

# ウェアラブル端末と心理入力アプリデータを統合した 感情予測アルゴリズムの開発

2025-06-13 進捗報告

森雄大

## 1. はじめに

### 研究の背景

過度な「怒り・イライラ・焦燥感・不安・抑うつ」は、放置すると心身の疾患や対人トラブル、生産性低下を招く。自他ともに感情変化を捉えにくいことが、早期ケアを難しくしている。

### 先行研究

- ・ 外部研究（在宅勤務者 30 名）：脳波・心電図・呼吸数・動作量と心理指標を同時取得し、ネガティブ感情と生体信号の相関を示した。
- ・ 中村教授による大規模データ計測：上記先行研究の知見を踏まえ、対象人数・測定項目を大幅に拡充。大規模データセットを新たに収集。

### 私の研究の立ち位置と目的

教授が収集した大規模データ（心拍数・呼吸数・活動量・皮膚電気反応など）を使わせていただき解析を行う。私が設定した目的は、ウェアラブル端末だけでリアルタイムにネガティブ／ポジティブの感情を高精度予測できるアルゴリズムを構築し、セルフケアに役立つ仕組みを提供すること。

### 期待される成果

- ・ 利用者が自分のメンタル状態を「見える化」できるセルフケアツール
- ・ 家族・職場との円滑な感情共有による早期サポート
- ・ 生理心理学とメンタルヘルス技術の発展への寄与

## 2. データ収集と解析の概要

### データの概要

提供いただいた生体計測データはすべて Empatica 製 E4 リストバンドによる計測データであった。各被験者には以下の 4 種類のデータファイルが用意されており、心理入力アプリで感情を入力する直前 15 分のデータが全入力分含まれていると思われる。

1. 運動データ（加速度）
2. 皮膚表面温度データ
3. 心拍数データ
4. 皮膚電位活動データ

併せて、感情状態の入力タイミングをまとめた Excel シートも提供されており、1 被験者あたり約 15 日間、1 日約 6 回（午前 9 時～午後 8 時）ほど入力が行われている。被験者数は資料上で約 150～270 名との記載あり。1 被験者あたりの入力回数は概ね 100 回前後とみられる。

感情状態の定量化には、

- A. 外山美樹氏による楽観・悲観性尺度アンケート（個人の心理傾向を測定）
- B. ラッセルの円環モデルに基づく 9 × 9 グリッドでのリアルタイム入力

の 2 手法が想定されているが、今回いただいたデータに含まれていたのは B の入力データのみであった。個人特性を考慮した汎用モデル構築には、A のアンケート結果の利用が必要と考えられるので、データが取得済みであればぜひ解析に使用させていただきたいです。

### データの形式

ID が 2 のデータを用いて、計測データの形式を確認した。

File	Dataset	Shape	Dtype
data_2_E4_act.h5	E4	(10, 28799)	float64
data_2_E4_eda.h5	E4	(1, 3599)	float64
data_2_E4_rri.h5	E4	(1, 3600)	float64
data_2_E4_temp.h5	E4	(1, 3599)	float64

例えば、皮膚表面温度データである temp は、4Hz でデータ取得をしている。約 3600 個のデータがあるということは、15 分間のデータである。つまり、このデータは被験者 1（S01）の 2 回目の感情入力タイミングの直前 15 分の皮膚表面温度の時系列データである。

## データの読み込み方法

以下のようなコードで h5 ファイルの読み込みを行う。

---

Listing 1: HDF5 ファイル読み込み関数

```
import os
import h5py

def list_and_preview_h5_files(directory="sample", preview_shape=(2, 5)):
    """
    指定ディレクトリ内のすべての .h5 ファイルを走査し、
    各データセットの形状と冒頭の一部データを表示します。
    """
    for fname in sorted(os.listdir(directory)):
        if not fname.endswith(".h5"):
            continue
        filepath = os.path.join(directory, fname)
        print(f"\n=== File: {fname} ===")
        with h5py.File(filepath, "r") as f:
            for dset_name, dset in f.items():
                data = dset[()]
                print(f"- Dataset '{dset_name}': shape={data.shape}, dtype={data.dtype}")
    )

if __name__ == "__main__":
    list_and_preview_h5_files()
```

---

## 全データの詳細

以下に、データの形状や値の単位などのメタデータを表形式でまとめる。

ファイル名	説明	データの長さ	単位
data_N_E4_act.h5	x 軸方向の生加速度 (32Hz*15min)	28799	不明
	y 軸方向の生加速度 (32Hz*15min)	28799	不明
	z 軸方向の生加速度 (32Hz*15min)	28799	不明
	身体運動ベクトルの大きさ (32Hz*15min)	28799	不明
	x 軸方向の加速度の重力成分 (32Hz*15min)	28799	不明
	y 軸方向の加速度の重力成分 (32Hz*15min)	28799	不明
	z 軸方向の加速度の重力成分 (32Hz*15min)	28799	不明
	x 軸方向の身体運動成分 (32Hz*15min)	28799	不明
	y 軸方向の身体運動成分 (32Hz*15min)	28799	不明
	z 軸方向の身体運動成分 (32Hz*15min)	28799	不明
data_N_E4_eda.h5	感情計測直前の皮膚電位データ (4Hz*15min)	(1, 3599)	不明
data_N_E4_rri.h5	感情計測直前の心拍間隔データ (4Hz*15min)	(1, 3600)	秒
data_N_E4_temp.h5	感情計測直前の皮膚温度データ (4Hz*15min)	(1, 3599)	摂氏温度

対応するタイムスタンプのデータは、「all-grid-data-extracted2025」から取得できる。計測ファイル名が「data-N-E4-act.h5」であれば対応するタイムスタンプはシートの N 行目を参照すればよい。

## 簡易解析

全被験者のデータをもとに簡単な解析を行う予定だったが、実行環境への全計測データのコピーが丸一日以上かかるため間に合わなかった。次回の ToDo とする。

## 次回の ToDo

- ・ データの傾向を理解するための簡単な解析（一日の計測頻度の分散等）
- ・ どんな回帰による予測が良いか、データの性質を踏まえて考察