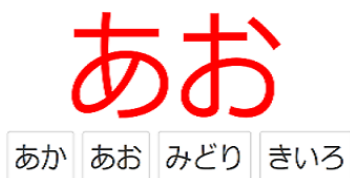


図7 ストループ課題の例

5.5.4 実験手順

実験手順としてはまず、TMT の練習を実施する。TMT の練習では本番で使用する TMT とは別に用意された、数字や文字の数が本番よりも少ないものを使用した。練習の終了後、脳波計の装着を行い、安静時間として5分間閉眼安

静を行ってもらう。安静時間の終了後、TMT-PartA, PartB を行う。この際、誤反応の指摘や用紙の切り替えなどを目的として TMT の実施中のみ監督者が実験協力者の後ろで監視を行った。また、TMT の検査用紙をテープにて固定を行い、TMT の検査用紙が滑らないようにした。TMT 終了後、認知負荷をかけることを目的としたストループ課題を10分間行った。ストループ課題の終了後、前半に行なった TMT とは別のパターン用の用紙に変更して、再度 TMT-PartA, PartB を行った。TMT の終了後、最後に安静時間として5分間の閉眼安静を行った。

5.6 認知負荷前後の EEG の比較

本項では認知機能推定モデル構築の際、データ数を増やすことを目的として、認知負荷を与える前後でデータの取得を行った。しかし、認知負荷の前後で認知機能や EEG に違いが見られ、認知負荷の前後両方とも有用なデータとして使用できるか不明であるため、認知負荷の前後である1回目と2回目の安静時間の EEG を用いてデータの一元配置分散分析を行った。今回、認知機能推定実験を行ったうちの1人の実験協力者のデータを使用して分析を行った。今回の分析では、各チャンネルで作成した脳波指標の中で、 δ (1-3Hz), θ (4-7Hz), α (8-12Hz), β (13-30Hz), γ (30-100Hz) を使用した。今回の分析では有意確率が0.05以下を有意差が現れたとした。分析の結果、AF3- β , F7- β , F4- β , O2- β , FC5- γ 以外の70個の脳波指標のうち65個の脳波指標で有意差が出る結果となった。

このことから、認知負荷の前後の結果について脳波指標全体で90%以上の脳波指標に有意差が認められたため、認知負荷の前後の安静時の EEG を同じ機械学習モデル内のデータとして使用した認知機能推定モデルの構築が有用であることがわかった。

6. 機械学習モデル精度比較

6.1 比較方法

データの前処理方法がモデル構築にもたらす影響の評価として、構築した機械学習モデルの精度にて比較を行った。比較の目的は、データの前処理手法の違いによる精度差を確認し、どの前処理手法が「感情推定」と「認知機能推定」、それぞれの精神課題に適した手法が明らかにするためである。精度評価手法としては、それぞれのデータセットごとにモデルを構築し、層化10分割交差検証にて精度評価を行った。モデル比較の評価として、それぞれの精神課題ごとに(1)正規化の有無、(2)正規化手法、(3)正規化の対象についての比較を行った。

6.2 比較結果

構築したモデルの精度評価結果を表5に示す。個人に対して正規化をした後で合わせたデータセットを(個人)、データ全体に対して正規化を行ったデータセットを(全体)とした。

表5 構築した精神課題モデルの精度評価結果

	感情(%)	TMT PartA(%)	TMT PartB(%)
正規化なし	92.53	100.00	100.00
min-max(個人)	99.93	99.92	100.00
min-max(全体)	92.92	100.00	100.00
Z-score(個人)	99.70	99.92	100.00
Z-score(全体)	93.34	100.00	100.00

「感情推定」では、正規化なしは92.53%、個人ではmin-maxが99.93%、Z-scoreが99.70%であり、全体ではmin-maxが92.92%、Z-scoreが93.34%となった。(1)正規化の有無について、正規化なしと全体の精度差を見た場合0.4-0.8%しか変わらない結果となった。一方で、正規化なしと個人ではmin-maxが99.93%、Z-scoreが99.70%となり、正規化なしや全体よりも約6~7%高くなる結果となった。(2)正規化手法については、個人のmin-maxとZ-scoreでは0.23%min-maxの方が高い結果に、全体のmin-maxとZ-scoreでは0.42%Z-scoreの方が高い結果となった。(3)正規化の対象では、min-maxについては個人の方が7.4%、Z-scoreについては個人の方が6.36%高い結果となった。このことから、「感情推定」においては正規化手法に関わらず「個人」で正規化を行った方が良いことがわかった。

次に「認知機能推定」では、TMT PartAだと、正規化なし100%、個人のmin-maxが99.92%、Z-scoreが99.92%であり、全体のmin-maxが100%、Z-scoreが100%となった。PartBは全てのモデルで精度が100%となった。(1)正規化の有無について、TMT PartAでは正規化なしと全体では正規化なしより0.08%低い結果に、個人では正規化なしと同じ100%となった。(2)正規化手法と(3)正規化の対象については、手法に関わらず個人では99.92%、全体では100%となる結果となった。このことから、認知機能に関しては正規化なしと全体の方がわずかに高い結果となったが正規化手法や正規化の対象によって精度がほとんど変わらないとわかった。

6.3 最適なモデルの検討

今回は、次の2点を課題とした。(1)正規化手法の差により影響があるのか。(2)正規化の対象が「個人」ごとのデータに対する正規化が良いのか、実験協力者「全員」分のデータを対象とした正規化が良いのか、明らかではない。そして、これらに対して、(1)代表的な2つのmin-max、Z-scoreの差の比較。(2)「個人」と「全体」の異なる対象に対する正規化の差の比較を行った。

まず、「感情推定」において(1)については、正規化手法による精度差はほとんど見られない結果となった。これに対して(2)については、「個人」ごとに正規化をした方がモデルの精度が高くなった。理由について述べる。「個人」ごとに正規化を行った方が、感情刺激への反応の個人差を残しつつ正規化を行っているため、個人差を考慮したモデ

ルが構築できたため高い精度のモデルが構築できたと考えられる。

次に、「認知機能推定」において、(1)の正規化手法、(2)の「個人」「全体」の異なる対象に対する正規化手法によるモデルの精度差がほとんど見られなかった。その理由として、感情と違い認知機能の決定の評価には、多様な個人の意思が関与していないため、個人ごとの特徴を残さずにモデル構築をしても、データセットによらず、同等の精度が出たと考えられる。

以上のことから、「感情推定」では、正規化手法に関係なく個人ごとに正規化を行った方が精度は高くなることがわかった。「認知機能推定」では、正規化なし、もしくは全体に正規化を行った方が良いが、正規化手法ごとによる違いはほとんどないことがわかった。

7. まとめと今後の課題

7.1 まとめ

本研究では、精神課題ごとの最適なモデルの構築手法の検討を目的として、感情と2つの認知機能、計3種類の精神課題を行い、それぞれ5つのデータセットを作成し、それぞれのモデルの構築、精度比較を行った。

データの预处理手法の違いがモデル構築に表す影響の比較を行った。感情推定モデルでは、データセット全体に対して正規化を行うよりも、個人ごとのデータに対して正規化を行った方が精度は6~7%高い結果となった。「認知機能推定」では、データセットの違いによる精度差はほとんど見られなかった。このことから、「感情推定」では個人ごとに正規化を行った方が良い結果に、「認知機能推定」では正規化手法の違いによる精度差は見られない結果となった。

7.2 今後の課題

1つ目として、感情推定モデルでは心拍変動指標と脳波指標を組み合わせた機械学習モデルにて高い精度が出ていたため、認知課題でも心拍変動指標が適応できるかの調査を行い、認知機能推定モデルに心拍変動指標の適応を行う予定である。

2つ目として、今回行った精度評価はテストデータと訓練データに同一人物のデータが入っている。そのため、実験協力者数を増加させ、テストデータと訓練データ内に同一人物のデータを含まない厳しい検証を行う予定である。

3つ目として、本研究のデータセット作成方法の比較として正規化のみを行っていたため、次元圧縮や移動平均など他の手法を用いてデータセットの作成を行い、脳の特徴を考慮したデータセット作成方法を検討する予定である。

参考文献

- [1] Narumon Jadram, Tipporn Laohakangvalvit and Midori Sugaya, "A Preliminary Evaluation of Comfortable Arousal using Biological Information Measurement for Autonomous Driving", Asia Pacific Conference on Robot IoT System Development and Platform (APRIS2020).

- [2] Soraia M, et al, Emotions Recognition Using EEG Signals A Survey, IEEE Transactions on Affective Computing, pp 1- 12 June 2017.
- [3] 柳澤 琢史, 畑 真弘, 福間 良平, 石井 良平, 吉山 颯次, 原田 達也, 池田 学, 貴島 晴彦. 安静時脳波・脳磁図に深層学習を用いた認知症診断. 日本脳神経外科認知症学会誌, 2021年1巻1号 p1-6.
- [4] 厚生労働省. AI懇談会報告書. <https://www.mhlw.go.jp/file/05-Shingikai-10601000-Daijinkanboukouseikagakuka-Kouseikagakuka/0000169230.pdf>, 2021年7月24日参照.
- [5] Yilong Yang, Qingfeng Wu, Ming Qiu, Yingdong Wang, Xiaowei Chen. Emotion Recognition from Multi-Channel EEG through Parallel Convolutional Recurrent Neural Network. International Joint Conference Networks. 2018.
- [6] Yuhei Ikeda, Midori Sugaya. "Estimate Emotion Method to Use Biological. Symbolic Information Preliminary Experiment," HCI. 2016, vol. 13, 332-340, 1994.
- [7] Russell, J. A. . A circumplex model of affect. Journal of Personality and Social Psychology, 39(6), pp1161-178, 1980.
- [8] Masaki Hayashi, Peeraya Sripian, Uma Maheswari Rajagopalan, Runqing Zhang, Midori Sugaya, "Investigation of Biological Signals under the Stimulation of Basic Tastes", , HCII 2020, Proceedings, Springer, pp.40-49, 2020.
- [9] Suzuki Kei, Laohakangvalvit Tipporn, Matsubara Ryota, Sugaya Midori, "Constructing an Emotion Estimation Model Based on EEG/HRV Indexes Using Feature Extraction and Feature Selection Algorithms", Sensors, 21(9), 2910, 2021.
- [10] NeuroSky. mindwave MOBILE2. <https://store.neurosky.com/pages/mindwave>, 2021年7月13日参照.
- [11] 竹原 繭子, 山本 泰豊, 長谷川 良平. 脳波スイッチによる認知機能評価システムの開発—トレイルメイキングテストとの結果比較について—. 日本感性工学会論文誌, 2021.
- [12] Konstantine K. Zakzanis, Richard Mraz, Simon J. Graham. An fMRI study of the Trail Making Test. Neuropsychologia, Volume 43, Issue 13, 2005, P1878-1886.
- [13] EMOTIV EPOC X. <https://www.emotiv.com/epoc-x/>, 2021年5月20日参照.
- [14] Tuomas Eurola, Jonna K. Vuoskoski. A comparison of the discrete and dimensional models of emotion in music. 2011.
- [15] Bradley, M. M., & Lang, P. J. Measuring emotion: The self-assessment Manikin and the semantic differential. Journal of Behavior Therapy and Experimental Psychiatry, 25, p49–59. 1994.
- [16] 千葉テストセンター 心理検査専門所, TMT-J Trail Making Test 日本版. <https://www.chibatc.co.jp/cgi/web/index.cgi?c=catalogue-zoom&pk=270>. 2021年2月1日参照.