Vol.2021-HCI-195 No.4
Vol.2021-UBI-72 No.4
2021/11/30

機械学習を用いた
手指のモーションキャプチャーデータの解析に関する研究

馮程静儀[†1]　坂井滋和[†1]

概要：本研究では，本研究室で製作されたグローブ式のハンドモーションキャプチャーを用いて指先の様々な動きを計測し，その計測データを機械学習させることで動作解析を試みた．ここで使用したモーションキャプチャーシステムは，近年小型・低価格化が進む IMU センサーを用いて，指の関節ごとの回転角度（Quaternion）を計測し，手を多数の関節を持つリンク構造として定義することで，各センサーからのデータを元にその動きや形態を再現することができる．本研究では，その基礎実験として，多様な行動を表す手の動きデータを大量に計測し，これを学習データとして使用して，手の動作に関する機械学習の有効性について検証を行った．具体的には，まず，手の静止ジェスチャーのデータの 2 分類問題を NN 手法，3 分類問題を DNN 手法で実装することにより，手指データを機械学習による解析の有効性を確認した．次に，手の似たような動きのデータを CNN・RNN によってそれぞれ学習を行い，中間層の増加と Dropout 層の導入によるモデルの精度を改善し，手指の動作を精度よく判別できる機械学習モデルを実装した．本稿では，手指データの収集，複数の機械学習モデルの実装の詳細と，手の動作データの解析結果について報告する．

キーワード：ハンドモーションキャプチャー，機械学習，手指データ，解析

Study on analysis of fingers motion tracking data using machine learning

CHENGJINGYI FENG [†1]　SHIGEKAZU SAKAI [†1]

Abstract: In this study, using the hand motion tracking glove produced by Sakai laboratory, a large amount of fingers motion data can be collected by this glove system without external sensor, we aimed to analyze the finger movement data accurately and efficiently using machine learning method. Specifically, first, by implementing a two-classification problem of hand gesture data by the NN model and a three-classification problem by the DNN model, I confirmed the usability of analyzing hand data using machine learning method. Next, two kinds of similar hand movement data were used to train a CNN model, a RNN model as well. By increasing the number of middle layers and the Dropout layers, I trained two kinds of machine learning models that can accurately discriminate hand movement data. This paper reports the method of collecting hand data, the details about the implementation of multiple machine learning models, also the analysis results of hand motion data.

Keywords: Hand motion tracking, Machine learning, Hand motion data, Analyze

1. はじめに

手は人間の様々な作業において大変重要な役割を果たすことは周知の事実である．従って手の動きを解析し，それについて知ることは，人体の仕組みに限らず，楽器演奏や描画などの芸術から，手を使う熟練工の技の解析，手と連動する脳の働き，スポーツなど，様々な分野において利用価値が高い．人間動作を計測するモーションキャプチャーシステムは現在，様々な分野で利活用が進んでいる．しかし，光学式のシステムでは手のような小さくて指同士が複雑に絡み合う動きを計測することは不可能であるため，手の動きに関する計測はあまり進んでいない．そこで本研究室では，近年小型化と高性能化かつ低価格化が進む IMU 方式のセンサーを複数用いて手指のモーションを細かく計測することが可能なモーションキャプチャーシステムを開発した．（図 1 に参照）

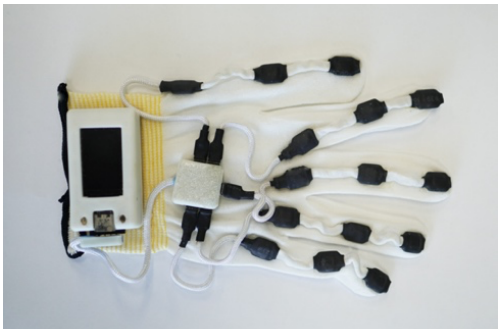

図 1　ハンドモーションキャプチャーグローブ

本研究では，この装置を利用して手の様々な動きを計測し，機械学習によってその動きを認識させるための基本実験を行った．データ解析では，まず DNN 手法を用いて，静止状態に置かれた手のジェスチャーデータを利用し

[†1] 早稲田大学
Waseda University

Vol.2021-HCI-195 No.4
Vol.2021-UBI-72 No.4
2021/11/30

て，2 分類問題と 3 分類問題を実装した．これによって手指データの機械学習による解析の有効性を確認することができる．次にピアノを演奏と PC のキーボード打鍵を対象にデータ計測を行い，CNN・RNN による学習を行った．ここでは中間層の増加とドロップアウト層の導入によるモデルの精度改善を行った．本論では，この 2 点の実験とその結果について報告する．

2. 関連技術

2.1 クォータニオン (Quaternion)

クォータニオンは四元数とも呼ばれる，イギリスの数学者ハミルトンにより 1843 年に考案された数体系である．クォータニオンは複素数を 3 次元空間に拡張したものであり，式 (2.1) のように定義され，その中の i, j, k は式 (2.2) の関係を満たす．

$$q = a + bi + cj + dk \quad (a, b, c, d \in \mathbb{R}) \tag{2.1}$$

$$i^2 = j^2 = k^2 = -1 \tag{2.2}$$

幾何学的には，クォータニオン Q(w, x, y, z)の各要素は，任意の軸 V(x, y, z)回り空間座標の回転(角度θ)を表し，w = cos(θ/2)，x = sin(θ/2) * Vx, y = sin(θ/2) * Vy, z = sin(θ/2) * Vz となる．V を回転軸ベクトル，θは回転角と呼ばれる．

3 次元物体の姿勢角度を表す際，XYZ の各軸に対して回転値を設定するオイラー角表現に比べ，クォータニオンは物体の回転を正確に表すのに行列よりも少ない情報量と計算負荷が軽いといった特徴があるため，3D の CG 分野など多くの場面で使用されている．

2.2 深層学習 (Deep Learning)

機械学習の手法の一種である深層学習はニューラルネットワークを多層に結合したものであり，近年画像処理や自然言語処理など様々な分野で大きな成果をあげている．

深層学習の特徴として，ある仮定のもとに任意の関数を任意の精度で近似できることと，中間層の構造と反復回数を調整することで汎化能力の向上を実現できること，サンプルデータのノイズに強いことなどが挙げられる．以下に深層学習の代表的なモデル CNN と RNN について述べる．

畳み込みニューラルネットワーク（CNN）は空間的な特徴を捉えることができる，空間的な特徴を持つ画像の分類に適したネットワークである．CNN は多くの層から構成されており，具体的には，畳み込みを行う畳み込み層と，サイズ圧縮を行うプーリング層がお互いに重なり構成されている．近年，発展された CNN は画像領域を超え，テキスト，動画，音声などの分類タスクにも幅広く利用されている．本研究では時系列データに対して適用したため，画像認識などで用いられる 2 次元の CNN ではなく，1 次元の CNN を用いた．

リカレンドニューラルネットワーク（RNN）は入力間の依存性を利用するニューラルネットワークの一種であり，テキスト，音声，時系列データなどの解析に適したネットワークである．本研究では長期依存性を学習できる RNN の亜種である LSTM（Long short-term memory）モデルを用いて実装した．LSTM は勾配消失問題をうまく回避できる，長期にわたる依存性をより効果的に学習するように設計されたネットワークである．

2.3 Keras ライブラリ

Keras[1]はニューラルネットワークを実装するためのライブラリであり，TensorFlow をバックエンドとして動かすことができる．Keras は迅速な実験を可能にすることに重点を置いて開発されたライブラリであり，事前に定義された様々な種類のニューラルネットワーク層，損失関数，活性化関数を組み合わせてネットワークを実装できるため，容易に素早くプロトタイプの作成が可能である．

3. 提案手法

本研究で提案する手指の動作認識手法では，まず製作されたハンドモーションキャプチャーグローブを用いて，手指の動作データを収集した後，複数の機械学習手法によって静止ジェスチャーと動く手の行動データをそれぞれ解析する．

3.1 リンク構造による手の数理モデル

本研究では，図 2 に示しているように手の骨格の数理モデルを定義した．緑マークは右手のジョイント（関節），黄色のラインは関節と関節の間のリンクを表している．その上で，手を開く際の向きを座標の z 方向として定義した．

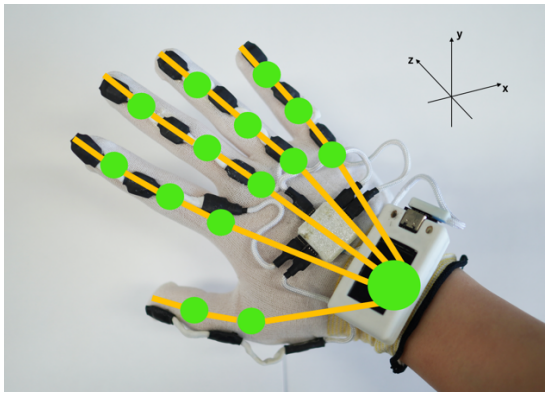

図 2　センサー装着位置

データ収集で用いるハンドモーションキャプチャグローブは，この手の骨格の数理モデルに基づき，手のリンクの部分と手のひらの部分に 16 個のセンサーを付け，姿勢角度データを 16 個のクォータニオンによって定義される．

そこで，1 つのクォータニオンは 4 つの実数値から構成されているため，1 つの静止ジェスチャーは 64 個の実数値

によって表現することも可能のため，データ収集の際に 64 個の実数値データで 1 つの静止ジェスチャーを定義することにした．5 組のデータを例として図 3 に示す．

	0	1	2	3	4	5	6	7	8	9	...	54	55	56	57	58	59	60	61	62	63
0	-12493	-8359	2774	5876	-13779	-6770	559	5683	15347	1364	...	372	-3931	-15507	-4143	3083	-1119	16001	3320	-1000	598
1	-12417	-8372	2783	6028	-13713	-6797	549	5818	15295	1429	...	406	-4314	-15443	-4233	3182	-1405	16011	3258	-988	752
2	-12334	-8395	2794	6167	-13619	-6828	536	6006	15224	1488	...	406	-4452	-15358	-4412	3284	-1551	16015	3223	-959	865
3	-12250	-8385	2808	6328	-13561	-6823	525	6135	15181	1476	...	552	-4950	-15197	-4561	3677	-1784	16010	3171	-979	1064
4	-12217	-8268	2707	6577	-13533	-6728	438	6302	15123	1373	...	703	-5160	-15100	-4656	3978	-1709	16015	3153	-970	1025

図 3　ジェスチャーデータの例

図に示しているように，一行のデータは 1 つの静止ジェスチャーを意味しており，これ以降 1 組のデータは 16 個のクォータニオンを構成する 64 個の実数値データを指すこととする．

4. 実装実験

本研究の実験は以下の流れに沿って実装を行った．

1) ハンドモーションキャプチャグローブを用いて，多様々な角度の静止ジェスチャーデータと動く手の動作データをそれぞれ収集する．
2) 静止ジェスチャーデータの前処理を行い，適応可能な DNN モデルを構築する．訓練データで DNN モデルの学習を行った後，テストデータによってモデルを評価する．この評価結果により，ハンドモーションキャプチャグローブで収集された手指データを機械学習による解析の有効性を確認する．
3) 動く手指の動作データの前処理を行い，適用可能な CNN・RNN モデルをそれぞれ構築する．訓練データで 2 種のモデルの学習を行った後，テストデータによってモデルの精度をそれぞれ検証する．

4.1 データの収集と前処理

本研究では，本研究室で製作されたハンドモーションキャプチャグローブを用いて，著者本人の静止ジェスチャーデータと動く手の動作データを多様々な角度から収集した．

具体的には，まず静止ジェスチャーにおいて，多様な角度からパー，グー，チョキ（図 4）のデータをそれぞれ 2 万組ずつ収集し，1 組ごとに 1 つのラベルを付けてサンプリングした．

図 4　ジェスチャー（左から：パー・グー・チョキ）

そして，手の動きのデータとしては，ピアノを弾く動作（図 5）とパソコンのキーボードを打つ動作（図 6）のデータを 3 万組ずつ収集した．ハンドモーションキャプチャグローブによって 1 秒で 30 組のデータが収集されるので，4 秒間で収集した 120 組のデータを 1 つの動きとして定義し，1 つのラベルを付けてサンプルデータとして用意した．

図 5　ピアノを弾く

図 6　パソコンのキーボードを打つ

4.2 DNN によるジェスチャーの認識

手の静止ジェスチャーの認識において，ジェスチャー「グー」であるかどうかの 2 分類問題と，パー，グー，チョキのどれかの 3 分類問題を DNN によって実装したが，ここでは 3 分類問題のタスクを中心に述べる．

まず，データの前処理として，パー，グー，チョキの 3 種のデータに 3 種のラベルを付けた後，訓練データとテストデータに分けた．訓練データとテストデータの詳細情報を表 1 に示し，設計した 5 層のニューラルネットワークのアーキテクチャを図 7 に示す．

表 1　ジェスチャー認識用のデータ数

	訓練データ数	テストデータ数
パー	15000	5000
グー	15000	5000
チョキ	15000	5000
総数	45000	15000

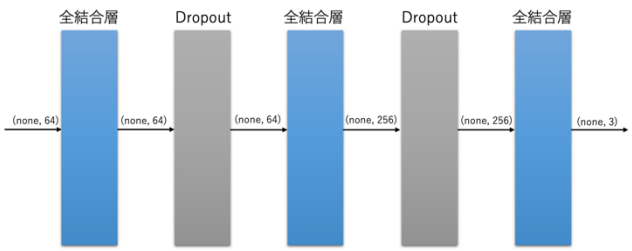

図 7　ジェスチャー認識に用いた DNN のアーキテクチャ

Vol.2021-HCI-195 No.4
Vol.2021-UBI-72 No.4
2021/11/30

ジェスチャーデータの認識に用いたニューラルネットワークは 3 層の全結合層（Fully Connected Layer）と 2 層のドロップアウト層（Dropout Layer）によって構成されている．ドロップアウト層では，全結合層を伝播する値を確立的に伝播させない処理を行う．ドロップアウト層の追加によって過学習を避けることができ，ネットワークの性能を向上できる．また，各層の活性化関数について，中間の結合層では ReLU を利用し，出力層では Softmax 関数を用いて 3 分類の出力を実現した．パラメータとして，学習回数を 20，バッチサイズを 256 に設定してモデルの学習を行った．

図 8 にジェスチャーの認識に用いた DNN モデルの学習曲線とテストデータによる評価曲線を示す．

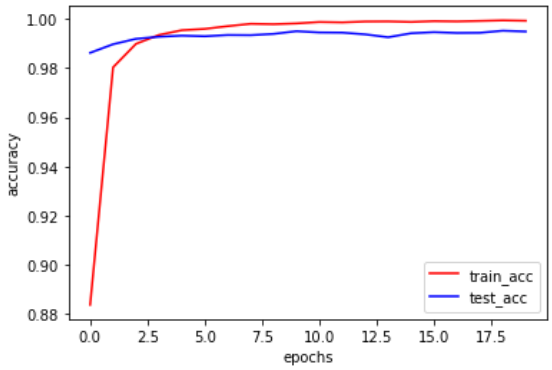

図 8　DNN によるジェスチャーの認識精度

図 8 のグラフから，学習の回数が進むにつれ，モデルによって認識した学習データの精度と評価データの精度が増加していく傾向と，過学習が起きてないことが分かった．この評価結果により，ハンドモーションキャプチャグローブで収集された手指のデータを機械学習による解析の有効性を確認できた．

4.3 CNN による手指の動作認識

次に，動く手の動作データを CNN によって解析する．本研究では時系列データに対して適用したため，画像認識などで用いられる 2 次元の CNN ではなく，1 次元の CNN を用いた．ここでは，1 次元の CNN でピアノを弾く動作とパソコンのキーボードを打つ動作の 2 分類問題を取り組んだ．

データの前処理として，データの収集で述べたサンプリング方法を用いて，6 万組のデータを 120 組ごとに抽出し，120 組に対して 1 つのラベルを付けて 1 つの動作として定義し，498 個の動作データを作成して用意した．そして，80%のデータを訓練データ，残りの 20%をテストデータとして分けた．データの詳細情報を表 2 に示し，設計した 5 層の 1 次元 CNN のアーキテクチャを図 9 に示す．

表 2　手の動作認識用のデータ数

	訓練データ数	テストデータ数
ピアノ	199	50
キーボード	199	50
総数	398	100

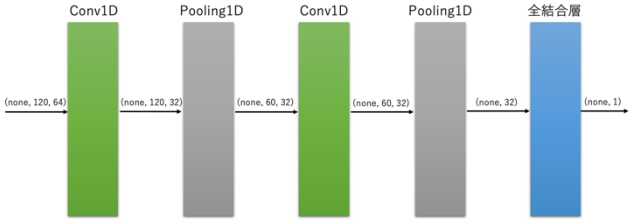

図 9　手の動作認識に用いた CNN のアーキテクチャ

手の動作データの認識に用いた CNN のニューラルネットワークは 2 層の Conv1D（1D Convolutional Layer）と 2 層のプーリング層（Pooling Layer），1 層の全結合層によって構成されている．また，各層の活性化関数について，中間の Conv1D 層では ReLU を利用し，出力層では Sigmoid 関数を用いた．パラメータとして，学習回数を 10，バッチサイズを 256 に設定してモデルの学習を行った．

図 10 に手の動作の認識に用いた CNN モデルの学習曲線とテストデータによる評価曲線を示す．

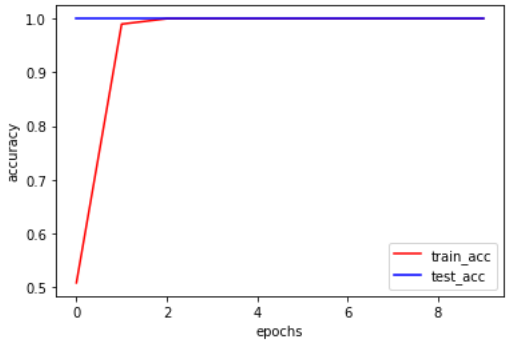

図 10　CNN による手の動作の認識精度

図 10 のグラフにより，サンプルデータの数が少ない場合でも，CNN モデルの訓練と未知のデータの認識を精度高く実現できることが分かった．ハンドモーションキャプチャグローブによって収集された手の関節のクォータニオンの特徴量での認識は，一般な手の画像や動画を用いる動作認識より，精度が高い認識を容易に実現できることも分かった．

4.4 RNN による手指の動作認識

最後は，入力データの依存性を利用するニューラルネットワークの一種である RNN によってピアノを弾く動作とパソコンのキーボードを打つ動作のデータを解析する．

サンプルデータは CNN での実装のサンプルデータを用いたため，データの詳細は表 2 に示す．図 11 に RNN のアーキテクチャを示す．

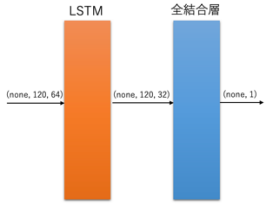

図 11　手の動作認識に用いた RNN のアーキテクチャ

本研究で用いた RNN のニューラルネットワークは比較的に簡単であり，1 層の LSTM 層と 1 層の全結合層によって構成されている．また，出力層では Sigmoid 関数を用いた．パラメータとして，学習回数を 50，バッチサイズを 256 に設定してモデルの学習を行った．過学習を防ぐために事前にデータの正規化も行った．

図 12 に手の動作の認識に用いた RNN モデルの学習曲線とテストデータによる評価曲線を示す．

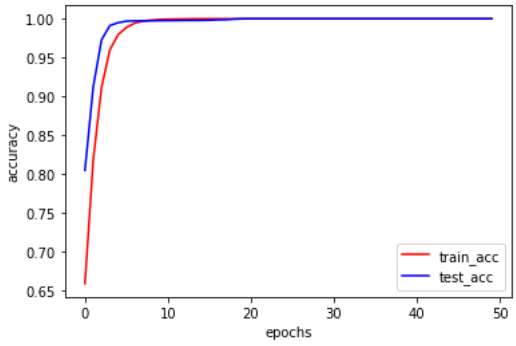

図 12　RNN による手の動作の認識精度

実行により，LSTM 層を用いる場合に学習速度がより速く，精度が高くなる傾向が見られ，その上でより細かいパラメータを調整することで正確に手の動作を認識することができた．

以上の結果により，RNN の動作認識モデルを 1 次元の CNN 動作認識モデルと比較した結果，いずれのモデルでも極めて高い認識精度のモデルを訓練できており，ピアノを弾く動作とパソコンのキーボードを打つ動作のような差が微妙な動作でも精度高く認識できることが分かった．

5. おわりに

本研究では，本研究室で製作されたハンドモーションキャプチャーグローブで手指の動作データを収集し，複数の機械学習手法によって手指の動作を解析するモデルをいくつか実装でき，迅速かつ正確に手の動作を解析できた．結果として，手の静止ジェスチャーデータは DNN モデルによって 99.4%の精度で認識できた．また，動く手の動作データは実装した CNN・RNN モデルによって 99%超えの認識精度で動作を認識できた．

今回の認識モデルでは，サンプルデータの種類が限られているため，認識できる手の動作の種類が限定された．今後の課題として，現実世界の課題解決に応用するために，より多くの手の動きのパターンへ同時に対応できるモデルと，更なる微妙な手の動きの差を認識できるモデルを実装することを目指す．

謝辞　本研究の過程において，終始懇切なる御指導と御鞭撻を賜り，本論文をまとめるに際して，親身な御助言と力強い励ましを頂いた，本研究室の坂井滋和教授に，心より感謝を申し上げます．これまでの研究過程において，有益な議論と情報など交換をして頂いた研究室の先輩方に感謝いたします．

参考文献

[1] “Keras”. https://github.com/keras-team/keras, (参照 2021-10-10).
[2] Tomas Simon and Hanbyul Joo and Iain Matthews and Yaser Sheikh: Hand Keypoint Detection in Single Images using Multiview Bootstrapping, CVPR (2017).
[3] Sijie Yan, Yuanjun Xiong, Dahua Lin: Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition, arXiv:1801.07455 (2018).
[4] 今野紀雄．四元数．森北出版, 2016, p. 12-74.
[5] Antonio Gulli, Sujit Pal. Deep Learning with TensorFlow 2 and Keras: Regression, ConvNets, GANs, RNNs, NLP, and more with TensorFlow 2 and the Keras API, 2nd Edition. Packt Publishing, 2019, p. 22-423.
[6] 瀧雅人．これならわかる深層学習入門．講談社, 2018, p. 19-37, p. 41-63, p. 65-83, p. 87, p. 114-130.
[7] 斎藤康毅．ゼロから作る Deep Learning−Python で学ぶディープラーニングの理論と実装．株式会社オライリー・ジャパン, 2018, p. 21-33, p. 39-81, p. 83-119, p. 123-163, p. 205-238.
[8] Ian Goodfellow. Deep Learning. The MIT Press, 2016, p. 132-243.