

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

Basic Analysis Toward Emotional Status Estimation During Artwork Appreciation

鳥垣 耀平[†] 諏訪 博彦^{‡,*} 安本 慶一^{‡,*} 松田 裕貴^{‡,*}
Yohei Torigaki Hirohiko Suwa Keiichi Yasumoto Yuki Matsuda

1. はじめに

美術館や博物館は教育、学術、文化の発展に寄与すると同時に、鑑賞者が展示物の鑑賞を通じて感じ、学び、理解することも重要である。今日の美術館や博物館には人々の学習要求の多様化・高度化や社会の発展・変化に対応し、さらに積極的な役割を果たすことが期待されている。2007年には、新しい時代の博物館に必要とされる役割の一つとして、来館者の知的要求に答えるためのコミュニケーションの活性化が掲げられている [1]。近年では博物館資料の収集、保管、展示および調査研究などを行う、博物館法で定められた学芸員に加え、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ーが登用されており、文化庁ではミュージアムエデュケーターを育成する研修が行われているなどといった取り組みが行われている [2]。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 [3]。ミュージアムエデュケーターは鑑賞者に対し、単に展示物に関する展示を行うだけではなく、鑑賞者との対話を通じて鑑賞者の心理状態を読み取り、鑑賞者の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、音声ガイド等貸出型端末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という事に留まっており、館内ガイドを電子化したものの域を出ていない。それらのガイドは主に健常者を対象としており、障害を持つ人々が利用することは困難なものも多い。このような現状から、音声ガイドシステムなどがミュージアムエデュケーターに代替するには多くの課題があると考えた。本研究では、展示物を鑑賞する人の心理状態（感情・満足度）の推定に基づいて、鑑賞者に

じて対話的・動的にコンテンツ提示（ガイド内容・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、奈良県立美術館 [4] で実施された2つの展示会における28点を対象とし、10代～70代の男女62名の被験者により、ガイドシステムを用いた作品鑑賞時における鑑賞者の心理状態のデータセット構築を行った [5]。構築したデータセットは、美術館ガイドアプリおよびボイスレコーダによって、センサデータ（慣性センサデータ・画面操作ログデータ・音声データ）およびアンケートデータ（事前・事後アンケート、作品ごとの心理状態ラベル）によって構成される。構築されたデータセットのうち、主に音声データと作品ごとの心理状態ラベルのデータを利用し、機械学習を用いた心理状態推定のための基礎分析を行った。

本稿では、機械学習を用いて音声データから心理状態推定を行う手法について述べるとともに、実際のデータを分析し得られた結果とその結果についての考察について報告する。

2. 関連研究

2.1 感情認識についての研究

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。

Reschらは、リストバンド型のウェアラブルデバイスからユーザの身体運動、ソーシャルメディアから収集されるデータを用いた感情収集システム（Urban Emotions）を提案している [6]。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 [7]。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ウェアな音声ガイドシステムを提案している [8]。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も取り組まれている [9, 10]。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。認識の性能向上のために、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提案されている。TzirakisらやGhalebらは対

[†] 岡山大学, Okayama University

[‡] 奈良先端科学技術大学院大学,
Nara Institute of Science and Technology

* 理化学研究所革新知能統合研究センター (AIP),
RIKEN Center for Advanced Intelligence Project

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 [11, 12]. 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 [13, 14] 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 [15, 16].

また, 著者らはこれまでに, 屋外における一般的な観光を対象とした心理状態推定手法の実現に向けた研究に取り組んできた [17, 18]. この手法では, 観光中の仕草データ (アイトラッキングデータおよび頭部の IMU センサデータ), および自撮り動画データ (音声データおよび映像データ) を用いて, 観光スポットごとの観光客の心理状態 (感情・満足度) の推定モデルを構築している.

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, 近年では貸出端末でのナビ (音声ガイド) が国内外問わず広く用いられている. これらの多くは, 館内案内や展示物案内をデジタル化したものとなっており, 個人や状況に合わせたガイドを提供するに至っていないのが現状となっている. 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が示されている [19] 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 [20].

2.3 これまでの取り組みと本稿の立ち位置

著者らは 2023 年 8 月と 2024 年 4 月の 2 度に渡り, 奈良県立美術館でデータセット構築のための作品鑑賞中の鑑賞者の状態のデータ収集実験を行っている. この実験に関して, 2024 年 3 月に行われた社会情報学会関西支部の研究会論文にて報告されている [5]. この実験で取得したデータセットの詳細については 3.1 節で述べる. 本稿では, このデータ収集実験で構築したデータセットを用いた, 心理状態推定モデルを構築した結果について報告する.

3. 美術作品鑑賞時の心理状態推定モデル

3.1 データセットの概要

本稿では, 心理状態推定モデル構築のためのデータセットとして, 奈良県立美術館の展覧会を対照として行われたデータ収集実験 [5] で構築されたデータセットを用いる. ここでは, 得られたデータセットの概要について説明する.

表 1: 心理状態ラベルの選択肢 (感情)

選択肢	ラベル
興奮した	0
幸せな・嬉しい	1
落ち着いた・リラックスした	2
特に変わらない	3
眠い・疲れた	4
つまらなかった・落ち込んだ	5
がっかりした	6
動揺した・イライラした	7
怖い・恐れた	8

3.1.1 データ収集の実験設定

データ収集の実験設定 (実験対象作品, 被験者数と内訳) について述べる.

2023 年 8 月 14 日 (月)・21 日 (月)・22 日 (火) の計 3 日間で富本憲一展の 15 作品を選定して実施した. 被験者は合計 40 名であり, 属性は以下の通りである. 実験参加の経緯は, 一般募集による参加者 31 名, 美術館での研修生 (大学生) 9 名であり, 7 割以上が一般市民となっている. 被験者はすべて健常者である. 性別については, 男性 16 名, 女性 22 名, 不回答 2 名であり, 年齢については, 20 代 16 名, 30 代 4 名, 40 代 7 名, 50 代 7 名, 60 代 3 名, 70 代 1 名, 不回答 2 名となっている.

3.1.2 分析に利用するデータ

ここでは, データセット [5] について, 基礎分析に利用したデータを抜粋して説明する.

作品鑑賞中の音声データ

被験者が作品やガイドに対して抱いた感想を鑑賞直後に口頭で 30 秒程度述べてもらい, その内容をボイスレコーダーを用いて記録することで音声データを取得している.

心理状態ラベル

被験者が美術作品を鑑賞した後の心理状態データとして, 感情と満足度を取得した. これらのデータは, 各作品の鑑賞後に鑑賞者自身がアプリ上で入力したものである. 感情については, Russell らが定義する 2 次元の感情円環モデル [21] を用いる. このモデルに基づき, 感情状態を表 1 に示すように 9 カテゴリに分割した. 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 (表 2). なお, 平常時 (開始前) の満足度は 3 としている.

3.2 心理状態推定モデルの構築

ここでは, 本研究における心理状態推定モデルについて説明する. 一般的に感情状態は, 声の大きさや高さ,

表 2: 心理状態ラベルの選択肢 (満足度)

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

抑揚などの声の抑揚に現れるということが考えられる。そこで本研究では、鑑賞者の感想の音声データから音響的特徴量を算出し、機械学習によって分類・回帰を行う心理状態推定モデルを提案する。なお、このモデルでは 3.1 章で説明した取得したデータセットに含まれるデータのうち、作品鑑賞中の音声データと、心理状態ラベルデータとして「感情」と「満足度」を利用した。

以下に心理状態推定モデルを構築する手順について説明する。

ステップ 1: 音声データからの特徴量抽出

素性値の算出には音声データから特徴量抽出を行うことができるオープンソースソフトウェアの OpenSmile [22] を利用する。OpenSmile では Feature_set と Feature_level をそれぞれ指定する必要がある。Feature_level を指定することで、どのような特徴量をセットするのか決めることができる。本稿では ComParE.2016 を使用する。また、Feature_level を指定することで、特徴量の計算レベルを指定することができる。本研究では、Low Level Descriptors (LLD) の統計値を計算する Functionals を利用する。このように設定すると一つの音声あたりの特徴量は 6373 次元となり、機械学習はこの特徴量を用いて行うものとする。

ステップ 2: 推定モデル構築

まず、感情推定モデルの構築について述べる。用いたデータセットはネガティブな感情が非常に少ないアンバランスなデータセットであるため、ラベルデータは表 1 のうち比較的多くのデータがある「興奮した」「幸せな・嬉しい」「落ち着いた・リラックスした」「特に変わらない」の 4 クラスを対象とすることとし、分類モデルを構築した。モデル構築に利用した音声データは計 314 個であり、うち 8 割を教師用データ、残り 2 割をテストデータとして利用した。機械学習アルゴリズムとしては、Random Forest と LightGBM の 2 種類の方法を用いた。

次に、満足度推定モデルについて述べる。ラベルデータは表 2 の満足度ラベルを用い、回帰モデルを構築した。モデル構築に利用した音声データの数は 340 個であり、うち 8 割を教師用データとして利用した。機械学習アル

表 3: 感情ラベルによる心理状態推定の評価結果

	Random Forest	LightGBM
Accuracy	0.492	0.460
Recall	0.288	0.278
Precision	0.248	0.214
F1	0.242	0.239

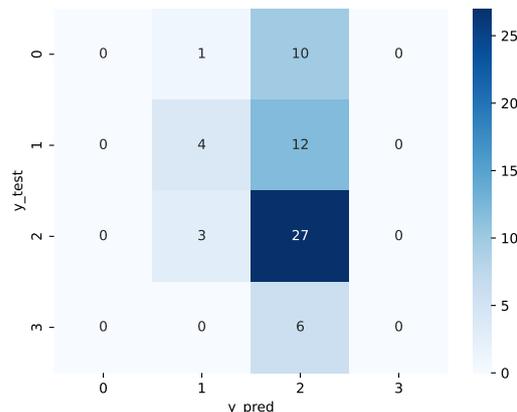


図 1: Random Forest の混同行列

ゴリズムとしては、重回帰分析とサポートベクター回帰 (SVR) の 2 種類の手法を用いた。

3.3 評価結果

3.3.1 感情ラベルによる心理状態推定の評価結果

分類モデルの評価指標として、正解率 (Accuracy), 再現率 (Recall), 適合率 (Precision), F1 値 (F1) を利用している。評価結果を表 3 に示す。また、Random Forest と LightGBM の混同行列をそれぞれ図 1, 図 2 に示す。

本手法の結果から、Random Forest, LightGBM のいずれを用いた場合でも、入力したデータの多くは「落ち着いた・リラックスした」に分類される結果となった。教師用データに最も多く含まれていたのは、「落ち着いた・リラックスした」のラベルを持つデータであったことは、このような結果を生じさせた 1 つの要因として考えられる。また、いずれの方法でも評価指標は低い値を示した。このことから、分類により有効な特徴量の抽出や、音声データの取得方法などにおいてさらなる検討が必要である。

3.3.2 満足度ラベルによる心理状態推定の評価結果

回帰モデルの評価指標として、決定係数 (R2), 平均絶対値誤差 (MAE), 平均二乗誤差 (MSE), 二乗平均平方根誤差 (RMSE), 対称平均絶対パーセント誤差 (SMAPE) を用いた。評価結果の詳細は図 4 に示す。ま

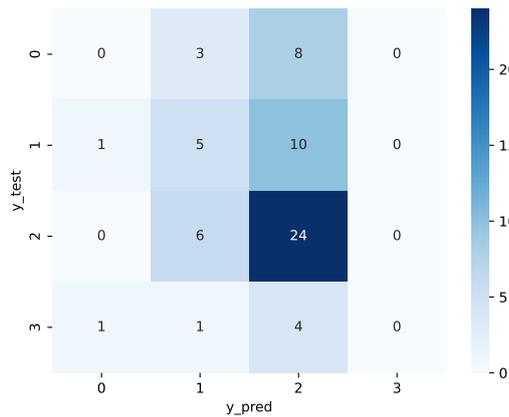


図 2: LightGBM の混同行列

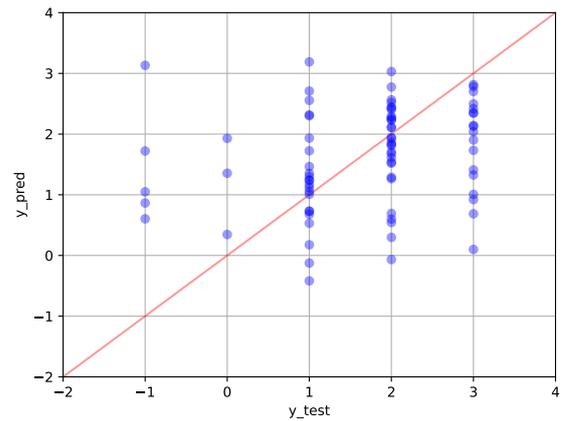


図 3: 重回帰分析の散布図

表 4: 満足度ラベルによる心理状態推定の評価結果

	重回帰分析	サポートベクター回帰 (SVR)
R2	-0.259	0.099
MAE	0.892	0.743
MSE	1.426	1.021
RMSE	1.914	1.010
SMAPE	65.190	47.814

た、重回帰分析と SVM の散布図をそれぞれ図 3、図 4 に示す。

それぞれの手法における評価指標では、重回帰分析よりも SVR がより予測精度が高いという結果が出た。しかしながら、図 4 を見ると、予測された満足度がラベル別に見てもあまり違いがなく、正しく予測できているとは考えにくい結果となっている。一方で図 3 を見ると、分散が大きいものの、特にデータ量の多かった満足度が「1」や「2」のデータについては正解に近い予測値が算出されたものも多く、この結果は音声データから満足度を予測するために精査すべき課題である。

4. おわりに

本研究では、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として、実際に美術館で得られた被験者の作品に対する感想を録音した音声データと作品を見た際の鑑賞者の感情と満足度を含むデータセットを構築し、機械学習による心理状態の推定精度の検証を行った。この検証により音声データのみで心理状態推定を行うことは現状では難しいという結果となった。その一方で、満足度の推定に関して一部のデータは正解に近い予測値を算出することができ、僅かではあるが心理状態推定に向け足掛かりが得られた結果となった。

今後研究を進めるにあたり、特に以下の 2 点の課題に

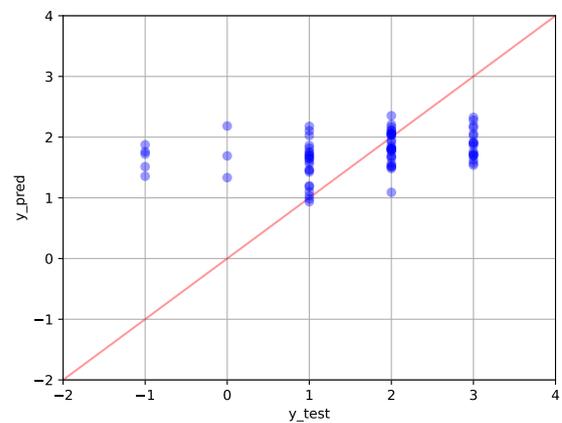


図 4: サポートベクター回帰 (SVR) の散布図

注力する。1 点目は、音声データによる感情認識の精度向上である。本研究では、音声データの特徴量の抽出や他の機械学習の手法の検討などにおいて課題が残る結果となった。今後心理状態推定を行うにあたり、この課題の検討は不可欠である。2 点目は、画像・映像データや生体信号データといったマルチモーダルなデータを用いた心理状態推定手法の検討である。本研究において、音声データのみによって心理状態を推定することは現状では難しいという結果となった。そこで、他のデータも含めた心理状態手法を検討することは必要であると考えられる。

謝辞 本研究の一部は、SCAT 研究費助成および JSPS 科研費 (22H03648) の助成を受けて行われたものです。

参考文献

- [1] 文部科学省. 新しい時代の博物館制度の在り方について. https://www.mext.go.jp/b_menu/shingi

- /chousa/shougai/014/toushin/07061901.pdf, 2007. Accessed on 18 July, 2024.
- [2] 文化庁. 博物館に関する研修. https://www.bunka.go.jp/seisaku/bijutsukan_hakubutsukan/kenshu/index.html. Accessed on 18 July, 2024.
- [3] 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/suii/. Accessed on 18 July, 2024.
- [4] 奈良県. 奈良県立美術館 奈良県公式ホームページ. <https://www.pref.nara.jp/11842.htm>. Accessed on 18 July, 2024.
- [5] 近藤亮介, 松田裕貴, Stefano Di Terlizzi, Dragan Ahmetovic, 福光嘉伸, 平良繁幸, 諏訪博彦, Sergio 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.
- [6] Bernd Resch, Anja Summa, Günther Sagl, Peter Zeile, and Jan-Philipp Exner. Urban emotions – geo-semantic emotion extraction from technical sensors, human sensors and crowdsourced data. In *Progress in Location-Based Services 2014*, pp. 199–212, 11 2014.
- [7] 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.
- [8] 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.
- [9] Heysem Kaya, Alexey A. Karpov, and Albert Ali Salah. Robust acoustic emotion recognition based on cascaded normalization and extreme learning machines. In *Advances in Neural Networks - ISNN 2016*, pp. 115–123, 2016.
- [10] Wei Y. Quack, Dong Y. Huang, Weisi Lin, Haizhou Li, and Minghui Dong. Mobile acoustic emotion recognition. In *2016 IEEE Region 10 Conference (TENCON)*, pp. 170–174, Nov 2016.
- [11] Panagiotis Tzirakis, George Trigeorgis, Michalis A Nicolaou, Björn W Schuller, and Stefanos 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.
- [12] Esam Ghaleb, Mirela Popa, and Stylianos Asteriadis. Multimodal and temporal perception of audio-visual cues for emotion recognition. In *8th International Conference on Affective Computing & Intelligent Interaction, ACII 2019, 2019*.
- [13] Wei L. Zheng, Bo N. Dong, and Bao L. Lu. Multimodal emotion recognition using eeg and eye tracking data. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5040–5043, Aug 2014.
- [14] Zhan Zhang, Yufei Song, Liqing Cui, Xiaoqian Liu, and Tingshao Zhu. Emotion recognition based on customized smart bracelet with built-in accelerometer. *PeerJ*, Vol. 4, p. e2258, 2016.
- [15] Niklas Rach, Yuki Matsuda, Stefan Ultes, Wolfgang Minker, and Keiichi Yasumoto. Estimating subjective argument quality aspects from social signals in argumentative dialogue systems. *IEEE Access*, Vol. 9, pp. 11610–11621, 2021.
- [16] Niklas Rach, Yuki Matsuda, Johannes Daxenberger, Stefan Ultes, and Minker Wolfgang Yasumoto, Keiichi. Evaluation of argument search approaches in the context of argumentative dialogue systems. In *The 12th International Conference on Language Resources and Evaluation (LREC '20)*, pp. 513–522, 2020.
- [17] Yuki Matsuda, Dmitrii Fedotov, Yuta Takahashi, Yutaka Arakawa, Keiichi Yasumoto, and Wolfgang Minker. Emotour: Estimating emotion and satisfaction of users based on behavioral cues and audiovisual data. *Sensors*, Vol. 18, No. 11, pp. 1–19, 2018.
- [18] Ryoya Hayashi, Yuki Matsuda, Manato Fujimoto, Hirohiko Suwa, and Keiichi Yasumoto. Multimodal tourists' emotion and satisfaction estimation considering weather conditions and analysis of feature importance. In *The 13th International Conference on Mobile Computing and Ubiquitous Networking (ICMU '21)*, pp. 145–150, 2021.

- [19] Jerry Chih-Yuan Sun and Shih-Jou 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.
- [20] Siyang Liu and Jian Guo. Smart museum and visitor satisfaction. *Journal of Autonomous Intelligence*, Vol. 7, No. 3, 2024.
- [21] James A Russell. A circumplex model of affect. In *Journal of Personality and Social Psychology*, pp. 1161–1178, 1980.
- [22] Florian Eyben, Martin Wöllmer, and Björn W. Schuller. Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM International Conference on Multimedia*, MM '10, pp. 1459–1462. ACM, 2010.