

あアラウド法における機械学習を用いた感情観測の初期検討

山岸 丈留* 安中 勇貴* 能丸 天志* 大見 亮介† 渡邊 恵太†

概要. 感情はインタラクショナルやエンタテインメントの評価指標であり、ゲーム体験の評価に重要である。川島らは、システム体験中の実験参加者に「あ」のみを発声させることで、リアルタイムかつ時系列な感情観測を可能にする「あアラウド法」のコンセプトを提案した。山岸らは、あアラウド法において Affect Grid を活用した分析手法を提案し、著者ら 2 名の分析結果からその利用可能性を示した。本研究では、山岸らが収集した発声データを使用し、機械学習による感情観測システムの初期検討を実施した。

1 はじめに

感情は、インタラクショナルやエンタテインメントシステムの主要な評価指標である [1][6][15]。そのため、インタラクティブ性やエンタテインメント性を伴うゲーム体験の評価では、体験者の時系列的な感情変容の観測が重要である。しかし、その観測は一般にゲーム体験型質問票 (GEQ)[12] やプレイヤー体験型満足度調査票 (PENS)[14] などの質問紙に頼っており、時系列な感情観測を達成できない [9]。他方で、生理反応や身体反応からゲーム体験を評価する取り組みもあるものの [16][7][11]、その数は少なく方法論の確立には至っていない。

川島ら [18] は、実験参加者のシステム体験中の感情をリアルタイムに得る実験手法として、思考発話法 [5] を応用し、「あアラウド法」を提案した。あアラウド法は、実験参加者を対象に、感嘆詞の「あ」のみの発声を用いてタスク中の感情を表現させ、その声の大きさや抑揚から感情観測を行う実験手法である。山岸ら [17] は、あアラウド法において Affect Grid [13] を活用した分析手法を提案し、著者ら 2 名の合議による分析結果からその利用可能性を示した。一方で、多くの実験者がこの分析手法を活用可能であるかは明らかでない。また、この手法はプレイ中の各秒に対してピアレビューを実施しており、分析に手間が掛かる課題がある。

そこで本研究では、山岸らが収集した発声データを使用し、機械学習による感情観測システム (図 1) の初期検討を実施した。実験参加者の「あ」の発声の大きさと高さに関する特徴量を抽出し、快度と覚醒度をそれぞれ 9 段階に分類するシステムを試作した。そして、学習モデルの性能評価から、課題と改善点を明らかにした。この研究を通じて、あアラウド法における分析の簡便化を目指す。

Copyright is held by the author(s). This paper is non-refereed and non-archival. Hence it may later appear in any journals, conferences, symposia, etc.

* 明治大学大学院 先端数理科学研究科

† 明治大学 総合数理学部

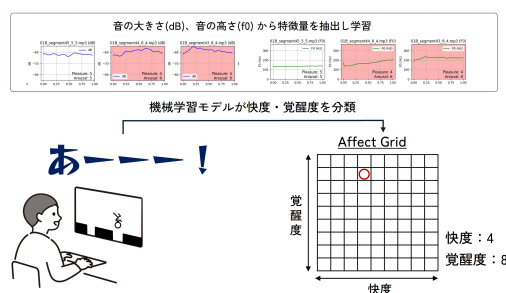


図 1. あアラウド法における機械学習を用いた感情観測。「あ」の発声の大きさと高さに関する特徴量から、快度と覚醒度をそれぞれ 9 段階に分類する。

2 機械学習を用いた感情観測の提案と検証

本章では、機械学習を用いて「あ」の発声から快度と覚醒度をそれぞれ 9 段階で評価するモデルを試作し、性能を評価した。

2.1 機械学習の設計

学習データとテストデータには、山岸ら [17] の研究で収集した実験参加者の発声データを使用した。この研究では、24 名の実験参加者が Sliding Penguin [10] のプレイ中、可能な限り「あ」を発声し感情を表現した。著者ら 2 名は、「あ」の発声とプレイ動画から、合議を通じて各秒の快度と覚醒度を 1~9 で評価した。一方、息継ぎなどで発声がなかった秒数は、快度と覚醒度ともに 0 とした。本研究では、各実験参加者の発声データを 1 秒ごとに分割し、この快度と覚醒度を正解ラベルとした。

学習の特徴量には、声の音量 (dB) と高さ (f0) に関するパラメータを選択した。あアラウド法は単母音「あ」の発声に限定されるため、フォルマントや MFCC などの特徴量は適さないと考えた。また、あアラウド法の先行研究 [4][17] において、多くの分析者が声の高さと大きさを指標に感情を分析したと述べた。これらの点を踏まえ、感情変容を直接的に反映する声の音量と高さのみに焦点を当てた。

声の音量と高さに関して、微細な振動成分による影響を排除し値を取得した。声の音量は、移動平均フィルタを使用した後、各秒の最大値と最小値を抽出した。声の高さは、各秒の第1四分位数と第3四分位数を抽出した。各秒で抽出した値そのものでなく、1秒前および5秒前の値との差分を特徴量とした。また、各秒の音量と高さのグラフに対し線形回帰を行い、その傾きを特徴量に加えた。

学習モデルは、Random Forest (RF) [2] を使用し構築した。RF は、複数の決定木を組み合わせて予測を行う教師あり学習アルゴリズムである。各木が独立に学習した結果の多数決をとることで、過学習を抑制しモデルの汎化性能を向上させる。

2.2 モデルの性能評価

構築した学習モデルを用いて、テストデータに対して快度と覚醒度を予測し、分類精度を確認した。

快度に関して、テストデータに対する精度指標と混合行列を図2に示す。快度の予測精度は、平均正解率が0.63、平均F1スコアが0.62、平均適合率が0.61、平均再現率が0.67であった。図2から、快度5のF1スコアは0.82、快度4では0.32であった。また、快度3、快度4はそれぞれ、快度4、快度5と予測する確率が最も高かった。一方で、快度6以上は2秒間しか正しく予測できなかった。

覚醒度に関して、テストデータに対する精度指標と混合行列を図3に示す。覚醒度の予測精度は、平均正解率が0.63、平均F1スコアが0.62、平均適合率が0.61、平均再現率が0.63であった。図3から、覚醒度5のF1スコアは0.78、快度6、快度7では0.35であった。また、覚醒度6、覚醒度7はそれぞれ、覚醒度5、覚醒度6と予測する確率が最も高かった。覚醒度4以下は、覚醒度5と予測する確率が最も高く、4以下と予測する確率の方が6以上と予測する確率よりも高かった。

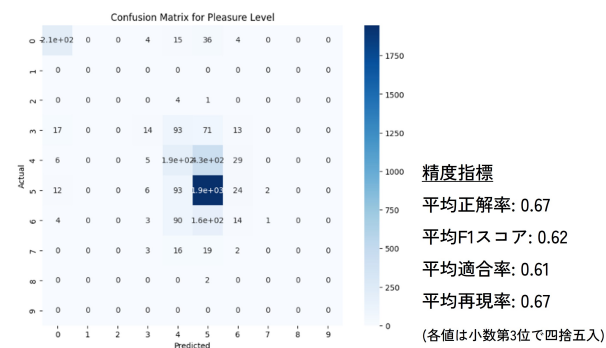


図2. テストデータに対する快度の混合行列と精度指標。快度3、快度4、快度5はそれぞれ、快度4、快度5、快度5と予測する確率が最も高かった。快度6以上は2秒間しか正しく予測できなかった。

3 議論

本章では2章の結果を踏まえ、現在の機械学習システムについて議論する。学習モデルにおける快度と覚醒度の平均F1スコアはいずれも0.6以上であった。また、図2、図3より、正解ラベルと予測が2以上外れる可能性が低かった。これらの結果は、学習モデルの一定の有用性を示唆している。

一方で、学習モデルの更なる精度向上には、発声データの再収集が必要であると考察する。快度5、覚醒度5の平均F1スコアは0.8以上であったが、その他のF1スコアは0.4以下であった。これは、快度と覚醒度が5のデータが多く、過学習が起きたと考察する。山岸らの合議によるラベル付けでは、機械学習への活用を前提としていなかったため、データに偏りが生じた可能性がある。また、著者ら2名の性格やゲームの特性が、ラベルの偏りに関与した可能性がある。今後は、より多くの感情発声を収集し、学習データとテストデータに使用する。

また、学習モデルを構築する過程にも、改善の余地があると考察する。本研究では、発声データを1秒ごとに分割し機械学習を実施した。より細かな分割や感情変化の瞬間に着目した分割により、ラベル付けの精度向上が期待できる。また、その他の音響特徴やゲーム画面の情報に関する特徴量を追加し、更なる精度向上が見込まれる。本研究では学習モデルにRandom Forestを使用した。LSTM[8]やXGBoost[3]など、時系列データの依存関係を考慮するモデルを用いることで、精度向上の可能性がある。また、それらのアンサンブル学習を取り入れることで、更なる精度向上が期待できる。

今後は、より高精度かつ多くのシステム評価に利用可能な学習モデルの構築を目指す。そして、あアラウド法におけるリアルタイム感情観測システムを開発し、広く実験者が活用できるようにする。



図3. テストデータに対する覚醒度の混合行列と精度指標。多くの場合で、覚醒度5以下は覚醒度5と予測した。覚醒度6、覚醒度7はそれぞれ、覚醒度5、覚醒度6と予測する確率が最も高かった。

参考文献

- [1] J. A. Bargas-Avila and K. Hornbæk. Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 2689–2698, 2011.
- [2] L. Breiman. Random forests. *Machine learning*, 45:5–32, 2001.
- [3] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.
- [4] 高山英里, 能丸天志, 安中勇貴, 山岸丈留, 渡邊恵太. 「あアラウド法」における感情観測手法の改善と分析手法の検討. エンタテインメントコンピューティングシンポジウム 2023 論文集, 2023:52–61, 2023.
- [5] K. A. Ericsson and H. A. Simon. *Protocol analysis: Verbal reports as data (Revised Edition)*. the MIT Press, 1993.
- [6] J. Forlizzi and K. Battarbee. Understanding experience in interactive systems. In *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, DIS '04, p. 261–268, New York, NY, USA, 2004. Association for Computing Machinery.
- [7] T. Hafeez, S. M. Umar Saeed, A. Arsalan, S. M. Anwar, M. U. Ashraf, and K. Alsubhi. EEG in game user analysis: A framework for expertise classification during gameplay. *Plos one*, 16(6):e0246913, 2021.
- [8] S. Hochreiter. Long Short-term Memory. *Neural Computation MIT-Press*, 1997.
- [9] D. Johnson, C. Watling, J. Gardner, and L. E. Nacke. The edge of glory: the relationship between metacritic scores and player experience. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play*, pp. 141–150, 2014.
- [10] T. Oka, T. Kawashima, D. Hayashi, and K. Watanabe. [Design and development of video games for ease of use and standardization of research] Kenkyū riyō shi yasuku hyōjun sei wo mezashita bideo gēmu no sekkei to kaihatsu. *Proceedings of Entertainment Computing Symposium 2021*, 2021:181–186, aug 2021.
- [11] M. Pasch and M. Landoni. Recognizing Bodily Expression of Affect in User Tests. pp. 264–271, 01 2011.
- [12] K. Poels, Y. de Kort, and W. IJsselstein. *D3. 3: Game Experience Questionnaire: development of a self-report measure to assess the psychological impact of digital games*. Technische Universiteit Eindhoven, 2007.
- [13] J. A. Russell, A. Weiss, and G. A. Mendelsohn. Affect grid: a single-item scale of pleasure and arousal. *Journal of personality and social psychology*, 57(3):493, 1989.
- [14] R. M. Ryan, C. S. Rigby, and A. Przybylski. The motivational pull of video games: A self-determination theory approach. *Motivation and emotion*, 30:344–360, 2006.
- [15] M. Thüring and S. Mahlke. Usability, aesthetics and emotions in human–technology interaction. *International journal of psychology*, 42(4):253–264, 2007.
- [16] E. N. Wiebe, A. Lamb, M. Hardy, and D. Sharek. Measuring engagement in video game-based environments: Investigation of the User Engagement Scale. *Computers in Human Behavior*, 32:123–132, 2014.
- [17] T. Yamagishi, Y. Yasunaka, E. Takayama, T. Nomaru, T. Komatsu, and K. Watanabe. Ah-Aloud Method to Comprehend Time-Series Emotion Observation During Gameplay: An Initial Investigation with Japanese Speakers. *Proceedings of the ACM on Human-Computer Interaction*, 8(CHI PLAY):1–23, 2024.
- [18] 川島拓也, 渡邊恵太. あアラウド法: 体験中の心理プロセスを「あ」の音声情報で評価する手法の提案と検証. エンタテインメントコンピューティングシンポジウム 2022 論文集, 2022:178–183, 2022.