

スケートボードトリック分類チャレンジの 解法

Solution for Motion Decoding Using Biosignals

データ①: 0.9155496 1st-place
データ②: 0.7104558 6th-place

Kot

2024 年 10 月 31 日

スケートボードトリック分類チャレンジの解法

SOLUTION FOR MOTION DECODING USING BIOSIGNALS

Kot^{*,A)},

Abstract

In this paper, we propose a model for classifying skateboard tricks using biosignals such as EEG and EMG. The proposed model is a basic encoder that combines convolution and pooling layers. The structure of the model was determined through hyperparameter tuning using Bayesian optimization. As a result, it was found that high-frequency features were important for classification in the raw signal data (Data ①), while low-frequency features were essential in the cleansed data (Data ②). This approach achieved scores of 0.9155496 for Data ① and 0.7104558 for Data ②.

1. はじめに

本コンテストは、スケートボードの動作を頭皮上の生体信号からポンピング、前向きキックターン、後向きキックターンの3つのトリックを分類することを目的としています。データは、生の生体信号（データ①）と生体信号に由来しない成分をクレンジング処理したもの（データ②）の2種類を用いて分析を行います。動作中に頭皮上から取得される生体信号データは多くのノイズを含むため、適切な前処理が重要です。実環境でのスケートボードの動作予測は、ブレイン・マシンインターフェースやヒューマノイドロボットの運動生成研究への応用が期待されます。スケートボードトリックの分類を通じて、最適なデータ前処理とモデル構築手法を探索しました。解法の要点は以下の通りです、

- **KFoldPerSubject:** 被験者を推定する分類モデルを構築したところ、データの種別に関係なく誤りなく分類できる結果が得られました。生体信号には被験者依存性が見られることから、ネットワークの学習は被験者ごとのデータを与えています。
- **前処理:** モデルが過学習しないようにするため、0.5秒間の短い信号からノイズを除去する適切な前処理を方法を探索しました。採用した前処理は、オフセット除去、駆動してないチャンネルの除去、バンドパスフィルター（2~125Hz）の単純なパイプラインです。
- **ネットワーク構造の最適化:** 学習では過学習が問題であり、汎化性を持たせるために適当なパラメータ数でモデルを構築することが重要でした。ベイズ最適化を用いたモデル構造とパラメータ数を探索させることでデータセットに適したモデルを構築しました。

2. データ処理の工夫点と考察

本章では、前処理と拡張の工夫した点をまとめます。検証は1次元畳み込みのUNetを使用しました。また、モデリングにあたって探索的データ解析で得られた知見をまとめます。

2.1 前処理

データ①の生体信号は、動作中のノイズが含まれるため、適切な処置が必要です。データ②を参考としてフィルター処理を検討しました。フィルターのカットオフ周波数は、0.5秒間で観測可能な最低周波数から2Hzとナイキスト周波数の1/2から125Hzを設定しました。この帯域は、脳波（EEG）と筋電位（EMG）を含みます。また、電源ノイズ（60Hz）の影響を考慮し、ノッチフィルターも試しました。フィルタリングの有無によるモデルの性能の比較を測定ファイル単位に交差検証（ $k=3$ ）で確認しました。検証にはニューラルネットワークを使用するため平均を0、標準偏差を1とする正規化をしました。結果をTable 1にまとめます。

Table 1 からバンドパスフィルターを使用することで、精度が向上しているのがわかります。一方、ノッチフィルターによる電源ノイズの除去は精度の向上が見られませんでした。したがって、生体信号に関する情報を最大限に活用しつつ、情報の欠損を防ぐために、前処理は最低限にしてバンドパスフィルターのみを採用しました。

2.2 データ拡張

ニューラルネットワークでは新しいデータに対する汎化性能が重要となります。電極配置の対称性を活用した拡張や信号の符号や振幅、位相などを変えるデータ拡張を検討しましたが、**データ拡張はモデルの精度向上に寄与しません**でした。特に、信号の振幅や符号をランダムに変更する手法は、モデル性能を低下させることがわかりました。

2.3 探索的データ解析

提供データについて、モデリング前に以下の3点を確認しました。

1. 訓練と評価のドメインが同一であるのか
2. 被験者ごとに依存性があるのか
3. 重要なチャンネルはあるのか

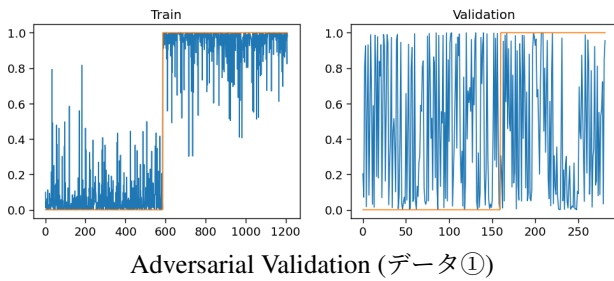
2.3.1 訓練と評価のドメインが同一であるのか

訓練データと評価データが同じドメインであるのか確かめるため、訓練データか評価データを推定する検証を実施しました。このとき、訓練と評価用のデータ数が同一となるように調整しています。学習結果をFig. 1に示す。オレンジの線はラベルを表し、0は評価データ、1は

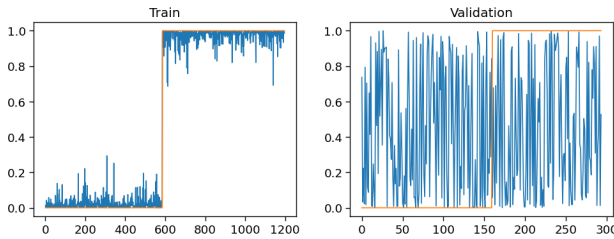
Table 1: 前処理の比較 (subject0)

Experiment	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5
Normalize	✓	✓	✓	✓	✓
Highpass Filter		✓			
Lowpass Filter			✓		
Bandpass Filter				✓	✓
Notch Filter					✓
OOF (Kfold) Score	0.788	0.811	0.788	0.847	0.832

訓練データです。青の線は活性度を表し、1に近いほど訓練データであることを表します。



Adversarial Validation (データ①)



Adversarial Validation (データ②)

Figure 1: Adversarial Validation Combined

Fig. 1 からは訓練データはラベルに沿った推論できているのに対して、検証データのばらつきは大きいことがわかります。検証データの正答率は5割程度でした。これは、訓練データが信号を記憶して、新規データでは有意な特徴を得られていないことを示します。よって、訓練用データと評価用データのドメインは同一であることがわかりました。また、モデルは過学習しないよう単純な構成がよい指針を得ました。

2.3.2 被験者ごとに依存性があるのか

データの被験者依存性を確認するため、入力信号から被験者を推定する検証を実施しました。分類結果の混合行列を Fig. 2 に示します。

Fig. 2 からはデータセット①とデータセット②ともに被験者を誤りなく分類できています。データセット②は生体信号以外のノイズを除去されたものではありませんでしたが、信号データに被験者依存性があることがわかりました。このことから、被験者ごとの違いを考慮し、被験者ごとにモデルを構築することが重要であることがわかりました。

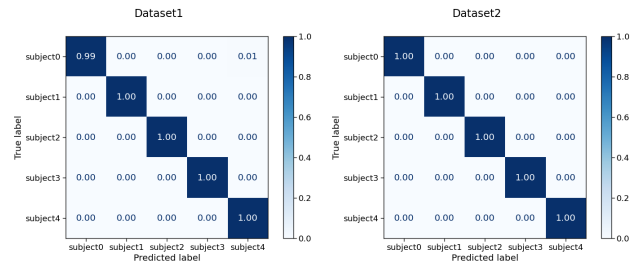


Figure 2: 被験者の分類結果

2.3.3 重要なチャンネルはあるのか

最後に、アテンションが有効であるか確認するため、Unet の各ダウンサンプル層に Squeeze-and-Excitation ブロックを挿入し、活性度を観察しました。Fig. 3 にダウンサンプルごとの活性度の変化を示します。

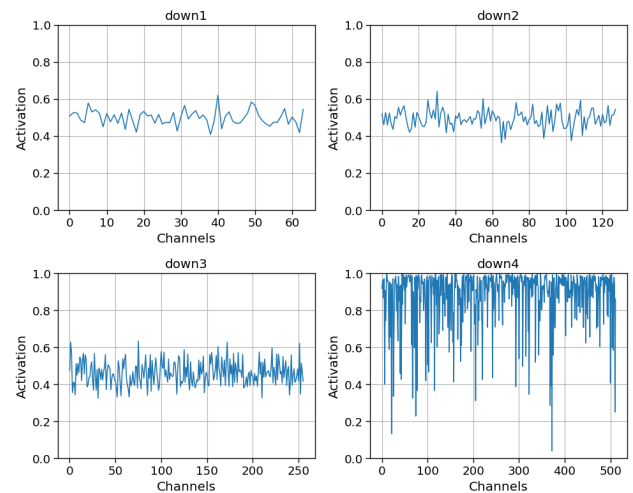


Figure 3: チャンネルごとの活性度

1 回目のダウンサンプルでは、活性度が 0.5 付近に集中しており、特定のチャンネルに着目してないことが分かります。チャンネル方向のアテンションが有効に働き始めるのは最終ブロックであり、この 4 ブロック目までに 8 層の畳み込みが使用されました。モデリングでは 8 層以下の単純なモデルにするため、チャンネル方向のアテンションを導入する必要はないと判断しました。同様の理由で、Transformer についても同様の理由で採用し

ていません。さらに、時系列方向のアテンションについても、サンプル数が少ないため実施しませんでした。

3. モデリングの工夫点と考察

本章では提案モデルを Feature Extractor、Pooling、および Head の 3 つのコンポーネントごとに検証した内容を説明します。1 次元畳み込み (Conv1D) に基づく U-Net アーキテクチャをベースラインモデルとして、最適なモデル構成を探索しました。

Feature Extractor

畳み込みで調整可能なパラメータとして、カーネルサイズ、パディング、ストライドがあります。カーネルサイズは特に重要で、受容野に相当します。例えば、低周波成分が重要な場合、小さいカーネルサイズだと受容野が不足する可能性があります。提供データは信号の中央部を切り出しており、端部のアーチファクトは 0 埋めしています。ストライドは 1 と固定し、全区間の信号を活用します。特徴抽出層として、以下の異なるアプローチを試行しました:

- **Conv1D:** 標準的な畳み込み演算
- **Depth-wise:** チャンネルごとに畳み込む
- **Separable:** 空間とチャンネルに分けて畳み込む

Pooling

プーリング層は特徴マップのダウンサンプリングを行うために使用します。次の手法を比較しました:

- **MaxPooling:** 最大値を取得して強い特徴を保持する。
- **AveragePooling:** 平均値を取得し、全体的な傾向を反映させる。

Head

出力部分は次元を削減した後、全結合層で 3 クラスの分類結果を出力します。以下のモジュールを試しました:

- **Convolution + Fully Connected Layers:** 畳み込み層で特徴を集約し、全結合層で最終出力を得る。

3.1 ネットワーク構造の探索

ハイパーパラメータの最適化には、Optuna を用いたベイズ最適化を行いました。探索したハイパーパラメータは以下の通りです:

- 隠れ層のチャンネル数: 128~1024
- カーネルサイズ: 3~19
- 畳み込み層の数: 2~3
- 畳み込みの種類: Conv1d、Depthwise、Separable
- プーリングの種類: Max、Average、Both
- 活性化関数: ReLU、SiLU、GELU、Leaky ReLU

カーネルサイズはデータセット①で 3、データセット②で 7 となりました。試行条件によって結果に変化はあるものの大小関係は維持されています。これはデータセット①では高波長の特徴が、データセット②ではより

低周波の特徴が有効であることを示します。詳細なパラメータは付録を参考ください。

4. 分析結果から得られたインサイトと考察

被験者依存性の強さ: 被験者ごとの EEG 信号の違いが顕著で、個別のモデルを作成することが精度向上に重要であることが確認されました。

アーチファクトを含むデータの方が精度が高い: クレンジング処理されたデータよりも、アーチファクトが含まれる生の EEG データの方が高精度であり、ノイズも有効な特徴として捉えられている可能性があります。

カーネルサイズの調整とデータの種別依存性: 生データ (データ①) では大きなカーネルサイズを使用することで低周波数成分を効果的に捉え、クレンジングデータ (データ②) では小さなカーネルサイズが有効でした。これは、データの種別に応じた異なる周波数帯域の特徴を抽出するため、カーネルサイズの調整が必要であることを示しています。

5. データ①②で作成したモデルの分類性能の違いに関する考察

生の生体信号 (データ①) の方が、クレンジングされたデータ (データ②) よりも高い分類精度を示しました。クレンジングによるノイズ除去が有効な特徴 (特に被験者固有の信号パターン) を除去している可能性があります。

6. 分析結果の社会課題への応用・展開に関する考察

リアルタイム処理: 提案モデルは計算量が比較的軽量であり、リアルタイムでの動作が可能です。特に FPGA などの低遅延ハードウェアと組み合わせた用途が有効です。

精度の改善余地と用途の検討: 現時点の分類精度では適用できるタスクが限定されますが、複数センサーのデータを組み合わせる形で精度を補完する形での利用が期待されます。

ハードウェアとの連携: 分類結果は自動制御システムや IoT デバイスとの連携に利用できます。分類結果の蓄積は将来のモデル改善や新たな応用に役立ちます。

7. 生体信号データの取得における課題に対する AI による解決策

ノイズ環境: 生体信号は外部ノイズやアーチファクトに汚染されやすいです。ノイズを許容した上でリアルタイムに AI によるリアルタイムにノイズを検知・除去する技術が考えられます。

長時間のデータ収集とリアルタイム処理: 長時間のモニタリングではデータの欠落やセンサーのズレが問題になります。AI によるデータ補完技術で欠損データの予測・補正が必要です。マルチモーダルに複数の信号を統合的に解釈し、精度の高い解析が求められます。

【付録】

1. データセット

5人の被験者(subject0～subject4)ごとに取得したデータがまとめられて提供されました。国際 10-10 法で決められた電極配置で取得した 72 チャンネルの電気信号、トリックの時刻と種別、チャンネルラベルなどがファイルに格納されています。電気信号は 500Hz でサンプリングされており、訓練データは全区間の信号、推論データは 0.5 秒間に切り出された形式で配布されました。

Table 2: データ①の前処理内容

項目	内容
駆動していないチャンネル	0 埋め

Table 3: データ②の前処理内容

項目	内容
バンドパスフィルタ	3～100Hz
電源ノイズ除去	60Hz
駆動していないチャンネル	除去
アーチファクト部分空間法	ASR
部分的に駆動していない信号の補完	-
レファレンス信号の平均化	-
独立成分分析	情報最大化基準
ダイポール推定法	脳波再構成
ICLabel 法	脳波成分分解
非「脳由来」成分の除去	閾値 50%

2. データ拡張

ニューラルネットワークでは新しいデータに対する汎化性能が重要となります。電極配置の対称性を活用した拡張や信号の符号や振幅、位相などを変えるデータ拡張を検討しましたが、いずれも**データ拡張はモデルの精度向上に寄与しません**でした。モデリング内で試行したデータ拡張の一覧を示す。

- Cutout
- Time Shift
- Amplitude Scaling
- Noise
- Roll
- Random Channel Flip
- Random Sign Flip
- Random Channel Mask

3. 最適なパラメータ設定

Optuna でモデルのハイパーパラメータを探索した結果を Table 4 と Table 5 にまとめます。Fig. 4 と Fig. 5 に可視化した結果を示します。

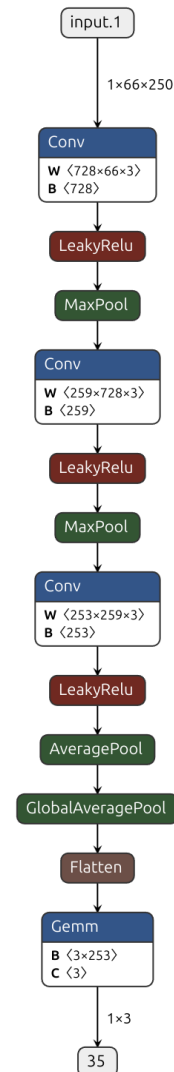


Figure 4: モデルアーキテクチャ：データ①

Table 4: 最適パラメータ（データ①）

Parameter	Value
hidden_channels	728, 259, 253
kernel_size	3
conv_type	conv1d
pooling_type	max, max, avg
activation	leaky_relu
num_layers	3

4. アテンション機構の試行

SE Block を Unet のダウンサンプルごとに挿入して学習したときの損失プロットを Fig. 6 に示す。学習の損失は下がっていて、精度 100% 近くまで出ていることがわかる。一方で検証データに対しては精度は 80% 程度、損失は安定していないことがわかる。このことから、レイヤー数は少なくなるようにモデルを調整した。

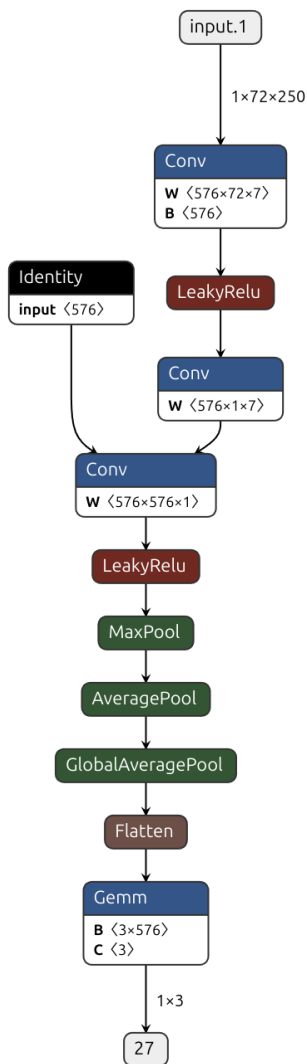


Figure 5: モデルアーキテクチャ：データ②

Table 5: 最適パラメータ（データ②）

Parameter	Value
hidden_channels	576, 576
kernel_size	7
conv_type	conv1d, separable
pooling_type	avg & max
activation	leaky_relu
num_layers	2

5. 様々なアーキテクチャの試行

画像分類で提案されているモデルアーキテクチャを参考に実装して、Trial ごとの交差検証を実施したときの分類精度を Table 6 にまとめる。

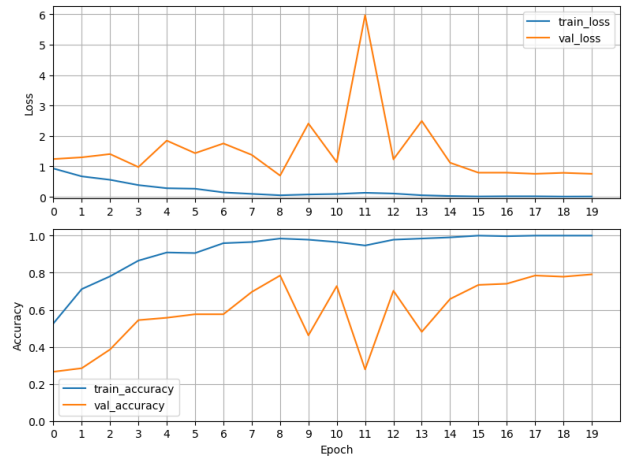


Figure 6

Table 6: Model Performance Comparison

Model	OOF Score
AlexNet	0.8008
DenseNet	0.6667
EfficientNetB0	0.5262
LeNet	0.7589
MnasNetA1	0.4969
MobileNetV1	0.4969
MobileNetV2	0.4382
MobileNetV3_small	0.5367
MobileNetV3_large	0.4969
ResNet50	0.6122
ShuffleNetV1_G3	0.4990
ShuffleNetV2	0.5241
SqueezeNet	0.3522
VGG19	0.6080
Xception	0.5828
ZFNet	0.8071
Encoder1D	0.8470