

ディープラーニングによるバスケットボールの テーブルオフィシャルズ支援

千田 了[†] 上田 芳弘[†] 坂本 一磨[†] 仙田 朋也[‡]

公立小松大学生産システム科学部[†] 公立小松大学大学院サステイナブルシステム科学研究科[‡]

1. はじめに

近年、様々な分野で人工知能が活用されるようになり、その活躍はスポーツ分野でも多くの事例が見受けられる。特にサッカーやテニスでは、会場内の多数のカメラを介して人の目では難しい判定を人工知能に手助けさせる AI 審判を試合中継などでよく目にする。それらの中には機械学習を行い、リアルタイムでカメラからの映像から情報を得て結果を出力するものがありそれらはスポーツ分野以外で手話の認識やジェスチャの認識などに活用されている。

バスケットボールには、テーブルオフィシャルズ（以下 TO と略す）という審判とともに試合を円滑に進行する役割を担う人物が複数人存在する。TO は、試合のスコアキーパーやタイマー、ファウル数の記録など、試合において重要な役割を果たしている。しかし、既存研究において、審判に着目した研究や選手のトレーニングやパフォーマンスの向上に関する研究は行われているが、TO に着目した研究は行われていない。関連研究には、Mediapipe, CNN を用いて、ASL の分類を行う研究がある。その TO にはファウルをした選手を記録するという役割があり、ファウルをした選手の背番号とその選手がしたファウルの種類を審判が TO にジェスチャで知らせる。本研究では、人工知能により審判が知らせるファウルのジェスチャを認識することを目指す。TO の役割は、現状人の手で行われており、プロの公式戦などを行う際は最低でも 4 人必要とする。人工知能を用いて TO の役割を支援することで、TO に割く人員を減らすことができ、少ない人数でも試合を行うことが可能となる。それに加え、試合中に起こる TO によるミスを削減し、試合の進行をより円滑に行うことが可能となる。

2. 提案手法

2.1 研究概要

本研究は、そのファウルのジェスチャをカメラで撮影し、MMpose[1]で 2 次元の骨格データを

Support for Basketball Table Officials using Deep Learning

[†]Ryo Senda, Yoshihiro Ueda and Kazuma Sakamoto · Faculty of Production Systems Engineering and Sciences, Komatsu University

[‡]Tomoya Senda · Graduate School of Sustainable Systems Science, Komatsu University

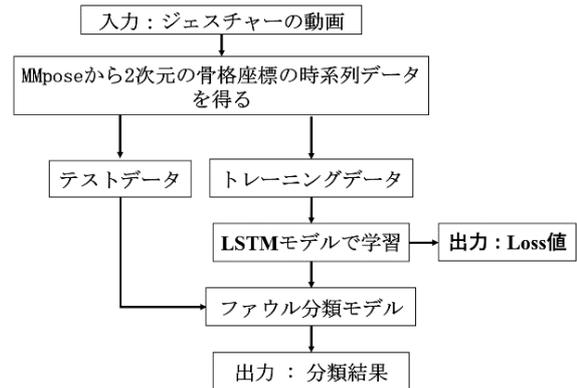
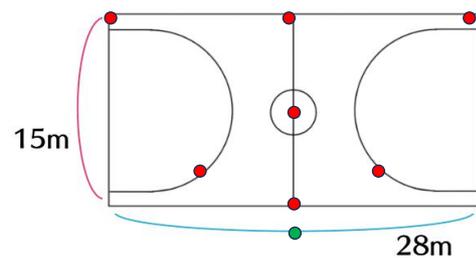


図1 システムの概要

一般のバスケットボールコート大きさ



●：審判がジェスチャを行う位置の例
●：カメラ位置

引用：<https://spojoba.com/articles/794>

図2 撮影位置

取得して、LSTM(Long Short Term Memory)[2]を用いて学習させる。

2.2 撮影方法

サイドラインとセンターラインの交差する点から 1.3m 離れた位置にカメラを設置し、図 2 に示す位置でジェスチャを行う。ジェスチャを行う際、カメラは、ジェスチャを行う審判を中心にとらえられるように設置し、設置位置は変えずに向きを変更するものとした。

撮影するジェスチャは、ルールに直接関係するものやジェスチャが似ているという観点から、15 種類中 5 種類のファウルに絞る。

3. 実験内容

本実験では、2 次元の骨格座標の時系列データから LSTM を用いて、「アンスポーツマンライクファウル」、「テクニカルファウル」、「ハン

ドチェックング」, 「ブロッキング」, 「ホールディング」の5種類のファウルのジェスチャを学習させる.

3.1 データセット

本研究では, 撮影したジェスチャの動画から得た2次元の骨格座標の時系列データを用いる. また, 各動画の長さ, 各ファウルのデータ数には差がある. 各ファウルのデータ数を表1に示す.

3.2 評価

本研究では, 学習を行う際の条件を変えることで, 各条件下で学習を行った際に得られる Loss 値で評価を行う.

3.3 学習条件

学習を行う際に, データ数, epoch 数, sequence の長さの3つの項目を変化させる.

データ数は, 審判役以外の人も映っている50個データを含む場合と含まない場合で条件分けた. なお, 審判役以外も映っている場合のデータ数は391であり, 映っていない場合のデータ数は341である. sequence の長さは, 各動画のframe数の違いから, 最大frame数である518とframe数の平均である253に条件分けする. ここで, frame数を最大値に設定した際はframe数が最大値以下のデータには0の値で補填し, frame数を平均値に設定した際はframe数が平均値以上のデータは平均値以上のframeを除外することとした. そして, この4つの条件を用いて学習を行う.

4. 結果と考察

学習によって得られた Loss 値を図3に示す. 図3より, どの条件においても Loss 値の変化が不安定であることがわかる. これは, データ数が少なかったことや各ファウルのジェスチャのデータの数に偏りがあることが原因となっているのではないかと考えられる. また, 図4に示すようにともに胸の前で手首をつかむジェスチャであり, 酷似している. このことからジェスチャの誤分類が起きている可能性も考えられる.

また, データ数が同じで sequence の長さが違う条件での Loss 値の変化を比較すると, どちらも Loss 値が1.4から1.6の間で変化し, どちらも最終的に Loss 値が1.5付近で変化が小さくなっている. このことから, sequence の長さによる学習への影響度はあまりないのではないかと考えられる.

それに対して, sequence の長さが同じでデータ数が違う条件での Loss 値の変化を比較すると, 審判役以外の人も映っているデータを含まない341個のデータ数より, 含む391個のデータ数の方が, 最終的な Loss 値がより小さくなっている

表1 各ファウルのデータ数

ファウルの種類	データ数
アンスポーツマン ライクファウル	62
テクニカルファウル	48
ハンドチェックング	118
ブロッキング	49
ホールディング	114

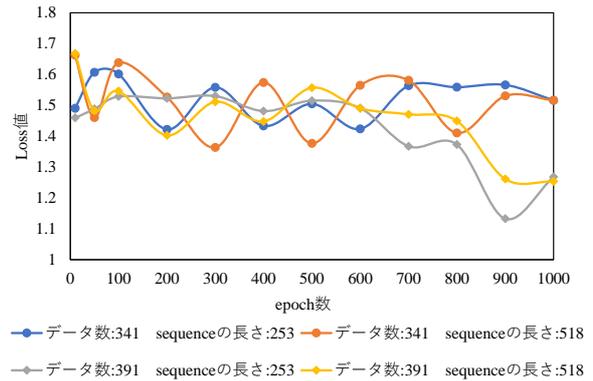


図3 epoch 数による Loss 値の変化



a)ホールディング b)ハンドチェックング

図4 酷似するファウルのジェスチャ

ことがわかる. これは, もともとのデータ数が少ないため, データ一つの学習への影響度が大きいのではないかと考えられる.

5. おわりに

今後はデータ数の調整し, 学習を続け, 分類モデルの精度を出していきたい.

また, 今回除外したファウルのジェスチャにも対応でき, ファウルをした選手の背番号を認識することができるシステムの開発を目指す.

参考文献

[1] MMPose Contributors, : OpenMMLab Pose Estimation Toolbox and Benchmark, 入手先 <<https://github.com/open-mmlab/mmpose>> (参照 2024-1-12).

[2] Hochreiter. S. and Schmidhuber. J. : LONG SHORT-TERM MEMORY, *Neural Computation*, Vol. 9, No. 8, pp. 1735-1780, (1997).

[3] Kumar. R., Bajpai. A., and Sinha. A. : Mediapipe and CNNs for Real-Time ASL Gesture Recognition, *arXiv preprint arXiv 2305.05296*, (2023).