

スケートボーダー重心位置予測チャレンジ 考察レポート

ユーザ名: DT-SN

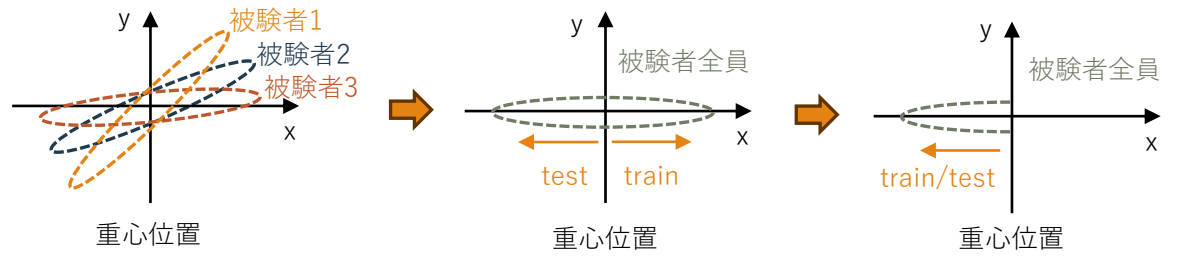
順位: 3位

スコア: 暫定評価0.7346682, 最終評価0.7513308

1. データ処理の工夫点と考察

■重心速度の補正

- 1. 被験者ごとにランプ長辺方向が異なっている状態のため下記補正を行った。
「速度が一定以上⇒進行方向はランプ長辺方向と一致する」と仮定し、
このときの進行方向の中央値とx軸が平行になるように回転変換した。
- 2. trainとtestでは重心の軌跡が真逆となっているためtest側に合わせた。

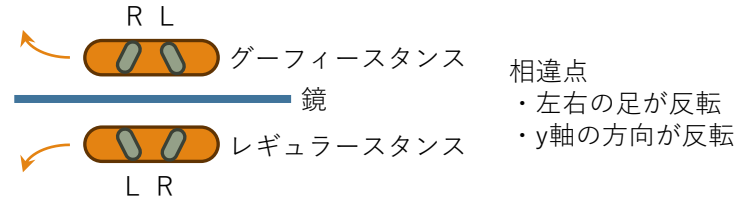


■EMGデータの補正

- 1. 筋電位の振幅と筋力には相関関係があるので、扱いやすいよう絶対値を取り被験者ごとに各信号の90パーセンタイルで割って値域を合わせた。
(ノイズによるスパイクの影響を避けるため最大値よりも少し下で割った)。
- 2. 目的変数と同じ周期(60Hz)のデータになるよう、33 or 34サンプル単位の平均と標準偏差に置き換えた。

■鏡像データ追加

ターンがパンピングの場合、その後の滑走は逆スタンスの状態となるが、trainではスロープ下りで逆スタンスになるのに対し、testではスロープ上りで逆スタンスになるためtrain/testで特徴に差異が生じる。そこで、鏡像データ(EMG左右反転とy軸速度反転したもの)を追加した。
鏡像データはスタンスを逆にしたものと解釈できるため、加えることでtrain/testを差異を緩和できる。

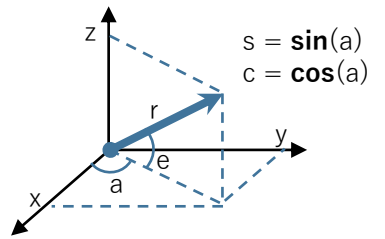


■加速度追加

筋電位から直接予測できるのは速度ではなく加速度であると考え加速度を目的変数に追加し、速度は加速度の時系列変化から予測することにした。
(最終的な目的変数ではないため厳密な加速度ではなく速度差分で代用している)

■球面座標系追加

スケートボードの操作は直進方向の加減速と方向回転であるので球面座標系と相性が良いと考えられる。
そこで目的変数である直交座標系の速度を球面座標系に変換し目的変数に加えた。
また、方位角は-180度から180度の間で不連続になってしまうのでsinとcosに置き換えた。



■平滑化

コース形状からターン時以外は単調な速度変化をすると考え、推論結果をSavitzky-Golayフィルタで平滑化する処理を入れた。

2.モデリングの工夫点と考察

■使用モデルと構成

使用モデル：LightGBM

下記の順で処理を実施(前段の結果は次段の特徴量として追加していく)：
z加速度の学習&予測→z以外の加速度の学習&予測→z速度の学習&予測
→z以外の速度の学習&予測→スタッキング→2種類の座標系の結果をマージ

■CV手法

trainのスロープ上りデータと下りデータを異なるfoldに分けることで
trainとtestの関係に近い状態のCVを行う。具体的には右図のように分割する
(上り側のみ5foldでCV+下り側のみ5foldでCV+リファレンスtrain側をholdout
+リファレンスtest側をholdout)。

train1 (vel_x > 10)				train2 (vel_x <= 10)				ref.train	ref.test
1	2	3	4	1	2	3	4	5	5
cross validation(5fold)				train					
train				cross validation(5fold)				train	
train								holdout	train
train									holdout



■Regressor Chainのオーバーフィット対策

Regressor Chain方式(前周期の予測結果を特徴量に加え次周期を予測) に下記対策を施すことでオーバーフィットが緩和され精度が向上した。
対策1：複数周期(5～15)まとめて学習/予測する
対策2：過去の予測結果をすべて追加する方式と1つ前の周期のみ追加する方式を併用する

3.分析結果から得られたインサイトと考察

■trainとtestでデータの特徴が異なる場合の対策について

下記2点に注意しながら設計を行うことで、trainデータとtestデータ間の差異に対する対策ができ精度向上に繋がった。
・試験条件からtrainとtestにどのような差異が生じているかを十分検討すること
・検討結果を元にできる限りtestデータに適合できるようなデータ処理とモデル設計をすること
だた闇雲にデータを処理しモデルを構築するのではなく、試験条件に合わせた設計を行うことが重要だと判った。

■説明変数から目的変数を直接予測するのが困難な場合の対処法

今回、筋電位から速度を直接予測するのは困難だと考え、まずは加速度(速度の変化量)を予測し、予測した加速度の時系列変化から速度を予測した。
同様の手法(目的変数の変化量を予測し、予測した変化量の時系列変化から目的変数を予測する)はどのような問題設定であれば有効か検討してみたい。