

深層学習による学習初期の運動パフォーマンスに基づく運動学習能力の推定

野尻 夏暉 【身体情報サイエンス研究室】

Estimation of Motor Learning Ability Based on Early-Stage Motor Performance Using Deep Learning

NOJIRI, Natsuki 【Brain-Muscle Coordination Lab.】

1 背景

ヒトは練習や経験を通じて運動技能を獲得し、熟練した動きを形成していく。この過程は運動学習と呼ばれ、その学習能力には個人差が存在する。こうした個人差は、学習初期の動作に既に表れている可能性があるが、その違いは微細であり、人間の視覚による判別は困難である。一方で、時系列データを解析する深層学習モデルである Recurrent Neural Network (RNN) や Long Short-Term Memory (LSTM) は、動作認識などの分野で高い性能を示しており、時間的変化を伴う動作特徴を捉えるのに適している [1]。さらに、近年では動画を入力とする深層学習モデルを用いた動作認識や行動分類の研究も多く報告されている [2]。これらのモデルを用いることで、これまで見逃されがちであった個人差を捉え、学習能力を推定できる可能性がある。

本研究では、ジャグリング技能獲得前の動作から、将来的な運動学習能力の推定を試みた。1つ目の手法として、骨格時系列データを用い、RNN および LSTM による2クラス分類を行った。Raw Skeleton Data, Euclidean Norm Data, PCA-Reduced Data の3種類のデータセットを作成し、技能獲得前の動作から成績上位群と下位群を判別できるか検証し、正解率を比較した。2つ目の手法として、背景差分処理を施した技能獲得前の動画を動画分類モデルに入力し、2クラス分類を行うことで、成績上位群と下位群を判別できるか検証した。

2 実験方法

2.1 ジャグリング課題

本研究では、運動学習課題としてツインワンハンドを用いたジャグリング課題を実施した。事前アンケートで右利きかつジャグリング未経験と判定された74名(平均年齢 20.80 ± 1.56 歳、女性 25 名)が参加した。

2.2 成績評価

ボールが落ちる、動作が止まる、片手で2球同時に保持するまでを1試行とし、250試行実施した。250試行中の成功回数の最大値をその被験者の成績とした。

試行データが欠損していた8名を除いた66名を解析対象とした。成績上位1/3(22人)および下位1/3(22人)

の2クラスに分類し、それぞれの最初の10試行分の動画に対してラベル付けを行った。

2.3 解析に用いる動画

解析に用いる動画として、投球開始から最初の1回目の投球が完了するまでの様子を記録した動画①、および投球開始からボールが頂点に達するまでの様子を記録した動画②の2種類を用意した。

3 解析方法

3.1 骨格時系列データを用いた解析

3.1.1 前処理

得られた2種類の動画データに対して MediaPipe を用い、33関節における3次元座標の骨格データを取得した。取得した骨格データに対し、前処理として、骨盤を原点とする位置合わせ、身長に基づくスケール、フレーム数の統一(動画①:30フレーム、動画②:15フレーム)を行った。

3.1.2 データセット

モデルに入力するデータセットとして、33関節の3次元座標を用いた「Raw Skeleton Data」、各関節のユークリッドノルムを特徴量とした「Euclidean Norm Data」、主成分分析により各関節の3次元情報を1次元に削減した「PCA-Reduced Data」の3種類を用意した。

3.1.3 推定モデル

骨格時系列データを用いた解析では、RNN と LSTM を用いて2クラス分類を行った。20回の解析を行い、正解率の平均をモデルの評価とした。

3.2 動画データを用いた解析

3.2.1 前処理・データセット

背景情報の影響を低減するため、OpenCV 実装のMOG2を用いた混合ガウスモデルに基づく背景差分処理を2種類の動画に適用し、得られた動画をデータセットとした。

3.2.2 推定モデル

動画データを用いた解析では、事前学習済の ResNet50 による特徴抽出と RNN および LSTM を組み合わせた深層学習モデルを用いて2クラス分類を行った。20回

の解析を行い、正解率の平均をモデルの評価とした。

4 結果

4.1 骨格時系列データを用いた解析

表1は、動画①を用いた解析におけるRNN, LSTMの最終スコア(20回の正解率の平均)である。チャンスレベル50%との1標本 t 検定を行った結果、すべてのモデル、データで正解率に有意な差が得られた($p<.01$)。

RNNにおいて、一元配置分散分析により、データ間の正解率に有意な差が得られた($p<.001$)。また、Bonferroniの多重比較により、Raw Skeleton Dataの正解率が他の2つのデータより有意に高いことが示された(いずれも $p<.001$)。

LSTMにおいて、一元配置分散分析により、データ間の正解率に有意な差が得られた($p<.001$)。また、Bonferroniの多重比較により、Raw Skeleton Dataの正解率が他の2つのデータより有意に高く(いずれも $p<.001$)、PCA-Reduced Dataの正解率はEuclidean Norm Dataの正解率より有意に高いことが示された($p<.05$)。

表1: 動画①における各推定モデルの最終スコア

Model	動画①		
	Raw Skeleton	Euclidean Norm	PCA-Reduced
RNN	81.25 ± 5.13	70.74 ± 4.66	68.58 ± 5.20
LSTM	87.61 ± 3.68	65.91 ± 4.73	68.75 ± 4.27

各値は平均 ± 標準偏差 (単位: %) で表す。

表2は、動画②を用いた解析におけるRNN, LSTMの最終スコア(20回の正解率の平均)である。チャンスレベル50%との1標本 t 検定を行った結果、すべてのモデル、データで正解率に有意な差が得られた($p<.01$)。

RNNにおいて、一元配置分散分析により、データ間の正解率に有意な差が得られた($p<.001$)。また、Bonferroniの多重比較により、Raw Skeleton Dataの正解率が他の2つのデータより有意に高く(いずれも $p<.001$)、Euclidean Norm Dataの正解率はPCA-Reduced Dataの正解率より有意に高いことが示された($p<.01$)。

LSTMにおいて、一元配置分散分析により、データ間の正解率に有意な差が得られた($p<.001$)。また、Bonferroniの多重比較により、Raw Skeleton Dataの正解率が他の2つのデータより有意に高いことが示された(いずれも $p<.001$)。

表2: 動画②における各推定モデルの最終スコア

Model	動画②		
	Raw Skeleton	Euclidean Norm	PCA-Reduced
RNN	84.09 ± 4.25	72.50 ± 5.20	68.92 ± 5.36
LSTM	85.17 ± 4.38	70.80 ± 3.75	72.05 ± 5.45

各値は平均 ± 標準偏差 (単位: %) で表す。

4.2 動画データを用いた解析

表3は、動画解析モデルの最終スコアである(20回の正解率の平均)。

ResNet50-RNNにおいて、チャンスレベル50%との1標本 t 検定を行った結果、動画①、動画②のいずれにおいても有意な差が得られた(いずれも $p<.001$)。

ResNet50-LSTMにおいて、チャンスレベル50%との1標本 t 検定を行った結果、動画①、動画②のいずれにおいても有意な差が得られた(動画①: $p<.05$, 動画②: $p<.01$)。

表3: 推定モデルの最終スコア

Model	動画①	動画②
ResNet50-RNN	60.06 ± 7.31	58.44 ± 7.45
ResNet50-LSTM	56.89 ± 12.00	58.56 ± 12.01

各値は平均 ± 標準偏差 (単位: %) で表す。

5 考察

骨格時系列データを用いた解析では、学習初期の動作から成績上位群と下位群を有意に判別できたことから、すべてのモデル、データにおいて運動学習課題の学習度を予測することができたと考えられる。このことから、学習能力の高い者は学習前の段階において、既に効率的な運動様式を部分的に備えている可能性が示唆された。また、Raw Skeleton Dataの正解率が他データの正解率と比較して有意に高かったことから、Raw Skeleton Dataは関節の3次元配置や相対関係を保持しているため、これらの情報が距離情報や次元削減後のデータと比べて、予測に有用であったと考えられる。

動画データを用いた解析においても、成績上位群と下位群を有意に判別できた。このことから、学習初期の動作に含まれる映像ベースの情報から、将来的な運動学習能力と関連する特徴を一定程度抽出できる可能性が示唆された。

6 まとめ

本研究では、ジャグリング技能獲得前の動作から、将来的な運動学習能力を推定できるかを検討した。その結果、骨格時系列データを用いた解析において、学習初期の動作から成績上位群と下位群を有意に判別できることが示された。特に、関節の三次元配置や相対関係を保持したRaw Skeleton Dataが高い予測性能を示した。

動画データを用いた解析においても、学習初期の動作から成績上位群と下位群を有意に判別できることが示された。

参考文献

- [1] Muhamad, et al., "Improved RNN Model for Real-Time Human Activity Recognition", *Iraqi Journal of Science*, Vol. 65, No. 9, pp. 5220-5238, 2024.
- [2] Gong, et al., "A Survey of Video Action Recognition Based on Deep Learning", *Knowledge-Based Systems*, Vol. 320, Article. 113594, 2025.