

スマートフォンを用いた行動認識における センサデータ欠損時の精度向上手法

後藤 広樹¹ Tahera Hossain¹ 井上 創造¹

概要: 厚生労働省によると, 2025 年には介護人材が 37.7 万人の不足となるとされ, 解決策の一つとして, 情報技術を活用した介護業務の効率化が期待されている [1]. 介護業務の効率化を目指す先行研究 [2] では, センサデータの欠損が問題である. 一般的に人の行動を推定するためのセンサデータに欠損が生じること, センサ行動認識において問題である. その問題解決のため, 本稿は欠損のあるデータに対して人の行動推定の精度を向上させる手法を提案する. 評価のため, 2つのデータセットを用いて実験を行い, データの一部を無作為に選択し, センサデータの欠損を発生させた. 結果として, シミュレーションされたデータセットを用いた評価実験において, データの欠損率 20%の場合, 推定精度は既存手法が 74.8%, 提案手法が 87.4%と向上した. HASC チャレンジデータセット [11] を用いた評価実験においては, データの欠損率 99%の場合, 推定精度は既存手法が 37.4%, 提案手法が 44.6%と向上した. HASC チャレンジデータセットによる評価実験では, 提案手法により推定精度の向上が確認できたが, シミュレーションデータによる評価実験ほどの結果が出なかったことを考察し, 今後の展開について述べる.

Improving Accuracy for Missing Sensor Data in Activity Recognition with Smartphones

HIROKI GOTO¹ TAHERA HOSSAIN¹ SOZO INOUE¹

1. はじめに

近年のスマートフォンやスマートウォッチといったウェアラブルデバイスの普及によって, デバイス内のセンサを用いたデータの収集が容易になり, 行動認識の技術が発展している. 本稿で扱う行動認識は, センサデータから人の行動を推定する技術である. 例として, 加速度センサを用いた行動認識による介護士の業務記録の自動化 [2] や在宅介護のみまもりのための行動認識 [3] がある. 本稿の行動認識では, 教師あり機械学習技術を用い, 教師データが入力された時に分類モデルを作成する. この分類モデルを作成する教師データのセンサデータを学習データとし, 分類モデルを作成する過程を学習フェーズと呼ぶ. また分類モデルを適用する先のデータをテストデータとし, 対象の種類を推定する過程を推定フェーズとする. しかし, 行動認識に必要な学習データとテストデータに欠損が含まれてい

ることがある [2]. 本稿は欠損を含むデータに伴い, 行動認識の推定精度が低下する問題への提案をする.

本稿の提案手法は, 故意的に欠損させた学習データから分類モデルを作成し, 欠損しているテストデータの推定を行う. 一般的にデータに欠損が含まれている場合, 欠損を他の値で補完する手法か欠損しているサンプルを削除する手法が用いられている. しかし, 提案手法は, 学習データを欠損させる手間が増えるが, テストデータに欠損がある場合において学習データを欠損させないとき, 欠損しているテストデータに対して学習データがより高い類似性が得られる.

評価用のデータセットは2種類用意する. 1つは, シミュレーションされたデータセットである. 異なる数式から作成された波形をセンサデータとし, その数式の違いの推定精度を求める. 1つは, HASC チャレンジデータセットである [11]. 実際に人の行動によって生成されたセンサデータで推定精度を求める. また提案手法の比較対象として, 学習データを欠損させず, 欠損を補完しない手法 (既存手

¹ 九州工業大学

法とする)の行動認識における行動推定も行う。実験方法は、テストデータの欠損率を0–100%まで上昇させて、5%毎に推定を行う。提案手法において推定を行うための分類モデルの作成に使用する学習データもテストデータに応じて0–100%まで欠損率を上昇させる。

推定精度を評価基準とし、各テストデータ欠損率において複数回提案手法を実行し、推定精度の平均値を求める。また、行動推定をする特徴量の重要度の平均値も求める。これらの結果から提案手法の性能を評価し、行動認識の進歩に貢献する。

2. 背景

厚生労働省の文献を参照すると、高齢者人口率が上昇することで、2025年に介護人材は供給不足であり、37.7万人の不足となる問題が予想されている[1]。この問題の解決のためにICT(Information and Communication Technology)やIoT(Internet of Things)の活用が計画されている[4]。この例として、加速度センサを用いた行動認識による介護士の業務記録[2]や在宅介護のための行動認識[3]がある。これらの研究のように、本研究は行動認識を用いて介護・看護分野の問題解決に貢献することを目的としている。

2.1 センサ行動認識

本節では、文献[5]をもとに、介護・看護の問題の対策例として出てきたセンサ行動認識の基本的な一連の流れを説明する。行動認識とは、機械学習技術を用いてセンサデータから人の行動を推定する技術であり、ウェアラブルセンサのデータから「歩いている」「寝ている」「立ち上がる」などという行動を推定する技術である。近年のスマートフォンやスマートウォッチといったウェアラブルデバイスの普及によって、デバイス内のセンサを用いたデータの収集が容易になった。センサ行動認識の一連の流れを図1に示す。

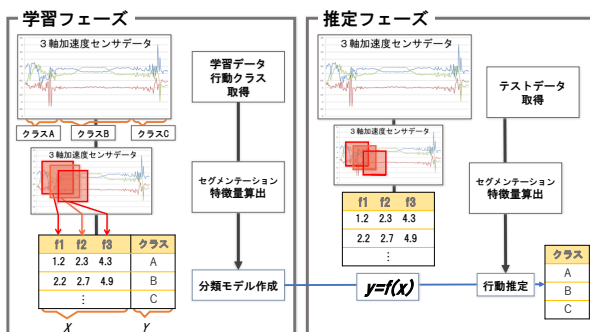


図1 センサ行動認識のデータ入力から行動推定の出力までの流れ

センサ行動認識は、過去のデータを学習データとして利用する教師あり学習を用いる。教師あり学習とは、目的変数である推定する対象の種類 y_i を説明変数であるデータ

x_i が対となったデータセットの集合を学習データとして用いる。教師あり学習の目的は、 x_i が入力された時に y_i を出力する分類モデル $y = f(x)$ を作成することである。この分類モデル作成過程を学習フェーズ、分類モデルを適用して対象の種類を推定する過程を推定フェーズとする。つまりセンサ行動認識の場合、学習フェーズは目的変数を行動の種類(以降行動クラスと呼ぶ。)、説明変数をセンサデータとして、センサデータが入力された時に行動クラスを出力する分類モデルを生成することが目的である。推定フェーズでは行動クラスの不明なセンサデータ x を説明変数として入力し、分類モデルをもとに目的変数の行動クラス y を出力することが目的である。

本稿は、この行動認識の過程、学習フェーズ中の分類モデルを生成する新たな方法を提案する。図1に示すように学習データは収集された時の状態のセンサデータではなく、センサデータを加工した特徴量である。学習データを機械学習に入力するまでの手順として、前処理・セグメンテーション・特徴量算出・次元削減がある。推定フェーズも学習フェーズと同様の手順を行い、テストデータを加工する。

3. センサデータ欠損の問題

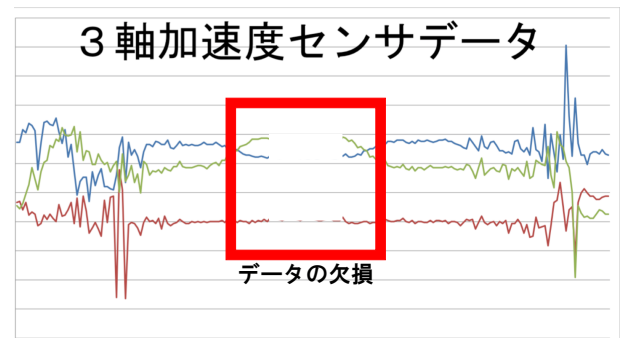


図2 センサデータの欠損の一例

先行研究[2]では、センサデータの欠損が大きな問題として挙げられている。この研究だけでなく、一般的にセンサ行動認識において、センサデータに欠損が生じることは、問題である。本稿で扱うセンサデータの欠損とは、一定の周期で取得しているセンサデータが一時的に欠落してしまう状態のことを指す。センサを用いてデータ収集した場合、意図しない要因によりデータが欠損するが、そのデータの欠損には大きく3種類の欠損データメカニズムがある[7]。1つ目のMCAR(Missing Completely at random)は、変数の欠損する確率が、当該変数の値及び他の観測されている変数の値に依存しないメカニズムである。2つ目のMAR(Missing at random)は、変数の欠損する確率が、当該変数の観測された値及び他の観測されている変数の値には依存するが、当該変数の欠損となった値には依存し

ないメカニズムである。3つ目の MNAR(Missing not at random) は、変数の欠損する確率が、その変数自体の値に依存するメカニズムである。データ収集の際に、MAR・MNAR で欠損する例を挙げるとセンサの接触不良、電力不足、通信障害などあるが収集の度にすべての解決を行うことは手間が大きくかかり、手間に対して得られるデータの相対的な価値が下がってしまう。そこで、欠損による問題の解決策として、センサデータの取得方法の改善でなく、手間と価値のバランスをとるためセンサ行動認識の過程において推定精度を向上させる方法が必要である。

センサデータに欠損があると機械学習の前段階である特徴量算出に問題が生じる。内閣府経済社会総合研究所によると、統計学として本来観測・記録されるべき値の一部が観測・記録されていない不完全なデータを利用した推定には推定精度の低下だけでなく、バイアスという問題も発生するとしている [7]。バイアスとは、推定結果に生じる偏りのことである。失われた一部のデータと観測・記録されたデータの異質性が大きいほどバイアスも大きくなる。欠損値に適当な値を代入してデータの補完を行う手法もあるが、その代入した値によってバイアスが生じる可能性がある。介護・看護分野における行動認識は豊富なデータセットがないため、個人に依存しない汎用性をもつ特徴量のためさらなる検証が必要とされている。

4. 関連研究

文献 [8] の研究は複数のセンサを使用しているが、消費電力を抑えるためにデータを定期的に欠損させる。その欠損値を補完し、行動認識をしている。補完方法は他のセンサデータとの相関から補完する手法など複数の手法を用いている。しかしセンサが単独の場合には用いることができないという問題がある。

文献 [9] の研究では、センサデータの欠損の発生原因を1つ以上のウェアラブルセンサが、通信切断や故障することとしている。また、故障したセンサのデータが永続的に途絶する場合を想定している。故障したセンサが永続的に途絶する場合、分類モデルの作成に使用したセンサデータの数よりも得られるセンサデータの数が減り、特徴量ベクトルの次元数が減少する。この問題に対して、欠損したセンサデータの直前のセンサデータから欠損データの補完を行うという手法で、行動認識することが目的である。この手法では、過去データによる影響を受けた欠損データの補完を行うことから過去データによるバイアスが大きいという問題がある。

このように、過去の研究では、テストデータの欠損に対して補完をするという方法が一般的であったが、本研究では、学習データをテストデータと同様の欠損率を与えるという新たな方法で欠損の問題に取り組む。

先行研究 [2] では、センサデータの欠損が発生する原因

を解決することで欠損を減らしている。主にデータ収集時のデバイスの不具合改修やネットワークの調整である。この実験では調整後にデータの収集率が大幅に改善されたが、1日毎のデータ量は一定でなく不安定である。つまり収集したデータに欠損が含まれ、原因の解決に尽力してもデータの欠損を減らす限界があると考えられる。

4.1 本研究の目的

本稿は行動認識におけるセンサデータ欠損時の認識精度向上のための手法を提案する。先行研究 [2] では、取得するセンサデータが欠損し、行動推定精度が低下する問題がある。つまり、行動認識は欠損したデータの量に応じて誤った推定をしやすくなる。介護施設におけるセンサの使用は、センサが人を傷つけるという危険性を伴うため、可能な限りセンサの数が少ないことが望ましい。よってセンサデバイスがスマートフォン単独の場合にもセンサデータの欠損に影響を受けにくい行動認識手法を提案する。

5. 提案手法

本稿の行動認識の手法は、欠損のあるテストデータに対して推定精度を高くすることが目的である。この手法について以下に詳細を記載する。提案手法の具体的適用先として、先行研究 [2] を念頭に置いているため、提案手法には適用する条件が存在する。先行研究 [2] はデータの収集において欠損が生じることは2章で述べた。実施場所は、介護付有料老人ホームであり、スタッフおよび入居者に事前に同意を得た上で実証実験は行われ、この施設における行動認識の導入条件として、センサをできる限り小さくすることが挙げられた。理由としては、恒例の被介護者が怪我をしやすいため、危険を含むものは身に付けることが好ましくないからである。この条件により、センサの大きさが制限され、電池の容量も同様に制限される。つまり、センサの種類・数が限られるということである。したがって、提案手法に関係する点として、単独のセンサデバイスのスマートフォンから行動認識をできる必要があると仮定した。また教師データセットについては事前に作成することとする。

行動認識の介護・看護応用における前提

- 単独のスマートフォンで行動認識ができること
- 学習データには欠損がないこと
- テストデータには欠損が含まれていること

5.1 提案手法の手順

一般的な行動認識手法と比較すると、本稿の提案手法は、故意的に欠損させた学習データから分類モデルを作成し、欠損しているテストデータの推定を行う点が異なる。提案手法は、学習データを欠損させる処理が増えるが、テストデータに欠損がある場合において学習データを欠損させな

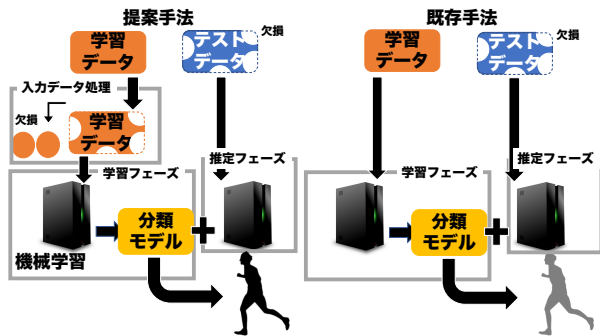


図 3 提案手法と既存手法の全体の流れ
(提案手法のみ、故意的に学習データを欠損させる手順がある)

かった場合より高い推定精度で行動認識ができると考えられる。全体的な手法の流れは図 3 に示す。以下に提案手法の概要をまとめ、次の節に詳細な記載の必要な部分について述べる。

5.1.1 入力データの欠損

入力データに提案手法の学習データの欠損を行う。

- (1) 欠損させたテストデータの欠損率で学習データの欠損量を設定する。
- (2) 学習データに欠損を与える。

5.1.2 学習フェーズ

学習フェーズでは、分類モデルの作成を行う。

- (1) セグメンテーションをする。
- (2) 欠損させた学習データの区画毎の特徴量を算出する。
- (3) 機械学習により欠損させた学習データから分類モデルの作成を行う。

5.1.3 推定フェーズ

推定フェーズでは、テストデータの行動推定を行う。

- (1) 欠損されたテストデータを時間窓毎に分割し、特徴量を抽出する。
- (2) 作成された分類モデルを使い、機械学習により欠損されたテストデータの行動推定を行う。
- (3) 推定できなかった行動ラベルは推定できなかったという行動クラスとして扱う。

5.2 詳細

以下の詳細をこの節において述べる。(1) 入力データ、(2) 欠損方法、(3) センサデータの加工法 (4) 機械学習

5.2.1 入力データ

入力データは、目的変数と説明変数の行列である。学習データは、センサから得られた値と行動ラベルがあり、テストデータは、センサから得られた値のみである。

5.2.2 入力データの欠損方法

学習データの欠損方法について、行動推定する予定のテストデータにおける欠損している箇所を数え、テストデータの総数で割ることで欠損率を算出する。算出された欠損率に従い、学習データ内の無作為に選択された箇所のデー

タを削除する。

5.2.3 センサデータの加工

学習データ・テストデータ共に同じ特徴量抽出をする。まず入力データにセグメンテーションを行う。セグメンテーション方法は移動時間窓を用いる [5]。

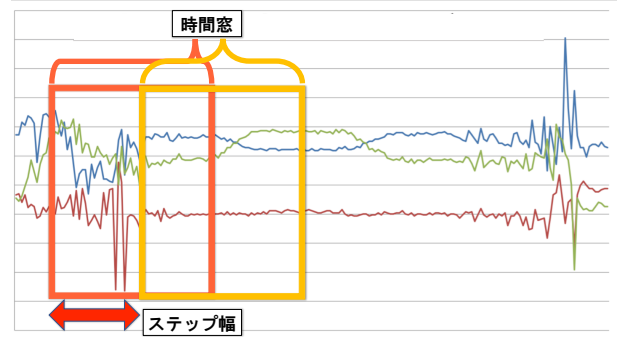


図 4 移動時間窓

その後、時間窓毎に統計量に基づく特徴量の計算を行い、機械学習に入力可能な形のデータとする。

5.2.4 機械学習

機械学習の 1 つである、RandomForest を用いる [10]. RandomForest は教師データセットの学習データの中から分類モデル作成に使うデータをランダム選択することで、教師データのノイズの影響を受けにくい学習機である。決定木を多数作成し、作成された決定木を組み合わせることで分類・推定を行う。

6. 評価実験

本章では、2 種類の実験データセットを用いて、提案手法の評価実験を行う。手順を説明した後に詳細な内容、実験データセット毎に結果の順に記載する。

6.1 目的

現在、介護・看護のデータセットは数が少ないため、提案手法には、代替の実験データセットを用いて、推定精度を求めて、提案手法を評価する。実験データセットは 2 種類用意する。1 つ、シミュレーションされたデータセット (以降シミュレーションデータセットと呼ぶ。) である。異なる数式から作成された波形をセンサデータとし、その数式の違いの推定精度を求めることで提案手法の分類モデルの評価をする。1 つ、HASC チャレンジデータセットである [11]. HASC とは人間行動理解のための装着型センサによる大規模データベース構築を目的とした研究団体 “Human Activity Sensing Consortium” であり、この目的のもとに作成されたデータセットが HASC チャレンジデータセットである。人の行動推定を行うことが提案手法の目的であり、人の行動は、個人毎に異なり、シミュレーションで作成することが困難であるため、HASC チャレ

ンジデータセットを利用する。また提案手法の比較対象として、既存手法の行動認識における推定精度も求める。シミュレーションデータセットを使用した実験をシミュレーション実験、HASC チャレンジデータセットを使用した実験を HASC 実験とする。

6.2 実験データセット

6.2.1 シミュレーションデータセット

提案手法のデータを分類する能力について評価をするためにシミュレーションデータを用いる。学習データは 300 行 3 列のベクトル行列である。各列が時間・説明変数・目的変数であり、説明変数は、正弦波・のこぎり波・方形波を組み合わせたセンサデータ値、目的変数は、正弦波・のこぎり波・方形波のクラスを用意する。テストデータは 250 行 3 列の行列であり、説明変数は、学習データで用意した説明変数を図 6 のように組み合わせたセンサデータとして、目的変数の推定を行う。

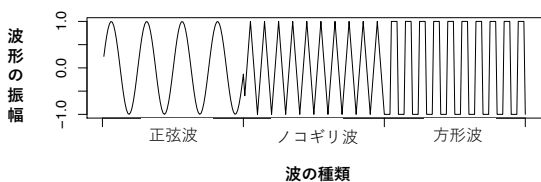


図 5 シミュレーションデータによる学習データ

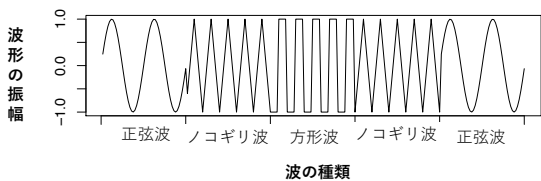


図 6 シミュレーションデータによるテストデータ

6.2.2 HASC チャレンジデータセット

本節において、行動認識をするためのデータセットとして HASC チャレンジデータセットを用いる。この HASC チャレンジデータセットには 540 人の被験者を含む 6700 以上の加速度データが収集されている。データセットには 6 つの行動（直立・歩行・ジョギング・スキップ・階段の登り・階段の下り）毎に分けられているもの（分別データ）と 6 つ行動が無作為に組み合わせられたもの（シーケンスデータ）の 2 種類がある。それぞれの行動毎のデータの 1 回の測定時間は 20 秒であり、無作為に組み合わせられたデータは 5 分である。またサンプリング周波数は 100Hz である。

6.3 実験方法

テストデータの欠損率を 0～100%まで上昇させて、5%毎に推定を行う。更に、HASC チャレンジデータセットの実験時、テストデータの欠損率 85～99%までの範囲で欠損率を 1%毎に推定を行う。推定を行うための分類モデルの作成に使用する学習データもテストデータに応じて 0～100%まで欠損率を上昇させる。今回の実験ではテストデータの欠損を、無作為に選ばれたデータの行を削除することで作成するため 10 種類のテストデータを用意した。それに応じて学習データも 10 種類作成したが、学習データとテストデータの欠損の仕方によって結果に影響を及ぼさないように配慮した。

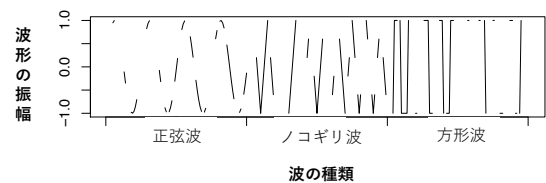


図 7 シミュレーション実験の学習データの欠損後 (欠損率 40%)

[h] シミュレーションではなく、実際のデータの欠損の仕方は、この図のほかにもあるが、実際の欠損は規則性が簡単には導くことができないことから欠損したセンサデータを作成する場合は、無作為にセンサデータの一部を欠損することで再現する。しかしそれは無作為ではあるが推定精度に影響を及ぼす可能性がないとは言い切れないため、無作為なセンサデータの欠損を 10 回試行する。

既存手法の学習データは欠損のないデータを使用する。提案手法の学習データは、テストデータに応じた欠損した学習データを使用する。既存の手法と提案手法の異なる点はこの学習データの欠損の有無のみである。

推定ができなかった時間窓については、目的変数が NA (欠損値) として返されるため、Unknown という行動クラスを作成し、置換する。

この手順を、用意したテストデータ 10 種類分繰り返す。
シミュレーション実験の方法

提案手法をシミュレーションデータに用いて行動認識の推定精度の確認をする。特徴量には、一般的な行動認識に用いられる平均と分散を使用する。欠損のない状態（以下完全な状態とする）のセンサデータと欠損した状態のセ

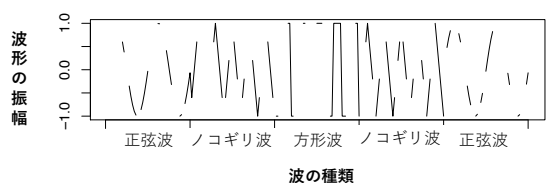


図 8 シミュレーション実験のテストデータの欠損後 (欠損率 40%)

ンサデータによる推定精度の確認をする。この実験では、時間窓幅を10行、ステップ幅を5行として使用し、特徴量の平均と分散を作成し、推定するための機械学習には、RandomForestを使用する。

HASC 実験の方法

提案手法をHASCチャレンジデータセットに用いて行動認識の推定精度の確認をする。今回使用するHASCチャレンジデータは、10人分の3軸加速度のデータであり、サンプリング周期は100Hzである。HASCチャレンジデータセットの分別データを学習データ、シーケンスデータをテストデータとして提案手法の評価実験を行う。つまり、クロスバリデーションを行うことにより、汎用的なモデルの作成を行う。クロスバリデーションとは、書籍[12]より、交差妥当化とも呼び、モデルの評価を行う場合にそのモデルの母数の推定に用いたデータは利用せずに、それとは別に得られたデータへの当てはまりの良さを利用する方法である。特徴量は、3軸加速度センサデータのX軸・Y軸・Z軸の平均と分散を抽出する。この実験では、文献[13]を参考に、時間窓幅を2.56秒、ステップ幅を0.64秒として使用し、特徴量の平均と分散を作成し、行動推定するための機械学習には、RandomForestを使用する。

6.3.1 評価方法

既存手法との比較をすることで有効性を確認する。比較する指標としては、行動認識の推定精度の平均値を用いることとする。環境はR言語を用いて、機械学習アルゴリズムはライブラリRandomForestを用いる。

6.4 実験結果

6.4.1 シミュレーション実験

シミュレーション実験の結果を図9に示す。図9は、各手法において得られた推定精度の結果を示している。横軸は、各テストデータの欠損率であり、縦軸は推定精度である。また、図9の折れ線グラフは試行したすべての場合で得られた推定精度の平均、各テストデータの欠損率における垂直線は、標準偏差を表している。図9の赤線が提案手法によるグラフ、青線が既存手法によるグラフである。ただし、提案手法において、テストデータの欠損率が100%のときは推定を行うことができなかったため、精度は0%としている。

既存手法と提案手法の平均推定精度は、テストデータの欠損率20%の場合で最大の差となり、既存手法が74.8%、提案手法が87.4%であった。これは、提案手法による影響で12.6%向上していると言える。また他のテストデータの欠損率においても、テストデータの欠損率が5–85%では提案手法の推定精度のほうが良いという結果であった。

6.4.2 HASC 実験

HASC実験の結果を図10に示す。図10は、各手法において得られた推定精度の結果を示している。横軸は、各テ

ストデータの欠損率であり、縦軸は推定精度である。図10の折れ線グラフは試行したすべての場合で得られた推定精度の平均、各テストデータの欠損率における垂直線は、標準偏差を表している。図10の赤線が提案手法によるグラフ、青線が既存手法によるグラフである。ただし、テストデータの欠損率100%では、テストデータがなく、推定が行えないため、精度は0%としている。

テストデータの欠損率85%以下では提案手法と既存手法の平均推定精度の差は1%未満であり、標準偏差内に含まれるため誤差と考えられる。テストデータの欠損率90%以上において提案手法の推定精度のほうが良く、最大の差が欠損率95%の場合で、46.8%から49.8%と2.9%向上した。

しかし、HASC実験では、テストデータの欠損率全体において、提案手法と既存手法の推定精度差が小さくなった。提案手法の有効性を確認するために平均推定精度差が現れたテストデータの欠損率の範囲において、テストデータの欠損率1%毎に平均推定精度を算出した。既存手法と提案手法の平均推定精度は、テストデータの欠損率99%の場合で最大の差となり、既存手法が37.4%、提案手法が44.6%であった。テストデータの欠損率85%–99%における推定精度の結果は、図11に図10と同様に示した。ただし、図11のx軸の範囲を、85%–99%とした。

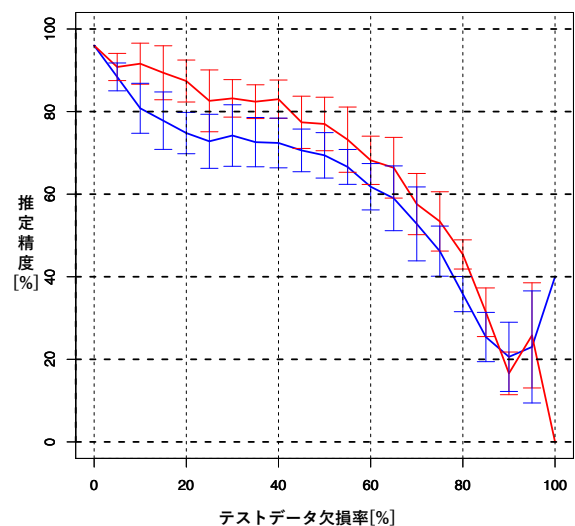


図9 シミュレーション実験における各テストデータの欠損率(0–100%)の推定精度の変化
(提案手法が赤線、既存手法が青線。)

7. 考察

シミュレーション実験から提案手法の有効性は認められる。しかし、HASC実験では、シミュレーション実験の結果のようにならなかったことから、HASC実験結果の改善点について考察する。

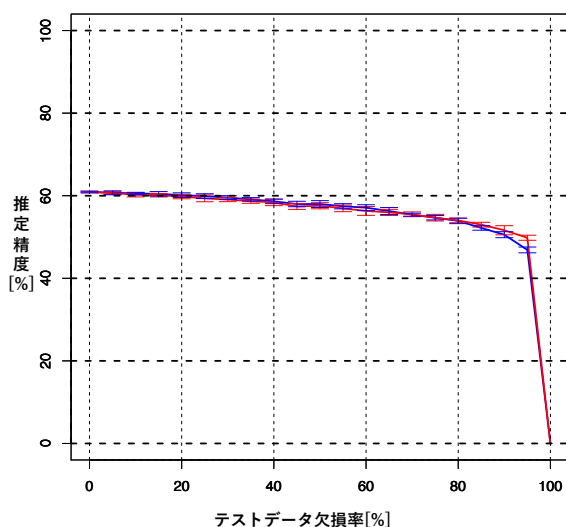


図 10 HASC 実験における各テストデータの欠損率 (0–100%) の推定精度の変化
(提案手法が赤線, 既存手法が青線)

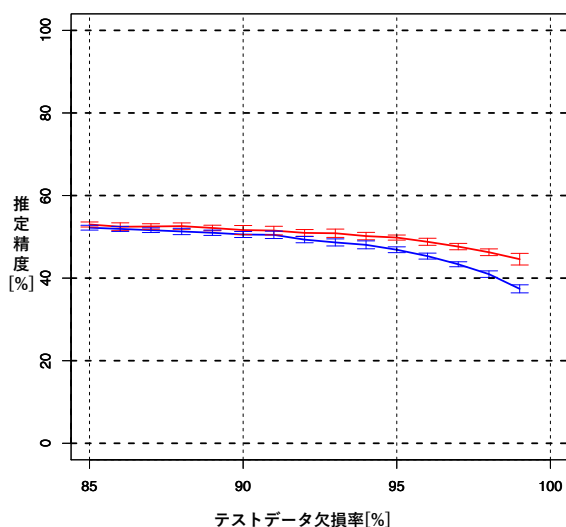


図 11 HASC 実験における各テストデータの欠損率 (85–99%) の推定精度の変化
(提案手法が赤線, 既存手法が青線)

7.1 実験の分析

2つの実験において、提案手法により推定精度が改善していることから、提案手法は有効であると考えられる。しかし、シミュレーション実験と HASC 実験の結果を比較すると、異なる点は2つあると考えられる。1つは、提案手法が有効であるとされるテストデータ欠損率の範囲が異なること。1つは、2つの手法における推定精度が全体的に低いことである。原因は、シミュレーション実験と異なるデータセットを使用したことだと考えられる。このデータの内容に

よって、データ欠損の有無という特徴量の効果が低くなったと考える。

1つ目の異なる点について考察する。文献 [14] によると、行動の種類が少ない場合、行動認識に必要なセンサのサンプリング周波数は $5Hz$ 以上であれば良いとされている。HASC 実験では、 $100Hz$ のサンプリング周波数で作成された加速度の値を使用していた。したがって、テストデータ欠損がテストデータの全体的に一律に加えられているとすると、提案手法が有効になったテストデータ欠損率の範囲では、センサのサンプリング周波数が $15Hz$ 以下であり、行動認識に必要なサンプリング周波数に近づいている。つまり、提案手法は既存手法の推定精度が変化しない場合には、有効性が低く、推定精度が減少する場合において有効性が高いと推測される。

2つ目の異なる点について考察する。文献 [13] によると、HASC チャレンジデータセットを使用した行動認識で、特徴量として、平均・分散を用いた場合、推定精度は60%前後になると結果が出ている。文献 [13] では、他に特徴量を7つ使用した場合や、8つ使用した場合の結果も出している。特徴量を増やした場合、推定精度は70%前後まで上昇するとされている。よって、原因は特徴量が少ないことであると推測される。

これら2つの点により、HASC 実験の結果がシミュレーション実験の結果と異なると考えられる。しかし、どちらの実験においても提案手法が既存手法より有効である欠損率の範囲はある。したがって提案手法の有効な環境をより詳細に把握するために、サンプリング周波数の減少・特徴量の増加を行い、さらなる検証が必要である。

7.2 実験データの欠損の現実性について

本節では、スマートフォンでの加速度取得において欠損の現実性を確認するため、スマートフォンでのセンサデータ取得を行った結果を示す。本稿では、無作為に選ばれたデータの一部を欠損させることで欠損センサデータを作成したが、現実の欠損と比較する必要がある。実際にスマートフォンを利用して加速度を取得するときに、データ欠損が生じる条件の1つとして、端末のメモリを圧迫することが挙げられる。そのメモリの圧迫を故意に発生させた場合、時間窓毎のデータの欠損率は図 12 のようになる。メモリの圧迫率のほかの明確でない要因にもデータの欠損を生じることがあり、すべての欠損が図 12 のようにはならないが、実際に発生した図 12 の場合でも提案手法が効果的であることを今後確認する必要がある。

8. おわりに

本研究は、先行研究 [2] におけるセンサデータ欠損によるスマートフォンを用いた行動認識の推定精度の低下問題の改善を目的として、新たな手法を提案した。提案手法

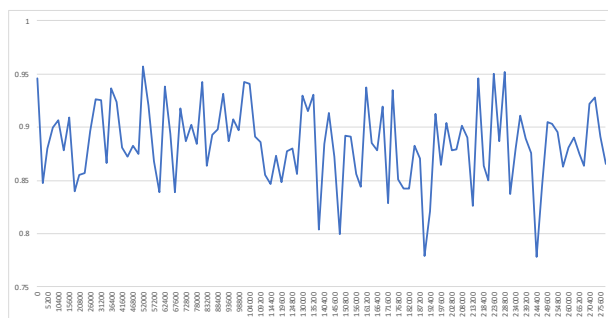


図 12 端末のメモリ圧迫による時間窓毎のデータ欠損率 (縦軸：一定時間窓あたりのデータ欠損率，横軸：時間)

の学習データに欠損を与えるという方法は、欠損のないテストデータの教師データを変更していることになり、推定精度が下がると考えられる。しかし、一般的に人の行動によって取得されるセンサデータから作成されるテストデータは欠損を含むことが多い。したがって提案手法は、教師データをより現実に近い状態に変更したことで、推定精度の向上に繋がった。評価実験の結果から、提案手法が無作為なセンサデータ欠損に対して有用であることが確認できた。評価実験において、欠損を無作為に作成することで現実におけるデータの欠損を想定した。もしデータの欠損の原因、もしくは欠損した値を補助する説明変数があれば提案手法以外の方法で欠損を補完することができる。

しかし、すべての欠損の原因解明や、補助する説明変数の情報を収集することは、手間が大きくかかる。したがって、本稿の提案手法を用いることで、手間を小さく推定精度を改善することが可能となる。有効なセンサのサンプリング周期の導出、より高度な特徴量の増加・選択を行うことで、提案手法による行動認識精度は、さらに向上するという可能性を示している。

本研究の今後の展開としては、無作為に欠損させたテストデータでなく、実環境で人の行動をセンシングする過程において欠損したテストデータに提案手法を適用することである。しかし、現在介護・看護に関するセンサデータは少ない。この実験データへ提案手法を適用し、センサデータの欠損による問題の解決を今後も引き続き、目標とするとともに、新たに再開した介護施設における実験のデータセット作成のため、センサデータの収集システムの構築の研究と結びつけることとする。

参考文献

- [1] 厚生労働省: 2025 年に向けた介護人材にかかる需給推計(確定値)について, <http://www.mhlw.go.jp/stf/houdou/0000088998.html> (2018.01.25)
- [2] 井上創造, 木村幸平, 内野百里, 大屋誠: 介護施設における介護スタッフの行動センシング実験, 研究報告ユビキタスコンピューティングシステム (UBI), Vol.2017, No.54, pp.1-8(2017)
- [3] 玉水一柔, 榊原誠司, 佐伯幸郎, 中村匡秀, 安田清: 変化時の

- 行動記録と環境センシングに基づく機械学習を用いた宅内行動認識, 信学技報, Vol.117, No.251, pp.101-106(2017)
- [4] 内閣府: 平成 29 年度版高齢社会白書, http://www8.cao.go.jp/kourei/whitepaper/w-2017/zenbun/29pdf_index.html(2018.02.02)
- [5] 井上創造: ウェアラブルセンサを用いたヒューマンセンシング, 知能と情報, Vol. 28, No. 6, pp. 170-186 <http://ci.nii.ac.jp/naid/40021082607/en/> (2016)
- [6] 勝手美紗, 内海ゆづ子, 黄瀬浩一: 物体と動き特徴を用いた行動認識, 電子情報通信学会技術研究報告.PRMU, パターン認識・メディア理解, Vol.111, No.430, pp.125-126(2012)
- [7] 内閣府経済社会総合研究所景気統計部: 欠測値補完に関する調査研究報告書 (2017)
- [8] 中島悠貴, 村尾和哉, 寺田努, 塚本昌彦: サンプリング制御とデータ補完による行動認識システムの省電力化手法, 情報処理学会論文誌, Vol.52, No.8, pp.2455-2466(2011)
- [9] 内田稜真, 堀野豊人, 大村廉: ARAR アルゴリズムによる欠損データ補完を用いたウェアラブルセンサ行動認識技術の耐故障性向上, 電子情報通信学会技術研究報告, Vol.112, No.495, pp.245-252(2013)
- [10] Sozo Inoue, Naonori Ueda, Yasunobu Nohara, and Naoki Nakashima. Mobile activity recognition for a whole day : Recognizing real nursing activities with big dataset. UbiComp 2015-Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp.1269-1280(2015)
- [11] 川口信夫, 小川延宏, 岩崎陽平, 梶克彦, 寺田努, 村尾和哉, 井上創造, 川原圭博, 角康之, 西尾信彦: HASC Challenge2010, 人間行動理解のための装着型加速度センサデータコーパスの構築, マルチメディア, 分散協調とモバイルシンポジウム 2011 論文集, Vol.2011, pp.69-75(2011)
- [12] 豊田秀樹: データマイニング入門, pp.15-17, 東京図書株式会社 (2008).
- [13] 小川延宏, 梶克彦, 川口信夫: HASC2010corpus を用いた被験者数と人間行動認識率の相関分析, マルチメディア, 分散協調とモバイルシンポジウム 2011 論文集, Vol.2011, pp.76-82(2011)
- [14] 倉沢央, 川原圭博, 森川博之, 青山友紀: センサ装着場所を考慮した 3 軸加速度センサを用いた姿勢推定手法, 情報処理学会研究報告ユビキタスコンピューティングシステム (UBI), Vol.2006, No.54, pp.15-22 (2006)