

Few-shot 弱教師あり学習を用いた ウェアラブル加速度センサによる動作カウント手法

西野 祐希¹ 前川 卓也¹ 原 隆浩¹

概要: 本研究では, Few-shot 弱教師あり学習を用いて, ウェアラブル加速度センサからワークアウトなどの人間の動作をカウントする手法について検討する. ターゲットユーザからのカウント対象の動作を含む弱ラベルが付与された少数のセグメントを学習データとして, 対象動作の繰り返し回数をカウントする WeakCounterF 手法を提案する. 入力データセグメントに含まれる対象動作の繰り返し回数のみを示す弱ラベルを利用することで, 学習用データセット収集のコストを低減することができる. まず, WeakCounterF は, データ拡張とラベル多様化を用いて, ターゲットユーザ以外のユーザ, すなわちソースユーザからの弱ラベルありデータから, 多様な学習データを生成する. さらに, ターゲットユーザの少数の弱ラベルありデータから, 多様な弱ラベルあり学習データを生成する. その後, Attention 機構を備える繰り返し動作カウントモデルを, ソースユーザの多様なデータで事前学習を行った後, ターゲットユーザからのデータでモデルのファインチューニングを行う. 9人の参加者による実験により, WeakCounterF の有効性を示した.

1. はじめに

近年普及しつつあるウェアラブルセンサデバイスにより, 健康状態や仕事管理を自動的に行うことができるようになりつつある. 特に, ウェアラブルコンピューティングの研究分野では, ウェアラブル慣性センサのスポーツや産業分野への応用が盛んに研究されている [1], [2].

本研究では, ウェアラブル加速度センサを用いて人間の動作の繰り返し回数をカウントする手法を提案する. 動作の繰り返し回数カウント手法により, ワークアウトの繰り返し回数を記録することで運動を管理することができたり, 工場行員の作業においてあらかじめ決められたタスクを行えているかを確認することができる [3], [4], [5], [6], [7]. 繰り返し回数カウントの先行研究では, ニューラルネットワークなどを用いた教師あり学習手法が主となっている. しかし, 先行研究の既存カウント手法では, 図1のように, カウント対象動作の繰り返しを含むセグメントを示すラベルと, 対象動作の開始時刻を指定するラベルが必要となり, ラベルのアノテーションに大きなコストが必要となってしまう.

この問題を解決するために, 著者らは動作の繰り返し回数カウントを行う弱教師あり学習手法 WeakCounter を提案した [8]. WeakCounter は, スマートウォッチから取得される3軸加速度セグメントを入力とし, この入力セグメ

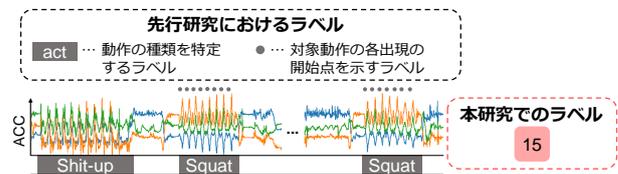


図1 先行研究と本研究で用いる教師あり手法のラベルの例. この例では, スクワット動作の繰り返し回数を数えることを想定している. 本研究では, スクワットを行った回数を示す数値のみを必要とする.

ントに含まれる対象動作の繰り返し回数を示すような弱ラベルが与えられることを想定している (図1). そのため, WeakCounter は先行研究のようなコストの大きいラベルを必要としない.

WeakCounter は, テストユーザのラベル付きセンサデータを学習用で使用しない leave-one-user-out 交差検証で評価が行われた. しかし, 被験者の身体的特徴の違いなどの要因により, 図2のような動作時の加速度波形の違いが生じ, 弱教師あり学習手法の性能が低下する可能性がある. この問題を解決するために, 本研究では, Few-shot 弱教師あり学習を用いた動作繰り返し回数カウント手法 WeakCounterF を提案する. WeakCounterF では, ラベルの弱い少数のテストユーザの学習データを用いて, 動作繰り返しカウントを行うニューラルネットワークを学習する. 本研究では, システムがテストユーザ (ターゲットユーザ) に対して, 対象動作を複数回繰り返し行うように指示することを想定

¹ 大阪大学大学院情報科学研究科

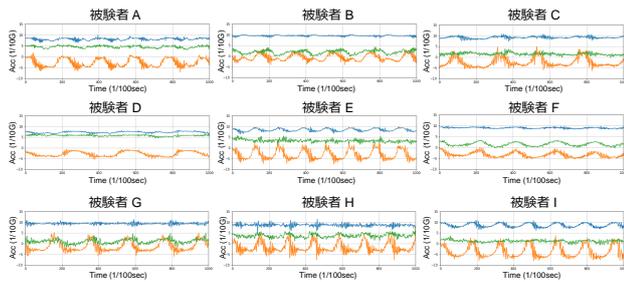


図 2 9名の参加者の腕立て伏せに関するスマートウォッチの加速度センサデータであり、センサデータの個人差を示す。青、オレンジ、緑の線はそれぞれ X 軸、Y 軸、Z 軸のセンサ値である。

する。例えば、加速度データの収集を開始し始めてから、ユーザは腕立て伏せを 3 回のみ行い、データの収集の停止するといった想定である。ユーザは取得した加速度データに対してラベル付けを行う必要がないため、データ収集とアノテーションのコストを大幅に削減することができる。本研究では、各動作に対して弱ラベルありの少ないセンサデータセグメントを活用して Few-shot 動作カウント手法を実現する。

繰り返し回数カウントの Few-shot 弱教師あり学習には、i) 同じテストユーザでも対象動作の 1 回分の波形が異なることがある、ii) テストユーザの動作から得られたデータには対象動作とは無関係な波形が含まれるため、少数の弱ラベルありデータから頑健なカウントモデルを構築することが困難である、iii) テストユーザの少数の弱ラベルありデータでは多様性が乏しく、カウントモデルを効率的に学習することが困難である、といった課題がある。

これらの課題を解決するために、まずテストユーザ以外の複数のユーザ（ソースユーザ）から得られた弱ラベルありセンサデータを用いてカウントモデルを事前に学習し、頑健なカウントモデルを構築を行う。そして、テストユーザから取得した少数の弱ラベルありセグメントを活用して、モデルのファインチューニングを行うことで、学習済みカウントモデルをテストユーザの動作に適応させる。ファインチューニングしたモデルの頑健性を高めるために、テストユーザの弱ラベルありセグメントから、自己相関係数を利用して、対象動作の 1 回分の発生に対応する短いセグメントを抽出し、抽出した短いセグメントを連結することでラベル（繰り返し回数）の異なる多様な学習セグメントを生成し、データ拡張を行い、モデルのファインチューニングを行う。

本研究の貢献は以下のとおりである。i) 本手法は動作繰り返しカウントにおける Few-shot 弱教師あり学習を初めて適用した手法である。ii) 提案手法は、テストユーザから取得する少数の弱ラベルありセンサデータを用いて頑健なカウントモデルを構築できる。iii) 9 人の被験者から収集した実センサデータを用いて本手法の有効性を検証した。

本論文は以下のように構成される、

2. 関連研究

2.1 加速度センサを用いた動作カウント手法

時系列センサデータを用いた繰り返しカウント手法はパラメトリック手法と教師あり手法の 2 つに分類される。パラメトリック手法には、対象動作に対応するセンサ値の振幅や連続する対象動作の間隔などのあらかじめ決められたパラメータに依存する [9], [10], [11], [12], [13]。これらのパラメータは動作の種類によって事前に決められる。近年では時系列データに対する教師あり学習手法は回帰型ニューラルネットワーク (RNN) や畳み込みニューラルネットワーク (CNN) などのニューラルネットワークが主となっている [14], [15], [16]。教師あり学習手法はあらかじめ定義されるパラメータを必要としないが、動作の出現を示すラベルを用意するのに膨大なコストを要する (図 1)。これに対して、提案手法は弱教師あり学習手法のためのエンドツーエンドの構成となっている。

3. 提案手法

3.1 想定環境

腕立て伏せのような対象とするワークアウトを含む複数の動作とワークアウトとは無関係な水を飲むなどの動作が含まれるような 3 軸加速度データセグメントをスマートウォッチのようなウェアラブルセンサから取得したとき、そのセグメントに含まれる対象動作の繰り返し回数をカウントすることが本研究の目的である。そのために、そのような行動を含む学習セグメントを用意する必要があるが、ターゲットユーザから十分な量の学習用セグメントを集めるのはコストがかかるため、ソースユーザから集めた学習用セグメントを用いる。さらに、異なるユーザ間で生じるセンサ値の差異の問題を解決するために、テストユーザに対象の動作を数回だけ行ってもらい、取得したデータと弱ラベル y_i 、つまり対象動作を繰り返し行った回数を追加の学習データとして用いる。そのような場合では、各対象動作ごとの繰り返しカウントモデルを用意し学習する。

すなわち、各対象動作に対して、テストユーザとソースユーザどちらからも少数の弱ラベルあり加速度セグメントを学習データとして用いる。弱ラベルありセグメントは対象動作が含まれており、対象動作の発生回数を示す弱ラベルがそのセグメントに紐づけられる。

3.2 手法概要

WeakCounterF の概要を図 3 に示す。提案手法は主に事前学習とファインチューニングの 2 つのフェーズから成り立つ。事前学習フェーズでは、対象動作の繰り返しカウントを行う Attention ベースのニューラルネットワーク WeakCounter-Net をソースユーザの学習データで事前に学習する。複数の対象動作をカウントする場合、各対象動作

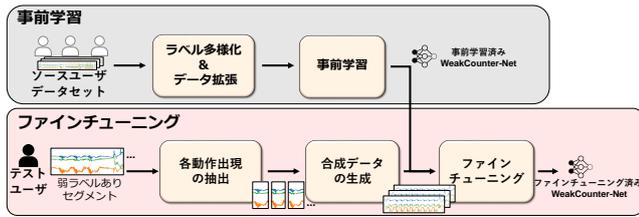


図 3 Few-shot 弱教師あり学習を用いた繰り返しカウント手法の概要

のカウントネットワークを用意する。効率的に事前学習を行うために、ラベル多様化とデータ拡張の技術を用いてソースユーザのセンサデータから多様な学習データを生成する。また、繰り返しカウントネットワーク WeakCounter-Net は、人間が数をカウントする方法を基に設計しており、高いカウント性能を実現する。

ファインチューニングフェーズは「ファインチューニング用データセットの生成」と「WeakCounter-Net のファインチューニング」のステップで構成される。「ファインチューニング用データセットの生成」のステップでは、事前学習したネットワークをターゲットユーザに適応するために、ターゲットユーザから収集した弱ラベルありセグメントからデータセットを構築する。まず、弱ラベルありセグメントから対象動作の各繰り返しに対応する短いセグメントを抽出する。抽出した 1 回分の動作が現れる短いセグメントを複製・連結することで、対象動作が複数回繰り返して出現する多様な合成セグメントを作成する。「WeakCounter-Net のファインチューニング」のステップでは、生成されたデータセットを用いて事前学習済みモデルのファインチューニングを行う。

3.3 事前学習フェーズ

3.3.1 ラベルの多様化とデータ拡張

まず、学習データのラベルのバリエーションを増やすためにラベル多様化を行う。ラベル多様化手法の考え方は次の通りである。いくつかのデータセグメントを連結すると、連結したセグメントの正解ラベル、つまり対象動作が行われる回数は連結する前のセグメントの正解ラベルの総和となる。このアイデアを基に、多様な弱ラベルありデータを生成することが可能である。

ソースユーザの弱ラベルあり学習データセグメントのペアを組み合わせてラベル多様化を行う。ここでラベル y_1 を持つセグメント s_1 とラベル y_2 を持つセグメント s_2 からなる組み合わせが与えられたと仮定する。 s_1 と s_2 を連結し、この連結セグメントにラベル $y_1 + y_2$ を与える。 $(s_1$ と s_2 は同じセグメントの場合もありうる。)

その後、ラベル付けされた各セグメントに対してデータ拡張を行い、行う動作の順序、動作の振幅、動作の速度、センサの向きなどの点において多様な N_a 個のセグメント

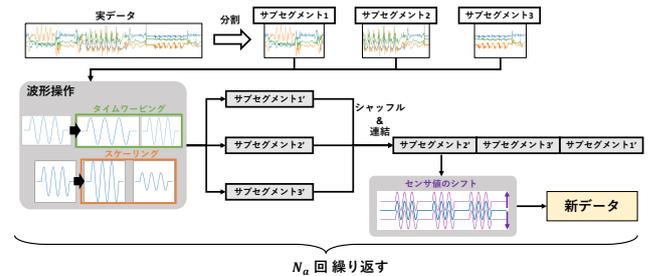


図 4 データ拡張の手順

を生成し、過学習を防ぐ。図 4 にデータ拡張の手順の概要を示す。まずオリジナルのセグメントをサブセグメントに分割する。ここで、セグメントの分割点は、対象動作が行われていない部分に設定する必要がある。そこで、スライディング時間窓内の分散を計算し、閾値より小さい分散を持つ窓の中心を分割点として設定する。その後、以下の手順を N_a 回繰り返す。Morris ら [17] や Yu ら [18] の先行研究に基づき、各サブセグメントに以下の操作を行う。

- スケーリング：一様分布 $U(1 - c_s, 1 + c_s)$ からサンプリングしたランダムな値をサブセグメントのセンサ値に乗算し、センサ値の振幅のスケールを変更する。
- タイムワーピング：線形補間を用いてサブセグメントをリサンプリングして、長さが $l \cdot U(1 - c_t, 1 + c_t)$ となるように各サブセグメントを伸縮させる。ここで、 l は元のサブセグメントの長さである。

その後、上記の操作によって波形の操作を行ったサブセグメントをシャッフルして連結して、1つのセグメントを生成する。最後に、各軸のセンサ値に一様分布 $U(-c_b, c_b)$ からサンプリングしたバイアス値を加算することで、センサ値をランダムにシフトして、センサの向きの違いを考慮する。

3.3.2 WeakCounter-Net

人がテーブルの上にあるリングの数を数えるとき、人はまずリングを一つ一つ認識し、認識した物体、つまりリングの数を数える。同様に、WeakCounter-Net を、Attention 機構を用いて対象の動作の出現を検出する検出ブロックと、検出された出現回数をカウントするカウントブロックの 2 つのブロックを備えるように設計する。検出ブロックは、対象動作の出現位置にピークが現れるような 1 次元時系列を出力し、その出力はカウントブロックに入力される。その結果、1 次元時系列内のピークの数を実数としてカウントすることができる。このネットワーク構造によって、学習ラベルのばらつきが少ない場合でも、任意の数のカウントを高い精度でカウントできると期待できる。

図 5 に示すように、提案カウントモデルは検出ブロックとカウントブロックで構成される。入力セグメント内には対象動作と無関係な動作が含まれる可能性があるため、対象動作に焦点が当たるようにするために、Atten-

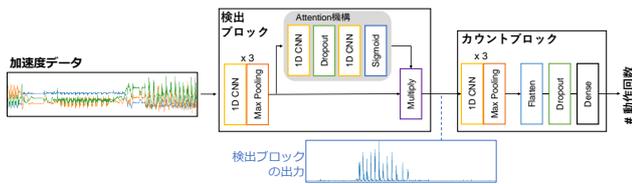


図5 WeakCounter-Netのネットワーク構成機構 [19], [20], [21] を検出ブロックに導入する. まず, 1次元の畳み込み層で特徴抽出をおこない, D_f 次元の時系列 D_f 次元の時系列 $F \in \mathbb{R}^{D_f \times L}$ を出力する. ここで, L はデータの長さを示す. そして Attention を示す a を $a = \text{sigmoid}(\text{Conv2}(F))$ によって求める. ここで, $\text{Conv2}(\cdot)$ は Attention 機構内のドロップアウト層を含む2層の1次元畳み込み層の計算を示す(図5). ここで, $a \in \mathbb{R}^{1 \times L}$ は F の各データポイントの重要度, つまり Attention を示す. F と Attention の時系列 a の要素ごとの積を求めた後, すべての次元の軸, つまり D_f の値の総和をとることで, 検出ブロックは, 対象動作の出現位置にピークが現れるような1次元時系列(図5の下部の青い波形)を出力する. カウントブロックでは, 1次元の畳み込み層, 最大値プーリング層, そして全結合層で検出ブロックの出力である1次元時系列を処理し, カウント推定値を出力する.

WeakCounter-Net はソースユーザからの弱ラベルありデータを用いて推定値と正解値の平均絶対誤差が小さくなるように学習を行う. 最適化アルゴリズムには Adam[22] を採用している.

3.4 ファインチューニングフェーズ

ターゲットユーザから得られた限られた学習データを用いて事前学習済みの WeakCounter-Net モデルのファインチューニングを行うための手法について説明する. 前述したように, テストユーザから対象動作が数回のみ繰り返されている数個のセグメントを取得する. これらのセグメントを用いて, 多様なデータセットを生成し, そのデータセットで事前学習済みの WeakCounter-Net のファインチューニングを行う.

3.4.1 ファインチューニング用データセット生成

3.4.1.1 対象動作の1回分のセグメント抽出

まず, 弱ラベルありセグメントから対象動作の各出現箇所を抽出する. ただし, このセグメントには行われた対象動作の繰り返し回数を示すラベルしか用いることができる情報がない. この弱ラベルのみを用いて, 対象動作の各出現箇所を抽出する. 対象動作の各出現箇所の抽出することは, すなわち各出現の開始時刻と出現時間の長さを求めることである. 図6の上部にテストユーザから取得するセンサーデータセグメントの一例を示す. この図では, 対象動作「腹筋」が3回行われていることが確認できる. しかし, セグメントの最初と最後には, 対象動作とは無関係な波形が

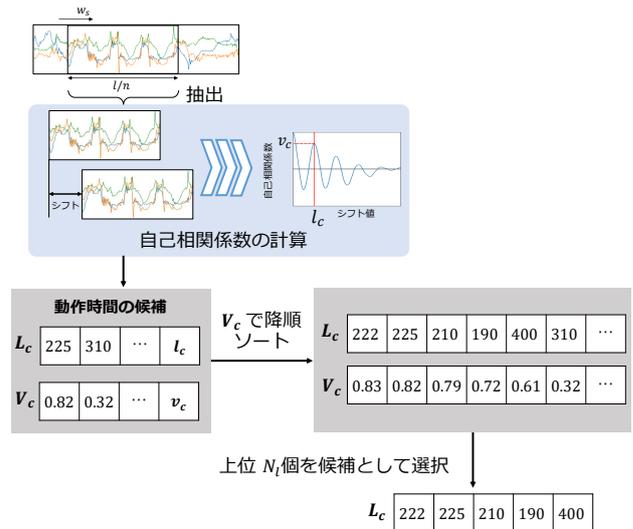


図6 自己相関係数を利用した対象動作が1回行われる時間の候補を決定

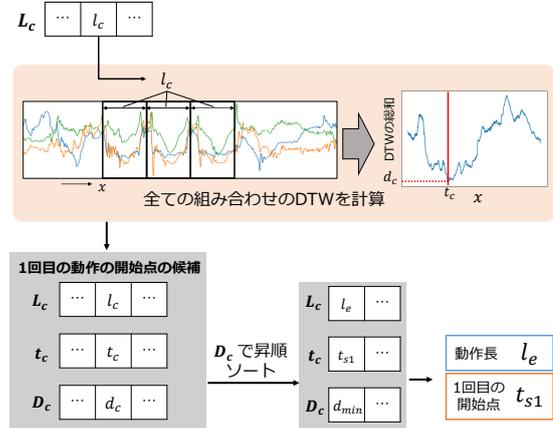


図7 DTW を用いて対象動作の1回目の開始点を決定

含まれる. これは, データ取得の開始と終了の際に関連する動作と対応している. 本手法では, まず自己相関係数などを利用して動作が行われる時間の長さとな出現の開始時刻を大まかに決め, その後, 最適化手法を用いて正確な開始時刻と長さを決定する.

(1) 対象動作が行われる時間の長さの候補決定: セグメント内には繰り返し行われる対象動作が含まれているため, セグメントの自己相関係数を計算することで, 対象動作の継続時間(長さ)を推定することができる. しかし, 図6で示されているように, 対象動作と無関係の波形がセグメントに含まれているため, 自己相関係数を用いるだけでは対象動作の継続時間を推定するのが困難である. そこで, 長さ l であるセグメントからスライディングウィンドウで長さ l/n となるサブセグメントを抽出し, そのサブセグメントについて自己相関係数を計算する. そうすることで, 対象動作

と無関係な波形が含まれないようなサブセグメントが抽出できることが期待できる。図6に抽出するサブセグメントの例と、自己相関係数を用いて対象動作が1回行われる時間の長さの候補を決定する手順を示す。図6の自己相関係数の計算の結果の図で l_c で最大のピークが現れており、これは対象動作が1回行われる継続時間の長さの候補が l_c であることを示している。そのときの自己相関係数の値を v_c とする。この手順によって得られた動作時間の候補の集合 L_c とそのときの値 V_c を V_c に従って降順に並び替える。並び替えられた L_c のうち、上位 N_l 個を動作の時間長の候補とする。

- (2) 対象動作の1回目の開始位置の決定：対象動作の時間の長さの候補が決定したら、その長さの候補を用いて、対象動作の最初の出現開始位置を大まかに推定する。ただし、セグメントには弱ラベル y_t 、つまり、対象動作の出現回数が与えられており、対象動作は連続的に行われている。そのため、1回目の出現開始位置を t_{s1} 、出現回数を y_t とすると、 $s(t_{s1}, l_e)$, $s(t_{s1} + l_e, l_e)$, $s(t_{s1} + 2l_e, l_e)$, ..., $s(t_{s1} + (y_t - 1)l_e, l_e)$ のセグメントのセンサ値の類似度が互いに高くなると考えられる。ここで、 $s(t, l)$ は図8のaのように時刻 t から始まり、長さ l の短いデータセグメントを示す。本手法では、これらの短いセグメント間の動的時間伸縮法(DTW)による距離が最小となる t_{s1} を見つけるようにセグメントを走査する。求めるDTW距離の計算は以下の式のように行われる。

$$\sum_{i,j \in N_{y_t}} \text{DTW}(s(t_{s1} + i \cdot l_e, l_e), s(t_{s1} + j \cdot l_e, l_e)), (1)$$

ここで、 $\text{DTW}(\cdot, \cdot)$ は2つのセグメント間のDTWの距離を計算する関数であり、 N_{y_t} は0以上かつ y_t 以下の整数の集合である。図8のaは推定される開始位置と動作時間によって特定される動作の出現の例を示している。図7に1回目の開始点を求める手順を示す。ある点 x から長さ l_c である y_t 個のセグメントのすべての組み合わせのDTW距離の総和を計算し、その値が最も小さくなる x を1回目の開始位置 t_{s1} の候補とする。すべての l_c についてDTW距離の総和の最小値を求め、その中でも最小となるものを t_{s1} とする。また、そのときの l_c を l_e とする。

- (3) 動作の各出現の開始位置と動作時間の決定：これまでの手順で、最初の1回目の動作の開始位置と動作時間を大まかに求めた。ただし、動作時間は各出現において同一であると仮定した。ここで、動作の各繰り返しで動作時間が異なる場合や、連続する2つの動作の出現の間に短いブランクがある場合があるため、最適化手法を用いて各出現ごとの開始位置と動作時間を求め

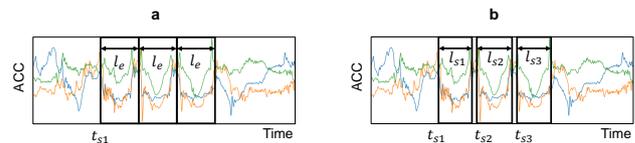


図8 腹筋のワークアウトの弱ラベルありセグメントにおける対象動作の各出現箇所の開始位置と動作時間の長さの決定。a) 最初の1回目の動作の出現の大まかな開始時刻 t_{s1} と動作時間 l_e を決定する。この場合、2回目、3回目の動作時間も l_e であると仮定する。そして、 t_{s1} と l_e に基づいて2回目と3回目の出現の開始時刻を決定する。

る。最適化手法を用いるための目的関数は以下のとおりである。

$$\sum_{i,j \in N_{y_t}} \text{DTW}(s(t_{s_i}, l_{s_i}), s(t_{s_j}, l_{s_j})), (2)$$

ここで、 t_{s_i}, l_{s_i} は i 回目の動作の出現の開始位置と動作時間をそれぞれ示す。最適化手法L-BFGS-B[23]を用いて、目的関数を最小化するような t_{s_i} と l_{s_i} を求める。最適化の際に、最終的に得られる最適化後の結果が前述の手順で得られた初期の推定値から大きく外れないようにするために、以下のような制約を加える。i) i 番目の動作の出現の推定開始位置と、前述の手順で推定された初期開始位置との差の絶対値は l_e より小さい。ii) i 番目の動作出現の推定動作時間は初期の推定値 l_e とあまり差がない($< 0.2l_e$)。最適化によって求められた動作の出現の例を図8bに示す。

3.4.1.2 合成データの生成

上記の手順によって、動作の各出現の開始位置と動作時間の推定値を取得する。そして、その推定値によって特定される動作の出現箇所に対応するセグメントを複製・連結することで、合成学習データを生成し、事前学習済みモデルのファインチューニングに用いる。ラベル y_i を持つ合成データを以下のように生成する。

- (1) 弱ラベル y_t をランダムに選択し、ユーザから取得したセグメントの中から対象動作が y_t 回行われているセグメントを選択する。上記の手順によって、選択したセグメント内の対象動作の各出現における開始位置と終了位置が得られる。カウントモデルの頑健性を向上させるために、動作の出現を抽出する際に開始/終了位置に r を加えることで、ランダムに開始/終了位置を前後させる($-0.1l_e < r < 0.1l_e$)。
- (2) 選択したセグメントから対象動作1回分の短いセグメントを y_i 回抽出し、そのセグメントを連結する。カウントモデルの頑健性を向上させるために、対象動作の出現の前後に対象動作と無関係なセグメントを挿入する。ここで、対象動作と無関係なセグメントとは、対象動作の出現として抽出されなかったセグメントの部分である。
- (3) 多様な学習データを生成するために、3.3.1節のデータ

表 1 本手法と検証実験で使用するパラメータ

パラメータ	値	説明
N_a	100	データ拡張で生成するセグメントの数
c_s	0.4	データ拡張でのスケーリングの大きさ
c_t	0.3	データ拡張でのタイムワーピングの大きさ
c_b	5.0	データ拡張でシフトされる最大絶対値
N_l	5	対象動作の動作時間の長さの候補の数
w_s	100	スライディングウィンドウでのストライド

拡張の技術を用いる。

3.4.2 WeakCounter-Net のファインチューニング

以上のように生成したファインチューニング用のデータセットを用いて、事前学習済みの WeakCounter-Net のファインチューニングを行う。ファインチューニングの際の学習率は 0.0001 と事前学習時の学習率 0.001 より低く設定する。

4. 評価実験

4.1 データセットと評価方法

9名の被験者の右手首に装着したスマートウォッチ ASUS Zen Watch 3 から 100Hz の 3 軸加速度データを収集した。入力セグメントのデータ収集セッションには、腕立て伏せ、スクワット、腹筋の 3 種類のワークアウトがランダムな順番で含まれる。各動作は 1 ~ 20 回の間でランダムな回数行われる。セッションには、水を飲む、汗を拭くなどのワークアウトとは無関係な動作も含まれている。また、異なる日に各動作を 2 ~ 5 回行うことで、ファインチューニングのための弱ラベルありセグメントを少数個取得した。1 人の被験者からファインチューニング用の弱ラベルありセグメントは各動作ごとに 4 つずつ取得できた。データセットの概要を表 2 に示す。

実験では、被験者 1 人ずつをターゲットユーザとして、残りの被験者をソースユーザとする leave-one-user-out 交差検証を行った。

提案手法の有効性を検証するために、以下の手法を用意した。

- WeakCounterF: 提案手法。ターゲットユーザから取得した弱ラベルありセグメントをすべて用いて事前学習済み WeakCounter-Net をファインチューニングする手法。
- WeakCounterF (one-shot): 提案手法。ターゲットユーザから取得した 1 つの弱ラベルありセグメントのみを用いて事前学習済み WeakCounter-Net をファインチューニングする手法。
- CNN: 提案手法において WeakCounter-Net の代わりに 1 次元畳み込みニューラルネットワークを用いる手法。このネットワークは 6 層の 1 次元畳み込み層、最大プーリング層と 1 つの出力層で構成される。
- 合成データのみ: ターゲットユーザのファインチューニング用の弱ラベルありセグメントから生成した合成

データのみを用いて WeakCounter-Net を学習する手法。ソースユーザのデータを用いた WeakCounter-Net の事前学習は行わない。

- WeakCounter: Few-shot 学習を用いない先行手法 [8]。ターゲットユーザの弱ラベルありセグメントを用いた WeakCounter-Net のファインチューニングは行わない。

評価指標には推定値と正解値の平均絶対誤差 (MAE) を用いた [10]。実験で用いたパラメータを表 1 に示す。

4.2 結果

4.2.1 カウント精度

表 3 にワークアウトの種類における各手法ごとの MAE を示す。この表からわかるように、提案手法はこれらのワークアウトにおいて最も低い MAE を達成した。WeakCounterF の平均 MAE は WeakCounter の平均 MAE より約 0.41 低く、少数の弱ラベルありセグメントを用いた WeakCounter-Net のファインチューニングの有効性を示している。また、WeakCounterF(one-shot) も WeakCounter と比較して低い MAE を達成していることから、one-shot 学習の有効性を示していることが分かる。CNN の平均 MAE は WeakCounterF の平均 MAE より約 0.60 大きく、Attention 機構を用いたカウントネットワークの有効性が示されている。また、合成データのみ MAE は他の手法と比較して非常に大きい。この結果から、ターゲットユーザの弱ラベルありセグメントのみを用いることは有効ではないことが分かった。これは、ターゲットユーザからの弱ラベルありセグメントのみを用いるだけでは、頑健なカウントネットワークを学習することが困難であるためと考えられる。

各被験者ごとの MAE を表 4 に示す。被験者 F のように WeakCounterF の MAE が WeakCounter の MAE より大きくなるような結果もあったが、ほとんどの被験者において WeakCounterF は MAE を減少させることができたといえる。また、WeakCounter と CNN の結果においては MAE が 3.0 よりも大きくなる被験者もいた。これは、被験者間のセンサデータの差異が大きいためであると考えられる。しかし、WeakCounterF では、ターゲットユーザの弱ラベルありセグメントから生成された合成データを用いることで、この問題に対処することができた。また、表 4 に示すように、CNN、合成データのみ、WeakCounter は、特定の被験者に対して非常に低い精度となることがある。一方、WeakCounterF は全ての被験者に対して安定した精度でカウントを行えている。

4.2.2 one-shot 学習による弱ラベルありセグメントに含まれる動作回数の効果

WeakCounter-Net(one-shot) は、ターゲットユーザから取得した弱ラベルありセグメントを用いる。ここでは、カ

表 2 データセットの概要

被験者	取得データの数	データ長平均 (s)	フィンチューニング 用に取得した セグメントの数	フィンチューニング用 セグメントの 長さの平均 (s)	年齢	身長 (cm)	体重 (kg)
A	40	135.1	12	20.4	23	182	85
B	20	134.0	12	20.1	24	176	65
C	20	145.1	12	15.0	23	164	68
D	5	156.2	12	19.4	22	179	67
E	5	136.6	12	17.1	22	172	58
F	5	90.0	12	13.0	22	171	69
G	5	151.0	12	15.2	22	176	66
H	5	94.8.0	12	16.2	22	176	70
I	5	99.8	12	13.8	22	167	49

表 3 各動作における各手法ごとの MAE

	腕立て伏せ	スクワット	腹筋	平均
CNN	1.02	2.81	1.02	1.62
合成データのみ	3.32	2.71	1.64	2.53
WeakCounter	1.37	2.03	0.88	1.43
WeakCounterF	0.71	1.35	1.01	1.02
WeakCounterF (one-shot)	0.93	1.32	1.01	1.09

ラベル性能と弱ラベルありセグメントに含まれる対象動作が行われる回数、すなわち y_t との関係について調査する。表 5 に y_t の値を変えた場合の WeakCounterF(one-shot) の結果を示す。 y_t が大きい場合、1 回の動作の抽出処理がより正確になると想定していたため、 y_t の値が大きいときは MAE が小さくなると思っていた。しかし、表 5 のように WeakCounterF(one-shot) の MAE は y_t の値に大きく影響されないことが分かった。これは、1 回の動作の抽出を行う際に、各発生の推定開始時刻と終了時刻にランダムなノイズを加えているため、各動作の出現の検出における小さな誤差に対して頑健であるためであると考えられる。

4.2.3 各出現の開始時刻と終了時刻の検出誤差

ここでは、各動作の出現の開始時刻と終了時刻を検出する際の我々の手法の性能について調査する。表 6 に本手法における 1 回の動作抽出における開始/終了時刻の検出誤差を示す。検出誤差は、各動作の出現の開始時刻と終了時刻を手動でアノテーションした正解値に基づいて計算した。今回用いたセンサのサンプリングレートが 100Hz であるため、25 サンプルの誤差は約 0.25 秒に相当する。その結果、 $y_t = 2$ の場合の MAE は他の場合よりも大きくなっていることが分かった。つまり、対象ユーザからの弱ラベルありセグメントに含まれる動作の出現回数が多い場合、対象動作の出現検出の推定はより頑健となる。

ここで、腕立て伏せ、スクワット、腹筋の各動作の平均動作時間は、それぞれ 177.8, 229.1, 269.7(サンプル数)であった。本手法では、動作抽出の性能を高めるために、各動作の出現を抽出する際に、推定された開始/終了時刻 $-0.1l_e < r < 0.1l_e$ となる r を加えて、ランダムに開始/終

了時刻を前後させている。ここで、 $0.1l_e$ は対象動作の推定動作時間の 10%に相当する。そのため、開始/終了時刻の検出の誤差は妥当であると考えられる。

5. おわりに

本研究では、ウェアラブル加速度センサを用いたワークアウトなどの人間の動作の回数をカウントする Few-shot 弱教師あり学習手法を提案した。本手法では、ターゲットユーザから対象動作を含む少数の弱ラベルありセグメントを利用し、正確な動作繰り返し回数のカウントを実現した。検証実験の結果、本手法における Few-shot, one-shot 学習を用いることで、動作繰り返し回数のカウント誤差が小さくなることが分かった。今後は、提案手法を改良し、Few-shot 弱教師あり学習を用いたリアルタイムな動作繰り返しカウント手法を実現する予定である。

謝辞

本研究の一部は JSPS 科研費 JP21H03428, JP21H05299, JP21K19769 の支援のもと行われました。

参考文献

- [1] Guo, X., Liu, J. and Chen, Y.: FitCoach: Virtual fitness coach empowered by wearable mobile devices, *IEEE Conference on Computer Communications (IEEE INFOCOM 2017)*, pp. 1–9 (2017).
- [2] Aehnelt, M., Gutzeit, E., Urban, B. et al.: Using activity recognition for the tracking of assembly processes: Challenges and requirements, *Workshop on Sensor-Based Activity Recognition (WOAR)*, pp. 12–21 (2014).
- [3] Lukowicz, P., Ward, J. A., Junker, H., Stäger, M., Tröster, G., Atrash, A. and Starner, T.: Recognizing workshop activity using body worn microphones and accelerometers, *International Conference on Pervasive Computing*, pp. 18–32 (2004).
- [4] Maekawa, T., Nakai, D., Ohara, K. and Namioka, Y.: Toward practical factory activity recognition: unsupervised understanding of repetitive assembly work in a factory, *The 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 1088–1099 (2016).

表 4 各被験者における各手法ごとの MAE

	A	B	C	D	E	F	G	H	I	Average
CNN	0.85	0.50	2.80	1.80	1.73	1.27	3.47	1.00	1.13	1.62
合成データのみ	1.30	4.85	3.50	2.60	2.07	1.80	1.93	2.13	2.80	2.55
WeakCounter	0.71	0.65	1.80	3.13	1.47	0.87	1.73	0.53	1.93	1.43
WeakCounterF	0.63	0.67	1.43	0.87	1.20	1.07	1.07	0.60	1.67	1.02
WeakCounterF (one-shot)	0.85	0.71	1.57	0.78	1.13	1.10	1.22	0.55	1.87	1.09

表 5 ラベル y_t の弱ラベルありセグメントでファインチューニングした WeakCounterF(one-shot) における各ワークアウトの MAE

y_t	腕立て伏せ	スクワット	腹筋	平均
2	0.86	1.51	0.99	1.12
3	0.84	1.16	1.00	1.00
4	1.13	1.23	1.20	1.19
5	0.84	1.38	0.83	1.02

表 6 y_t のそれぞれの値に対する各動作出現の発生開始/終了時刻の検出における MAE (サンプル数) の平均値

y_t	腕立て伏せ	スクワット	腹筋	平均
2	23.53	28.14	53.42	35.03
3	18.77	26.73	26.56	24.02
4	17.64	24.31	34.74	25.56
5	18.41	18.45	38.84	25.23

- [5] Xia, Q., Wada, A., Korpela, J., Maekawa, T. and Namioka, Y.: Unsupervised factory activity recognition with wearable sensors using process instruction information, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 3, No. 2, pp. 1–23 (2019).
- [6] Xia, Q., Korpela, J., Namioka, Y. and Maekawa, T.: Robust Unsupervised Factory Activity Recognition with Body-worn Accelerometer Using Temporal Structure of Multiple Sensor Data Motifs, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 4, No. 3, pp. 1–30 (2020).
- [7] Morales, J., Yoshimura, N., Xia, Q., Wada, A., Namioka, Y. and Maekawa, T.: Acceleration-based Human Activity Recognition of Packaging Tasks Using Motif-guided Attention Networks, *IEEE Int'l Conf. on Pervasive Computing and Communications (PerCom 2022)*, pp. 1–12 (2022).
- [8] Nishino, Y., Maekawa, T. and Hara, T.: WeakCounter: Acceleration-based Repetition Counting of Actions with Weakly Supervised Learning, *2021 International Symposium on Wearable Computers*, pp. 144–146 (2021).
- [9] Bian, S., Rey, V. F., Hevesi, P. and Lukowicz, P.: 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).
- [10] Skawinski, K., Roca, F. M., Findling, R. D. and Sigg, S.: 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).
- [11] Skawinski, K., Roca, F. M., Findling, R. D. and Sigg, S.: 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).
- [12] Seeger, C., Buchmann, A. P. and Van Laerhoven, K.: myHealthAssistant: a phone-based body sensor network that captures the wearer's exercises throughout the day, *BodyNets*, pp. 1–7 (2011).
- [13] Sundholm, M., Cheng, J., Zhou, B., Sethi, A. and Lukowicz, P.: Smart-mat: Recognizing and counting gym exercises with low-cost resistive pressure sensing matrix, *The 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 373–382 (2014).
- [14] Soro, A., Brunner, G., Tanner, S. and Wattenhofer, R.: Recognition and repetition counting for complex physical exercises with deep learning, *Sensors*, Vol. 19, No. 3, p. 714 (2019).
- [15] Mortazavi, B. J., Pourhomayoun, M., Alsheikh, G., Alshurafa, N., Lee, S. I. and Sarrafzadeh, M.: Determining the single best axis for exercise repetition recognition and counting on smartwatches, *11th International Conference on Wearable and Implantable Body Sensor Networks*, pp. 33–38 (2014).
- [16] Pernek, I., Hummel, K. A. and Kokol, P.: Exercise repetition detection for resistance training based on smartphones, *Personal and Ubiquitous Computing*, Vol. 17, No. 4, pp. 771–782 (2013).
- [17] Morris, D., Saponas, T. S., Guillory, A. and Kelner, I.: 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).
- [18] Yu, X., Wu, X., Luo, C. and Ren, P.: Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework, *GI-Science & Remote Sensing*, Vol. 54, No. 5, pp. 741–758 (2017).
- [19] Zeng, M., Gao, H., Yu, T., Mengshoel, O. J., Langseth, H., Lane, I. and Liu, X.: Understanding and improving recurrent networks for human activity recognition by continuous attention, *The 2018 ACM International Symposium on Wearable Computers*, pp. 56–63 (2018).
- [20] Yu, X., Vu, N. T. and Kuhn, J.: 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).
- [21] Xiao, T., Xu, Y., Yang, K., Zhang, J., Peng, Y. and Zhang, Z.: 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).
- [22] Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).
- [23] Zhu, C., Byrd, R. H., Lu, P. and Nocedal, J.: Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization, *ACM Transactions on Mathematical Software (TOMS)*, Vol. 23, No. 4, pp. 550–560 (1997).