

# 表情や仕草データを用いた美術鑑賞者の心理状態推定手法の提案と評価

鳥垣 耀平<sup>†</sup> 松田 裕貴<sup>†, ††, †††</sup>

† 岡山大学 〒700-0082 岡山県岡山市北区津島中3丁目1-1

†† 理化学研究所革新知能統合研究センター 〒103-0027 東京都中央区日本橋1-4-1

††† 奈良先端科学技術大学院大学 〒630-0192 奈良県生駒市高山町8916-5

E-mail: †yohei.torigaki@cocolab.jp, ††yukimat@okayama-u.ac.jp

**あらまし** 近年の美術館の役割として、人々の知的 requirement に答えるためのコミュニケーションの活性化が求められており、博物館・美術館における学びの専門家であるミュージアムエデュケータの登用が行わっているが、人員は不足している状況が続いている。音声ガイドなどによるシステムが代替する役割を果たすことが期待されているが、人間によるガイドとはまだ大きな隔たりがある。そこで本研究では、鑑賞者の興味・関心に合わせた動的なガイドの提供を目指し、鑑賞者の心理状態を推定する手法を提案する。心理状態推定の手法確立に向けて、新たに開発したセンシング機能を有するガイドシステムを用いて、大原美術館での30名の被験者を対象としたデータ収集実験を実施した。得られた表情や仕草のデータを用いた心理状態推定モデルの構築・評価を行った結果、ポジティブ・ネガティブの2クラス感情分類モデルについて精度が64%となった。一方で、満足度・理解度・困惑度の回帰モデルについては有効な結果は得られなかったが、一部である程度の相関が確認されるなど、今後の改善にむけた知見を得ることができた。

**キーワード** 美術館、心理状態推定、機械学習、ガイドシステム、データセット、マルチモーダルセンシング

## Proposal and Evaluation of a Method of Estimating Emotional State during Art Appreciation based on Facial and Behavioral Data

Yohei TORIGAKI<sup>†</sup> and Yuki MATSUDA<sup>†, ††, †††</sup>

† Okayama University Tsushima-naka 3-1-1, Kita-ku, Okayama, 700-0082 Okayama, Japan

†† RIKEN Center for Advanced Intelligence Project Nihonbashi 1-4-1, Chuo-ku, 103-0027 Tokyo, Japan

††† Nara Institute of Science and Technology Takayama-cho 8916-5, Ikoma, 630-0192 Nara, Japan

E-mail: †yohei.torigaki@cocolab.jp, ††yukimat@okayama-u.ac.jp

**Abstract** In recent years, museums have been increasingly expected to facilitate communication in order to meet the intellectual needs of visitors. Although museum educators—experts in learning within museums and art galleries—have been employed to fulfill this role, there remains a persistent shortage of personnel. While systems such as audio guides are anticipated to serve as substitutes, a significant gap still exists between these systems and human guidance. Therefore, this study proposes a method to estimate the emotional state of museum visitors, aiming to provide dynamic guidance tailored to individual interests and concerns. To establish a method for emotional state estimation, we conducted a data collection experiment using a newly developed guidance system equipped with sensing capabilities at the Ohara Museum of Art, involving 30 participants. Based on the collected data of facial expressions and gestures, we constructed and evaluated a emotional state estimation model. As a result, the binary emotion classification model for positive and negative emotions achieved an accuracy of 64%. On the other hand, regression models for satisfaction, comprehension, and confusion did not yield effective results; however, some degree of correlation was observed, providing insights for future improvements.

**Key words** Museum, Emotion estimation, Machine learnig, Guidance system, Dataset, Multi-modal sensing

## 1. はじめに

美術館や博物館では鑑賞者が展示物の鑑賞を通じて感じ、学び、理解することが重要である。近年では学芸員と協力しながら博物館の教育方針やプログラムを企画・実施するミュージアムエデュケータが登用されている。しかしながら、2008年時点では約3,000人であった学芸員数は2018年時点では約3,600人と20%程度しか増加しておらず、また1館あたりの学芸員数は4人に満たない状況（相当博物館・類似博物館を含めると1.5人程度）であり、依然として人員が不足している状況が続いている[1]。ミュージアムエデュケータは鑑賞者との対話を通じて鑑賞者の心理状態に合った解説を行うことで、より鑑賞者の展示物に対する興味・関心を引き出すが、この技術は形式化されていない。

近年、貸出型端末等による音声ガイダンスシステムの活用も進んでいるが、多くは一方的に作品の解説をすることに留まっており、館内ガイドを電子化したものの域を出ていない。このような現状から、音声ガイダンスシステムなどがミュージアムエデュケータに代替するには多くの課題がある。本研究では、展示物を鑑賞する人の心理状態の推定に基づいて、鑑賞者に応じて対話的・動的にコンテンツ提示（ガイダンス内容・順序の編纂）することで、この問題を解決することができるのではないかと考えた。その実現に向け、多様な仕草や表情などをセンシングする機能を有する音声ガイダンスシステムを新たに構築し、大原美術館<sup>(注1)</sup>における30名の被験者を対象とした作品鑑賞時における鑑賞者の心理状態のデータ収集実験を行った。構築したデータセットは映像データ（鑑賞者の表情・鑑賞中の作品）、加速度・角加速度データ（スマートフォン・頭部）および心理状態ラベルデータ（作品ごとに回答するアンケート回答）で構成される。このデータセットを用いて、心理状態を推定する機械学習モデルを構築・評価を行った。評価の結果、ポジティブ・ネガティブの2クラス感情分類モデルについて精度が64%となった。一方で、満足度・理解度・困惑度の回帰モデルについては有効な結果は得られなかったが、一部である程度の相関が確認されるなど、今後の改善にむけた知見を得ることができた。

## 2. 関連研究

### 2.1 心理状態推定に関する関連研究

心理状態推定、特に感情認識に関しては、これまでに様々な手法が提案されている。

古澤らは、学習者の状態に応じた教育学習支援の実現のため、生体情報から学習者の心的状態を深層学習モデルを用いて推定する手法を検討している[2]。森らは、対話の中で発せられる相槌や復唱、笑いなどの社会的シグナルから聞き手の理解度を推定し、それに応じて発話タイミングを変える聞き手アウェアな音声ガイダンスシステムを提案している[3]。また、対話システムを介して収集した音声データを基に感情推定を行う研究

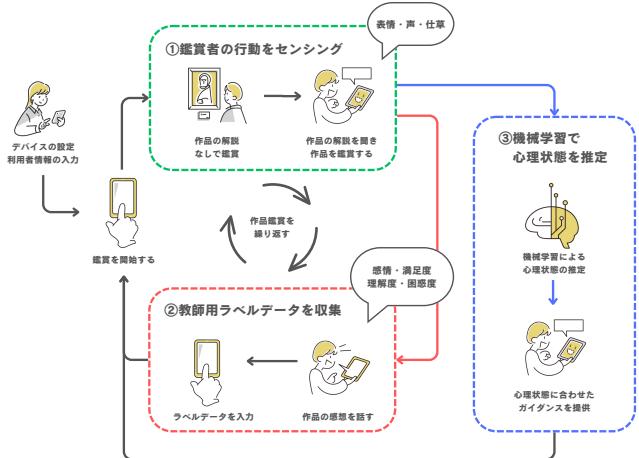


図1：データ収集と心理状態推定の流れ

も取り組まれている[4], [5]。近年では、音声対話による機器の操作などが現実的なものとなってきており、観光を含むより多くの場面において音声の利用が期待できる。さらなる認識性能の向上のために、複数のセンサデータを組み合わせることによるマルチモーダルな感情認識手法についても研究が進んでいる。Tzirakis らや Ghaleb らは対話を介して収集した音声に加え映像データを組み合わせる手法を提案している[6], [7]。ほか、視線や身体の動きのセンシングに基づく感情推定を行う研究[8], [9]も提案されている。

また、感情以外の心理状態として、Rach らは議論的対話システムを対象として、興味度、納得度、理解度、関連度といった異なる尺度の心理状態を取り扱う研究を提案している[10], [11]。

### 2.2 美術館におけるシステムの研究

美術館来館者の理解を助けるためのツールとして、近年では貸出型端末やスマートフォンアプリでのナビゲーション（音声ガイダンス）が国内外問わず広く用いられている。

Sun らは、個人の学習スタイルに合わせて美術館ガイドの情報提示方法を変更することで、認知コストの軽減や学習効果が高まることを報告している[12]など、その重要性については注目されているといえる。また、Liu らは、美術館の評価基準を確立し、その評価基準を用いて美術館の面白さと来場者の満足度の関係性を導出する研究を行っている[13]。

また、近藤らは奈良県立美術館での鑑賞者の心理状態推定に向けた鑑賞中の仕草や音声のデータセット構築を行っており[14]、著者らは構築したデータセットのうち、音声を利用した心理状態推定に向けた基礎分析を行っている[15]。

## 3. 提案手法

### 3.1 提案手法のワークフロー

美術作品を鑑賞した際に得られる様々な心理状態は、鑑賞中の鑑賞者の表情や仕草などに表出すると考えられる。そこで、鑑賞中の鑑賞者をセンシングすることに基づく心理状態推定手法を提案する。

提案手法のワークフローを図1に示し、それぞれのステップについて以下に述べる。

(注1) : <https://www.ohara.or.jp/>

**ステップ 0：事前準備** 鑑賞者は初めに、美術館やガイダンスシステムの利用経験、パーソナリティなどの情報を入力し、その後鑑賞を開始する。

**ステップ 1：作品鑑賞** 鑑賞者は任意の作品を選択・移動し作品の鑑賞を行う。鑑賞方法は、解説を聞かずに作品を鑑賞する場合と、作品の音声ガイダンスを聞きながら作品を鑑賞する場合が存在することを想定する。作品鑑賞中は、ガイダンスを提供するスマートフォンアプリケーションやウェアラブルデバイスなどによって、鑑賞者の表情・声・仕草などの行動データをセンシングする。

**ステップ 2：ラベル入力（モデル構築フェーズのみ）** 作品鑑賞が終わったら、鑑賞者は作品に関する感想を口頭で述べ、スマートフォンアプリケーション上で心理状態推定モデルのラベルデータ（正解データ）を入力する。

**ステップ 3：心理状態推定（モデル運用フェーズのみ）** 作品鑑賞が終わったら、心理状態推定モデルが直近の鑑賞者の心理状態を推定するとともに、推定結果に基づいてガイダンスを動的に変更、提供する。以降、ステップ 1～3 を繰り返す。

### 3.2 モデル構築に使用する特微量、正解ラベル

心理状態推定は、収集したデータから特微量を抽出し、それを入力値、ラベルデータを出力値として鑑賞者の心理状態を推定するモデルを構築することで実現する。

作品鑑賞中に鑑賞者をセンシングすることによって得られるデータは、スマートフォンのインカメラ（鑑賞中の表情など）およびリアカメラ（鑑賞中の視線方向の様子）で撮影された映像データ、スマートフォンおよび頭部装着モーションセンサから取得する加速度・角加速度データである。各センサデータについての特微量抽出方法について以下に述べる。

**映像データ** 映像の各フレームについて画像データから顔のランドマーク検出、視線推定、頭部の姿勢推定、Action Units (AUs) の推定を行うことができるライブラリである OpenFace [16] を利用してフレームごとの特微量を取得する。得られた時系列の特微量から、以下の特微量を抽出し使用する。

- **鑑賞時間（秒）**：作品の鑑賞に要した時間を算出する。
- **継続作品鑑賞時間の時間幅別回数（回）・アプリ閲覧回数（回）**：視線推定データ (`gaze_angle_x, -y`) を用い、連続して作品の方向に視線が向いていた時間（鑑賞時間）が設定した時間幅  $W_g$  (= 5, 10, 20, 30, 40, 50, 60, 90, 120 秒) を超えた回数、および、その合間にスマートフォンの方向に視線が向いた回数を算出する。加えて、頭部の姿勢推定データ (`pose_Rx, Ry`) を用い、同様の特微量を算出する（時間幅  $W_h$  は 0.5, 1, 2, 3 秒）。
- **瞬き回数（回）、瞬き間隔の時間幅別回数（回）**：Action Units に含まれる瞬きイベント (`AU45_c`) を用い、瞬きの回数、および、瞬きイベントの間隔（瞬きしていない時間）が設定した時間幅  $W_b$  (= 5, 10, 20, 30, 40, 50 秒) を超えた回数を算出する。
- **Action Units の統計量**：Action Units の強度データ `AUx_r` ( $x$  は AUs のアクション ID) について、要約統計量（平均、標準偏差、最小値、最大値、中央値、第 1 四分位、第 3 四分位）を算出する。

表 1: 心理状態ラベルと設問

ラベル	設問
感情	作品を見てどのような感情が湧き上がりましたか？
満足度	作品を見てどれくらい満足しましたか？
理解度	作品の内容や意味が理解できたと感じますか？
困惑度	作品の何を見ればいいのかわからなくなった瞬間はありましたか？

表 2: 感情ラベルの選択肢

選択肢	ID
興奮した・驚いた	0
幸せな・嬉しい	1
落ち着いた・穏やかな	2
特に変わらない	3
眠い・疲れた	4
惨めな・悲しい	5
退屈な	6
困惑した・イライラした	7
怖い・不安な・怒った	8

表 3: 満足度ラベルの選択肢

選択肢	点数
非常に満足できた	6
満足できた	5
やや満足できた	4
どちらとも言えない	3
あまり満足できなかった	2
満足できなかった	1
全く満足できなかった	0

表 4: 理解度ラベルの選択肢

選択肢	点数
非常に理解できた	6
理解できた	5
やや理解できた	4
どちらとも言えない	3
あまり理解できなかった	2
理解できなかった	1
全く理解できなかった	0

表 5: 困惑度ラベルの選択肢

選択肢	点数
非常にあった	6
あった	5
ややなかった	4
どちらとも言えない	3
あまりなかった	2
なかった	1
全くなかった	0

位、最頻値) を算出する。

加速度・角加速度データ 加速度・角加速度の合計 6 軸の時系列モーションデータに対して、以下の特微量を抽出し使用する。

- **外れ値イベント回数（回）**：外れ値検知のアルゴリズム（対象データと過去 N サンプルの平均との差の絶対値が、過去 N サンプルの標準偏差の 2 倍を超える場合に外れ値とみなす）を施し、外れ値イベントが発生した回数を算出する。
- **加速度・角加速度データの統計量**：加速度・角加速度データについて、要約統計量（平均、標準偏差、最小値、最大値、中央値、第 1 四分位、第 3 四分位、最頻値）を算出する。

心理状態ラベル（感情・満足度・理解度・困惑度）は、表 1 の設問文を提示し、それぞれ表 2, 3, 4, 5 の選択肢から一つを選んだデータを用いる。感情ラベルは、Russell らが定義する 2 次元の感情円環モデル [17] に基づき、感情状態を表 2 に示すように 9 カテゴリに分割したものを使用する。満足度については、国土交通省の観光満足度調査に利用される 7 段階リッカート尺度を採用した。理解度・困惑度についても同様の尺度を採用している。心理状態推定モデルの構築にあたっては、感情ラベルは分類モデル、それ以外のラベルに対しては回帰モデルを構築する。

以降では、5. 章にて、データ収集および表情・仕草データによる心理状態推定モデルの構築について述べる。

## 4. 大原美術館でのデータ収集実験

本研究では、センシング機能を有する音声ガイダンスシステ



図 2: 骨伝導スピーカに取り付けたモーションセンサ

ムを新たに開発し、実際の美術館においてデータ収集実験を行った。

以降では、実験の概要、使用したデバイスと収集したデータ、およびデータ収集の実験手順について述べる。

#### 4.1 実験概要

本研究では2024年12月23日（月）大原美術館の館内に展示されている30作品を用いて、30人の被験者に対して、鑑賞中の行動をセンシングし、データを収集した。

詳細なデータ収集の手順については、4.3節に記載している。使用するデバイスと取得するデータの詳細については4.2節に記載している。

#### 4.2 使用デバイスと収集データ

本実験で使用したデバイスは以下の通りである。

**スマートフォン：** iPhone12 Pro 5台と iPhone14 Pro 5台（いずれもApple社製）を使用する。被験者に音声ガイドを提供しながら、鑑賞中の映像・音声データ、スマートフォン自体の加速度・角加速度データ、作品鑑賞後のアンケートデータを収集できるアプリケーションを開発した。アバター<sup>(注2)</sup>が音声ガイドを提供するという画面構成となっている。常に被験者の表情と鑑賞中の作品の映像を記録できるように、首掛けスマートフォンスタンドを利用して被験者の身体に固定する。

**ヘッドセット：** OpenRun<sup>(注3)</sup> (Shokz社製)を利用する。使用するスマートフォンと接続し、アプリケーションを鑑賞中の被験者に対し、作品のガイド音声を提供する。

**モーションセンサ：** MetaMotion S<sup>(注4)</sup> (MBIENTLAB社製)をモーションセンサとして利用した。ヘッドセットの左側に取り付けることで、被験者の頭部の加速度・角加速度データを取得する。モーションセンサの取り付け方については図2に示す。

これらのデバイスで取得するデータは以下の通りである。

**作品鑑賞中の映像データ：** 作品鑑賞中の被験者の表情をスマートフォンのフロントカメラから映像データとして記録した。鑑賞中の作品の様子について、スマートフォンのリアカメラから映像データとして記録した。また、これらの映像のフレームレートはいずれも30fpsである。

**加速度・角加速度データ：** 頭部に取り付けたモーションセンサと、鑑賞中身に着けているスマートフォンに内蔵されている

(注2) : CG-CA Gene (c) 2023 by Nagoya Institute of Technology, Moonshot R&D Goal 1 Avatar Symbiotic Society, <https://github.com/mmdagent-ex/gene/>

(注3) : <https://jp.shokz.com/products/openrun>

(注4) : <https://mbientlab.com/metamotions/>



(a) 鑑賞開始 (b) 音声ガイド (c) 感情の入力 (d) 満足度の入力  
図 3: ガイダンスアプリの画面



図 4: 実験の様子

モーションセンサからそれぞれ取得する。それぞれ周波数はスマートフォンによって取得したデータは20Hz、モーションセンサによって取得したデータは100Hzである。

**心理状態ラベル：** 作品鑑賞後にアプリケーションから表1に示した4個の設問に回答する。ラベルは、感情、満足度、理解度、困惑度の4種類であり、感情は9種類の選択肢から一つ、満足度、理解度、困惑度は7段階から選択する方式を取っている。

#### 4.3 データ収集の手順

ここでは、音声ガイダンスアプリを用いた作品鑑賞に沿って、データ収集の流れについて説明する。

##### 手順1：データ収集実験に関する説明

まず被験者は、実験の趣旨、実験の流れ、アプリの使用方法について説明を受け、事前アンケートに回答する。この際に被験者は身体にスマートフォンを固定し、加速度センサと一緒にした骨伝導スピーカを装着する。

##### 手順2：アプリケーションを用いた美術作品鑑賞

次に被験者は、図3のようなガイダンスアプリを利用して美術作品の鑑賞および、心理状態ラベル付けを行う。

鑑賞の開始にあたり、被験者は鑑賞する作品の前に立ち、作品の付近に掲載されている作品番号をアプリ内で入力する。作品番号を入力し、作品鑑賞画面に移ると、被験者はまず解説なしで作品を鑑賞する(図3(a))。数十秒作品を鑑賞したのち、音声で解説を聴きながら作品を鑑賞する(図3(b))。その後、被験者は作品に関する感想を30秒程度話し、図3(c),(d)のような画面で心理状態ラベルとなる設問に回答する。ここまでが作品鑑賞の1セッションであり、他の作品を鑑賞する際もこの一連の流れを繰り返す。実験の様子については図4に示す。作品鑑賞時間は45分間であり、この間被験者は館内の実験対象作

**表 6: 感情ラベル推定の評価結果**

	9 クラス	2 クラス
Accuracy	0.14 ± 0.08	0.64 ± 0.09
Recall	0.19 ± 0.09	0.64 ± 0.09
Precision	0.19 ± 0.09	0.64 ± 0.09
F1	0.19 ± 0.09	0.64 ± 0.09

**表 7: 満足度・理解度・困惑度ラベル推定の評価結果**

	満足度	理解度	困惑度
R2	-0.21 ± 0.12	-0.21 ± 0.15	0.05 ± 0.22
MAE	1.05 ± 0.13	1.24 ± 0.14	1.56 ± 0.20
RMSE	1.32 ± 0.14	1.55 ± 0.14	1.82 ± 0.21

品の鑑賞を行う。

### 手順 3: 事後アンケート

作品鑑賞が終了した後、被験者は事後アンケートに回答する。

## 5. 映像・加速度データを用いた心理状態推定

ここでは、4. 章で説明した実験で取得したデータを用いて心理状態モデルの構築し、評価した結果について考察する。

### 5.1 心理状態推定モデルの構築

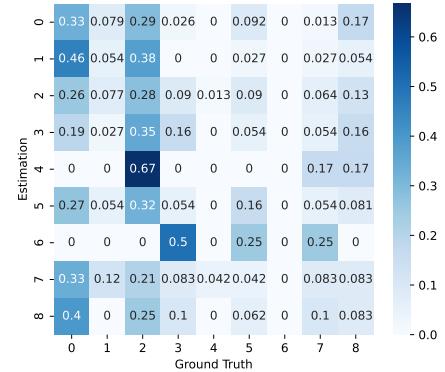
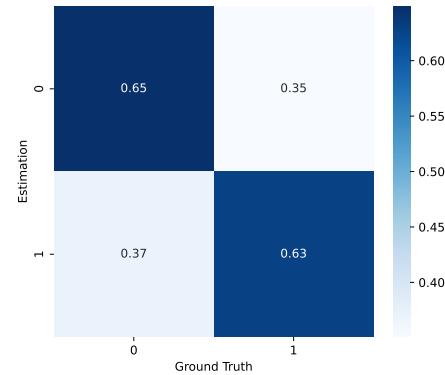
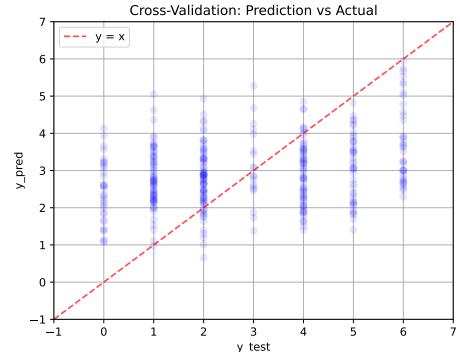
まず、収集されたデータから3.2章に示す特徴量をそれぞれ抽出する。なお、音声データについてはデバイス接続の問題によりデータ取得に失敗しており利用できなかった。

次に、推定モデルの構築と評価を行った。感情ラベル推定については表2に示す9つのラベルを分類する9クラス感情分類モデルと、それらをグルーピングしポジティブ感情(0:「興奮した・驚いた」「幸せな・嬉しい」「落ち着いた・穏やかな」)とネガティブ感情(1:「眠い・疲れた」「惨めな・悲しい」「退屈な」「困惑した・イライラした」「怖い・不安な・怒った」)を分類する2クラス分類モデルを構築した。機械学習アルゴリズムはLightGBM Classifier<sup>(注5)</sup>を用い、ハイパーパラメータの自動最適化フレームワークであるOptuna<sup>(注6)</sup>を用いたハイパーパラメータ調整を行った。満足度・理解度・困惑度ラベル推定においては回帰モデルを構築した。機械学習アルゴリズムはLightGBM Regressor<sup>(注7)</sup>を用い、同様にOptunaを用いたハイパーパラメータ調整を行った。また、構築した推定モデルの評価にあたっては、ラベル数が不均衡であることを考慮し、層化10分割交差検証(Stratified 10-Fold Cross Validation)を利用した。

### 5.2 評価結果と考察

感情ラベルの推定を行う分類モデルの評価結果を表6に示す。

感情ラベルについて、9クラス分類を行った結果を図5に、2クラス分類を行った結果を図6に示す。図5の各軸の数値(0~8)は、表2のIDに対応し、図6は0:ポジティブ、1:ネガティブに対応する。9クラス分類では精度が14%と低いものに

**図 5: 感情ラベルの推定結果（9 クラス）****図 6: 感情ラベルの推定結果（2 クラス）****図 7: 困惑度ラベル推定結果（全体）**

なっていたが、2クラス分類では64%となっており、ある程度の精度で分類を行うことができた。

9クラス分類では高い推定精度を得られなかった一方で、positiveとnegativeの2クラスの分類においてはある程度の精度を得られた。その理由として、2クラス分類では9クラス分類に比べ1クラスあたりのデータ量が多く確保できたことが考えられる。また、今回の実験では一つの作品の鑑賞全体に対する感情を回答してもらったため、その感情状態が作品に与えられたものか、ガイダンスによって与えられたものか、あるいはそれ以外の要因によるものかが分離できていない可能性がある。今後は、事前説明の際により明確に回答基準を示すなど、回答のばらつきを抑制する対策をとる必要がある。

満足度、理解度、困惑度の推定を行う回帰モデルの評価結果

(注5) : <https://lightgbm.readthedocs.io/en/latest/pythonapi/lightgbm.LGBMClassifier.html>

(注6) : <https://optuna.org/>

(注7) : <https://lightgbm.readthedocs.io/en/latest/pythonapi/lightgbm.LGBMRegressor.html>

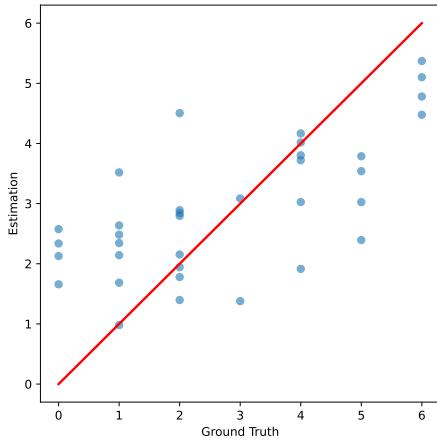


図 8: 困惑度ラベル推定結果（一部）

の詳細は表 7 に示す。満足度ラベル、理解度ラベルの推定については、いずれも傾向が見られなかった。そのため今回用いた特徴量と手法ではこれらを推定することは難しいことが明らかとなった。困惑度ラベルの推定についても、図 7 に示すように全体としては相関は見られなかった。しかしながら、交差検証を行ううちの 1 回のデータに着目すると図 7 のような結果が得られている場合があり、このデータにおいては R2 スコアが 0.45 とある程度の相関を持っている結果となった。困惑度ラベルが満足度ラベルや理解度ラベルとは異なり、一部において傾向が見られた理由として、困惑度ラベルは他のラベルに比べてクラスバランスによる偏りが少なかったと考えられる。また、一部でのみ傾向が見られた理由として、取得したデータと特徴量の個人間の差が大きかったと考えられる。そのため、他のラベルにおいての推定精度を向上させるためには、1) クラスバランスを可能な限り均等にすること、2) 個人間の差を排除する特徴量抽出を行うことが必要であると考察できる。また、個人変動を考慮した特徴量を追加するなど、個人差を抑制する方法について模索する必要がある。

## 6. 結 論

本研究では、大原美術館内で実際に作品を鑑賞する被験者の行動をセンシングし、映像データや加速度・角加速度データ、アンケートデータを収集する実験を行った。また、実験で得たデータを用いて作品鑑賞時の心理状態推定を行うモデルを構築し、その結果から分析を行った。

映像データと加速度・角加速度データを利用した推定手法について、感情ラベルの推定においては 64% の精度を得ることができた。しかしながら、満足度・理解度・困惑度については相関が見られなかった。困惑度の推定においては一部の推定結果においてある程度の傾向が見られた。

本研究における今後の展望として、1) 推定精度向上に有用なデータや特徴量抽出手法の模索、2) 他の機械学習手法を利用した推定モデル構築の検討が挙げられる。これらを行うことで実用化に向けた推定精度の向上を目指す。

## 謝 辞

本研究の一部は、SCAT 研究費助成および JSPS 科研費 (JP22H03648, JP24K20763) の助成を受けて行われたものです。また本研究の実施にあたり、公益財団法人大原美術館の協力の下で実施されたものです。ここに感謝の意を申し上げます。

## 文 献

- [1] 文化庁，“博物館数・入館者数・学芸員数の推移,” [https://www.bunka.go.jp/seisaku/bijutsukan\\_hakubutsukan/shinko/suui/](https://www.bunka.go.jp/seisaku/bijutsukan_hakubutsukan/shinko/suui/). Accessed on 18 July, 2024.
- [2] 古澤嘉久, 田和辻可昌, 松居辰則, “生体情報による学習者個人の心的状態推定モデルの精度評価とラベリングコスト低減に関する実験的検討,” 人工知能学会論文誌, vol.37, no.2, pp.C-L66\_1–10, 2022.
- [3] 森 大毅, 森本洋介, “適応的に発話タイミングを変える聞き手アウェアな音声ガイダンス,” 人工知能学会論文誌, vol.39, no.3, pp.IDS6-B\_1–10, 2024.
- [4] H. Kaya, A.A. Karpov, and A.A. Salah, “Robust acoustic emotion recognition based on cascaded normalization and extreme learning machines,” Advances in Neural Networks - ISNN 2016, pp.115–123, 2016.
- [5] W.Y. Quck, D.Y. Huang, W. Lin, H. Li, and M. Dong, “Mobile acoustic emotion recognition,” 2016 IEEE Region 10 Conference (TENCON), pp.170–174, Nov. 2016.
- [6] P. Tzirakis, G. Trigeorgis, M.A. Nicolaou, B.W. Schuller, and S. Zafeiriou, “End-to-end multimodal emotion recognition using deep neural networks,” IEEE Journal of Selected Topics in Signal Processing, vol.11, no.8, pp.1301–1309, 2017.
- [7] E. Ghaleb, M. Popa, and S. Asteriadis, “Multimodal and temporal perception of audio-visual cues for emotion recognition,” 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), pp.552–558, 2019.
- [8] W.L. Zheng, B.N. Dong, and B.L. Lu, “Multimodal emotion recognition using eeg and eye tracking data,” 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp.5040–5043, Aug. 2014.
- [9] Z. Zhang, Y. Song, L. Cui, X. Liu, and T. Zhu, “Emotion recognition based on customized smart bracelet with built-in accelerometer,” PeerJ, vol.4, p.e2258, 2016.
- [10] N. Rach, Y. Matsuda, S. Ultes, W. Minker, and K. Yasumoto, “Estimating subjective argument quality aspects from social signals in argumentative dialogue systems,” IEEE Access, vol.9, pp.11610–11621, 2021. <https://ieeexplore.ieee.org/document/9321384>
- [11] N. Rach, Y. Matsuda, J. Daxenberger, S. Ultes, and M.W. Yasumoto, Keiichi, “Evaluation of argument search approaches in the context of argumentative dialogue systems,” The 12th International Conference on Language Resources and Evaluation (LREC '20), pp.513–522, 2020.
- [12] J.-C.-Y. Sun and S.-J. Yu, “Personalized Wearable Guides or Audio Guides: An Evaluation of Personalized Museum Guides for Improving Learning Achievement and Cognitive Load,” International Journal of Human-Computer Interaction, vol.35, no.4-5, pp.404–414, 2019.
- [13] S. Liu and J. Guo, “Smart museum and visitor satisfaction,” Journal of Autonomous Intelligence, vol.7, no.3, pp.1–13, 2024.
- [14] 近藤亮介, 松田裕貴, S.D. Terlizzi, D. Ahmetovic, 福光嘉伸, 平良繁幸, 謙訪博彦, S. Mascetti, 安本慶一, “美術館における鑑賞者の心理状態推定に向けたマルチモーダルデータセットの構築,” 社会情報学会関西支部研究会 (SSI), pp.1–10, 2024.
- [15] 鳥垣耀平, 謙訪博彦, 安本慶一, 松田裕貴, “美術館における鑑賞者の心理状態推定に向けた基礎分析,” 2024 年度 情報処理学会関西支部 支部大会 講演論文集, no.G-16, pp.1–6, 2024.
- [16] T. Baltrušaitis, P. Robinson, and L.-P. Morency, “Openface: An open source facial behavior analysis toolkit,” 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), pp.1–10, 2016.
- [17] J.A. Russell, “A circumplex model of affect,” Journal of Personality and Social Psychology, pp.1161–1178, 1980.