

配布データと応募用ファイルの作成方法について

本コンペティションで配布されるデータと応募用ファイルの作成方法について説明する。

配布データ

配布するデータは以下の通り。

- `README.pdf`
- `train.zip`
- `test.zip`
- `sample_submit.csv`
- `tutorial.zip`

README

このファイル、最初に参照するドキュメント。`README.pdf`として与えられる。

学習用データ

モデルを学習するための系列データ。`train.zip`として配布されている。解凍すると以下のようなディレクトリで展開される。

```
train
├── negative
│   └── ...
└── positive
    └── ...
```

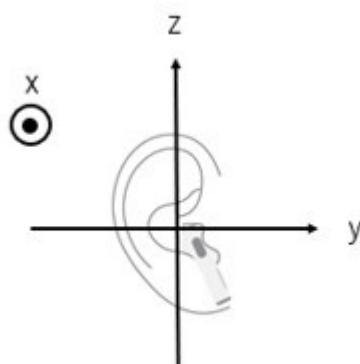
頸関節症の自覚症状がある(=`positive`)群とない(=`negative`)群がディレクトリ名として分けられている。それぞれの配下に被験者ごとに取得された系列データが被験者ID名別にディレクトリでまとめられている(`00`, `01`, ...). データはcsv形式で, `{セット数}_{動作}_{左右}.csv`という名前となっている。セット数は'set01'~'set20'の合計20回、動作の種類は"自力最大開口"('01'), "側転運動"('02')の2種類で、左耳('left')と右耳('right')から取得されたデータがある。例えば`set01_02_left.csv`という名前であれば、セット1回目の側転運動において左耳から取得されたデータという意味である。データは左右の耳にセンサー(AirPods)を付けた状態で様々な運動をして取得された。被験者は全部で21人である。

取得されたデータの説明は以下の通り。項目名が各ファイルのカラム名となっている。

項目名	物理的な 意味	単位	解説
QuaternionX, QuaternionY,	姿勢(ク オータニ オン)	なし(正 規化数 値)	計算用の姿勢データ。ジンバルロック(回転の不具合) を避けて姿勢を計算するための4次元ベクトル。
QuaternionZ,			
QuaternionW			

項目名	物理的な意味	単位	解説
AttitudePitch	ピッチ (うなずき)	ラジアン(rad)	X軸周りの回転. 首を「縦に振る(Yes)」動作の角度. 上や下を向いている度合いを示す.
AttitudeRoll	ロール (首の傾げ)	ラジアン(rad)	Y軸周りの回転. 首を「肩に近づける(かしげる)」動作の角度. 画面が斜めになっているかどうかの判定に使う.
AttitudeYaw	ヨー(左右の首振り)	ラジアン(rad)	Z軸周りの回転. 首を「左右に振る(No)」動作の角度. コンパスの方角(北や南など)や, 左右のどちらを見ているかを表す.
AccelerationX, AccelerationY, AccelerationZ	ユーザー加速度	G	動きの勢い(衝撃). 歩行時の振動, ジャンプ, 急停止, 頭を素早く振った時の「G」が記録される. 静止している時はほぼ 0 になる.
RotationX, RotationY, RotationZ	角速度 (Rotation Rate)	rad/s(ラジアン毎秒)	回転のスピード. 例えば RotationX の数値が大きい場合, 「猛スピードでうなずいている(ヘッドバンギングしている)」状態. ゆっくり動くと数値は小さくなり, 止まると 0 になる.
GravitationalAccelerationX, GravitationalAccelerationY, GravitationalAccelerationZ	重力ベクトル	G	「下」がどっちかを示す. ユーザーが静止していても, 常に合計で約1.0Gの値が出力される. これを見てることで, ユーザーが寝転がっているのか, 立っているのか, 逆立ちしているのかがわかる.
Timestamp	時刻	s	ある時点を基準として取得した時刻

座標系は以下の通り。



評価用データ

精度評価の対象となる系列データ. [test.zip](#)として配布されている. 解凍すると以下のようなディレクトリで展開される.

```
test
└── 000
    └── ...
        └── 279
```

各ディレクトリは(被験者, セット数)のIDを表していて、その配下にデータが格納されている。IDは全部で280個ある。データはcsv形式で、`{動作}_{左右}.csv`という名前になっており、例えば`01_right.csv`というファイルは自力最大開口運動における右耳からのデータという意味である。なお、被験者は学習用データとは別である。各csvファイルのフォーマットは学習用データと同じ。

応募用サンプルファイル

投稿用のサンプルファイル。`sample_submit.csv`として配布されている。このフォーマットに従って応募用ファイルを作成する。ヘッダーなしのcsvファイル。

カラム番号 説明

0	サンプルID。評価用データのディレクトリ名(被験者, セット数のID)に対応する。
1	対応するサンプルに対する頸関節症の自覚症状がある確率値。値は0~1の実数値。

チュートリアルノートブック

配布されたデータを用いて応募用ファイルを作成するまでの一連の流れを解説したノートブックで、`tutorial.zip`として配布されている。解凍すると`tutorial.ipynb`というファイルができる。必要に応じて分析やモデリングを行う際の参考にされたい。

応募用ファイルの作成方法

上記の[応募用サンプルファイル](#)のフォーマットに従って作成する。ヘッダーなしのcsvファイルで、すべての評価用ID(000~279)を含めること。

注意事項

応募用ファイルを投稿する前に、正しいフォーマットになっているかを確認すること。特に下記の点に注意。

- ヘッダーなしになっているか
- IDと確率値が正しく対応しているか
- IDに過不足はないか
- ファイルの拡張子が`.csv`になっているか