

骨格推定に基づくダーツトレーニングシステムに関する研究 A Study on a Dart Trainer System Based on Human Pose Estimation

吉田 理音[†] 樽松 理樹[†]
Rion Yoshida Masaki Kurematsu

1. 研究の背景

スポーツトレーニングにおいて正しい動きを身につけることは、スポーツのパフォーマンス向上や怪我の予防に不可欠である。正しい動きを身につけるには知識や経験があるスポーツ指導者から学ぶのが望ましい。しかし、優れた指導者を見つけるのは簡単ではない。公益財団法人日本スポーツ協会による学校運動部活動指導者の実態に関する調査[1]では、学校運動部活動の指導者が、担当教科が保健体育でないかつ現在担当している部活動の経験なしの割合が40%を超えている。また、日本スポーツ協会による公認スポーツ指導者の登録状況調査[2]では、マイナースポーツにおける指導者数はサッカーなどの有名なスポーツにおける指導者数に比べて少ない。スポーツの指導者不足はスポーツ文化の衰退につながり、大きな課題である。

このようなマイナースポーツの1つとしてダーツがある。ダーツは2人以上のプレイヤーがダーツボードに向かってダーツを投げる競技であり、体格や年齢、性別を問わず誰でも楽しめるスポーツである。しかし、競技者数や指導者数が少なく練習方法は、書籍や動画を使った独学になることが多く、指導者からの指導を受けにくいのが現状である。

以上のような現状の解決策として、近年盛んになっているスポーツにおけるAIによるトレーニング支援が挙げられる。これにより、VRや姿勢推定技術のツールを用いることで選手の動きを分析し、効率的なフォーム指導の実現が期待できる。

このようなダーツのトレーニングシステムに関してはすでに研究が取り組まれている。永田[3]の研究では、姿勢推定技術を用いて被験者の投擲動作を可視化し、骨格点の位置の違いについて指摘している。しかし、このシステムでは、ダーツは身長によるリリースの高さがスコアに影響するため、被験者の特徴を考慮していない。また、柴尾[4]の研究では、VR技術を用いたダーツのトレーニングシステムの提案をしている。VR空間内での手本の観察や、ユーザのフォームを手本と定量的に比較することによって運動パターンの習得を目的としている。しかし、器具が高価なため簡便性に欠ける。

以上の背景から、本研究ではダーツにおける競技者へのフォーム矯正・上達支援を目的とし、体格を考慮したダーツトレーニングシステムの開発を行う。

2. 関連研究

永田の研究では、提示した手本とユーザの動きを重ねて表示、リアルタイムのフィードバック、視覚による教師信号の提示、力覚による教師信号の提示を踏まえたダーツにおける初心者へのスキル支援システムの構築を提案している。そのための予備実験として、Kinectの姿勢推定を用いて被験者の投擲フォームを計測している。この実験では同

じ動作を行おうとしても、骨格点の位置の違いが大きいことが示されている。

柴尾の研究では、近年スポーツの技能向上において注目されているVR技術を用いたダーツのトレーニングシステムを提案している。具体的にはVR空間内での手本の観察や、ユーザのフォームを手本と定量的に比較することによって運動パターンの習得を目的としている。予備実験として、システムを使用して練習する場合とシステムを使用せずに練習する場合とを比較し、より正確な技術の獲得が可能であるかを検証している。検証の結果、提案システムを使用したグループは、トレーニング後の方がトレーニング前よりダーツを中心付近に投擲できたことを示している。

Yamaguchi[5]の研究では、ダーツにおけるリリースポイントに着目し、動作軌道、筋活動、投擲精度について検証を行っている。検証の結果から、熟練者は時間的変動が小さくリリースのタイミングが固定されており、リリースのタイミングは肘の速度が最大になる直前であることを示している。

3. 提案手法

3.1 提案手法の概要

図1に本提案手法の全体図を示す。本提案手法では、最初にユーザの動作を撮影した動画からユーザの骨格情報を抽出する。次に前処理として、動作主体の抽出、欠損値補完、相対変換、標準正規化、ゼロパディングを行う。続いて前処理したデータに3.3節で述べる予測モデルを用いて分類を行い、分類されたラベルから教師動画選択を行う。最後に選択された教師動画に基づく骨格情報とユーザの骨格情報とを比較し、アドバイスを生成する。

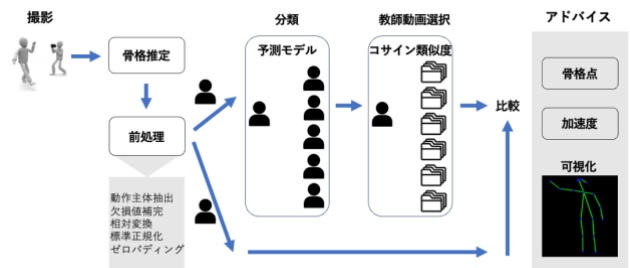


図1 提案手法の全体図

3.2 骨格推定

本研究では、骨格推定にHao-Shu Fangら[6]により開発されたAlphaPoseを用いる。AlphaPoseは画像や動画中の人物の骨格座標を推定することができる。骨格座標データは人物の手首、肩、肘といった合計18種類の骨格座標が出力される。出力形式は(x座標, y座標, 信頼値)である。

本研究では、予測モデルに適切な値を入力させる為、前処理を行う。

[†]岩手県立大学 Iwate Prefectural University

初めに AlphaPose より出力された骨格座標から動作主体の骨格座標を取得する。1 枚の画像に複数人物が写っている場合、AlphaPose は複数人の骨格推定をする為、動作主体のみの骨格情報を抽出する。

次に、欠損値補完を行う。激しい動作などに骨格推定モデルが骨格情報を認識漏れする可能性がある為、信頼度が 0.3 以下の場合に欠損区間の直前の推論結果をそのまま欠損区間に代入する方法を用いる。

最後に、相対座標への変換、標準正規化、ゼロパディングを行う。

3.3 予測モデル

本研究では、予測モデルの構築に LSTM(Long Short Term Memory)[7]を用いる。LSTM とは過去の情報を利用して現在および将来の入力に対するネットワークの性能を向上させるディープラーニングの構造である RNN (Recurrent Neural Network)[8]の 1 種である。RNN の特徴は、ネットワークに隠れ状態とループが含まれている点であり、ネットワークにループ構造を用いると、過去の情報を隠れ状態で保存し、シーケンスで処理することができる。しかし、RNN では、より長期の依存性の学習において問題が生じる。RNN は、通常、逆伝播により学習し、その際に勾配が消失する問題が生じることがある。この問題を改善した手法が、LSTM ネットワークである。LSTM ネットワークは、ゲートを増やすことで、隠れセル内のどの情報を出力や次の隠れ状態に利用するかを制御する。これにより、データに含まれる長期の関係を、ネットワークがより効果的に学習できるようになる。そのため、本研究では、3.2 節で述べた骨格推定結果をもとに骨格座標を列に、時間経過を行って持つ時系列データを構築し、LSTM の学習に用いる。

3.4 教師動画選択

予測モデルで分類されたラベルからユーザに最も近い動作を抽出する為にコサイン類似度(Cosine Similarity)を用いる。コサイン類似度とは、ベクトル空間における 2 つのベクトルの類似度を測定するための指標である。本提案手法においては、各時点における肘のベクトルのコサイン類似度を、肘から手首へのベクトル x 、肘から肩へのベクトル y の 2 つのベクトルコサイン類似度の計算式を式(1)で求める。

$$\cos \theta = \frac{\langle x, y \rangle}{\|x\| \|y\|} \quad (1)$$

ユーザの各時点での肘のコサイン類似度と、各動画の肘のコサイン類似度の差の合計が最も小さい動画を、教師動画として選択する。

3.5 アドバイス生成

本研究では、次の 3 つの視点からユーザに助言する。

1 つ目は骨格点の差である。各骨格点のユーグリッド距離を求め、それを教師動作と比較する。閾値以上であれば骨格点を表示する。

2 つ目は加速度である。手首の加速度を求めフレーム間の合計平均加速度を表示する。合計平均加速度とは、各フレームの加速度を求め、加速度の合計をフレーム数で割った値である。この値を教師動作の合計平均加速度と比較し、

閾値より速い場合は“速い”，遅い場合は“遅い”，閾値の範囲内であれば“適切”と表示する。

3 つ目は動作の可視化である。ユーザの骨格点を基に骨格線で表示し、ユーザに投擲動作を確認してもらう。可視化の生成例を図 2 に示す。図 2 において、点は骨格推定に基づいて取得された骨格座標から得られる骨格点を示し、線はその骨格点を結んだ骨格線を示している。

これらのアドバイスを通し、ユーザの技術向上を目指す。

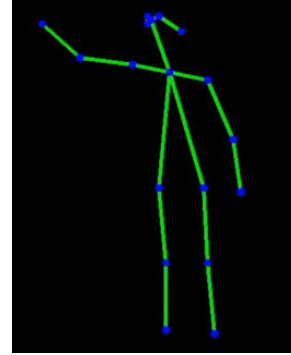


図 2 動作の可視化

4. 予備実験

本研究で提案した手法の有用性を検証する為に予備実験を実施した。

4.1 動画撮影

予備実験では、ソフトダーツのルールに従い、ダーツボードの中心から床までを 173cm、ダーツボード表面真下からスローライン先端までの距離を 244cm とした。撮影位置はダーツボードに向かってダーツを投げる実験参加者の視点から見て、ダーツボードの左側 45 度の場所から撮影した。撮影設定を図 3 に示す。

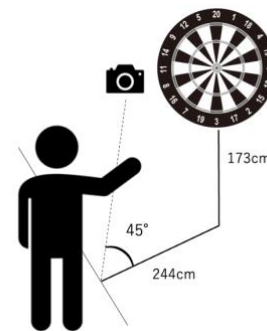


図 3 撮影設定

4.2 取り扱う骨格について

骨格推定で取り扱う骨格点は、以下の通りである。

鼻、首、右肩、右肘、右手首、左肩、左肘、左手首、右腰、左腰の計 10 点とし、動作の可視化は上半身のみで行う。

4.3 データセット

本研究では、YouTube に投稿されている 5 人のプロ選手の投擲動作動画を予測モデルの構築に使用した。動作開始をリリース開始時点とし、動作終了を投げ放った後に腕をまっすぐ伸ばしている状態とする。5 人の身長はそれぞれ

150cm, 160cm, 170cm, 180cm, 200cm である。学習データは 1 選手あたり 70 回の投擲動作, 合計 350 回を取得した。これらの動画から得た, 骨格推定から得られた 0.076 秒ごとの骨格キーポイントの座標と加速度を説明変数, 人物を目的変数とし, 学習を実施した。学習においては, 訓練データとテストデータに 8 対 2 に分け, 5 回の交差検証を行った。

4.4 予測モデル

LSTM の構造は入力層 128 次元, 中間層は 64 次元と 32 次元の 2 層, 出力層が 5 次元の 4 層構造である。入力層と出力層の間には 3 つのドロップアウト層を追加し, 前半の 2 つの中間層で正則化を行った。バッチサイズは 64 とし, 損失関数には交差エントロピー誤差, 活性化関数には softmax を用いた。最適化アルゴリズムには学習率を 0.01 に設定した Adam を用い, エポック数を 300 とした。

本研究では, Python 3.10.4, pandas 2.0.0, scikit-learn 1.2.2, および TensorFlow 2.13.0 を使用してモデルを構築した。

4.5 アドバイス生成の閾値

今回の実験において, 3.5 節で述べたアドバイス生成の閾値は, 下記の通りとした。

骨格の差については, 2.0 以上であれば, 骨格点の表示を行う。

加速度については, 教師動作との差分が 0.05 以上の場合には“速い”, -0.05 以下の場合には“遅い”, それ以外は“適切”とした。

4.6 評価方法

実験参加者は身長 174cm の男性 1 名とし, 投擲回数を 10 回, 1 回の投擲終了ごとにアドバイスを行う。得点の付け方を図 4 に示す。赤色が 3 ポイント, 黄色が 2 ポイント, 白色が 1 ポイント, 枠外は 0 ポイントである。評価方法は, 1 回の投擲について, トレーニング前後で得点が向上したかどうかを確認する。

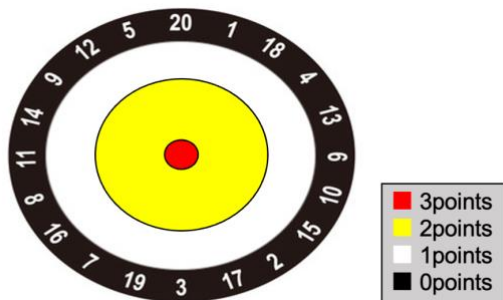


図 4 得点の付け方

4.7 実験結果

予測モデルの評価精度を表 1 に示す。表 1 では, 学習させた予測モデルに対してテストデータを用いて評価した際の指標を示している。Accuracy は式(2), Precision は式(3), Recall は式(4), F1-score は式(5)でそれぞれ求める。ここで, T_p , T_n , F_p , F_n はそれぞれ真陽性, 真陰性, 偽陽性, 偽陰性である。真陽性は正と予測し正であること, 真陰性は正と予測し偽であること, 偽陽性は偽と予測し正であること, 偽陰性は偽と予測し偽であることである。

表 1 予測モデルの評価精度

評価指標 (Evaluation Metrics)	値 (Value)
Accuracy	95.7%
Precision	95.9%
Recall	95.7%
F1-score	95.6%

$$\text{Accuracy} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad (2)$$

$$\text{Precision} = \frac{T_p}{T_p + F_p} \quad (3)$$

$$\text{Recall} = \frac{T_p}{T_p + F_n} \quad (4)$$

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

上記の予測モデルを用いて実験参加者の投擲動作を分類した結果, 実験参加者に身長が最も近い人物を抽出した。

実験参加者の得点と加速度の評価結果を表 2 に示す。得点は 4.6 節で述べた基準に従う得点, 加速度は 3.5 節で述べた実験参加者の手首の合計平均加速度及びコメントである。得点に関して, 前半には向上が見られるが, 後半では下降している。加速度に関して, 徐々に加速度が教師動作に近づいている。またアドバイスに関して, 右肘と首の骨格点に対するアドバイスが最も多かった。

表 2 得点と加速度の評価結果

回数	得点	合計平均加速度	加速度
1	2	0.24(+0.07)	速い
2	0	0.21(+0.04)	適切
3	1	0.19(+0.02)	適切
4	2	0.20(+0.03)	適切
5	2	0.18(+0.01)	適切
6	0	0.15(-0.02)	適切
7	1	0.14(-0.03)	適切
8	2	0.19(+0.02)	適切
9	0	0.20(+0.03)	適切
10	1	0.14(-0.03)	適切

5. 予備実験に基づく評価

5.1 効果がある点

予備実験の実験結果を踏まえ, 効果があると評価した点としては, 次の点が挙げられる。

効果があると判断した点として, 1 つ目は予測モデルの分類結果から実験参加者と体格が最も近い人物に分類された点である。これは体格によってフォームが異なる為であると考えられる。

2 つ目は合計平均加速度及びコメントの表示である。実験結果から, 投擲動作の速さが徐々に改善していることが示された。これは合計平均加速度を表示することで, 実験参加者は投擲動作の速さを調整することができたと考えられる。

5.2 改善点

1 つ目の改善点としては骨格点の指摘である。ユーグリッド距離が大きい骨格点を表示しても、実験参加者は改善方法が分からず必要な改善を行うことができなかった為、具体的なアドバイスが必要である。

2 つ目の改善点は動作の可視化である。実験参加者から見て投擲動作が分かりづらいことが指摘された。これは 2 次元表記の骨格線の表示のみであり、立体的な可視化がされていないこと、上半身のみの可視化であることが原因だと考えられる。

3 つ目の改善点は、コメントやアドバイスの柔軟性不足である。現在あるデータのみからコメントなどを構築するため、制限がある。これはデータ数とラベル数を追加することでより柔軟に分類できるシステムになると考える。

5.3 改善案

5.2 節で挙げた改善点に対し、次のような改善案を検討している。

現在、ユーザの骨格点を基に骨格線で表示していたが、今後は人体ベースの 3D モデルを提示する。現在、PyMAF[9]や ROMP[10]を用い、動画からメッシュモデルの作成を試みている。ROMP での投擲動作の可視化例を図 5 に、Blender[11]の可視化例を図 6 に示す。3D モデルを用いることで、2 次元の骨格線の表示に比べて視覚的な理解が容易になり、またオープンソースで開発されている 3D コンピュータグラフィックスソフトウェアである Blender を用いることで様々な角度からの可視化が可能である。しかし、これらの手法では、ダーツのリリース動作時の手首などの角度を正確に再現できず、正常に手を検出することができていない。



図 5 ROMP の可視化例

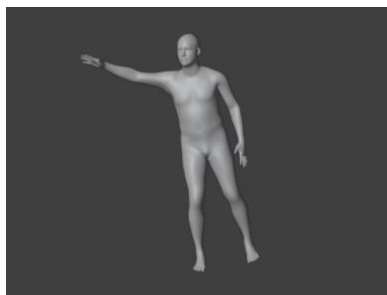


図 6 Blender の可視化例

この問題を改善するために、現在 Diogo Luvizon[12]らが実証した Scene-Aware 3D Multi-Human を用いている。Scene-Aware 3D Multi-Human は静止カメラで録画された単一の RGB ビデオから、シーン内の複数の人間の 3D 位置、および体の形状と関節を推定し、SMPL モデルとして出力

が可能である。この技術を用いて投擲動作の可視化を試みる。

また、予備実験では 1 人のみを対象に検証を行ったが、システムの評価をより充実させるために、今後は 3 人以上を対象に評価実験を行う。評価方法に関しても、柴尾の評価方法を参考に、1 回の投擲ではなく、複数回の投擲結果がトレーニング前後でどれだけ向上したかを検証する。

6. おわりに

本研究ではダーツにおける競技者へのフォーム矯正・上達支援を目的とし、体格を考慮したダーツトレーニングシステムの提案を行った。予備実験では、予測モデルが実験参加者と体格が最も近い人物に分類されることで実験参加者の特徴を考慮している点と、加速度の表示により実験参加者は投擲速度を調整することができ、教師動作に近づけることができた。しかし、実験参加者が教師動作との違いを比較するのが困難であった為、アドバイスの効果が十分に発揮されなかったことが課題である。この課題に対して、動作の可視化に着目し、立体的なモデルで動作を可視化する。今後の予定として、アドバイスの改善とデータ量の追加を行い、評価実験を実施し、本システムの向上を図る。

謝辞

本研究で使用した実験データは岩手県立大学の学生の協力を得て収集したものである。ここに謝意を表す。

参考文献

- [1] 学校運動部活動指導者の実態に関する調査:https://www.japan-sports.or.jp/Portals/0/data/katsudouuisishin/doc/R3_houkokusho_hika_ku.pdf(最終アクセス: 2024 年 4 月 16 日)
- [2] 日本スポーツ協会公認スポーツ指導者登録状況:https://www.japan-sports.or.jp/Portals/0/data/katsudouuisishin/doc/20221001_tourokusha_events.pdf(最終アクセス: 2024 年 4 月 16 日)
- [3] 永田 雄樹: ダーツを例とした初心者への運動スキル支援システム, 28th Fuzzy System Symposium (Nagoya, September 12-14, 2012)
- [4] 柴尾 啓太: VR 技術を用いたダーツにおけるスローイング動作のトレーニングシステムの開発, 情報処理学会インタラクシオン 2024, 2A-02
- [5] Hiroshi Yamaguchi: Throwing Darts Training Support System Based on Analysis of Human Motor Skill, Intelligent Autonomous Systems 12 pp 469-478
- [6] Hao-Shu Fang, Jiefeng Li, Hongyang Tang, Chao Xu, Haoyi Zhu, Yuliang Xiu, Yong-Lu Li, Cewu Lu: "AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time", arXiv:2211.03375.
- [7] Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, Neural Computation, 9(8), pp.1735-1780, 1997.
- [8] David E. Rumelhart, Geoffrey E. Hinton, Ronald J. Williams: Learning representations by back-propagating errors, Nature, Volume 323, Issue 6088, pp. 533-536 (1986).
- [9] Zhang, H., Tian, Y., Zhou, X., Ouyang, W., Liu, Y., Wang, L., & Sun, Z. (2021). PyMAF: 3D Human Pose and Shape Regression with Pyramidal Mesh Alignment Feedback Loop. ICCV 2021, Oral paper. arXiv:2103.16507.
- [10] Sun, Y., Bao, Q., Liu, W., Fu, Y., Black, M. J., & Mei, T. (2021). Monocular, One-stage, Regression of Multiple 3D People. ICCV 2021. arXiv:2008.12272.
- [11] Blender Foundation. 2024. Blender. バージョン 4.1.:<https://www.blender.org>. (最終アクセス: 2024 年 4 月 30 日)
- [12] Luvizon, D., Habermann, M., Golyanik, V., Kortylewski, A., & Theobalt, C. (2023). Scene-Aware 3D Multi-Human Motion Capture from a Single Camera. Eurographics 2023. arXiv:2301.05175.