

弱教師あり学習による加速度センサを用いた 動作カウント手法の提案

西野 祐希¹ 前川 卓也¹ 原 隆浩¹

概要：近年のセンサ技術の発展に伴い、ウェアラブルセンサから得られる加速度データを用いた人間の行動認識の研究が盛んに行われている。また、ワークアウトなどの応用においては、動作の認識だけではなく、動作の繰り返し回数をカウントする研究も行われている。既存のカウント手法では、各動作の出現時刻等の情報を含むラベルあり学習データが大量に必要であり、その取得コストが膨大であるという問題がある。そこで本研究では、弱教師あり学習により、与えられた加速度データに含まれる特定の動作をカウントするための手法を提案する。本研究では、加速度データに含まれる動作の数のみをラベルとする学習データが少量のみ与えられる条件を想定する。このような弱いラベルが付与された少量学習データで効果的に動作カウントモデルを学習するため、提案手法ではカウントモデルを動作の出現時刻を特定する Attention 機構と、特定された出現の回数をカウントする方策を学習する機構から構成する。出現の回数をカウントする方策を学習することができれば、学習データのラベル（動作の回数）のバリエーションが少なくとも、学習データに含まれていない回数のカウントを行うことができる。さらに提案手法では、Data Augmentation や、他のユーザから得られた学習データを用いたモデルの事前学習により少量学習データの問題に対処する。評価実験では、スマートウォッチを右手首に装着した3人の被験者がワークアウトを行った際に得られた加速度データを用いて、提案手法の有用性を確認した。

Investigation on Motion Counting Method Using Accelerometer by Weekly Supervised Learning

YUKI NISHINO¹ TAKUYA MAEKAWA¹ TAKAHIRO HARA¹

1. 序論

近年、コンピュータの小型化や軽量化が進み、加速度センサなどの様々なセンサが搭載されたスマートウォッチなどのウェアラブルセンサの普及が進んでいる。ウェアラブルセンサは健康管理や、ワークアウトの支援など人々の生活の質の向上に役立っている。ウェアラブルセンサの普及に伴い、ウェアラブルセンサから得られる加速度センサデータを用いて人間の行動を認識する研究が行われてきた [3]。例えば、スマートウォッチの加速度センサから得られたデータから、機械学習などを用いてユーザの行っているワークアウトの種類を認識するような研究などが行われている [2], [6]。また、行われているワークアウトの種類を認識するだけでなく、そのワークアウトを何回行ったかを

カウントするような研究も行われている [1], [9], [10]。

しかし、既存のカウンティング手法では特定の1種類の動作のみが含まれる加速度データを入力とするものが多く、複数の動作や対象とする動作以外の動作が含まれる加速度データの中から、特定の1つの動作のカウントを行うことはできない。実際にワークアウトを行う際、複数の動作（腕立て伏せ、腹筋、スクワットなど）を連続もしくは、交互に行うことは多い。また、ワークアウトの途中では、汗を拭く、水を飲むなどのカウント対象とならないような動作も行われる。このような加速度データの入力に対して、それぞれの動作が加速度データのどの部分に対応するかを認識してから、特定の動作が行われた回数をカウントする既存手法も存在する。この手法では、動作繰り返しの開始位置を検出し、動作繰り返しの開始位置をカウントすることで動作回数をカウントしている。この手法の場合、加速

¹ 大阪大学大学院情報科学研究科

度データに対して動作認識を行うためのラベリングに加えて、1回の動作の開始位置のラベリングを行う必要がある。このようなラベリングには、データ収集時のビデオ録画など多大なコストを必要とし、大量の学習データを用意するのは困難である。

そこで本研究では、複数の動作が混在している加速度データを入力として、対象の動作が行われた回数の情報のみをラベルとする弱教師あり学習を行うことで、end-to-endで動作のカウントを行うニューラルネットワークモデルを提案する。提案モデルの学習において、入力データに対するラベルはカウント対象となる動作の実行回数のみであるため、データ取得の際に対象の動作を行った回数を記録しておくだけでラベリングができる。そのため、加速度データに対する動作が行われた時刻のラベリングが必要なく、既存手法と比べてラベリングコストを低減できる。また、機械学習モデルの学習には一般的に大量の学習データが必要となる。本研究では大量の学習データを用意するコストを削減するため、少ないデータセットでも学習を行えるような弱教師あり学習モデルを設計する。

少量学習データによる弱教師学習を実現するため、本研究では人間によるカウント方法のロジックを模してネットワークを構成する。人間はカウントの対象の位置を特定し、特定した回数を数えてカウントを行っている。このようなロジックを模して、ネットワークをカウント対象動作の出現位置の特定を行うような特徴抽出ブロックと、特定された出現位置をカウントする方策を備えるカウントブロックの2つのブロックで構成する。カウントブロックでカウントの方策（数を数える行為）を学習できれば、学習データに含まれていない動作回数を認識することができると期待される。

具体的には、特徴抽出ブロックでは、Attention機構を用いることで、ラベルである対象の動作の実行回数からその対象の動作のセグメントに注目して、その動作の出現位置で極大値（ピーク）を示すシーケンスを出力する。カウントブロックでは、特徴抽出ブロックの出力シーケンスを入力とし、対象動作の出現回数をカウントする。動作の出現位置でピークを示すシーケンスをカウントブロックの入力とすることで、カウントブロックはピークの数を実数としてカウントするように学習されると期待されるため、学習データに含まれていない数のカウントを行うことができるようになる。そのため、カウントを行いたい回数すべてのラベルを持つ学習データを集めなくても学習が可能となる。しかし、学習に用いるデータが極端に少ないと特徴抽出ブロックとカウントブロック双方の学習を行うことが不可能となる。そこで、実際の加速度データに対してデータ拡張を行い、少量の学習データから多様な学習データを生成する。画像分類などのタスクにおいて、学習データが少ない際にデータ拡張を行う手法はよく用いられている [8]。本研究

では、加速度データに対して振幅を変化させたり、分割して順序を入れ替えたりすることで元の加速度データと同じラベルを持つ新たな加速度データを生成し、学習データを拡張する。また1つの加速度データをコピーして連結することで、ラベル（動作回数）が2倍のデータを作成してラベルのバリエーションを増やす。さらに、他のユーザの学習データを用いてモデルを事前学習しておくことで、対象とするユーザに必要とされる学習データを削減する。

2. 関連研究

ウェアラブルセンサを用いた動作の繰り返し回数をカウントする既存手法とデータの数を拡張する既存研究について紹介する。

2.1 ウェアラブルセンサを用いた動作カウント手法

Soroらが提案した動作カウンティング手法 [10] では、センサデータに対してどの部分でどの動作が行われているかを認識し、カウント対象の動作が行われている部分に対して繰り返し回数をカウントする。この手法では、動作の認識を複数の Convolution 層から構成される CNNで行っている。動作の認識の結果から得られた1種類の動作のみが含まれるデータセグメントをウィンドウに分割し、ウィンドウごとに繰り返し動作の開始部分が含まれているかどうかを判定するようなネットワークを学習する。ウィンドウに対して繰り返し動作の開始部分が含まれていれば1を返し、含まれていなければ0を返すことで、最終的に1と0の二値シーケンスを出力として得る。シーケンス中に含まれる1の数が動作繰り返し回数となる。この手法においては動作を認識するためのラベリングと繰り返し動作の開始時間のラベリングが必要となる。

他にもウェアラブルセンサを用いた動作カウント手法が提案されている [5], [9] が、どの手法も入力に対する動作の種類を認識結果から得られた1種類の動作のみが含まれるデータセグメントに対してカウントを行っている。動作の種類を認識する手法には、Soroらの手法と同様にCNNが採用されているが、カウントにはデータセグメントに対するピークの検出を行うことでカウントしている。

Skawinskiらが提案している手法 [9] では入力加速度データに対して繰り返し部分を強調し、ピークを検出することでカウントを行なっている。まず、入力に対して主成分分析 (PCA) [7] を行い、第一主成分のみを抽出することで、3次元の加速度データから1次元の時系列データに変換する。第一主成分のみを抽出することで繰り返しの波形に対して動作が行われた部分を強調することができる。次に1次元の主成分から閾値を設定してピークの検出を行う。ピーク検出には手動で設定したピーク間の最小距離、ピークの最小値などのパラメータを実際のデータを参照して手動で設定する必要がある。

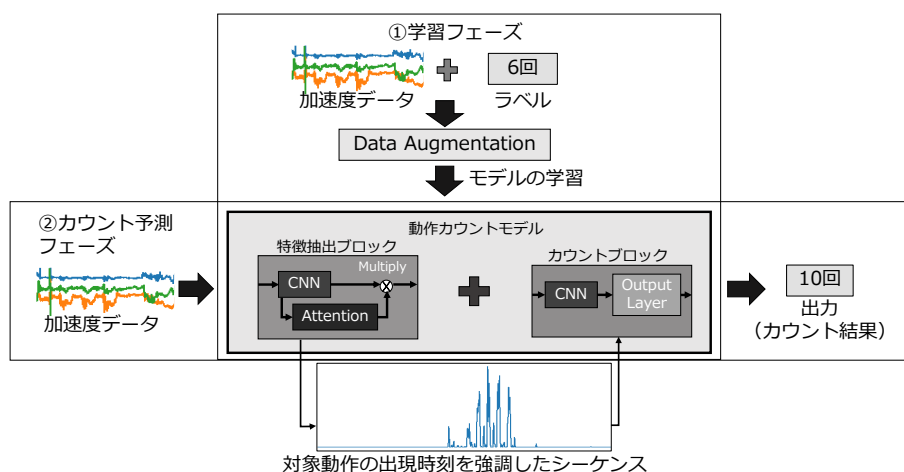


図 1 提案手法の概要

上記の既存手法はどれも入力データに対してどの部分でどの動作が行われているかを認識した後に、対象動作のセグメントからカウントを行っている。そのため、データを収集する際に、動作の開始時刻や終了時刻のラベリングを行う必要があり、多大なラベリングコストを必要とする。本研究では、ラベリングコストを低減するために、動作の認識を行わずに与えられた入力データから動作の繰り返し回数を直接カウントするネットワークを提案する。

2.2 データ拡張

本研究では、学習データの不足を解消するために学習データの拡張を行う。センサデータを用いた行動認識における Data Augmentation の研究を紹介する [11], [14]。Umらは、以下に示すようなノイズ付加、切り抜き、次元の入れ替えなどを用いてデータ拡張を行った [11]。

Jittering センサデータに対してノイズを付加

Scaling ウィンドウごとにランダムな値を掛け、センサデータの振幅を変化

Rotation センサ配置をランダムに入れ替え（次元の入れ替え）

Permutation ウィンドウを1から5個のセグメントにスライスし、順序をシャッフル

MagWarp ウィンドウに対して滑らかな曲線で畳み込み演算を行い振幅を変化

TimeWarp 時間間隔を滑らかに歪ませてセンサデータの時間的位置を変化

Cropping センサデータの切り抜き

Umらは上記の手法を多様に組み合わせ、評価実験を行っており、Rotation, Permutation, Timewarp, これらを組み合わせた手法が最も分類精度の向上に寄与することを明らかにした。彼らの手法以外にもセンサデータに対してダウンサンプリングを行ったり、データをランダムにスライスしてシャッフルするような手法も提案されている [14]。

本研究ではこれらの手法を参考にして、動作の繰り返し回数のカウントのタスクに支障を及ぼさないように Data Augmentation を行う。本研究で用いた Data Augmentation の詳細な手法については3章に記す。

3. 提案手法

3.1 想定環境

本研究では、ウェアラブルセンサを装着してワークアウトを行うことで3次元の加速度データを取得し、それを解析することで動作のカウントを行う。特定の長さを持つ1つのセッションに複数の動作（腕立て伏せ、腹筋など）が含まれており、カウント対象の動作が繰り返し行われていることを想定する。セッションの中にはカウント対象の動作を切り替える際に発生するカウント対象外の動作（水分補給など）も含まれる。本研究ではラベリングコストの低減を目的の1つとしているため、学習用のラベルを対象動作の実行回数のみとする弱教師あり学習を行う。そのため、得られた加速度データにはネットワークの学習用ラベルとして、カウント対象動作の実行回数のみを付与する。また、対象とするユーザから取得する学習データは少量であるため、ラベルのバリエーションが不十分であると想定する。対象ユーザとは異なるユーザからは、対象ユーザと比べて多量の学習データが得られるものとする。ただし、人によって同じ動作でも得られる加速度データには差異がある。

3.2 提案手法の概要

本研究の弱教師あり学習で問題となるのは、複数の動作が混在する加速度データの中でカウント対象の動作の出現位置の情報が無い点である。そのため、カウント対象の動作の出現位置を捉え、動作の実行回数をカウントできるネットワークを作成する必要がある。本研究では、複数の動作が混在している加速度データから対象とする動作のカ

ウントを少ない学習データで行うために、特徴抽出ブロックとカウントブロックの2つのブロックで構成されるネットワークモデルを提案する。図1に提案手法の概要を示す。特徴抽出ブロックは対象動作のセグメントを認識し、その出現時刻に極大値（ピーク）をもつシーケンスを出力する。そして、カウントブロックでは特徴抽出ブロックで出力されたシーケンスを入力としてピークの出現回数をカウントすることで、対象の動作の繰り返し回数をカウントする。ネットワークを2つのブロックに分けて構成しているのは、カウントブロックに動作の出現位置を示すシーケンスから動作出現回数をカウントするという行為を学習させることで、正解値のバリエーションが少ない学習データで学習を行っても、学習データに含まれない出現回数のカウントを行えるようにするためである。

ただし、学習データが少ないとそもそも学習を行うことができないため、Data Augmentation を用いてラベルあり学習データから学習に十分な量のデータを生成する。提案手法では、まず他のユーザのラベルありデータを拡張して、モデルの事前学習を行う。その後、対象とするユーザの少量のラベルありデータを拡張し、事前学習したモデルの再学習を行う。

3.3 ネットワーク構成

図2に提案ネットワークモデルの構成を示す。提案ネットワークを構成する2つのブロックの詳細について説明する。

3.3.1 特徴抽出ブロック

特徴抽出ブロックでは、加速度データを用いた行動認識でしばしば用いられるCNNを基に特徴抽出を行う。本研究では入力データが3次元加速度データであるため、時間方向の畳み込みを行うような1D Convolution層を含むCNNを用いる。図2のように3層の1D Convolution層とMax Pooling層を交互に重ねてCNNを構築する。特徴抽出ブロックでは、カウント対象の動作に注目する必要がある。そこで、タスクに重要な部分系列に注目するAttention機構を用いる。Attention機構は自然言語処理や画像分類などのタスクにおいて、重要な単語に注目したり、画像の中で分類する対象のオブジェクトに注目するために用いられている[12], [13]。特徴抽出ブロックの出力 M は式(1)のように計算される。

$$M = F \otimes A \quad (1)$$

$$A = \sigma(F') \quad (2)$$

$$F' = \text{Conv2}(F) \quad (3)$$

⊗ は要素ごとの積を表しており、Attention機構の出力 A とCNNから得られる特徴 F を掛け合わせるような構造とすることで、入力データの重要な部分を注目するように学

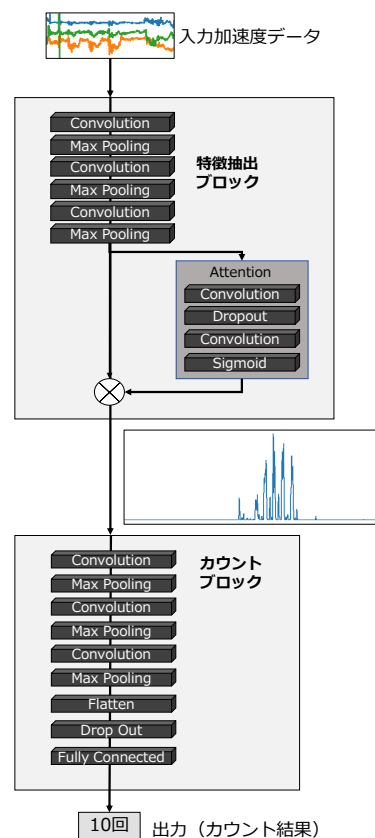


図2 提案ネットワークの構成

習される。Attention機構 A の出力は式(2)のように求められる。 $\sigma(\cdot)$ は Sigmoid 関数を表しており、Attention機構に含まれる Convolution 層から得られる特徴 F' の各時刻の値を 0~1 の値に変換して出力 A を得る。そうすることで Attention 機構の出力 A は時系列入力データに対して重要な部分には 1 に近い値、そして重要でない部分に 0 に近い値となるように学習される。Attention機構が重要な部分を注目するように学習されることで特徴抽出ブロックの出力はカウント対象の動作に注目し強調したような時系列データとなる。また、 $\text{Conv2}(\cdot)$ は Attention 機構を構成する2層からなる 1D Convolution 層に入力した出力を表す。CNN の Convolution 層のフィルタの数を 8 としているため、 M は 8 次元の時系列データであり、各次元を足し合わせて 1 次元にしたものをカウントブロックへの入力とする。図2で示されている時系列データは 1 次元化した特徴抽出ブロックの出力の例である。

3.3.2 カウントブロック

カウントブロックでは与えられた 1 次元時系列データからピークの出現のカウント回数を出力するため、出力を実数の値とする必要がある。本研究では出力を実数値とするために、活性化関数には入力値に重みを掛け合わせてそのまま出力値とするような linear 関数を用いる。図2のように3層の 1D Convolution 層と Max Pooling 層、平滑化を行う Flatten 層、過学習を防ぐための Dropout 層、そして

出力を1つの値にまとめる全結合層（活性化関数に linear 関数を用いる）でカウントブロックを構成する。

3.4 学習フェーズ

対象および他のユーザから得た学習データに対して、Data Augmentation を用いて学習用データを拡張する。そして、他のユーザの拡張データで事前学習を行い、対象ユーザの拡張データで再学習を行う。

3.4.1 Data Augmentation

提案ネットワークモデルの学習データを拡張する Data Augmentation について説明する。Data Augmentation では1つの加速度データ系列から複数の新しい系列を生成する。まず、加速度データ系列を動作が行われていない位置でスライスして複数のセグメントに分割する。分割したそれぞれのセグメントに対して時間方向の伸縮、振幅の増減、加速度値のシフトなどの操作を行い、最終的にセグメントをランダムに並び替える。加速度データ系列をスライスする際、カウント対象の動作が途中で途切れてしまうことを防ぐために、動作が行われていない位置でスライスする必要がある。動作が行われている位置では加速度の値が大きく変化する一方、動作が行われていない位置では加速度の値の変化が小さい。そこで、加速度データ系列のウィンドウごとの標準偏差を計算し、標準偏差が小さいウィンドウを見つけることで動作が行われていない位置を見つける。図3は加速度データ系列に対してスライスを行う位置の例である。赤く色掛けされている部分はカウント対象の動作が行われている部分を示しており、黒い太線はスライスを行う位置を示している。この図からカウント対象の動作が行われている部分ではスライスが行われないことが分かる。繰り返される毎回の動作は、そのスピードやフォームが若干異なり、得られる加速度データも異なる。そのような多様性を考慮して拡張を行うために、加速度データ系列のスライスによって得られた各セグメントに以下の処理を順に行う。

時間方向に伸縮

1回の動作を行う時間に多様性を持たせるために、ダウンサンプリングやアップサンプリングなどを用いて時間方向の伸縮を行う。線形補間を行いながらランダムにサンプリングレートを変換させることで、各セグメントの時間方向における伸縮を行う。

振幅の増減

加速度データ中の1回の動作によって得られる波形の振幅に多様性を持たせるために振幅の増減を行う。セグメントごとにランダムな値を加速度の値に掛け合わせることで振幅の増減を行う。

加速度値のシフト

加速度センサの装着方向のずれにより、得られた加速度データの値に加速度のシフトが生じる場合がある。

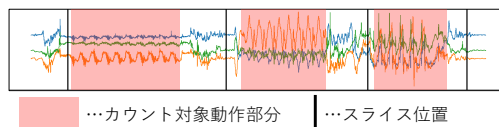


図3 加速度データを動作が行われていない部分でスライス

表1 3人の被験者に関する身体情報および取得加速度データ数

	取得加速度系列数	性別	身長 (cm)	体重 (kg)
被験者 1	40	男性	182	85
被験者 2	20	男性	176	65
被験者 3	20	男性	164	68

このような値のずれを考慮して学習を行えるように加速度の値をシフトしたようなデータを生成する。セグメントごとに加速度データの値にランダムな値を加算、もしくは減算を行うことで加速度の値をシフトする。

以上の操作を行ったデータセグメントをランダムに並び替えることで動作の順番が異なるデータを作成することができる。これらの操作を1つの加速度データ系列に対して複数回繰り返し、繰り返した回数分のデータを生成する。

上記の手順では、拡張元と同じラベル（動作回数）をもつ拡張データしか生成できない。提案手法では、さらに、拡張元の加速度データ系列をコピーして連結して上記の操作を行い、ラベルが2倍の拡張データを生成する。

3.4.2 ネットワークの学習

カウントを行う対象の動作ごとにモデルの学習を行う。カウント対象の動作1種類の動作回数を正解値として弱教師あり学習を行う。損失関数には平均絶対誤差 MAE (Mean Absolute Error) を用いる。optimizer には Adam[4] を用いる。

まず、提案モデルをカウントの対象とするユーザ以外の複数人のデータで事前に学習を行う。その後、学習済みモデルに対して実際に対象ユーザの学習データで学習率を下げて再学習を行う。これにより、他のユーザで学習されたモデルを対象ユーザに適応させることができる。

4. 評価実験

4.1 データセット

3人の被験者の右手首にスマートウォッチを装着した状態で、ワークアウトを行ってもらい、100Hzの加速度データを取得した。スマートウォッチには ASUS の ZenWatch3 を用いた。

加速度系列には腕立て伏せ、スクワット、腹筋の3種類の動作が含まれている。動作の種類を切り替える際は適当な時間のインターバルを挟んだ。1つの加速度系列にはそれぞれの動作に対して動作を行った合計回数をラベルとして記録した。図4に取得した加速度データに対する動作回数を表すラベルの例を示す。1つの加速度系列における各動作の実行回数の範囲は1~20回とした。表1に各被験者

加速度データ	ラベル		
	腕立て伏せ	スクワット	腹筋
	6回	4回	7回
	5回	18回	9回
⋮	⋮	⋮	⋮

図 4 加速度データに対する動作回数のラベルの例

の情報，加速度系列の取得数を示す。

4.2 評価方法

評価には交差検証を用いる。本来の k 分割交差検証では k 個にデータセットを分割し，1つをテスト用，残りを学習用とするのだが，本研究では少量の学習データでの動作カウントを目的とするため1つを学習用，残りをテスト用とする交差検証を用いる。少量の学習データを想定するため，学習データ数を5とする。つまり被験者1ではテスト用データ数35での検証を8回，被験者2，3ではテスト用データ数が15での検証を4回行う。

他のユーザのデータで事前にモデルの学習を行う場合，カウント対象とならない2人の被験者の取得したデータセットの75%，つまり被験者1では30，被験者2，3では15のデータを用いて事前に学習を行い，その後上記の交差検証を行う。

また，学習データはData Augmentationにより1つの加速度系列を1000個に拡張する。得られたデータセットの中で最も長い時間の1/2以下の時間の加速度データ系列に対してラベルが2倍となる拡張データを同じく1000個生成する。

4.2.1 評価指標

評価指標には予測値と正解値の絶対誤差の平均を示すMAE (Mean Absolute Error) を用いる。MAEは式(4)のように求められる。

$$MAE = \frac{1}{n} \sum_{k=0}^n |p_k - y_k| \quad (4)$$

p_k, y_k は k 番目のテストデータに対する予測値，正解値をそれぞれ指す。 n はテストデータの数を表す。

4.2.2 比較手法

ネットワークモデルを6層のConvolution層からなるCNNと比較する。またData Augmentationの有無による精度の比較，事前に他のユーザのデータで学習を行わない場合との比較も行う。これらの比較手法をまとめると以下のようなになる。

- CNN : 6層のConvolution層からなるCNN (他ユーザデータでの事前学習あり)
- CNN (w/o DA) : Data Augmentationによるデータ拡張なし
- CNN (w/o pre-train) : 他ユーザデータでの事前学習なし

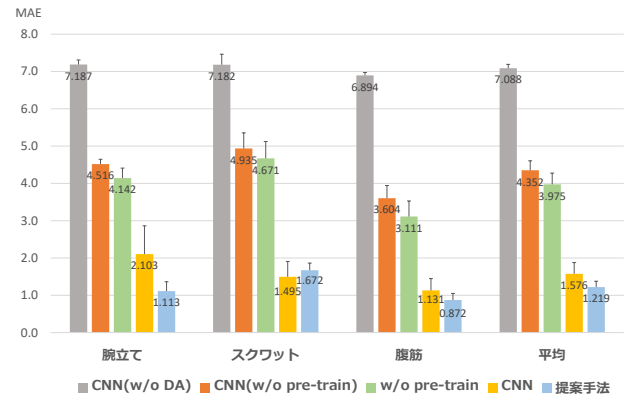


図 5 動作ごとにカウント誤差を平均した結果

- w/o pre-train : 提案ネットワークモデルで他ユーザデータでの事前学習なし
- 提案手法 : 提案ネットワークモデルかつ他ユーザデータでの事前学習あり

4.3 評価結果

動作ごとに3人の被験者から得られた結果を平均した結果と全ての動作でのカウント誤差を平均した結果を図5に示す。スクワット以外の動作では提案手法におけるカウント誤差が最も小さいという結果が得られた。また，動作全体の平均の結果をみると，提案手法は比較手法CNNよりも21%カウント誤差を低減していることが分かる。それぞれの手法の結果に対して標準誤差を求めた。その結果，比較手法CNNの標準誤差は大きく，カウント誤差のばらつきが大きいことが分かった。一方で，提案手法における標準誤差は比較手法CNNと比較すると小さく，安定して小さい誤差でカウントが行えた。スクワットの動作におけるカウント誤差の平均は比較手法CNNが1.495，提案手法では1.672と比較手法よりも誤差が大きくなってしまっているが，標準誤差が比較手法CNNでは0.412であるのに対して，提案手法では0.192と比較手法CNNよりも安定してカウントを行えていることが分かる。

また，CNN(w/o DA)，CNN(w/o pre-train)，w/o pre-trainにおけるカウント誤差はCNNや提案手法におけるカウント誤差と比較すると，かなり大きいことが分かる。この結果により，Data Augmentation，他のユーザによるモデルの事前学習の有効性が示された。

4.4 評価結果に対する考察

4.4.1 誤差の妥当性

提案手法におけるすべての動作のカウント誤差は平均約1.2であった。動作のカウントに関する既存研究における検証結果と比較して本研究におけるカウント誤差1.2の妥当性を確認する。Bianら[1]の研究では，おおよその動作でのカウント誤差は約1となっている。Soroら[10]の研究

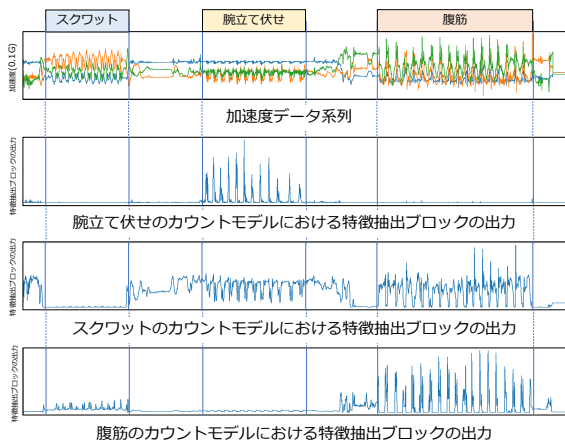


図 6 被験者 1 における各動作のカウントモデルにおける特徴抽出ブロックの出力の可視化を行った例

では、カウント誤差が動作によって大きく異なり、カウント誤差が腕立て伏せでは 1.22、スクワットでは 1.82、腹筋では 0.02 であった。本研究では少量のデータを用いた弱教師学習を行っているにも関わらず、これらの既存研究におけるカウント誤差と近いカウント誤差でカウントを行えた。しかし、ユーザや動作によってカウント誤差のばらつきがあるため手法をさらに改善する必要がある。同じユーザで同じ動作においても、カウント誤差が大きく異なる場合があった。例えば、被験者 1 の腕立て伏せの動作での交差検証において、MAE は約 0.7 から約 2.5 にまでばらついていた。このようなばらつきの原因として、学習に用いるラベルのバリエーションが考えられる。被験者 1 における腕立て伏せの検証において、MAE が約 0.7 となったセットのラベルは $[1, 3, 5, 12, 20] + [2, 10]^*$ であるのに対して、MAE が約 2.5 となったセットは $[8, 9, 10, 13, 16] + [16, 18, 20, 26, 32]$ だった。今回の検証において動作カウントを行う範囲は 1～20 回となっており、MAE が約 0.7 となるセットのラベルはこの範囲に満遍なく分布しているのに対して、MAE が約 2.5 となるセットでは全てのデータにおいてラベルのバリエーションが 2 倍になるデータが生成されているにもかかわらず、1～7 回の範囲のラベルが含まれていない。このようなラベルの分布が学習に大きく影響するため、ラベルのバリエーションを増やす手法を見直す必要がある。

4.4.2 Attention の可視化

提案ネットワークの特徴抽出ブロックに含まれる Attention 機構によってカウント対象の動作の出現位置に注目できていたのかを特徴抽出ブロックの出力を可視化することで確認した。被験者 1 における各動作のカウントモデルにおける特徴抽出ブロックの出力を図 6 に示す。腕立て伏せと腹筋のカウントモデルの特徴抽出ブロックの出力は対象動作の出現時刻でピークを示すような系列を出力していることが分かる。対象動作以外の時刻では値がほぼ 0 になっ

*1 括弧内はラベルが 2 倍になることで追加されるラベル

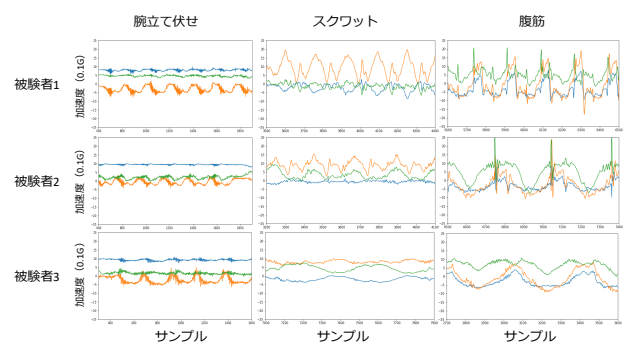


図 7 3 人の被験者による 3 種類の動作の加速度データの比較

ており、カウントに関係ない要素を無視していることが分かる。しかし、スクワットのカウントモデルにおいては対象動作の出現位置において出力値がほぼ 0 となっており、注目に失敗していることが分かる。事前にスクワットのカウントモデルを被験者 2 と被験者 3 の学習データで事前学習する際、被験者 1 のスクワットと被験者 2, 3 のスクワットから得られる加速度データの差異が大きいことが原因であると考えられる。図 7 に 3 人の被験者の各動作から得られた加速度データ系列を比較した様子を示す。腕立て伏せや腹筋に比べてスクワットから得られる加速度は被験者ごとに差異が大きいことが確認できる。この差異は、被験者 2 と被験者 3 におけるスクワットのカウントにも影響すると考えられるため、提案手法におけるスクワットのカウント誤差は他の動作に比べて大きくなっていると考えられる。しかし、スクワットのカウント誤差は事前学習を用いない手法と比較すると、かなり小さい。そこで、被験者 1 のスクワットカウントモデル内の全結合層への入力を可視化することでネットワークの出力の判断要因を確認した。図 8 はその例を示したものである。この図からスクワットの動作が行われている時刻に特徴が現れ、スクワット以外の時刻では特徴が 0 となっていることが分かる。つまり、カウントブロックのみでスクワットのカウントを行うように特徴を学習している。事前学習時にユーザによる動作の差異を埋められる学習方法の工夫と、ユーザごとの差異があっても再学習時に特徴ブロックが正確に対象動作の出現位置を注目できるようにネットワークの構成を見直すことが今後の課題となる。

5. 結論

本研究では、少量かつ低ラベリングコストの学習データを用いた弱教師あり学習によって人間の動作のカウントを行う手法を提案した。動作回数をラベルとした弱教師あり学習を可能とするために、モデルはカウント対象の動作の出現位置を特定する特徴抽出ブロックと、特定された出現位置をカウントする方策を備えるカウントブロックの 2 つのブロックで構成した。Data Augmentation を用いること

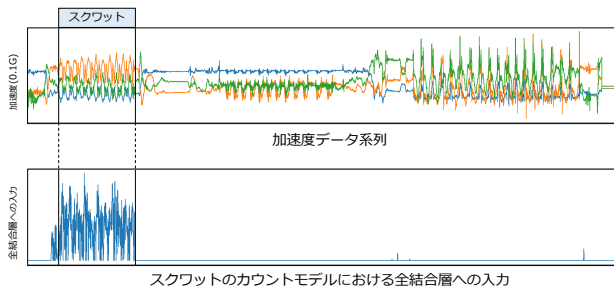


図 8 被験者 1 のスクワットモデルにおける全結合層へ入力される特徴を可視化した例

で加速度データの拡張を行い、学習データから多様なデータを大量に生成したり、カウント対象としないユーザーのデータで事前にモデルを学習しておくことで、カウント対象ユーザーの少量の学習データによるモデルの学習を可能とした。

評価実験により、提案手法が少量の弱いラベルの学習データで高精度なカウントを行えることを確認した。しかし、動作やユーザ、学習データのラベルによってカウント誤差にばらつきがあったため、今後は安定して小さい誤差でカウントを行える手法を目指す。またアプリケーション実用の観点からカウント誤差を 1 未満にまで低減することを目指す。

謝辞

本研究の一部は JST CREST JPMJCR15E2, JSPS 科研費 JP16H06539, JP17H04679 の助成を受けて行われたものである。

参考文献

- [1] S. Bian, V.F. Rey, P. Hevesi, and P. Lukowicz, "Passive capacitive based approach for full body gym workout recognition and counting," 2019 IEEE International Conference on Pervasive Computing and Communications (PerCom), pp.1–10, 2019.
- [2] X. Guo, J. Liu, and Y. Chen, "FitCoach: Virtual fitness coach empowered by wearable mobile devices," IEEE INFOCOM 2017-IEEE Conference on Computer Communications, pp.1–9, 2017.
- [3] W. Jiang and Z. Yin, "Human activity recognition using wearable sensors by deep convolutional neural networks," The 23rd ACM International Conference on Multimedia, p.1307–1310, 2015.
- [4] D.P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.
- [5] D. Morris, T.S. Saponas, A. Guillory, and I. Kelner, "Recofit: using a wearable sensor to find, recognize, and count repetitive exercises," The SIGCHI Conference on Human Factors in Computing Systems, pp.3225–3234, 2014.
- [6] I. Pernek, G. Kurillo, G. Stiglic, and R. Bajcsy, "Recognizing the intensity of strength training exercises with wearable sensors," Journal of Biomedical Informatics, vol.58, pp.145–155, 2015.
- [7] M. Ringnér, "What is principal component analysis?,"

- Nature Biotechnology, vol.26, no.3, pp.303–304, 2008.
- [8] C. Shorten and T.M. Khoshgoftaar, "A survey on image data augmentation for deep learning," Journal of Big Data, vol.6, no.1, p.60, 2019.
- [9] K. Skawinski, F.M. Roca, R.D. Findling, and S. Sigg, "Workout type recognition and repetition counting with CNNs from 3D acceleration sensed on the chest," International Work-Conference on Artificial Neural Networks, pp.347–359, 2019.
- [10] A. Soro, G. Brunner, S. Tanner, and R. Wattenhofer, "Recognition and repetition counting for complex physical exercises with deep learning," Sensors, vol.19, no.3, p.714, 2019.
- [11] T.T. Um, F.M. Pfister, D. Pichler, S. Endo, M. Lang, S. Hirche, U. Fietzek, and D. Kulić, "Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks," The 19th ACM International Conference on Multimodal Interaction, pp.216–220, 2017.
- [12] T. Xiao, Y. Xu, K. Yang, J. Zhang, Y. Peng, and Z. Zhang, "The application of two-level attention models in deep convolutional neural network for fine-grained image classification," The IEEE Conference on Computer Vision and Pattern Recognition, pp.842–850, 2015.
- [13] X. Yu, N.T. Vu, and J. Kuhn, "Learning the Dyck Language with Attention-based Seq2Seq Models," The 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, pp.138–146, 2019.
- [14] X. Yu, X. Wu, C. Luo, and P. Ren, "Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework," GIScience & Remote Sensing, vol.54, no.5, pp.741–758, 2017.