

人体姿勢推定技術を用いたカーリング動作解析の一考察

Consideration of Curling Motion Analysis Using Human Pose Estimation Technology

黒田 英慈¹ 林 将寛¹ 榎井 文人² 柳 等²

Eiji Kuroda¹, Masahiro Hayashi¹, Fumito Masui², and Hitoshi Yanagi²

¹株式会社アイエンター

ai-enter Corporation, Ltd.

²北見工業大学冬季スポーツ科学研究推進センター

² Research Center for Winter Sports Science, Kitami Institute of Technology

Abstract: We aim to develop a system for judging the delivery form of curling using human pose estimation technology in the image processing field. As a preliminary study, we investigated the feasibility of 3D human pose estimation from camera images of delivery form.

1. はじめに

カーリングは、冬季スポーツの1つであり、ストーンと呼ばれるハンドルの付いた石を数メートル離れた同心円上の的(ハウス)に目掛けて投球を行うスポーツである。またカーリングは、氷上のチェスとも呼ばれるように、戦術が勝敗に関わるスポーツである。戦術を遂行するためには、ストーンを狙った場所に投石しなければならない。カーリングの投石はデリバリーと呼ばれており、陸上競技のスターティングブロックのような足掛け(ハック)を蹴ることで勢いをつけて氷上を滑り、ストーンに回転をつけて石を離す一連の動作。デリバリー動作においては、正確な投石が求められるため、投石精度を高めるためのトレーニングが重要である。

しかしながら、従来のデリバリートレーニングは目視による確認が一般的であり、経験の浅い競技者には判断が難しかったり、目視判断の結果の一般化が難しいといった問題があった。

一方、近年野球やサッカーといったスポーツでは、フォームをAIで分析フィードバックするようなコーチング方法も取り入れつつある。このような取り組みを行うことにより、人材不足の改善や競技力の向上につながる。

AIによるフォームの分析では、主に(1)センシング機器での取得や、(2)画像解析による方法が考えられる。(1)は精度が良いものの、導入にコストがかかることや競技者にセンシング機器を取り付けることでプレーに支障が出る可能性がある。

一方、(2)は、精度がセンシング機器に劣るものの、低コストであることやカメラ撮影のみで処理可

能であることから、通常のパフォーマンスでプレーを行うことができる。

画像解析による人体の骨格情報解析方法として、姿勢推定がある。この技術を用いることにより、カメラ映像や画像から被写体となる人物の骨格情報の取得が可能となり、効率良く動作解析が行える。

我々は、カーリングのデリバリー動作に着目し、被写体となるカーリング競技者を撮影して、その選手のデリバリー動作の良し悪しを判定するシステムの開発を目指している。本稿では、事前研究として、姿勢推定技術を用いてカーリング競技者の骨格位置の取得が可能であるか、動作解析が可能であるかどうかの実現性調査を行ったので報告する。

2. 関連研究

本章では、関連研究としてスポーツの動作解析に関して、センシング機器や姿勢推定を用いた例について述べる。

桂ら[1]は、ボルダリングを対象に姿勢推定ライブラリを用いて被写体がクライミングウォールを登り終えたどうかの完登判定の研究を行っている。クライミングウォール上のホールド(手をかける石)を配列に登録し、ゴールとなるホールドに推定された骨格が範囲内に入ることによって判定している。研究の課題として、手首をゴール判定のトリガーとしているため、ホールドに手が付いている状態でも手首の座標がゴールの範囲内に入らず失敗判定となることや、胴体などで対象となる骨格が遮断され推定が行えず失敗することあったと言及している。中村ら[2]は、サッカーのシュート動作を対象として、姿勢

推定の精度向上について研究をしている。サッカーのシュート動画に対して一度姿勢推定を行い、目視で推定結果から失敗した動画フレーム画像を抽出して、抽出した画像に正解データを付与して、再度学習を行うことで、精度向上を目指している。失敗フレームには、やはり正解を付与したい骨格が隠れていることや、服の色などで付与が難しいといったことを言及している。上記に2件について、姿勢推定を行う上で対象となるスポーツの特色を把握した上で学習データやそのスポーツに特化した学習モデルの生成が必要であることが窺える。

スポーツの動作解析においては、その競技の未経験者、経験者に分類して分析されることがしばしばある。清水ら[3]は可視光センシング機器を用いてバドミントンのスマッシュの解析を行っており、佐藤ら[4]も、可視光のセンシング機器を用いて、サッカーのリフティング動作についての研究を行っている。上記のような研究では、そのスポーツの動作で気を付けることが明確になるため、カーリングのデリバリ動作の動作解析でも応用したい。

一方カーリングのデリバリ動作について、姿勢推定を用いた動作解析の研究は事例が見当たらず、新規性があると言える。

3. 人体姿勢推定について

本章では、画像解析分野の人体姿勢推定について概略を述べる。なお具体的なアルゴリズムや考え方が知りたい場合には、こちらのサーベイ論文を参照されたい[5]。

人体姿勢推定とは、画像解析の一分野であり、画像上の被写体の肩や肘といった骨格点を AI で認識する技術である。姿勢推定には画像上の被写体の骨格を画像上の座標位置で推定する 2D 姿勢推定と、3次元空間上での座標位置を推定する 3D 姿勢推定がある。

2D 姿勢推定には、Top-Down アプローチと、Bottom-Up アプローチの2つの手法が存在する。前者は、姿勢推定を行う前処理として画像中の人物の位置を特定するために物体検出を行う必要がある。物体検出後、人物のみが写った画像を姿勢推定 AI に入力することで被写体の画像中の骨格座標を得ることができる。後者は、物体検出で画像中から被写体の位置を取得する前処理が不要であり、画像中に存在するそれぞれの肩や肘、膝といった骨格の座標を推定し、その後それぞれの骨格位置から人物の骨格を再構成することで全身の骨格情報が得られる。両者を比較すると、Top-Down アプローチは、精度が良いものの、画像上の人物位置の特定が必要であり、

Bottom-up アプローチは、ワンショットでの推定や、複数人推定に強い一方、Top-Down アプローチに比べ精度が落ちる。

3D 姿勢推定にも、大まかに2種類の推定手法が存在する。前者は、予め推定を行った 2D 骨格座標を用いて 3D 骨格を推定する手法で、前処理が必要となるが、精度自体は良いとされている。後者は、画像上を直接入力して 3D 骨格を推定する手法で、ワンショットで推定できるが、2D 骨格から 3D 骨格を推定する手法に比べ精度が低いとされる。

本研究においては、精度を重視するため、2D 姿勢推定では、Top-Down アプローチの human-pose-estimation.pytorch というライブラリ[6, 7]と、3D 姿勢推定では、2D 骨格から 3D 骨格の推定を行う GAST-Net というライブラリ[8, 9]を選択した。

4. 実験

本章では、姿勢推定を用いてカーリングのデリバリ動作について骨格情報の取得が可能であるかどうかの実現性調査のため、いくつかの学習データを用意し、学習を行い予め撮影したデリバリ動画に対して精度向上が見込めるか確認を行う。

4.1 多方向からの姿勢推定アプローチ

一般的な姿勢推定においては、単眼カメラで撮影した RGB 画像、映像に対して被写体の骨格推定を行う。しかし、被写体の映る角度によっては体の一部が死角となり精度低下を招く可能性がある。

この点の解決策として、3D の姿勢推定について2方向からの撮影を行い、2つの骨格情報を入力して1つの 3D 姿勢推定結果を出力する新たなネットワークを検討した。図1に処理の流れのイメージ図を示す。これが実現できれば、姿勢推定で2台のカメラが必要となるコストは発生するが、死角がなくなることにより精度が向上すると仮定している。

以後上述したネットワークを多方向姿勢推定と呼称する。

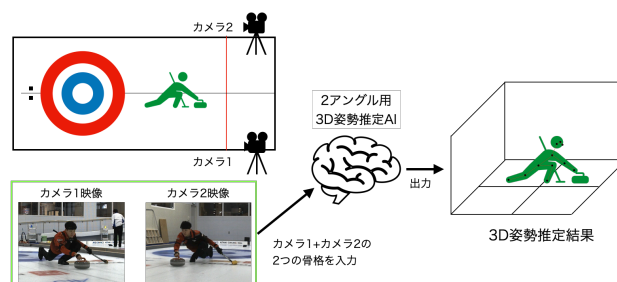


図1：多方向姿勢推定のイメージ

4.2 取り扱う骨格について

姿勢推定で取り扱う骨格は、以下の通りである。2D は一般的な姿勢推定データセットである coco-dataset[10]形式に倣い、鼻、左目、右目、左耳、右耳、左肩、右肩、左肘、右肘、左手首、右手首、左腰、右腰、左膝、右膝、左足首、右足首の計 17 骨格を扱う。3D もまた、一般的な 3D 姿勢推定データセットである human-3.6,m[11]形式に倣い、腰の中心、右腰、右膝、左足首、右腰、左膝、左足首、背骨、胸骨、首、鼻、頭頂部、右肩、左肘、左手首、左肩、右肘、右手首の計 17 骨格を取り扱う。

4.3 評価指標

本節では 2D、3D 姿勢推定の精度についてそれぞれ記述する。

2D 姿勢推定の精度評価には、姿勢推定分野で一般的に用いられる PCK (Percentage of Correct Keypoints) を選択した。PCK は、推定された骨格と正解骨格との差を算出し、任意の閾値以内であれば、正しく推定が行えているとする手法である。本実験では、人体矩形の対角線を 0.05 倍した値をシビアな閾値 (@0.05) とし、0.1 倍した値をルーズな閾値 (@0.1) とした。

3D 姿勢推定については、LIU ら[8]の手法に従い、P-MPJPE (Procrustes analysis MPJPE) で評価を行う。元となる MPJPE (Mean Per Joint Position Error) は 2D の評価と同様に推定結果と正解データの差の平均値を算出するが、3D の推定では、正解データ必ずしも出力結果の角度や出力座標が一致するとは限らないため、回転角や並行移動などが考慮された P-MPJPE を用いる。算出される数値は、精度が高いほど 0 に近い値となる。

4.5 実験設定

本節では 2D と 3D の姿勢推定の学習を行う上で用意したデータについて述べる。

2D 姿勢推定用学習データ

2D の学習データは、インターネット上からダウンロードしたデリバリー画像と、インターネット上とそれから実際に撮影を行った動画をフレームに分割した画像、計 2707 枚用意した。上記画像に対して、人手で骨格情報を付与するアノテーション作業を行い、学習用 2437 枚、検証用 270 枚として学習させた。また coco データセットを学習した事前学習済みモデルをベースラインとして比較対象とした。

3D 姿勢推定用学習データ

3D 姿勢推定の学習データを表 1, 2 に示す。

データセット 1 は、3D 姿勢推定分野で一般的に用いられる Human3.6m データセットを学習させたベースラインとなる学習モデルである。データセット 2, 3 は Noitom 社のパーセプションニューロンプロ [12] というセンシング機器を用いて収集する。パーセプションニューロンは、無線式のモーションキャプチャーであり、15 個のセンサーを体に装着することで、3D のモーションデータを取得することが可能となる。データセット 2 は、パーセプションニューロンを用いて実際に氷上でデリバリー動作を行った際に得られた 3D のモーションデータを元に GAST-Net 用に整形したデータで学習を行ったモデルである。データセット 3 については、データセット 2 と同様にパーセプションニューロンを用いて取得したモーションデータであるが、こちらは氷上での動作ではなく陸上でデリバリーの動作を疑似的(素振り)に再現した動作である。

上記 3 つのデータセットをそれぞれ学習させて生成した学習モデル A~C と、多角方向 GAST-Net 用学習モデル D~F に対して、事前に撮影した動画で性能の評価を行った。

表 1 : 3D 姿勢推定の学習データ一覧

学習モデル名	データセット概要
学習モデル A	データセット 1
学習モデル B	データセット 2
学習モデル C	データセット 3

表 2 : 多方向姿勢推定の学習データ一覧

学習モデル名	データセット概要
学習モデル D	データセット 1
学習モデル E	データセット 2
学習モデル F	データセット 3

※

- データセット 1 : ベースライン
(一般的な 3D 姿勢推定データセット, Human 3.6m)
- データセット 2 : パーセプションニューロンによる、
デリバリーのデータセット(299 ショット分)
- データセット 3 : パーセプションニューロンによる、
疑似的なデリバリーのデータセット(177 ショット分)

5. 結果

本章では、上述した実験の結果について述べる。2D 姿勢推定の精度検証の結果を表 3 に、推定され

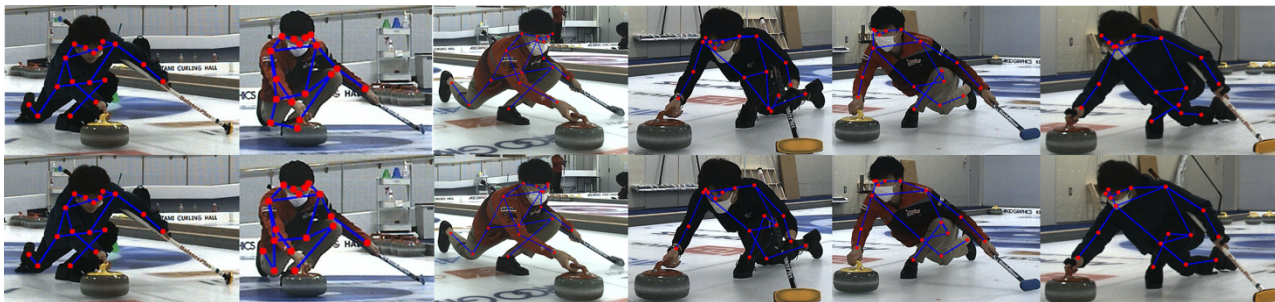


図 2：学習モデル別 2D 姿勢推定結果例

画像上部は、事前学習済みモデルの推定結果、画像下部は、カーリング画像で学習を行ったモデルの推定結果。足以外は精度に大きな差は無いが、足の推定結果は大きく異なり、事前学習済みモデルは推論が困難であるが、カーリングモデルではデリバリー動画を学習したことにより正確に推論が行えている。

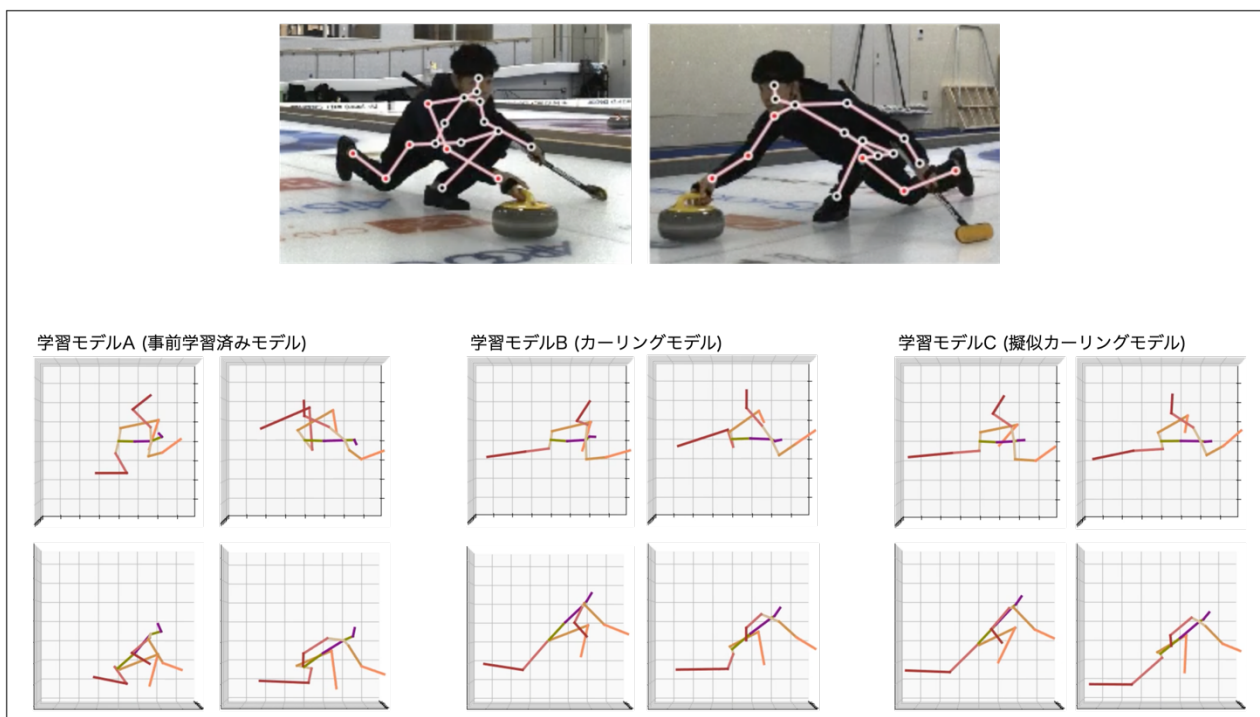


図 3：学習モデル別 3D 姿勢推定結果例

画像上部は精度検証用の動画の 1 フレームを抜き出し、2D 姿勢推定結果をそれぞれの学習モデルに入力し、出力された 3D 骨格座標を画像下部に示している。画像下部の 3 つのモデルに対して、上部が上からのアングル、下が真横からのアングルとなっている。また、ブロック左側が画像上部左の画像、ブロック右側が画像上部右の画像に対応する

た画像の例を図 2 に示す。

まず、デリバリー画像を学習したカーリングモデルと事前学習済みモデルの平均精度を比較すると、@0.05 は 15.7%、@0.1 では 8.1% と共に精度が向上していることが分かる。特に右足、左足に関しては、大きく精度が向上していることが確認できる。図 2 上部の推定例を確認すると、事前学習済みモデルでは両足の推定がうまく行えておらず、カーリングモデルでは正しい場所にプロットされている。カーリングのような一般的なデータセットに内包されない

特殊な動作については、新たに学習を行うことで姿勢推定が可能になることが示唆される。

続いて 3D 姿勢推定の精度検証の結果を表 4 に、推定された画像の例を図 3, 4 に示す。3D の姿勢推定については、ベースラインである学習モデル A の P-MPJPE が 0.1549 に対して、学習モデル B は 0.0906、学習モデル C は 0.0894 となり、数値では擬似的なデリバリー動作を学習させた学習モデル C が最も高性能となった。図 3 を確認すると、学習モデル A については、デリバリフォームの特徴とも言える右足が全

表 3：2D 姿勢推定の学習結果の骨格別精度評価

学習モデル	顔 (%)	左腕 (%)	右腕 (%)	左足 (%)	右足 (%)	平均 (%)
事前学習済み モデル @0.05	89.5	72.2	74.7	45.6	38.7	64.1
事前学習済み モデル @0.1	94.9	88.8	89.4	69.7	67.9	82.2
カーリング モデル @0.05	89.4	79.6	79.8	76.0	74.4	79.8
カーリング モデル @0.1	92.2	90.1	90.9	89.6	88.8	90.3

表 4：3D 姿勢推定のモデル別 P-MPJPE 値

学習モデル	カメラ 1	カメラ 2
学習モデル A	0.1549	0.1794
学習モデル B	0.0906	0.0838
学習モデル C	0.0894	0.0791
学習モデル D		0.2551
学習モデル E		0.0746
学習モデル F		0.0753

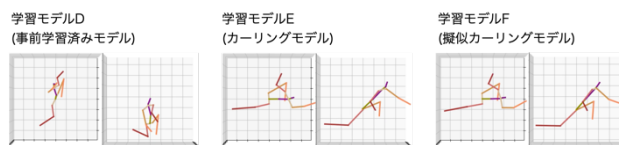


図 4：多方向姿勢推定の結果

く伸びていない。これは、2D の事前学習済みモデルと同様の問題として考えられる。学習モデル B, C については、右足が伸びており学習が成功していると判断できるが、B のカメラ 2 の右足においては想定し得ない方向に存在した。

多角の推定についても、同様に学習モデル D よりデリバリーモーションを学習させたモデル E, F の方が精度が良い結果となった。総合的に判断すると、学習モデル E, F についていずれも単眼のモデルより P-MPJPE 値が低くなり、多角での学習の有用性が証明できた。また、カーリングのデリバリー動作については、学習を行うことで姿勢推定を行うが確認できた。これに加え、仮にその他のスポーツの姿勢推定を行うことを考慮すると、カーリングと同様に専用の学習モデルを作成する必要があることが示唆された。

6. 考察

本章では、実験に対する考察と、本実験を踏まえ

た上でのデリバリーの動作解析についての検討を行う。

まず、学習モデル B と C について、氷上での動作と擬似的な動作の学習結果に差異が生じた点について考察を行う。本実験ではデータ収集にパーセプションニューロンを用いていたが、この機器は電波や、機器による電波干渉を受けやすい性質があり、またデリバリー動作は移動を伴う動作であるため、少なからずモーションデータに悪影響を及ぼしてしまう。対して、擬似的な動作については、移動を伴わずに動作を収集できるため、電波影響を受けづらいと言える。上記が影響したため、単眼の 3D 姿勢推定モデルでは実際のデリバリー動作を用いた学習モデル B より擬似的な動作を用いた C の精度が良くなったと考えられる。また、多方向姿勢推定においては、モデル E と F で同等の推定精度となったため、場所や時間に囚われずデータ収集が行えることが可能となり、素振りといった類似したデータでの学習の有用性も示唆された。さらに、より高精度なモデルを作成するためには、コストがかかるが無線式のモーションキャプチャではなく光学式のものを用いてモーションを取得して学習を行うことで、さらに精度が向上するのではないかと考えている。

本実験では、3D 姿勢推定の精度指標として P-MPJPE を用いたが一般的な姿勢推定ではなくデリバリー動作に特化した姿勢推定を行う上では、指標として不十分な可能性がある。デリバリー動作は一般的にストーンを持つ手と、左足の爪先か踵、伸ばした脚が進行方向に対して真っ直ぐな状態が正しい姿勢とされている。一方本実験の推定結果を確認すると、画像では真ん中の線に従ってストーンと両足が一直線に伸びているが、推定結果は上記 3 点が一直線上に存在しない。そのため、現状の学習モデルの状態では十分な精度があるとは言えない。他にもカーリングにおけるフォームの特性が存在する可能性があるため、その点の調査を行い、精度評価の検討や学習損失として用いることで精度の向上が期待できる。

次に動作解析についての検討を行う。

本実験の評価指標として用いた P-MPJPE は本研究の最終目的である良し悪し判定にも適用できるのではないかと考えられる。例えば、初級中級者であれば、目標としている選手の動作を正解データとして自分の推定結果との P-MPJPE を算出することで、どのくらい似ているフォームを行えているが数値として確認することができる。また、上級者であれば、自分の中で一番良いプレーを正解データとして姿勢推定を行えばその日のプレーの良し悪しが判断できると考えられる。さらに、P-MPJPE を時系列とする

ことで時間軸上での動作の良し悪しの判断も可能となる。上記を実現するためには、デリバリー姿勢推定の精度の更なる向上はもちろんであるが、他選手との姿勢推定結果との比較の際は、デリバリーの時間が異なるため、時系列を揃える手段を検討が必要になる。

7. 画像解析システム

本章では、2020年10月にオープンしたアルゴグラフィックス北見カーリングホールに導入された画像解析システムについて述べる。

画像解析システムでは、2方向のカメラを用いてカーリング競技者の動作を推定するシステムである。クライアントPCで2方向のデリバリー動画を撮影し、姿勢推定サーバへ送信。サーバからの解析結果をクライアントPCで受け取り結果を可視化できる。可視化できる内容として、2Dの姿勢推定結果と、3Dの姿勢推定結果については、360°から骨格の確認が可能である。また、それぞれの推定された骨格の時系列データをグラフとして出力しているため、不自然な動作等あれば動画と照らし合わせいろいろ確認できる。さらに、2つのデリバリー動画の比較を行うことができ、3D骨格を用いて比較が行える。

現時点(2021年10月23日)では、目視での解析しか行えていないため、解析方法を検討して結果の出力を行えるようにしたい。

8. おわりに

本稿では、カーリングのデリバリー動作の良し悪しを判定するシステムの構築を目指し、予備調査として、画像処理技術の一分野である人体姿勢推定を用いてデリバリー動作の動画から骨格の推定を行えるかどうかの実現性調査を行った。結果として、用意した姿勢推定モデルについて2Dと3Dのデリバリー学習データを用いてデリバリーに特化した学習モデルを作成することで骨格の推定が行えることが確認できた。

今後は、考察で検討した内容を踏まえて動作解析を行い、客観的にデリバリーがショットに影響があるかの調査や、主目的である良し悪し判定のアルゴリズムの検討を行いたい。

参考文献：

- [1] 桂大地, et al. 単眼カメラを用いたボルダリング完登判定システムの提案. エンタテインメントコンピューティングシンポジウム 2019 論文集, 2019, 2019: 145-153.
- [2] 中村拓馬; 森裕一; 矢入郁子. 動画像からのサッカーシュート動作の姿勢推定の精度改善. In: 人工知能学会全国大会論文集 第34回全国大会 (2020). 一般社団法人 人工知能学会, 2020. p. 3M5GS1204-3M5GS1204.
- [3] 清水幹弥, et al. バドミントン競技経験者と未経験者のスマッシュ動作中の運動連鎖に関する研究. 東海大学スポーツ医科学雑誌, 2021, 33: 7-14.
- [4] 佐藤幸一郎, et al. 熟練度の違いによるサッカーボールリフティングの運動制御機構. バイオメカニズム学会誌, 2021, 45.2: 95-101.
- [5] DANG, Qi, et al. Deep learning based 2d human pose estimation: A survey. Tsinghua Science and Technology, 2019, 24.6: 663-676.
- [6] XIAO, Bin; WU, Haiping; WEI, Yichen. Simple baselines for human pose estimation and tracking. In: Proceedings of the European conference on computer vision (ECCV). 2018. p. 466-481.
- [7] GitHub - microsoft/human-pose-estimation.pytorch: The project is an official implement of our ECCV2018 paper "Simple Baselines for Human Pose Estimation and Tracking(<https://arxiv.org/abs/1804.06208>)" (<https://github.com/microsoft/human-pose-estimation.pytorch>) (2021/10/15)
- [8] LIU, Junfa, et al. A Graph Attention Spatio-temporal Convolutional Network for 3D Human Pose Estimation in Video. arXiv preprint arXiv:2003.14179, 2020.
- [9] GitHub - fabro66/GAST-Net-3DPoseEstimation: A Graph Attention Spatio-temporal Convolutional Networks for 3D Human Pose Estimation in Video (GAST-Net) (<https://github.com/fabro66/GAST-Net-3DPoseEstimation>) (2021/10/15)
- [10] LIN, Tsung-Yi, et al. Microsoft coco: Common objects in context. In: European conference on computer vision. Springer, Cham, 2014. p. 740-755.
- [11] IONESCU, Catalin; LI, Fuxin; SMINCHISESCU, Cristian. Latent structured models for human pose estimation. In: 2011 International Conference on Computer Vision. IEEE, 2011. p. 2220-2227.
- [12] PN Pro Landing Page (<https://neuronmocap.com/ja/node/5199>) (2021/10/17)

ボールジャグリング練習映像に基づいた 言語教示推薦システムの構築と応用

Building the Instruction Recommendation System Based on Ball Juggling Practice Movie and Application

岡田健嗣^{1*} 脇坂崇平¹ 高原慧一¹ 荒川陸^{1,2} 稲見昌彦¹

Taketsugu Okada¹ Sohei Wakisaka¹ Keiichi Takahara¹ Riku Arakawa^{1,2} Masahiko Inami¹

¹ 東京大学

¹ The University of Tokyo

² Carnegie Mellon University

Abstract: One of the methods to acquire motor skills is to receive verbal instructions from a skilled person. While they have been used in various situations, they are based on the subjective judgment of a skilled person who observes the movement of a novice. Therefore, it has not been clearly expressed how the instructions act on the indexes of the motor movement and the structural relationship between the instructions. In this study, we focused on ball juggling learning. We constructed a system “Skill Sapuri” that recommends instruction according to the skill state expressed calculated from the ball juggling video. We also discussed the possibility of using this system to aid in the understanding of embodied knowledge.

1 序論

1.1 はじめに

ヒトが獲得した技や身体運動のコツのことを、身体知と呼ぶ。[1, 2]. これはスポーツや楽器演奏などに限らず、歩く、座る、などの日常的な動作、ひいては音楽やアート鑑賞などの感性、お笑い芸人のボケなども身体知に含まれる。初心者が新しい身体知を獲得する際には、インストラクターから受ける指導がしばしば重要な役割を果たす。例えば接触の伴う指導（いわゆる『手取り足取り』）や、模範的動作の呈示、言語教示などがそれに当たる

（ここで教示による指導とは、“言語によって問題の解決について一定の方向を示唆すること [3], ”を指す）。

最近では、ICT 技術を援用して身体スキル指導を自動化する試みが広く行われているが、その際パーソナリゼーションを考慮する必要がある。すなわち、インストラクターによる対人指導では普通に行われるような、練習者の習熟状態、癖、特性を考慮した上での教示をどのよ

うに実現するか、という課題である。

例えば、ある運動スキルを同程度に実現していたとしても、適性や状況によって身体の使い方に個人差が存在する（スキルの”多様性”）[4]。それは、練習者の癖や特性、あるいは獲得してきた身体知によるが、そのような個人差を考慮せずに一律的に教示を与えると、効果的な教示にならないだけでなく、場合によっては矯正すべき癖の固定が促進されるなど、学習をむしろ妨げることもありえる。異なる練習理論、アプローチに基づく教示を混ぜる場合も、その効果に注意を払う必要がある。対人指導における指導では、その性質上こういった課題がある程度解決されているが、自動教示推薦システムにおいては、まだこれからの課題であるといえる。

個人の癖や特性を考慮せず、理想的状態に近づけることを目的とした教示手法については、実用化されたアプリケーションが存在する（[5, 6] 等）。そのようなアプリケーションは、身体スキルの種類および学習局面においては、十分に効果的であると期待できる。しかし、理想的状態（モデル）を前提として、そこに向かう教示を自動生成するという手法と、パーソナリゼーションに対応した教示生成の間には、現時点では相当隔たりがあるといえる。

* 東京大学情報理工学系研究科システム情報学専攻

〒153-8904 東京都目黒区駒場 4-6-1

E-mail: taketsugu.okada@star.rcast.u-tokyo.ac.jp

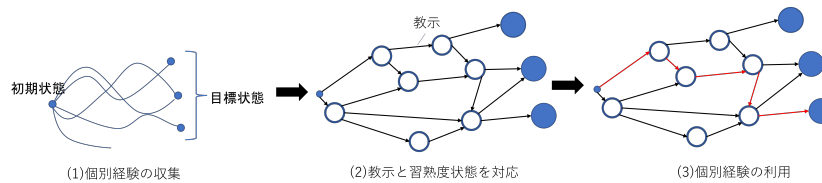


図1 経験サプリメント概略. (1) スキル習熟状態には個人差が存在する. (2) それぞれの習熟状態と対応する解決策 (本研究の場合は教示) を, 経験バンクとして蓄積する. (3) 新規ユーザーが身体スキルを習得する際, 習熟状態に応じた教示を利用する.

1.2 本研究の目的

本稿では, (対人指導で見られるような) パーソナリゼーションを組み込んだ自動教示推薦システム『スキルサプリー』の提案とプロトタイプを紹介を行う. 一言でいうと「対人指導データを大量に収集し, 再利用する」というシンプルな発想であり, このアプローチは黄瀬らが提案した『経験サプリメント』に基づく ([7, 8]. 概要図を図1に示す). より具体的には, 以下の手続きをとる.

教示収集

- 特定の身体スキルの練習動画を撮影し, サーバにアップロード
- 教示者が動画を見て, 教示を行う
- 同時に, 練習動画から練習者の習熟状態を計算
- 以上のループを回して, 習熟状態, 教示の対をデータベースに蓄積

教示利用

- 新規ユーザーが練習動画をアップロードする
- ユーザーの習熟状態を計算
- 教示データベースとユーザーの習熟状態に基づき, 推薦システムが教示を出力

本アプローチを実現するためには, (1) 対象となる身体スキルを決定, (2) 習熟状態を計算可能な指標として数値化, (3) 推薦アルゴリズムの構築 (4) アプリケーション開発, の四点が必要となる. 本論文では, 以上について説明するが, 今後データ収集に伴い, 構築したシステムの妥当性検証, そして収集された習熟状態および教示に対する構造分析をすすめる予定である.

2 スキルサプリー

2.1 対象とする身体スキル: ジャグリング

本研究では, 学習対象となる身体スキルとして, ボールジャグリングの基本技である3ボールカスケード

(3BC) を採用した. 市川らが指摘するように [9], ジャグリングスキル分析には,

- 多くの身体運動には周期的運動が含まれ, ジャグリングのボールの投げる/受けるという運動も同様に周期的である
- 自己組織化の特徴が現れる運動である
- 評価がキャッチ回数という容易な手法で行える
- ジャグリングを習得している人が少ないため被験者実験が容易

という利点がある. これに加え, ジャグリング自体の運動負荷があまり大きくないこと, 広い場所を要しないため気軽に練習を行えること, 習熟状態を表現する指標に関する研究も行われているという利点もある. 更に, 重要な要件である, 「短い練習映像を見ることにより, 教示者が教示を行うことができること」を満たしている. 以上より, スキルサプリーの対象として適切だと判断した.

2.2 ジャグリングの習熟状態を表す指標

本節では, BC の習熟状態を表す指標について記述する. Beek ら [10] は, 3ボールカスケード (3BC) の習熟状態を示す指標として, ボール保持率 k を導入している ($k = \text{ボールが手にある時間} / (\text{ボールが手にある時間} + \text{ボールが保持していない時間})$). Hashizume, 市川によって [11, 9], 実際に熟練者は, 初心者に比べて統計的に k が小さいことが報告されている一方, 学習の初期過程ではこの値が大きい群と小さい群に二分されるという指摘もある [11, 12]. そのため, ボールキャッチ位置安定性に関する研究 [11] や, ボールの軌道に関する研究 [13] がある. また, 高原ら [14] は, k のみでは個人ごとの性質を十分にとらえた習熟状態を表現できないと考え, 肘やボール軌道頂点の位置の安定性をもとに, 習熟状態の表現を試みた.

本論文では, Beek による指標 k , 関節 (肘等) やボール軌道の安定性, 対称性, キャッチ回数をスキルサプリーの習熟状態指標として暫定的に採用している. 後述する推

薦アルゴリズムは、指標の追加が可能なものとなっている。また、習熟状態指標をベクトルとしてまとめたものを習熟状態ベクトル (Skill State Vector:SSV) とする。

2.3 アプリケーション実装及び推薦アルゴリズム

2.3.1 ユーザーの利用フロー

スキルサプリユーザーは、3BC を数キャッチ以上成功できる初心者想定している。ユーザーは、映像を撮影し、ウェブ経由でアクセスするサーバーに動画をアップロードすることで、教示を得る。教示は (1) 熟練者によって直接付与される方法と、(2)(1) のプロセスで収集した教示データを利用した推薦システムによって教示が付与される方法、の 2 通りが存在する。データベース収集初期は、(1) のみを行う。ユーザー側の操作の流れを図 2 に示す。動画をアップロードし教示が付与されると、メールが届くようになっており、スキルサプリ上で教示内容を確認することができる。

2.3.2 スキルサプリ構成

この章ではユーザーに教示を提示する、スキルサプリの構成について概観する (図 3)。入力された映像をもとに、骨格検出、ボール座標検出を行い、データの後処理を行ったのち SSV を計算する。(1) の、熟練者によって直接言語教示が付与される場合は、SSV は言語教示とセットにしてデータベースに保存される。(2) の、データベースに基づいて言語教示が付与される場合は、SSV とデータベースを用いて言語教示推薦が行われる。

SSV を計算するために、ユーザーの骨格、ボール座標の検出を行う必要がある。ユーザーの骨格検出には、OpenPose[15] を用いた。また、ボール座標の検出には、Åkerlund による畳み込みニューラルネットワークを用いたアルゴリズム [16] を使用した。これらの値は、動画の解像度、ユーザーの身長によって補正されたのち、正規化される。

ユーザーの骨格や、ボール座標に対し、欠損値計算、平滑化などの前処理をおこなったのち、習熟状態指標計算を行う。本論文執筆時点では、習熟状態を表現する指標として、Beek の指標、高原らが検討した、肘の平均、標準偏差、ボール軌道頂点位置座標の平均、標準偏差、ボールキャッチ回数を扱い、SSV とした。このベクトルと教示を対応させるが、後からでも習熟状態を追加できるようにスキルサプリの設計を行った。

熟練者による言語教示は、自由入力する方法と、過去に用いた教示の中から選択する方法がある。過去に用いた教示と異なる教示をしたいときのみ自由入力で教示を

してもらおう。

スキル習熟状態に基づいた教示を自動で利用可能にするため、ユーザーベース協調フィルタリングの推薦アルゴリズムの手法の一つである、Factorization Machines[17] を用いた推薦を行う。Factorization Machines は、

- 推薦に用いる指標の追加が容易
- ドメインに関する知識がなくてよい
- 習熟状態指標同士の交互作用を用いて推薦可能
- スパースなデータに対応

という利点がある。概要図を図 4 に示す。ユーザーと教示、教示の評価値からなる行列に対して、教示の評価値が高いものを推薦するアルゴリズムである。この行列の値をより正確に表現するために、評価値の行列を、ユーザーの習熟状態を含む、図 4 の右図のような形式に変換することで、ユーザーの習熟状態を利用することができる。また、Factorization Machines には、Ribeiro らによって提案された Local Interpretable Model-agnostic Explanations:LIME[18] を用いることにより、出力 (推薦結果) にどの SSV の要素が効果的に働いたかを確認することができるという利点がある。また教示と習熟状態、習熟状態の交互作用の対応関係を定量的に観察することも可能である。さらに教示同士の構造についても考えることができる。例えば、類似した SSV が複数の異なる教示を推薦した場合、それらの教示は (1) 類似した効果をもつ、もしくは (2) 異なるアプローチの教え方に属する、といった解釈がありえる。習熟状態間の構造、教示間の構造の具体的分析については、今後解析に必要となるデータ数が蓄積した後に行う予定である。

推薦スコア予測値 \hat{y} は、次の式で計算される。

$$\hat{y}(\mathbf{x}) := w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j \quad (1)$$

\hat{y} : 教示の推薦スコア予測値

x : 特徴ベクトル (ユーザー, SSV, 教示)

w : 重み

\mathbf{v} : 潜在ベクトル

w, \mathbf{v} が学習パラメータである。

3 考察

3.1 スキルサプリとメタ認知

メタ認知自体は 1979 年に Flavell によって [19] 提唱された、自分の認知活動を見直したり調整したりするメカニズムに意識的に気づくという概念である。Nelson ら



図2 スキルサプリーユーザー側の操作動画をホームページ上にアップロードし、技の種類、ボール数、練習している時間を指定する。教示が付与されるとメールが届く。ユーザーからは教示がシステムによって付与されたものか、熟練者が付与したものかはわからない。

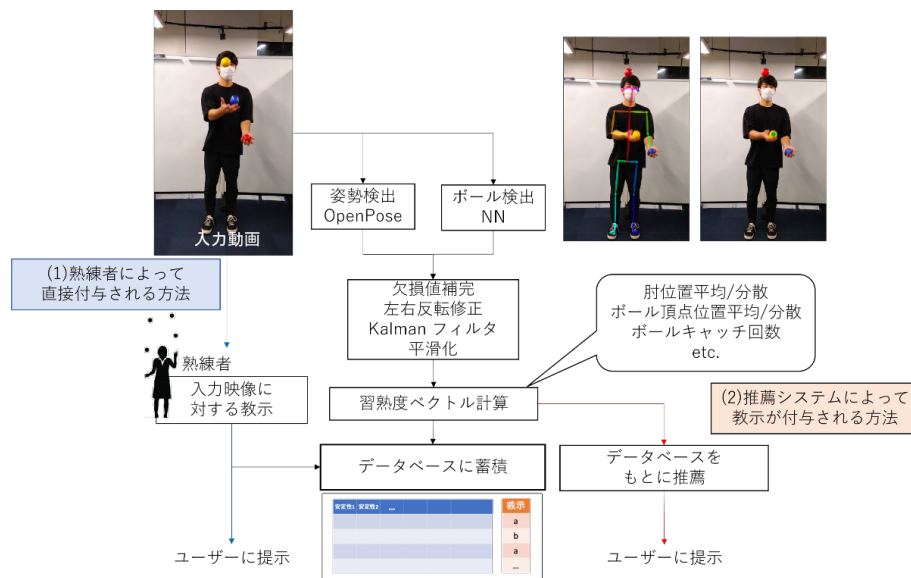


図3 スキルサプリー構成図。入力された動画は、教示付与の方法にかかわらず、習熟状態指標が計算される。熟練者によって言語教示が付与される場合、習熟状態指標と教示がデータベースに格納され、システムによって言語教示が付与される場合、習熟状態指標とデータベースに基づいて言語教示が推薦される。

によって [20] メタ認知知識とメタ認知制御によるメタ認知プロセスがモデル化されている。身体スキルに関するメタ認知は、メタ認知言語化として、諏訪ら [1] によって提唱されてきた。これは、スキル習熟が進行するにつれ、環境や道具との関わりや、身体部位の使い方が漸進的に変化するが、それを言語的に表出させることで、身体と環境のインタラクション自体を発展させることを指している。メタ認知言語化によって、それまで意識していなかった変数に対する意識が生じる。メタ認知言語化はあくまで練習者が一人称的、主体的に行うことが重要であり、単純に教示（コーチング）を与えるのみでは”

受け売り”になってしまうという指摘があったが、スキル習熟に応じた教示を与えることができれば、メタ認知言語化を促進できる可能性がある。

メタ認知言語化に基づき、ジャグリング習熟プロセスにおける言語化の内容と、ジャグリングのパフォーマンスを関連させることで、ジャグリング習熟中に意識する内容がどのように変化するのかを調査した研究が行われてきている ([21],[9],[22])。これらの研究は、スキル習熟プロセスにおいて、主観的な視点での言語化に基づいて分析を行っている。ジャグリングにおける言語化内容の質と、スキル習熟状態に対し一定の関連があることが示

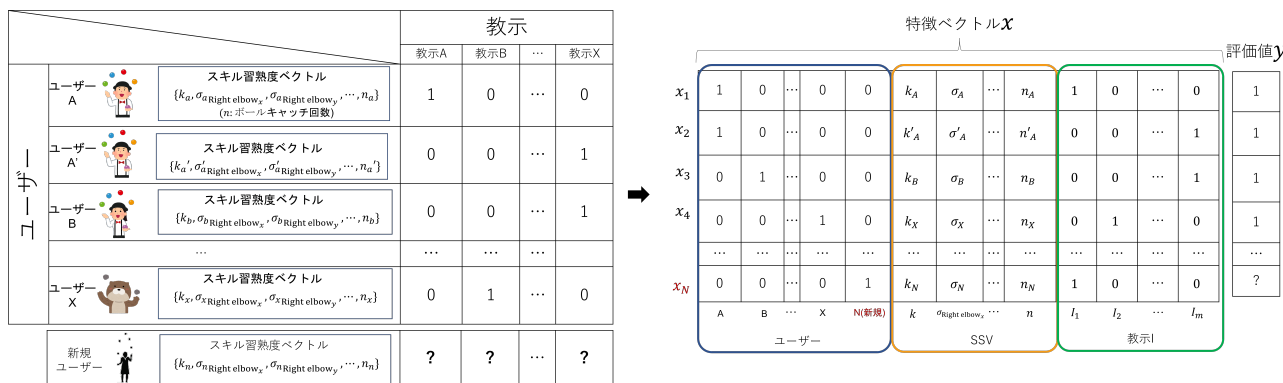


図4 Factorization Machines の概要図. 左表のようなユーザーと教示の組のみを用いて教示推薦を行うのではなく、右表のような、ユーザー、SSV、教示を組み合わせた特徴ベクトル x を用いて教示推薦を行う。同一のユーザーに対して、SSV が変化していれば、新しく特徴ベクトルとして加える。SSV を構成する各指標の説明は省略。

唆されている。スキルサプリア自体は熟練者からの教示を与えるだけであるが、これらの研究をもとに、自分でメタ認知言語化した情報を入力することができるようになったり、教示を受けた際に連鎖的に発見したことを記入することができるようになれば、他者と言語化した情報の共有を行えるようになる可能性がある。また、スキルの習熟過程において、習熟状態がどのような変遷をするのかわかれば、それを初心者に提示することで、メタ認知を促進しうる。

3.2 教示者自身への効用

教示者はスキルサプリアの教示収集モードを介してユーザーに対し教示を与えるが、これは教示者自身のスキル習熟も促進すると期待している。Duran らによると [23], 「教えることで学ぶ」ことは以下の段階に分けることができる。すなわち、(1) 保有する知識を他者に伝搬する教授 (telling knowledge), (2) 他者により積極的に理解を促すために知識構築 (building knowledge) を伴う解説, (3) 質疑応答などのあるやりとり (interaction), (4) 教える側と教えられる側による共同作業 (collaboration) の4段階である。この分類に従えば、スキルサプリアでは、通常に教えるだけで得られる (1) だけの効果だけでなく、(2) のプロセスも含めることができ、教示を通して大きな学びの効果を期待することができる。

3.3 流派, 方法論の整理

コミュニティや流派, 方法論によって、教え方が異なる教示と習熟状態の対応を調査することで、方法論の差異を明確にし、初心者が方法論を選択して学習できる。

3.4 ライブストリーミング映像への対応

スキルサプリアでは、撮影した動画を手動でアップロードするという手法を用いている。アップロード作業に最短でも1分ほどかかることに加え、教示出力計算に2分程度かかる。すなわち、動画を撮影してから教示を得られるまでに、ある程度の時間が経過してしまうという問題が存在する。習熟状態や教示内容によっては、できる限り早く教示しないと効果が出ない場合も考えられる。そのため、手動アップロードする方式に加えて、ストリーミング方式 (練習中撮影していると、適宜教示が与えられる) を追加実装することが望ましい。

3.5 汎用性

本稿では3BCに特化したアプリケーションを開発した。ただし、映像をみて教示が可能、習熟状態を数値指標化できること、などの要件を満たすならば、他のスキルを扱うこともできる。例えば5ボールジャグリングは、メタ認知言語化や個人の状況にあわせた教示が特に難しい。そういったスキルではスキルサプリアの恩恵がさらに明確に現れると期待している。さらにジャグリングにかぎらず、応用先についても検討する。

4 結論

スキル習熟プロセスにおける、教示と身体動作の特徴量を対応付けることで、教示構造を明らかにするためのツールとして、スキルサプリアを作成した。また、スキルサプリアを用いることで分析できることについて議論を行った。習熟状態に応じ教示を利用することを目的として

スキルサブリを構成したが、教示に関する様々な知見を得るのに利用できる可能性がある。今後は、実際にスキルサブリを用い、データベースを構築し、教示に関する分析や、ユーザー体験の分析を中心に行っていく。

参考文献

- [1] 諏訪正樹. 身体知獲得のツールとしてのメタ認知的言語化 (< 特集 > スキルサイエンス). 人工知能, Vol. 20, No. 5, pp. 525–532, 2005.
- [2] 諏訪正樹. 身体が生み出すクリエイティブ. 2018.
- [3] 小玉耕平, 中山正吉. 体育における言語的教示に関する研究. 島根大学教育学部紀要 (教育科学), Vol. 23, No. 1, pp. 17–24, 1989.
- [4] 古川康一. スキルサイエンス入門—身体知の解明へのアプローチ. 人工知能学会, 2009.
- [5] NTT docomo. Golf ai. <https://golfai.jp/>.
- [6] 小坂田光, 多田昌裕. Android 端末を用いたリアルタイム安全運転アドバイス提供システムの試作. 映像情報メディア学会技術報告メディア工学, pp. 11–14. 一般社団法人映像情報メディア学会, 2016.
- [7] 国立研究開発法人科学技術振興機構. 経験サブリメントによる行動変容と創造的協働. https://www.jst.go.jp/kisoken/crest/project/1111083/1111083_09.html.
- [8] Sohei Wakisaka, Atsushi Hiyama, and Masahiko Inami. Transmission of experiences with augmented human techniques. In *UbiComp*, 2017.
- [9] 市川淳, 三輪和久, 寺井仁. 運動計測と言語報告に基づく身体スキル獲得に関する実験的検討. 人工知能学会論文誌, Vol. 30, No. 3, pp. 585–594, 2015.
- [10] PJ Beek and AAM van Santvoord. Learning the cascade juggle: A dynamical systems analysis. *Journal of Motor Behavior*, Vol. 24, No. 1, pp. 85–94, 1992.
- [11] Ken Hashizume and Tomoyuki Matsuo. Temporal and spatial factors reflecting performance improvement during learning three-ball cascade juggling. *Human Movement Science*, Vol. 23, No. 2, pp. 207–233, 2004.
- [12] 山本耕太. 3 ボールジャグリング学習過程における時空間的変数の変化. 3.
- [13] Pamela S Haibach, Gregory L Daniels, and Karl M Newell. Coordination changes in the early stages of learning to cascade juggle. *Human Movement Science*, Vol. 23, No. 2, pp. 185–206, 2004.
- [14] 高原慧一, 脇坂崇平, 荒川陸, 檜山敦, 稲見昌彦. 経験バンクに基づくジャグリング学習支援システムの提案. エンタテインメントコンピューティングシンポジウム 2020 論文集, Vol. 2020, pp. 199–203, 2020.
- [15] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*, 2017.
- [16] Rasmus Åkerlund. Real-time localization of balls and hands in videos of juggling using a convolutional neural network, 2019.
- [17] Steffen Rendle. Factorization machines. In *2010 IEEE International conference on data mining*, pp. 995–1000. IEEE, 2010.
- [18] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. ” why should i trust you?” explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135–1144, 2016.
- [19] John H Flavell. Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, Vol. 34, No. 10, p. 906, 1979.
- [20] Thomas O Nelson. Metamemory: A theoretical framework and new findings. In *Psychology of learning and motivation*, Vol. 26, pp. 125–173. Elsevier, 1990.
- [21] 田中彰吾, 小河原慶太. 身体知の形成: ボールジャグリング学習過程の分析. 人体科学, Vol. 19, No. 1, pp. 69–82, 2010.
- [22] 内山光太, 伊藤毅志ほか. ジャグリングの熟達における思考過程の分析-3 ボールカスケードの事例より-. *SIG-SKL*, Vol. 22, No. 07, pp. 34–41, 2016.
- [23] David Duran. Learning-by-teaching. evidence and implications as a pedagogical mechanism. *Innovations in Education and Teaching International*, Vol. 54, No. 5, pp. 476–484, 2017.