

ウェアラブルデバイスを用いた行動の自動分類システムの構築

佐藤慶太¹ 千田正史¹ 早川吉弘² 小林秀幸²
千葉慎二² 高橋晶子² 藤木なほみ²

概要: ウェアラブルデバイスに内蔵されたセンサから取得したデータで解析、分類を行い、人の行動を自動で特定する手法を提案し、その能力を検討する。その手法の実現可能性を探るため、限定された行動を扱うスポーツ競技を対象に解析を行う。特に、筆者にとってデータ収集が行いやすい卓球競技を採用し、その中で打球フォームの分類の自動化を検討する。

Automatic Behavior Classification System Using Wearable Device

KEITA SATO¹ MASAFUMI CHIDA¹ YOSHIHIRO HAYAKAWA²
HIDEYUKI KOBAYASHI² SHINJI CHIBA²
AKIKO TAKAHASHI¹ NAHOMI M. FUJIKI²

1. 序論

近年、物体(モノ)に通信機能を持たせ、インターネットを経由し、センサ情報を取得する、IoT と呼ばれるシステムを利用した研究が進んでおり^[1]、人の行動認識を用いた様々なサービスの可能性が見えてきた。従来の研究^[2]における行動認識はモデルとなる行動データと任意の行動データの特徴量を比較し、パターンマッチングを用いて行動の推定を行うものが多くみられる。しかし、本研究では、収集した行動データを機械学習によって分類を行った後、分類したデータの意味付けを行う。さらに、機械学習で分類を行わせることで、人が直感的に特徴を判別出来ないような多量のデータから詳細な動作の特定ができると期待される。しかし、センサデータのみで人物などの行動の意味づけを行うことは困難であり、システム構築時には動画像など行動の意味づけが容易なものと組み合わせて行動を定義する必要がある。本研究では、センサデータと合わせて画像データを記録し、さらにそれを機械的に分類し、総合的に行動を定義することで容易な行動特定が可能であると考えている。また、動画像から行動を分類する際、その画像を低次元で表現することが出来れば、センサデータとの比較もしやすく、センサでの行動特定法の確立を行う際に活用できると考えている。この時、階層型ニューラルネットワークの学習機能を用いることでデータの低次元化を試みる。この階層型ニューラルネットワークの層を多層にしたディープラーニングを使用することで特徴を抽出することに成功しており、「Google の研究開発によってコンピュータが猫を認識できるようになった」^[3]というニュースが話題となった。

本研究で用いる砂時計型の階層型ニューラルネットワークを使用し、恒等写像学習を行わせると、絞り込まれた中間層に抽象化された低次元の特徴空間が作られ、そこに入力データの特徴が写像される。例えば、手書き文字を学習することで文字の傾き具合などの連続的特徴を低次元の特徴空間で表現できる^[4]。

以上より本研究では、対象となる人物がウェアラブルデバイスを装着し、取得したセンサデータを機械学習の手法を用いて解析する。センサデータ取得と並行して、対象となる人物の行動が含まれる画像を同時に録画し、そのデータを合わせて解析することで、それぞれの解析から人物の行動を特徴空間上に写像し、抽出された特徴を利用して行動の特定を行う手法の確立を目的とする。特に本発表では、センサデータと画像データの両面からの行動特定手法の確立を目指すため、限定された行動を扱うスポーツ競技を対象とした。具体的な行動として卓球ラリーのフォームを採用し、行動特定手法の検討を行う。

2. 提案手法の説明と調査方法

(1) 本実験の調査方法

本実験では、卓球のラリーを行うプレイヤーのラケットを持つ手にウェアラブルデバイスを装着し(図 2.1)、3方向の手の動きに対応した加速度データを取得する。同時に固定されたビデオカメラを用いて真横から撮影を行った。卓球のフォアハンドとバックハンドから成るラリー50回を対象の行動とし、1回のスイングは返球された球が卓球台についてから再び台に着くまでと定義する。

1 仙台高等専門学校 専攻科
2 仙台高等専門学校



図 2.1 ウェアラブルデバイス装着時

(2) センサデータの解析手法

今回は、polar 社製の Android Wear(m600)に内蔵されている加速度センサを使用し、Android Wear を基準とした 3 軸データを 100 分の 1 秒毎に取得する。この時、加速度の向きは図 2.2 に示す通りである。



図 2.2 ウェアラブルデバイスの加速度方向

このようにして取得されたセンサデータを動画と対比させ、スイングにかかる時間を特定し、タイムスタンプに基づいて各スイングにおける加速度データに分割を行う。1 スイングに含まれる加速度データを 1 つのパターンベクトルに変換し、高次元空間上で 1 つの点として扱う。そこで、サンプリング時間毎に取得した各成分を並べていき、ベクトルを作成する。このベクトルを使用してクラスタリングを行う。この際、用いた手法は k-means 法であり、複数のクラスに分類を行い、スイングの特定を行う。分類結果に対応する動画を確認することで、抽出されているフォームの特徴を視覚的に確認できることになる。

(3) 画像データの解析手法

(2)の加速度データの取得と並行して SONY のハイスピードカメラ(アクションカム HDR-AS300, 800×480)で撮影を行った動画(240fps)のフレーム画像(100×100)を入力とし、階層型ニューラルネットワークを用いてディープラーニングを行う。この時、ネットワークのモデルは図 2.3 に示すような砂時計型のもを用いる。砂時計型とは入力層

から中間層へ向かうにつれニューロン数を減少させ、中間層から出力層へ向けてまたニューロン数を増加させる、中間層で一度絞り込まれた構造である(近年では、オートエンコーダと呼ばれている)。入力値と教師データに同じ画像データを用いて、教師有り学習を行うことで、中間層の一番少ないニューロン数の層へ入力データを圧縮し、その特徴を抽出するのが今回のディープラーニングである。ディープラーニングの学習には、出力と教師データとの誤差を下層へ伝搬させる、誤差逆伝搬法(バックプロパゲーション)^[3]を用いる。このネットワークモデルでは中間層のニューロン数が入力層や出力層のニューロン数よりも少なくなっているため出力に入力データと同じ値を出すような恒等写像を学習することが出来れば、中間層の一番少ないニューロン数だけで入力データの値を表現できたことになる。恒等写像学習で特徴空間内に写像された卓球のラリーにおける各スイングの軌道を調査し、スイングの分類を行う。

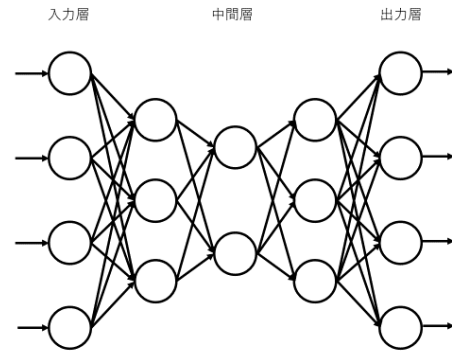


図 2.3 階層型ニューラルネットワークのイメージ

3. センサデータによる解析

(1) 1 スイングの特徴調査

Android Wear を腕に装着して行った卓球のフォアハンドとバックハンドから成るラリーの収集データから 1 スイングの特徴を調査した。タイムスタンプに基づいて加速度(センサ)データを切り出し、個々のスイングの加速度変化に対応する動画と比較する事で解析を行った。図 3.1(a)から(c)はフォアハンド、図 3.2(a)から(c)はバックハンドにおける加速度の各成分をグラフ化したものである。横軸に時間、縦軸に加速度を取っている。用いたセンサの分解能は 0.01 秒であるためフォームとしての情報を加速度で正確に表現していると言える。そこで、フォアハンドとバックハンドにおける加速度の変化を比較した。図 3.1, 図 3.2 より、x, y成分はどちらの加速度も同様の変化をしているが、z成分に着目すると、ピークが逆向きになっている。今回、対象とする卓球のスイングでは、ボールを打つ瞬間の加速度が最も大きいことに加え、フォアハンドとバックハンドでは打つ瞬間のz方向の向きが反対となる。したがって、z成分

によりフォアハンドかバックハンドかの判定をすることが出来る。

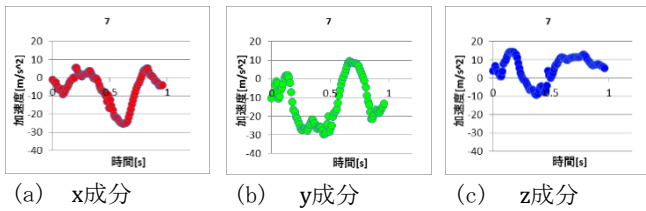


図 3.1 フォアハンドの加速度変化

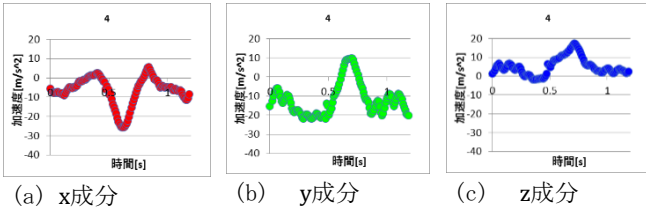


図 3.2 バックハンドの加速度変化

(2) 行動の自動分類

3. (1)の結果からz成分でフォアハンドかバックハンドを判定できると考えられるため、スイング毎に分割したセンサデータのz成分から作成したベクトルを使用して3つのクラスに分類を行った。映像を基に図 3.7 の結果を検証すると、フォアハンドとバックハンドの2つにデータを分類することが出来た。クラス1、クラス2がフォアハンド、クラス3がバックハンドとなっている。分類結果に対応したスイング画像の一例を図 3.4(a), (b)に示す。フォアハンドとバックハンドではラケットを振る向きがz方向において反対であるため、このように分類できたと考えられる。

次に、同じフォアハンドの中でクラス1とクラス2に分類される違いを明確に調べるため、同時に録画を行った動画画像を確認し、行動の意味づけを行った。その際、対象となるスイングは分類した各クラスの中心に最も近いデータをそのクラスの代表点として定義し、クラス1, 2の代表点におけるフォームを視覚的に確認した。その画像を図 3.5(a), (b)に示す。その結果、同じフォアハンドの中でクラス1は返球した後、次のスイングに備えるため、ラケットを元の位置に戻すまでの動作に遅い傾向が見られ、クラス2ではラケットを戻す動作が速い傾向が確認できた。さらに、クラス1には比較的返球されたボールが遅く、プレイヤーがボールを返球するタイミングも遅い傾向にあり、クラス2には相手の返球が速く、プレイヤーの打つタイミングも早い傾向にあった。このことから、相手の返球が比較的に遅い場合、相手のスピードに合わせて返球をしたため、フォームが乱れたと考えられ、相手の返球が比較的に速い場合は平常時のフォームで返球を行ったため、ラケットを戻す動作が速かったと考えられる。一方でいくつかの例外となるケースも存在した。それは、本データの多くは腕の動きが静止するタイミングが存在するが、返球されたボールのコースにより、プレイヤーが立っていた場所から大幅にず

れた場所で返球行った場合、腕は静止しているが、体が動いてしまっているためにセンサに体の移動の加速度情報が入力されてしまい、誤った分類がされたものと解釈される。

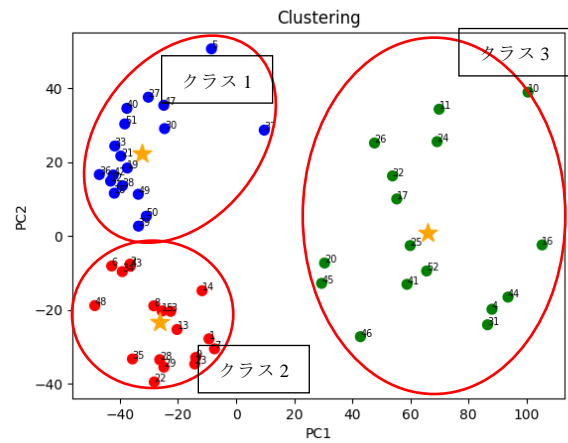


図 3.3 z成分の分類結果

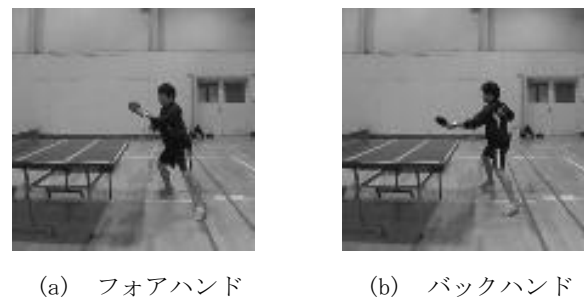


図 3.4 分類結果

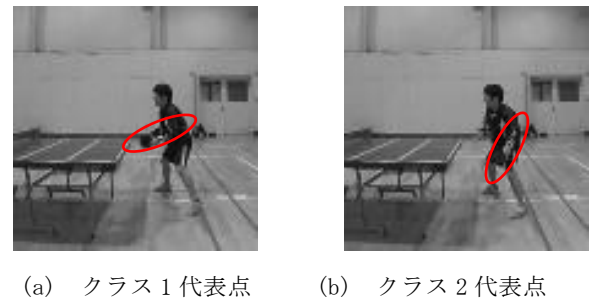
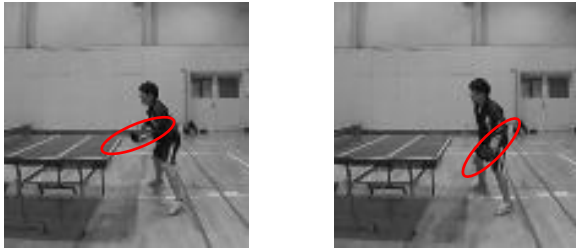


図 3.5 代表点の画像

各クラスの他のデータも比較してみると、代表点と同様の特徴が見られた。その例を図 3.6(a), (b)に示す。さらに再現性を高めるため同様のフォームからなる別のラリー50回についても解析を行い、等しい分類がされるかを検証した。別のラリーにおいても同様にフォアハンドとバックハンドに分類を行うことが出来た。



(a) クラス 1 (b) クラス 2

図 3.6 クラス内の他データ

4. 動画データによる解析

(1) スイングの軌道生成

本実験では図 4.1 のネットワークモデルを使用し 20 回目までのスイングを学習させ、学習データを含んだ 50 回のフォアハンドとバックハンドからなるラリーをテストデータとして処理を行った。ここでの学習とは、ネットワークの出力と入力との 2 乗誤差 E を減少させるように各ニューロン間のシナプス荷重値 w_{ij} を更新していくことである。学習データが n 個の時の誤差の計算式を以下に示す。ここで t_i は i 番目のニューロンの出力に提示する正解データの値、 o_i はネットワークに入力した際の出力データの値を表している。

$$E = \frac{1}{2} \sum_{i=0}^n (t_i - o_i)^2$$

この誤差を用いて荷重値を更新し、出力データを正解データに近づけていく。学習 k 回目の荷重値の更新式を以下に示す。荷重値の変化量を計算し、それを加算することで荷重値を更新していく。ここで η は学習係数と呼ばれ、最荷重値をどの程度更新するかを制御する値である。また α は安定化定数と呼ばれ、前回の荷重値の更新量を用いて振動を抑える効果がある。

$$\Delta w_{ij}(k) = -\eta \frac{\partial E}{\partial w_{ij}} + \alpha \Delta w_{ij}(k-1)$$

入力には図 3.8 のような 100×100 のモノクロ画像を 2 値化データに変換し、全体で約 12000 枚の画像を用いる。学習には確率的勾配降下法^[4]を用い、学習係数 η を 0.001、モメンタム α を 0.8 にして 100,000 回の学習を行った。その時の特徴空間内に写像される軌道について調査を行った。出力となる中間層のニューロン数を 3 つとし、それぞれの出力を neuron1, neuron2, neuron3 と定義する。

フォアハンドとバックハンドの軌道で特徴が顕著に見られた例を以下に示す。図 4.2, 図 4.3 からわかるように、フォアハンドとバックハンドの軌道では大きく違うことが確認できる。また、画像から軌道の分類を行ったところ、フォアハンドの特徴として図 4.2 のように 8 の字を描くよ

うに軌道が表現される傾向がみられた。バックハンドの特徴としては図 4.3 のように一度、内側を通るように軌道を描く傾向が多くみられた。

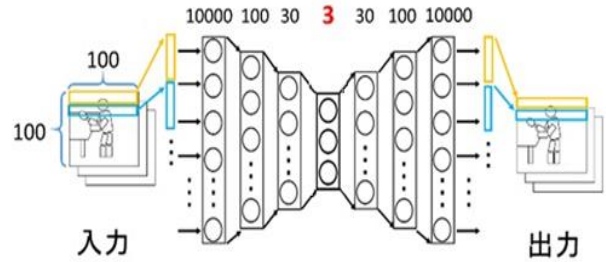


図 4.1 使用したモデル

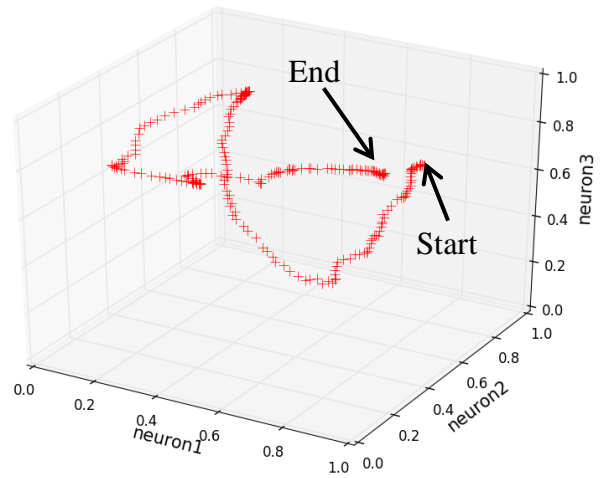


図 4.2 フォアハンドの軌道

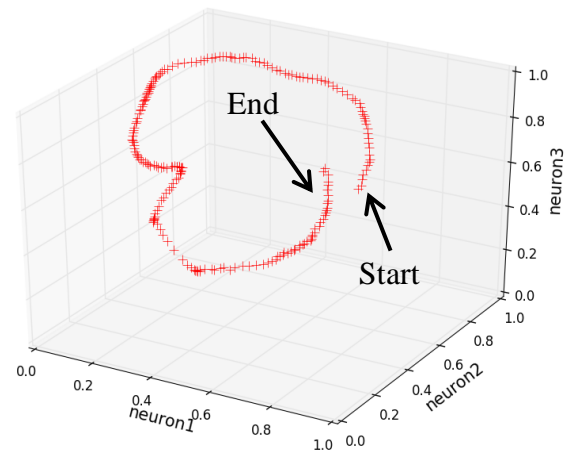


図 4.3 バックハンドの軌道

(2) 行動の分類

各スイングの画像データを 3 つの要素に圧縮して得られた軌道の図を、さらに実際の画像と参照しながら詳しく解析を行った。フォアハンド軌道の中で 8 の字を描くような傾向がみられなかったものに対し分類を行ったところ、図

4.4 のようなバックハンドを打った後にフォアハンドを打っているときの軌道は前述のフォアハンドと比べ **neuron3** における出力の変化が小さい傾向があることが見て取れる。また、ラリーを横から撮影した際に画面の奥でフォアハンドを打っているときの軌道(図 4.5)は図 4.2 の軌道と比べると小さく、**neuron2** における出力の変化が小さいことが分かる。このような結果から、フォームの順番や球を打った際のプレイヤーの位置などで特徴が表れていると考えられる。しかしながら、センサデータと同様に返球されたボールのコースにより、プレイヤーが立っていた位置から大幅にずれた場所で返球行った場合、フォアハンド、バックハンドの両方で図 4.6 に示すように大きな特徴が軌道に表れず、分類が困難であった。これは、テストデータが予め学習させたデータではなかったためこのような軌道を描いたのだと考えられる。

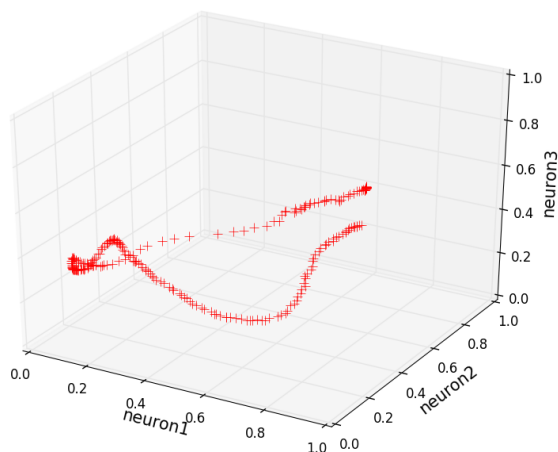


図 4.4 バックハンド後のフォアハンド軌道

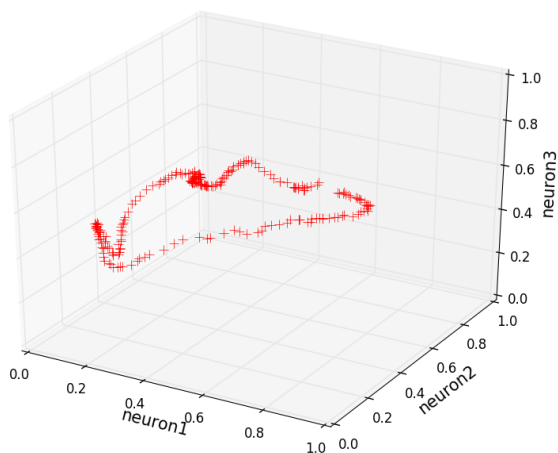


図 4.5 画面奥でのフォアハンド軌道

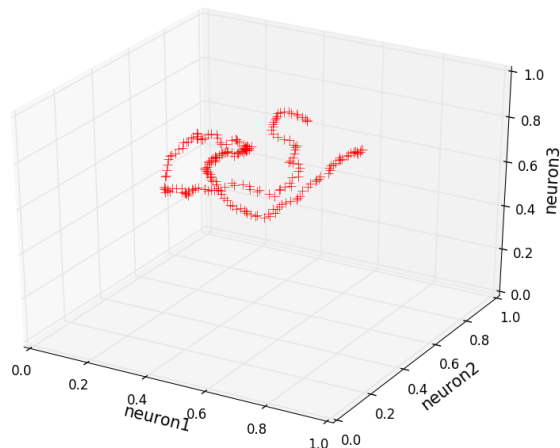


図 4.6 大幅にずれた位置からのフォアハンド軌道

5. おわりに

本論では、卓球のフォアハンドとバックハンドからなるラリーを対象に機械学習を用いて解析を行うことでその結果からフォームの分類の可能性を確認した。3 軸加速度データにおける解析は腕に装着した **Android Wear** を通じてラケットを振る動作に関しての各クラスの特徴を得ることができた。今回は、腕のみにセンサを装着して実験を行ったが、今後はより実用を想定した複雑な行動への適用を考える必要がある為、複数台のセンサを装着し体全体のフォームに関する、より高度な分類が必要であると考えられる。

一方、画像データにおける解析はフォアハンドの中でもフォームの順番や返球時の立ち位置で特徴空間内の軌道に変化が見られ、同じフォーム内の違いも軌道から分類ができる可能性を確認した。現段階では、軌道を用いてデータの分類を行い、人間の目で画像を確認し、行動の定義を行っている。今後は、センサデータの分類と画像データの分類を複合的に用いてより正確な動作特定を行っていく予定である。

参考文献

- [1] 吉田尚平, 村田嘉利, 鈴木彰真, 佐藤永欣, 「ウェアラブルデバイスによる歩行携帯判別システムの提案」, DICOM2016 シンポジウム論文集, pp.301-307, 2016
- [2] 川添恭平, 中居新太郎, 守井清吾, 青木功介, 「スマートウォッチを用いたモーション認識システムの開発」, INTEC TECHNICAL JOURNAL 第 17 号, pp.62-67, 2016
- [3] Quoc V. "Building High-level Features Using Large Scale Unsupervised Learning" ICML 2012
- [4] 大沼峻徳, 早川吉弘, 「手書き文字の連続的特徴表現の抽出」, 電子情報通信学会基礎・協会ソサイエティ/NOLTA ソサイエティ大会講演論文集, p.367, 2016.
- [5] 平野廣美, 「C++とJava でつくるニューラルネットワーク」, パーソナルメディア株式会社出版, pp.25-33, 2008
- [6] 岡谷貴之, 「深層学習」, 講談社, 2015
- [7] 佐藤慶太, 千田正史, 大沼峻徳, 小林秀幸, 高橋晶子, 千葉慎二, 藤木なほみ, 早川吉弘, 「画像とセンサデータを組み合わせた卓球のフォーム解析 I, II」, 第 22 回高専シンポジウム in Mie 講演要旨集, F-09, F-10, 2016