第 27 回日本バーチャルリアリティ学会大会論文集（2022 年 9 月）

# 身体動作予測のための多様な動作データ取得手法

板井俊樹 [1)]，牧野泰才 [1)2)]，篠田裕之 [1)]
Toshiki ITAI, Yutaro TOIDE, Yasutoshi MAKINO and Hiroyuki SHINODA

1) 東京大学 新領域創成科学研究科 (〒 277-8561 千葉県 柏の葉 5-1-5, t.itai@hapis.k.u-tokyo.ac.jp)
2) JST さきがけ

**概要**: VR 空間内でアバタを表示する際に生じる描画遅れを解消するために, 動作を予測し表示する手法が提案されている. 既存の予測手法では，学習データセットに含まれないような動作を行った場合，予測に失敗してしまうことがある．この問題を解消するため，適切なデータセットが必要であると考えた．本稿では予測結果をヒトに見せながらそれを欺くように動作させることで，敵対的にデータセットを計測し，この課題の解消に取り組む．

**キーワード**： 身体動作予測，ニューラルネットワーク

## 1. 序論

短期的な未来の身体動作を予測が可能になると，VR 空間上でリアルタイムに身体動作に基づくコミュニケーションを行う際に，通信による遅延やレンダリング時の計算に起因する遅延を解消する手段として用いられることが期待される．

これまでに人の身体動作情報から，短期的な未来の動作を予測する研究が行われている．例えば，Martinez ら [1] や Chiu ら [2] の研究では，Human3.6M[3] のデータに対し，回帰型ニューラルネットワークを用いて未来の動作を予測した．また，Horiuchi[4]，倉井らによる研究 [5] では，順伝搬型ニューラルネットワークを用いて過去 10 フレームの骨格情報から，0.5 秒程度先の未来の位置と姿勢を予測し，歩行やジャンプのような全身運動において，重心位置の平均二乗誤差が数 cm 程度で予測できるということを示した．Wu らの研究 [6, 7] や須田らの研究 [8] では Sports での応用を提示している．

これまで多くの研究では，公開されているものや，自分で取得した決められた種類の動作を学習し，その動作の種類の傾向を学習して予測を行っていることが多い．これらの予測器に対して，訓練データに含まれていない動作を入力すると誤差が大きい予測をしてしまうことがある．一方で，ヒトはコミュニケーションやスポーツを行う場合，多様な動作を行う．そのため，VR 空間上に短期的な未来を予測する予測器を実装しようとすると，多様な動作に対して予測が可能である必要がある．

本研究では，予測結果を見せながらその予測精度が下がるような動作をしてもらうことで，データセットの多様性を確保する手法を提案する．具体的には，被験者に特定の動作を指示をして取得したデータを用いて予測器を学習する．その予測器による予測結果を見せながら，予測された重心位置が実際の位置と離れるような動作をしてもらうことによって，予測器が現状予測できない動作を敵対的に取得する．

## 2. 入力に用いるデータセットの取得

横方向への移動について，データの多様性を確保しながら取得する．計測は 2 段階に分けて行った．1 段階目は可能な限り多様な動作を取得できるような指示をして取得し，2 段階目に関しては，1 段階目に取得したデータを用いて学習した予測器の結果を見ながらそれを欺くように動作することで，1 段階目のデータの学習だけでは対応しきれなかった動作データを取得する．計測は Azure Kinect を用いて行い，被験者は高さ 1.1 m に設置した Kinect から 2.9 m 離れ，3.4 m の範囲で横方向に動作した．

### 2.1 指示をしてデータセットを取得

可能な限り多様な横方向の動作データを取得するため以下のように指示をして計測を行った．

1. 被験者は Kinect に正対し，正面にあるモニタを見る．
2. モニタ上に左右の矢印の画像をランダムな間隔（0.5 s〜1.2 s）で交互に 25 回ずつ提示する．
3. 被験者は指示をされた方向に左右に動き続ける．このとき，指示をされた方向にできる限り素早く切り返すようにし，矢印の切り替えスピードに体がついて行かない状況が生じたとしても矢印が描画される限りは動き続ける．

このとき，Kinect の計測エリア外に被験者が出ていかないようにするため，被験者が計測エリアの端に近づいた場合，被験者が中央に戻るまでは矢印の方向を切り替えないようにした．20 代の男性 7 人女性 2 人の被験者について，一人の被験者に対して上記の計測を 3 セット繰り返し，指示された横方向の動作データを 30 fps で取得した．

### 2.2 敵対的にデータセットを取得

指示をされた動作データを用いて学習された予測器により，図 1 のようなゲームアプリケーションを作成し敵対的に動作計測を行った．このアプリケーションは予測した骨格の重心座標にボールが引っ張られるように動作する．つまり，予測に成功している場合には，現在のヒトの動作から未来の骨格の位置を予測し，ヒトが動こうとしている方向にボールが先に動き出す．適切な粘性を設定することで，未来の重心位置に対して遅延を持って移動することになるため，現在の重心と予測値との間の適当な位置にボールが常に存在し続けるようになる．一方，予測に失敗した場合には，ヒトが動こうとしている方向にボールが動かなかったり，あるいは逆に動きすぎたりする．そのため，予測が成功している状況下においてはボールがアバターの近くに存在し続けるが，予測に失敗すると，アバターからボールが離れてしまうことがある．

予測に失敗した動作が予測器が対応できない動作であるため，この動作を効率良く取得できるようにゲームアプリケーションを設計した．被験者には，ボールがアバターから離れるように行動してもらいたいため，ボールが現在の骨格の重心から 0.6 m 離れた位置に引かれているラインより外に出ると，ゲームのスコアが加算されるようにした．被験者には，なるべく高いスコアを狙うように指示することで，制限時間の間，予測が失敗するような動作を試行錯誤して可能な限り行ってもらうようにした．

また，ゲームアプリケーションは以下のように実装した．

1. 被験者の骨格情報をリアルタイムに反映したアバターを描画する．
2. 学習器を用いて予測した 0.5 秒後の骨格の重心座標を描画する．(図 1 の緑の点)
3. アバターの足元に 0.02 kg，直径 33 cm のボールを用意する．ボールの $x$ 座標と 0.5 秒後の重心の $x$ 座標にばね定数 1 N/m のばねと比例定数 0.016 N・s/m のダンパを付ける．この数字は予備実験により経験的に決定した．
4. 現在の骨格の重心座標から $x$ 軸上に ±0.6 m 離れた距離にラインを引く (図 1 のピンクの 2 本のライン)．このラインよりボールが外に出ている経過時間に応じてスコアを加算する．
5. 左上に 30 秒間のタイマー，右上に現在のスコアを表示する．

2.1 と同様の被験者について，一人の被験者ごとにゲームを 30 秒 ×3 セット繰り返し，現在の予測器では対応できない動作データを取得した．

## 3. ネットワークと学習方法

既存の研究 [4][5] では, 順伝搬型ニューラルネットワークに過去 10 フレーム程度の骨格座標データを入力し，0.5 秒先の骨格データを出力として，身体動作予測を行っていた．

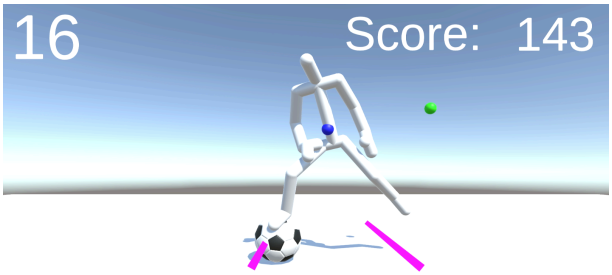


**図 1: 敵対的にデータセットを取得する際に用いたゲームアプリケーション**

本研究では，入力フレーム数を 15 フレーム（0.5 秒）とし，入力に用いた最後のフレームから 0.5 秒後の 32 点関節情報を出力とした. ネットワークは，図 2 の構造を利用した. 正解となる骨格は 32 箇所の 3 次元骨格であり，96 次元となる（図 2 の最終段の $n = 1$ のとき）．ニューラルネットワーク (NN) の誤差関数に平均二乗誤差 (MSE)，活性化関数に ReLU，オプティマイザーに Adam を用いた．バッチサイズは 20，エポック数は 100 とした．学習は，2.1 で指示をして取得したデータ (Instruction) だけを用いたものと，Instruction に 2.2 で敵対的に取得したデータ (Adversary) を追加して行ったものの 2 通り行った．

取得したデータのまま学習すると，動作エリアの端に被験者が来た後，動作エリアの外には出ないことが学習されてしまうため，訓練データに取得したデータの x 座標に 2.55, 1.7, 0.85, -0.85, -1.7, -2.55 m を足したものをそれぞれ追加し，計測エリアの位置に応じた被験者の行動パターンを学習できないようにした．

予測器は，個人ごとの動作の癖を学習させないようにするため，9 人分のデータのうち 8 人分を訓練データとして学習し，取り除いた 1 名のデータでテストすることとした．すなわち，評価時の予測器にテストの被験者の動作データは一切含まれない．この除外する 1 名を変えながら，同じ検証を計 9 回行った．テストデータは，Instruction と Adversary でそれぞれ評価した．

ゲームアプリケーションに用いた学習器に関しては，2.1 で取得した 9 人分のデータを訓練データとして用いた．

## 4. 結果

図 4.(a) は Instruction を用いて学習した予測器に対して，Instruction を用いて評価を行った 0.5 秒後の予測と実測との誤差をフレームごとにカウントしたヒストグラムである．誤差 1cm ごとに一つのビンとしてまとめてあり，9 回のテストで，用いた被験者（学習モデル）によって色を分けている．縦軸は全データに対する出現頻度の割合を表している．

Instruction を用いて学習した予測器では，誤差の平均値は 14.0 cm，中央値は 11.8 cm となった．また，図 4.(b) は予測器は変えずに Adversary 条件で計測されたデータを用いて評価を行った 0.5 秒後の予測と実測との誤差ヒストグ

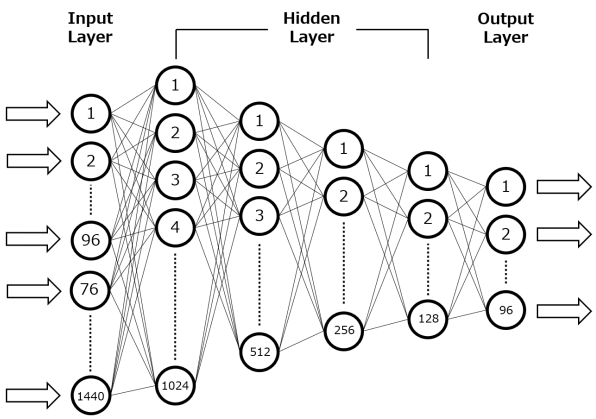


**図 2: 学習に用いたネットワーク**

ラムである．誤差の平均値は 29.2 cm，中央値はは 22.8 cm となっており，図 4.(a) に比べて誤差が増加していることが確認できる．

図 4.(c) は Instruction のデータと Adversary のデータを合わせて学習した予測器に対して，Instruction を用いて評価を行った 0.5 秒後の予測と実測との誤差分布である．誤差の平均値は 15.3 cm，中央値は 12.7 cm となっており，図 4.(a) に比べて誤差が増加していることが確認できる．図 4.(d) は予測器は変えずに Adversary を用いて評価を行った 0.5 秒後の予測と実測との誤差分布である．誤差の平均値は 15.4 cm，中央値は 13.0 cm となっており，図 4.(b) に比べて誤差が減少していることが確認できる．また．テストに用いた被験者ごとの誤差の中央値に対して，(a) と (b)，(c) と (d) ではウィルコクソンの符号順位検定，(a) と (c)，(b) と (d) に関してはウィルコクソンの順位和検定を行った．(a) と (b)($p = 0.0002$)，(c) と (d)($p = 0.0788$)，(a) と (c)($p = 0.0136$)，(b) と (d)($p = 0.0007$) となり，(c) と (d) 以外に関しては 5%の有意水準で有意差があることが確認できた．本研究で期待した通り，予測結果を見せながら，それを欺くような動作をさせ，その時の動作情報を利用することで，予測性能が向上することが確認された．

## 5. 考察

動作を指示をして取得したデータを用いて多様な動作に対応した予測器を作成する際には，その指示に応じた癖に特化した予測器が作成されてまう．そのため，自由に動く動作を予測しようとすると，誤差が大きくなる予測をしてしまう状況が存在する．そこで，指示をしたデータで学習した予測器を見ながらその予測を欺くように動作を計測することで，現在の予測器が苦手とする動作を取得することができ，そのデータを用いて学習することで予測の癖が汎化されてより多様な動作を予測可能な予測器が作成することができる．

結果を確認すると，図 4.(b) は図 4.(a) に比べて誤差が中央値で 11.0 cm 増加し，指示をして取得したデータだけで学習した予測器が苦手とするデータが収集できたことが確認できる．また，そのデータを追加して学習すると図 4.(c) は図 4.(a) より誤差が中央値で 0.9 cm 増加し，図 4.(d) は図 4.(b) より誤差が中央値で 9.8 cm 減少しているため，指示をして取得したデータに特化してしまっていた予測器の癖が汎化され，より多様な動作に対応できる予測ができていることが確認できる．

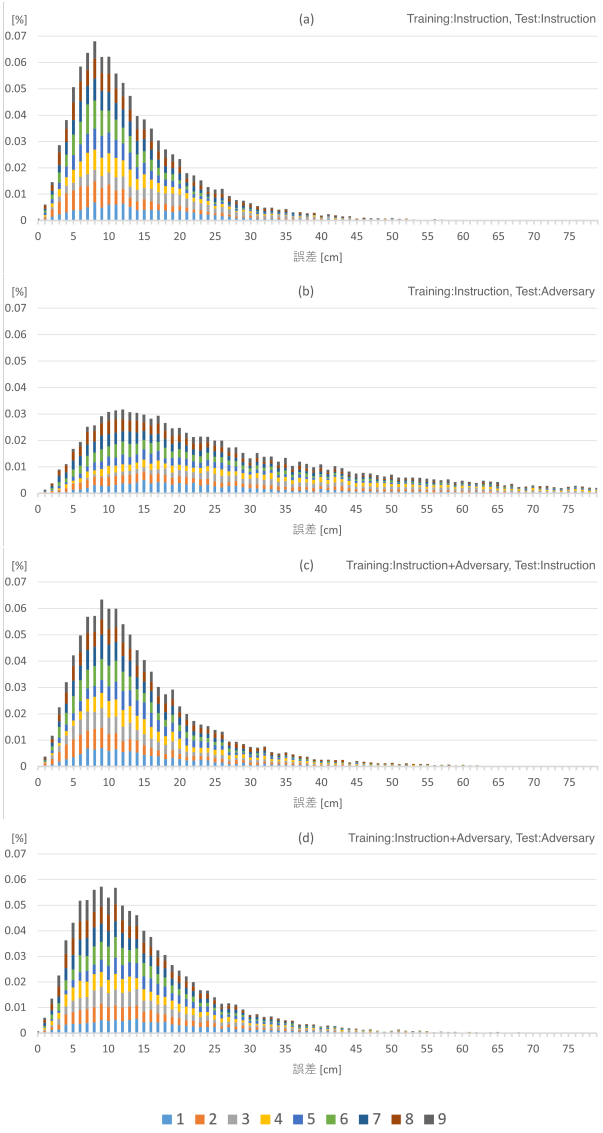


**図 3: 0.5 秒後の予測と実測との誤差のヒストグラム (a) 訓練データ：Instruction，テストデータ：Instruction，(b) 訓練データ：Instruction，テストデータ：Adversary，(c) 訓練データ：Instruction＋Adversary，テストデータ：Instruction，(d) 訓練データ：Instruction＋Adversary，テストデータ：Adversary**

## 6. 結論

本研究では，指示をして取得したデータを用いて学習した予測器を見ながら，それを欺くように動作してもらうことによって，予測器の穴を付くような動作データを敵対的に取得する方法を提案した．また，それらのデータを追加して再度学習することによって，多様な動作に対する予測

精度が向上することを確認した.

今後の研究では，学習とその結果を見て敵対的にデータを取得することを繰り返すことで，より多様な動作に対応した予測器作成を目指す.

**謝辞**

本研究は科研費（基盤 B）21H03479 の支援を受けた.

## 参考文献

[1] J. Martinez, M. J. Black, and J. Romero, "On human motion prediction using recurrent neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2891–2900, 2017.

[2] H.-k. Chiu, E. Adeli, B. Wang, D.-A. Huang, and J. C. Niebles, "Action-agnostic human pose forecasting," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1423–1432, IEEE, 2019.

[3] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 7, pp. 1325–1339, 2013.

[4] Y. Horiuchi, Y. Makino, and H. Shinoda, "Computational foresight: Forecasting human body motion in real-time for reducing delays in interactive system," in *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*, pp. 312–317, 2017.

[5] T. Kurai, Y. Shioi, Y. Makino, and H. Shinoda, "Temporal conditions suitable for predicting human motion in walking," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 2986–2991, IEEE, 2019.

[6] E. Wu and H. Koike, "Futurepose-mixed reality martial arts training using real-time 3d human pose forecasting with a rgb camera," *in 2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1384–1392, IEEE, 2019.

[7] E. Wu, F. Perteneder, and H. Koike, "Real-time table tennis forecasting system based on long short-term pose prediction network," *in SIGGRAPH Asia 2019 Posters*, pp. 1–2, 2019.

[8] S. Suda, Y. Makino, and H. Shinoda, "Prediction of volleyball trajectory using skeletal motions of setter player," *in Proceedings of the 10th Augmented Human International Conference 2019*, pp. 1–8, 2019.