

環境センサから推定される時系列行動尤度ベクトルによる行動シーケンスの推定

西 高史[†] 櫛 惇志[†] 宮崎 純[†]

[†] 東京工業大学情報理工学院情報工学系 〒152-8550 東京都目黒区大岡山 2-12-1

E-mail: [†]{nishi,keyaki}@lsc.cs.titech.ac.jp, ^{††}miyazaki@cs.titech.ac.jp

あらまし ユーザの行動を環境センサのみからの情報により学習器で推定し、学習器が出力する尤度の低い行動のシーケンスを扱うための手法を提案する。行動推定時、学習器は各行動の尤度の計算を行う。従来の行動シーケンスの推定手法では最も尤度が高い行動の組み合わせを行動シーケンスとして推定していた。しかし環境センサの値は誤差が生じやすく学習器の推定性能が低くなるため、正解となる行動の尤度が最も高くない場合があり、正答率が下がるケースがある。そこで、行動間の関係性に着目し、各行動の尤度を一つの尤度ベクトルとして纏める。また尤度ベクトルの挙動に着目し、行動ラベル付き尤度ベクトルを行動毎にクラスタリングを行う。本研究では、クラスタと行動ラベル不明の尤度ベクトルの距離により行動を推定、時間窓による補完を行い、最終的な推定結果の組み合わせを行動シーケンスとして出力するシステムを提案し、その有効性の評価を行う。

キーワード 環境センシング, センサデータ, 尤度, クラスタリング, 行動認識

1 はじめに

近年、無線通信技術やセンサ技術の発展に伴い、スマートフォンやウェアラブルセンサなどの小型電子デバイスに温度センサや加速度センサなどの様々なセンサを搭載することが可能となった。これにより、センサデータによるビッグデータ収集や、遠距離にあるモノの状態を把握するといった事が可能となり、IoT (Internet of Things) 時代に活用できる技術として、近年盛んに研究が行われている。その分野の一例として、センサから人間の行動を認識する技術が注目されており、行動や行動シーケンスを把握することにより、独居老人のケアやヘルスケアなどのサービスの向上を行う事が期待されている。例えば、ウェアラブルセンサによる行動認識 [5]、スマートフォンに搭載されたセンサを用いた行動認識 [13] など様々な研究がなされてきた。本研究では温度センサや湿度センサなどの空間内の物理的な環境を数値化するセンサ（環境センサ）を部屋に設置し、室内における行動シーケンスを推定することを試みる。環境センサによる行動認識は、動画像による行動認識より通信データ量とプライバシーに配慮しており、ウェアラブルセンサのようにセンサを持ち運びする必要がない点で優れている。

センサデータを利用して行動認識を行うにあたって、機械学習を用いることが多い [11]。学習時にはセンサの値で構成されたベクトル（データベクトル）に行動ラベルを付け学習器の作成を行う。推定時には行動ラベル不明のデータベクトルを学習器に入力し、行動を推定する。しかし、環境センサによる行動認識は、数種類のセンサで行うためデータベクトルに外れ値が入る頻度が高くなり、かつ類似する行動を正確に判別することが難しいため、学習器の推定性能が低くなる。これにより、既存の手法では正解となる行動が推定されにくく、正しい行動シーケンスを推定する事は困難である。そこで本研究では、類似する行動の尤度情報も行動ラベル推定の材料になると仮定し、

各行動の尤度を一つの尤度ベクトルとして纏める。尤度の高さのみに依存しない推定手法として、行動ラベル付きの尤度ベクトルを行動毎にクラスタリングを行う。生成されたクラスタには尤度が最大でない行動ラベルが割り当てられる事もあり、かつ類似する行動の尤度情報も含まれている。行動ラベル不明の尤度ベクトルと各クラスタの距離を測ることにより行動ラベルを推定し、連続した推定結果を時間窓による平滑化などの後処理を行い、最終的な行動ラベルの組み合わせを行動シーケンスとして推定することを試みる。

本稿では本手法の実装を行い、その性能について検証を行う。評価実験では、最尤法によって出力された行動シーケンスと、提案手法によって出力された行動シーケンスを比較し性能を評価する。

2 関連研究

環境センサを用いた行動認識に関する研究は、動画像による行動認識やウェアラブルセンサによる行動認識技術と比べると少ない。鈕らはセンサボックスと呼ばれる小型コンピュータを一つの部屋に設置し、室内における行動推定を [15]、Van らは家の各所にセンサの設置し、屋内における行動推定を行っている [9]。しかしこれらの研究では、学習器による行動単体での推定のみを行い、行動シーケンスを取り出すまでのことは行っていない。

センサによる行動認識で尤度を用いた研究として、杉浦らは尤度のシーケンスから特定の行動パターンを発見する手法を提案した [12]。しかしこの研究では、パターンマッチングに焦点を当てており、行動シーケンスの推定は行っていない。本研究では、行動単体での推定や行動パターンの発見ではなく、行動シーケンスを推定することを試みる。

一般に時系列データから頻出パターンを発見することが行われる。時系列データから、出現順序を維持しつつ特定の閾値を

満たす部分列を発見する手法を時系列パターンマイニングと呼び、様々な手法が提案されている [4]。本研究で推定した行動シーケンスに時系列パターンマイニングを適用することで、頻出する行動パターンを取得することができ、独居老人のケアなどに応用することが考えられる。

3 準備

3.1 環境センサを利用した行動認識

環境センサを用いた行動認識の長所として、動画像による行動認識 [14] よりも通信データ量が少なく、かつプライバシーに配慮しており、ウェアラブルセンサやスマートフォンによるセンサ行動認識技術 [9] のようにセンサを持ち運びする手間がない。その一方で、他の行動認識技術よりも推定性能が悪く、行動を推定できる場所が限られるといった短所がある。環境センサを用いた行動認識の推定性能が悪い理由として、複数の環境センサを用いることによって外れ値が入る頻度が高くなることと、類似する行動の判別が難しいことが挙げられる。後者の例として、「椅子に座ってテレビを見る」行動と「椅子に座って本を読む」行動が挙げられる。現在の環境センサでは、「椅子に座って何かしらのことをしている」という事を推定できる可能性は高いが、どちらかを正確に判別する事は難しく推定結果に偏りが出やすい。例えば、独居老人のケア等が目的である場合は、ユーザが正常な生活ができているかを判断することが重要なため、「外出」「睡眠」のような健康に関わる行動は個別に認識すべきだが、「テレビを見る」「読書」など、健康状態に大きく関与しない行動に関しては、正確に識別する必要がない場合も多い。本研究では、このような正確な識別を要しない行動を許容し、これをオーバーラップと呼ぶ。

3.2 機械学習と最尤法

一般的な教師あり機械学習 [10] では、センサの値や画像の各データの組であるデータベクトルを用いて学習や推定が行われる。学習時には正解のラベルを付けたデータベクトルを集め、機械学習のアルゴリズムに入力することで学習器が作成される。推定時には学習器にラベル不明のデータベクトルを入力し、ラベルを推定する。推定時、各ラベルの尤度の計算が行われ、尤度が最も高いラベルを推定結果として出力している。本研究では、推定結果を直接用いるのではなく推定時における各行動の尤度の情報を用いる。行動ラベルを推定する際、オーバーラップする行動の尤度も行動推定の材料となると考えるため、各行動の尤度を一つの尤度ベクトルとして纏める。

機械学習によるラベル推定を連続で行う場合、連続した推定結果を並べることで、ラベルのシーケンスを取得することができる。機械学習の推定結果からラベルのシーケンスを取得する手法として、推定時の尤度が最大となるラベルを繋ぎシーケンスとして取得する手法（最尤法）があり、多くの分野 [6], [8] で用いられている。本研究では、提案手法と最尤法による行動シーケンス推定との性能比較を行う。

3.3 クラスタリングと所属判定

クラスタリング [3] とは、分類対象の集合を内的結合と外的分離が達成されるような部分集合に分割することである。クラスタリングを行う事で、データ間の距離が近いものを同じクラスタとして纏め、データ間の距離が遠いものを別のクラスタとして分けることができる。本研究では、尤度ベクトルにより行動を推定する手法として、尤度の高さのみに依存しない推定を行うために行動ラベル付きの尤度ベクトルを用意し、行動毎にクラスタリングを行う。

クラスタリングの後、行動ラベル不明のデータがどのクラスタに所属するかを判定する（クラスタ所属判定）。本研究では、行動ラベル不明の尤度ベクトルを各クラスタとのマハラノビス距離を測定することによりクラスタ所属判定を行う。マハラノビス距離 [2] は、クラスタの平均ベクトルを $\vec{\mu}$ 、クラスタの分散共分散行列を Σ 、行動ラベル不明の尤度ベクトルを $\vec{x}$ とした時に以下で定義されるクラスタの平均ベクトルと行動ラベル不明の尤度ベクトルとの距離 d である。

$$d = \sqrt{(\vec{x} - \vec{\mu})\Sigma^{-1}(\vec{x} - \vec{\mu})}$$

マハラノビス距離の特徴として、多次元におけるデータの分散を考慮していることが挙げられる。

クラスタ内の各点が p 次元正規分布をしていると仮定した場合、マハラノビス距離の二乗は自由度 p のカイ二乗分布に従う性質を持つ。カイ二乗分布は確率分布の一種であり、この性質を利用してクラスタ内部に含まれるデータの割合を計算することが可能となる。本研究では、行動ラベルの候補数を p として、各クラスタ内の点が p 次元正規分布をしていると仮定する。

4 提案手法

3.1 節で述べたように、環境センサを用いた行動認識の問題点として、正解となる行動の尤度が最も高くならない場合があることが挙げられる。その結果、最尤法では正しい行動シーケンスを推定することが困難となる。そこで本研究では、類似する行動の尤度情報も行動ラベル推定の材料になると仮定し、各行動の尤度を一つの尤度ベクトルとして纏める。尤度ベクトルを用いた行動ラベル推定のため、行動ラベル付きの尤度ベクトルを用意し、行動毎にクラスタリングを行う。生成されたクラスタには尤度が最大でない行動ラベルが割り当てられる事もあり、かつ類似する行動の尤度情報も含まれている。行動ラベル不明の尤度ベクトルとクラスタの距離を測ることにより行動ラベルを推定し、連続した推定結果を時間窓による平滑化と矛盾するオーバーラップの解消を行い、行動シーケンスの取得を行う。

4.1 全体のアーキテクチャ

本研究で提案するシステムのアーキテクチャの全体図を図 1 に示す。システムのフローには、学習フェーズと行動シーケンス推定がある。

学習フェーズ

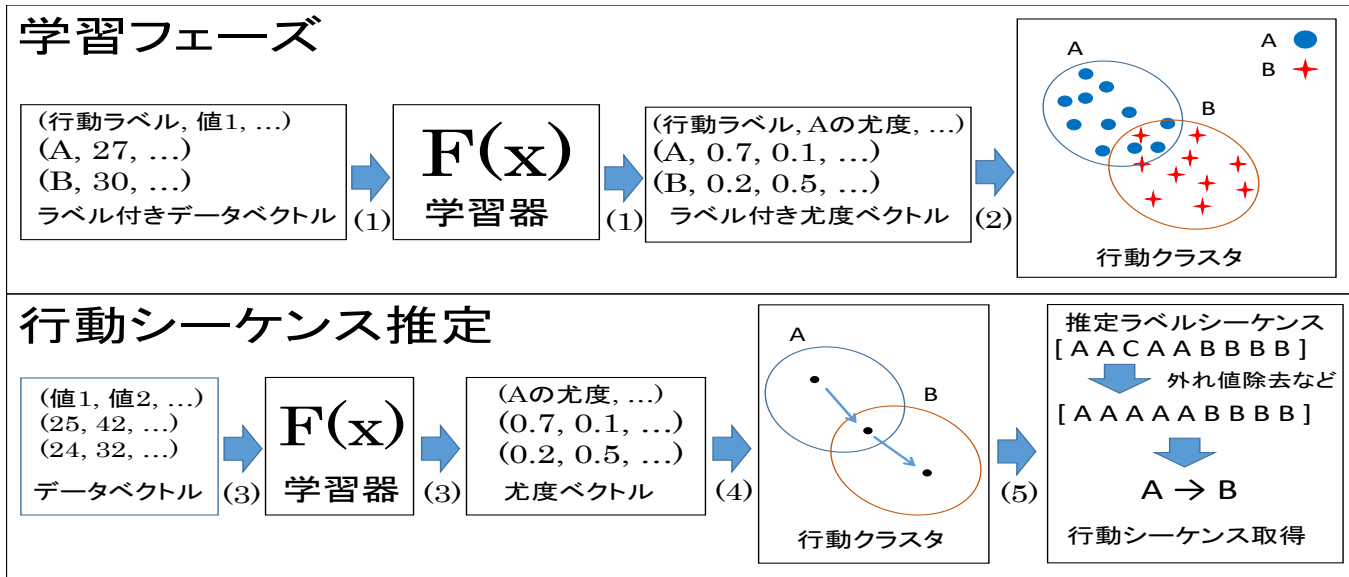


図 1 全体のアーキテクチャ

- (1) 行動ラベル付きのデータベクトルを学習器に入力し，行動ラベル付き尤度ベクトルを取得
 - (2) 尤度ベクトルを行動毎にクラスタリング
- 行動シーケンス推定
- (3) 行動ラベル不明のデータベクトルを学習器に入力し，行動ラベル不明の尤度ベクトルを取得
 - (4) 行動ラベル不明の尤度ベクトルと各行動のクラスタの距離を測り，どのクラスタに属しているかを判定
 - (5) 判定後に時間窓を用いた平滑化と矛盾するオーバーラップの解消を行い，行動シーケンスを取得

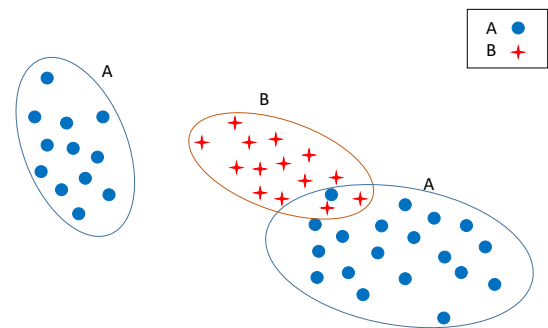


図 2 尤度ベクトルのクラスタリング

4.2 学習フェーズ

4.2.1 尤度ベクトルの取得

類似する行動の尤度情報を用いるため，尤度ベクトルの取得を行う。学習器にデータベクトルを入力し，学習器によって計算された各行動の尤度を一つの尤度ベクトルとして取得する。学習フェーズでは，行動名のラベルが付いたデータベクトルを用いて，行動名のラベルが付いた尤度ベクトルを取得する。

4.2.2 クラスタリング

尤度ベクトルによる行動ラベル推定を行うために，行動名のラベルの付いた尤度ベクトルを集め，行動毎に尤度ベクトルによるクラスタリングを行う。生成したクラスタ (行動クラスタ) は行動ラベル不明の尤度ベクトルとの距離関係により行動の推定を行うために用いる。図 2 では尤度によって構成された空間上に行動 A と行動 B の行動ラベル付き尤度ベクトルがプロットされており，クラスタリングによって距離が近い点を同じクラスタとして纏めている。行動クラスタを作る際，行動ラベル不明の尤度ベクトルと各行動クラスタとのマハラノビス距離を測定するために，クラスタの平均ベクトルと分散共分散行列を求め，その値を保持する。

4.3 行動シーケンス推定

4.3.1 尤度ベクトルの取得

行動ラベル不明のデータベクトルを学習器に入力し，行動ラ

ベルが不明な尤度ベクトルを取得する。

4.3.2 クラスタ所属判定

行動ラベル不明の尤度ベクトルと各行動クラスタとのマハラノビス距離を測定する。行動クラスタとの距離が一定の値以下であれば，行動ラベル不明の尤度ベクトルはその行動クラスタに所属しているとみなす。図 3 では，行動ラベル不明の尤度ベクトルと各行動クラスタとのマハラノビス距離を測定しており，A と B のクラスタに所属していることを表している。

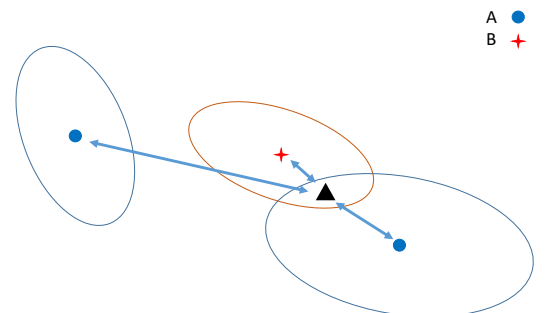


図 3 クラスタ所属判定

このクラスタ所属判定の手法では，複数のクラスタに所属していると判定される可能性がある。提案手法では，例えば健康

状態に大きく関与しない行動に関しては、正確に識別する必要がない場合も多く、このような正確な識別を要しない行動をオーバーラップとして許容する。

4.3.3 行動シーケンス推定

クラスタ所属判定を連続で行い、判定結果を時系列順に並べることにより、行動ラベルシーケンスを取得することができる。ただし、環境センサの外れ値による誤った推定結果をそのまま用いると正しい行動ラベルシーケンスを推定することが困難になる。そこで本研究では時間窓を利用した平滑化を行う事で、外れ値の処理を行う。また、時間窓による処理時に矛盾するオーバーラップの解消を同時に行い、最終的な行動ラベルシーケンスを推定行動シーケンスとして出力する。

4.3.4 矛盾関係

行動のオーバーラップを許容した場合の問題点として、例えば睡眠と外出という通常では同時に起こりえない行動ラベルがオーバーラップする可能性がある。そこで本研究ではそのような関係を矛盾関係とよび、矛盾関係の行動どうしがオーバーラップしている場合は、その候補を出力の候補から破棄する。矛盾関係はアプリケーション依存であり、本研究では、健康に大きく関わらず、かつ行動の位置や体勢などの状況が似た行動どうしは矛盾関係とせず、それ以外の行動どうしは矛盾関係とした。(表 1 を参照)

4.3.5 時間窓による平滑化

時間窓中の行動ラベルを考慮して行動の再推定を行い、局所的に起きる外れ値などを除外する。平滑化では時間窓中で出現頻度の高い行動を優先的に残すため、アプリアリアルゴリズム [1] を利用し、最小支持度以上かつ矛盾関係でない行動の組み合わせのうち、長さが最長かつ支持度が最大となる行動ラベルの組み合わせを計算し、これを時間窓中の行動とする。この平滑化で出力された行動ラベルのシーケンスを行動シーケンスとして出力する。

5 評価実験

本節では、実データを用いて提案手法によりどの程度行動が正しく取得できているか、また、正解行動シーケンスと推定行動シーケンスがどの程度一致しているか、最尤法を用いた行動シーケンス推定手法を比較対象として性能評価を行う。また、行動ラベル毎のヒット率や平均のオーバーラップ数を調査し、その考察も行う。

5.1 実験環境

学習器作成に関して、Java による機械学習ソフト Weka を用いた。学習用アルゴリズムとして、Weka の学習アルゴリズムのうち、評価用のデータセットを用いた試行実験において、最尤法により最も推定精度が高くなったロジスティック回帰を用いた。

クラスタリングに関して、本研究では尤度ベクトルをクラスタリングすることによって生成されるクラスタの数やクラスタを分割するために用いる距離を事前に予測することが難しいと判断したため、クラスタ数を自動推定してクラスタリングを行

う X-means [7] を用いた。

クラスタ所属判定を行う際のマハラノビス距離として、過剰なオーバーラップを防ぐために、自由度 8 (行動の種類数より) のカイニ乗分布の 50% 信頼区間内、クラスタの中心からマハラノビス距離 2.71 以内にある点をクラスタに所属しているとみなす。

時間窓における平滑化において用いる時間窓の長さは 30 に、最低支持度は、過剰なオーバーラップを防ぎつつ、各組み合わせが一つでも最低支持度を上回るような可能性を考慮したため、最低支持度を $2/3$ とした。

5.2 データセット

実験で用いたデータセットは、ある部屋にて 8 種類のセンサ (温度、湿度、音、光度、赤外線、距離、人感、UV、ダスト) を搭載したマイコン (Arduino) を 4 個設置し、1 秒ごとに行動名とセンサデータが紐づいたデータベクトルを収集した。行動はあるユーザが行ったものであり、行動名は以下の 8 種類である。

- テレビを見る (TV)
- 睡眠 (Sleep)
- 外出 (Out)
- 運動 (Workout)
- 本を読む (Read)
- タブレットを操作する (Tablet)
- 食事をする (Eat)
- PC 作業をする (PC)

これらの行動と各マイコンのそれぞれセンサ値、それらセンサ値の 1 分間ごとの平均、マイコン間の各センサ値の差、計 127 個のデータを一つのデータベクトルとして扱う。学習器作成に 12129 個、クラスタ作成に 12128 個、評価用に 58089 個のデータベクトルを用いた。評価用のデータセットでは、1 日弱の連続した行動のシーケンスデータを用いる。各行動一回以上、最低 1 分以上行った。行った行動はのべ 24 となる。また、本実験での矛盾関係を表 1 に示す。True の場合は出力される候補から除外する。

表 1 矛盾関係表

	TV	Sleep	Out	Workout	Read	Tablet	Eat
Sleep	True	-	-	-	-	-	-
Out	True	True	-	-	-	-	-
Workout	False	True	True	-	-	-	-
Read	False	True	True	True	-	-	-
Tablet	False	True	True	True	False	-	-
Eat	False	True	True	False	False	False	-
PC	False	True	True	True	False	False	False

5.3 評価指標

「正解行動におけるヒット率」「各行動ラベルの個別評価」「正解行動シーケンスとの距離」により提案手法の性能評価を行う。共通する指標として、ある時間における正解行動ラベルが同時刻の推定結果に含まれていた場合、推定成功とみなす。推定行動ラベルシーケンスにおける推定成功の割合をヒット率

とする。

以後、行動ラベルを l , t 種類の行動ラベルの集合を $\Lambda = \{l^1, \dots, l^t\}$, 時間窓処理後のオーバーラップを含む推定行動ラベルを L (ただし $L \subseteq \Lambda, L \neq \emptyset$), L を時系列で並べた長さ n の推定行動ラベルのシーケンスを $S = \langle L_1, L_2, \dots, L_n \rangle$, S に対応する正解行動の行動ラベルシーケンス $S' = \langle l'_1, l'_2, \dots, l'_i, \dots, l'_n \rangle$ (ただし $l'_i \in \Lambda, 1 \leq i \leq n$) とする。

推定行動ラベルのシーケンス S は、評価に用いるために全体の推定行動ラベルのシーケンスから指定された箇所のシーケンスを取り出したものである。

5.3.1 ヒット率

推定行動ラベルシーケンス中にどの程度推定成功があるかを示す値であり、推定行動ラベルシーケンスの性能評価に用いる。ヒット率が高いと性能が高いとみなす。正解行動ラベルが同時刻の推定結果に含まれていた場合、すなわち $l'_i \in L_i$ のとき、推定成功とみなす。この時、推定行動ラベルシーケンス $S = \langle L_1, L_2, \dots, L_n \rangle$, 正解の行動ラベルシーケンス $S' = \langle l'_1, l'_2, \dots, l'_n \rangle$ として、推定成功の割合をヒット率 $H(S, S')$ は以下の式で表される。

$$H(S, S') = \frac{1}{n} \sum_{k=1}^n f(k) \quad (1)$$

$$f(k) = \begin{cases} 1 & (l'_k \in L_k) \\ 0 & (l'_k \notin L_k) \end{cases} \quad (2)$$

式 (2) は、正解行動ラベルの時刻と対応する推定行動ラベルにおいて、正解行動ラベルが、推定行動ラベルに含まれているかどうかを確かめ、含まれていた場合には 1 を、そうでない場合は 0 を返す。式 (1) では、各推定行動ラベルにおいて推定に成功した数を、全体の推定行動ラベルの総数で割る。

5.3.2 正解行動におけるヒット率

評価用の正解の行動ラベルシーケンスは、行動開始から次の行動に移移するまで、同じ行動ラベルが続く。このようなシーケンス全体を 1 単位の行動として、それぞれどの程度ヒット率があるかを調査する。正解の 1 行動から見て、どれだけ正確に推定ができているかを確かめる指標とし、提案手法と最尤法による行動シーケンスの性能比較を行う。

5.3.3 各行動ラベルの個別評価

各行動ラベルに対して、オーバーラップした行動ラベル数 $|L|$ の平均とヒット率を集計する。提案手法では行動ラベルのオーバーラップを認め、複数の行動を許容する状態を考えるため、提案手法により、各行動ラベルがどの程度オーバーラップをしているかを確かめる。本研究では、「Sleep」「Out」が健康状態に大きく関わる行動と判定し、「オーバーラップを認めない行動」として扱う。オーバーラップを認めない行動は矛盾関係の処理によりオーバーラップ数が 1 となる。また、各行動の行動ラベルのヒット率について、提案手法と最尤法の比較と推定成功の内訳の調査を行った。

5.3.4 正解行動シーケンスとの距離

推定行動シーケンスの評価を行う。推定された行動は、推定

行動ラベルシーケンスにおいて、時間窓の長さ以上の同一の行動ラベルシーケンスを纏めて一つの行動とみなし、それ以外の推定行動ラベルのシーケンスは外れ値として除外する。これにより取得される推定行動シーケンスと正解行動シーケンスを、Dynamic Time Warping に類似した対応付けより評価を行う。正解行動の行動ラベルを時系列に並べた行動ラベルのシーケンスを $\Sigma' = \langle l'_1, \dots, l'_i, \dots, l'_n \rangle$ として、 l'_i と時刻が一部でも一致する推定行動に対応する行動ラベルのシーケンスを $S_i = \langle L_1^i, \dots, L_k^i, \dots, L_m^i \rangle$ とし、 S_i を正解行動のシーケンスに対応付けしたものを $\Sigma = \langle S_1, \dots, S_i, \dots, S_n \rangle$ とする。ここで l'_i と S_i の編集距離を d_i とした時、正解行動シーケンスと推定行動シーケンスとの距離 $D(\Sigma', \Sigma)$ を以下の式で定義する。

$$D(\Sigma', \Sigma) = \sum_{k=1}^n h(i) \quad (3)$$

$$h(i) = \begin{cases} 1 - e^{-\alpha d_i} & (\exists k, l'_i \in L_k^i, 1 \leq k \leq m) \\ 1 & (otherwise) \end{cases} \quad (4)$$

式 (4) では、正解行動 l'_i に対応する推定行動シーケンス $S_i = \langle L_1^i, \dots, L_k^i, \dots, L_m^i \rangle$ において、一つでも正解行動 l'_i が含まれていた場合は、 l'_i と S_i の編集距離を用いて 0 以上 1 未満の距離を出力する。値が 0 であった場合は、正解行動と完全一致であり、値が 1 に近づくにつれて、正解行動との距離が長くなる。また、推定行動の各推定行動ラベルのシーケンス S_i に一つも l'_i が含まれていなかった場合は、完全間違いとし、距離 1 を返す。 α は定数であり、本実験では $\alpha = \log 2$ とした。式 (3) は各正解行動において、式 (4) によって計算された距離の和を求めたものである。 $D(\Sigma', \Sigma)$ の値が低いほど、正解行動シーケンスと近い推定行動シーケンスと見なし、性能が良い。また、完全間違いの数の調査も行った。

図 4 に正解行動シーケンスとの距離算出の例を示す。図 4 において、正解行動【B】の時刻において、【B】【A】【B,C】が推定行動として出力され、正解行動【A】の時刻において、【B,C】【B】が推定行動として出力されている。図 4 の正解行動【B】では、推定行動に【B】を含む行動があり、かつ編集距離が 2 となるため、式 (4) より正解行動【B】と推定行動【B】【A】【B,C】との距離は $1 - e^{-2\alpha}$ となり、 $\alpha = \log 2$ とした場合 $1 - e^{-2\alpha} = 0.75$ となる。図 4 の正解行動【A】では、推定行動に【A】を含む行動がないため、式 (4) より正解行動【A】と推定行動【B,C】【B】との距離は 1 となる。この距離を合計した値が、正解行動シーケンスと推定行動シーケンスの距離となる。

5.4 正解行動におけるヒット率の評価

図 5 に各行動におけるヒット率を示す。図 5 において絶対誤差を正解行動の 0.05 以内とした場合、提案手法と最尤法と比較して、誤差範囲内にある行動が 10 個、提案手法の方が高い行動が 8 個、最尤法の方が高い行動が 6 個となった。最尤法でヒット率が 0.9 を超える 15 行動【1,5,6,7,8,11,12,14,15,16,17,18,22,23,24】

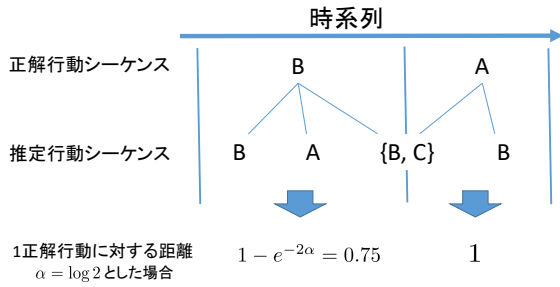


図4 正解行動シーケンスとの距離算出の例

のうち、誤差範囲内にある行動が9個、最尤法の方が高い行動が6個となった。この行動のうち、最尤法によるヒット率の平均が0.987、提案手法によるヒット率の平均が0.901となり、差は0.086である。最尤法でヒット率が0.1を下回る7行動【2,4,10,14,20,21,22】のうち、誤差範囲内にある行動が1個、提案手法の方が高い行動が6個となった。この行動のうち、最尤法によるヒット率の平均が0.06、提案手法によるヒット率の平均が0.698となり、差は0.692である。それ以外の2行動【3,9】は提案手法の方が高い結果となった。この行動のうち、最尤法によるヒット率の平均が0.509、提案手法によるヒット率の平均が0.826となり、差は0.317である。

【10:Out】に関しては、どちらの手法もヒット率が低い結果となった。ヒット率が低い原因として、学習器の学習不足やクラスタリング失敗の可能性が挙げられる。

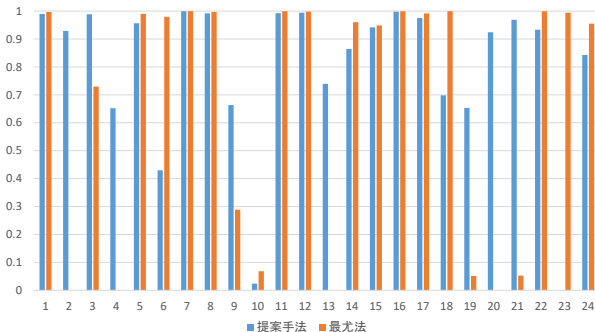


図5 各行動におけるヒット率

5.5 各行動ラベルの個別評価

図6に行動ラベル毎のヒット率を示す。提案手法の各行動ラベルのヒット率の平均は0.893、最尤法は0.671となり、全体の平均としては提案手法の方がヒット率が高い結果となった。個別の行動ラベルについて、最尤法による行動ラベルのヒット率の絶対誤差を0.05以内とした場合、全8種類の行動のうち、誤差範囲内にある行動は2行動「Table」「PC」、最尤法の方が高い行動が2行動「TV」「Sleep」、提案手法の方が高い行動が4行動「Out」「Workout」「Read」「Eat」となった。特に「Workout」「Read」に関しては、それぞれ0.929、0.73の差があることが分かった。最尤法の方が高い行動「TV」「Sleep」について、最尤法のヒット率の平均は0.997であり、提案手法は

0.935であり、差は0.062であった。提案手法の方が高い行動「Out」「Workout」「Read」「Eat」について提案手法のヒット率の平均は0.84であり、最尤法は0.375であり、差は0.465であった。

図7に図6における推定成功の内訳を示す。「両手法」は提案手法と最尤法の両手法で推定成功した割合を示す。「クラスタリング」は最尤法で推定に失敗し、提案手法でオーバーラップを認めない場合に推定成功した割合を示す。これはクラスタリングの導入のみによって得られた効果を表している。「オーバーラップ」はオーバーラップを認めた場合に推定成功した割合を示す。これは「クラスタリング」に加えてオーバーラップを認めたことによって得られた効果を表している。「最尤法」は提案手法で推定に失敗し、最尤法で推定成功した割合を示す。

「Workout」と「Read」の内訳は、「Workout」は「クラスタリング」のみ、「Read」は「クラスタリング」と「オーバーラップ」で占められている。これより、「Workout」は尤度ベクトルのクラスタリングを用いた手法により推定成功する割合が増え、「Read」はこれに加え、オーバーラップを認めることで更に推定成功する割合が増えたことがわかる。

図8に各行動において、正解行動ラベルシーケンス中にどの程度オーバーラップがあったのかを示す。「Sleep」「Out」は矛盾関係の処理により平均オーバーラップ数は1になる。「TV」「Tablet」「PC」においては、1に近い値であり、ほぼオーバーラップをしていないと見なせる。「Read」「Eat」に関しては1.7付近であり、詳細を調査したところ高々2つのオーバーラップであり、過剰なオーバーラップではないと考えられる。「Workout」の平均オーバーラップ数が2に近い結果となったが、「Workout」の詳細を調査したところ、全体の90.9%が「Eat」とオーバーラップをしていた。「Workout」と「Eat」は、室内の同じ位置で異なる動作を行っており、位置情報が同じである点において、行動に類似性がある。これにより識別が困難になったと考えられる。しかしながら、表1より、「Workout」と「Eat」は正確に識別する必要がないとしているため、このケースでは問題はないと考えられる。

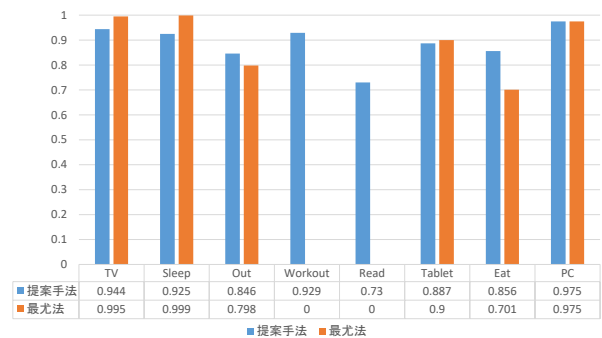


図6 行動種別のヒット率

5.6 正解行動シーケンスとの距離の評価

提案手法による推定行動シーケンスと正解行動シーケンスとの距離は11.75、最尤法による推定行動シーケンスと正解行動

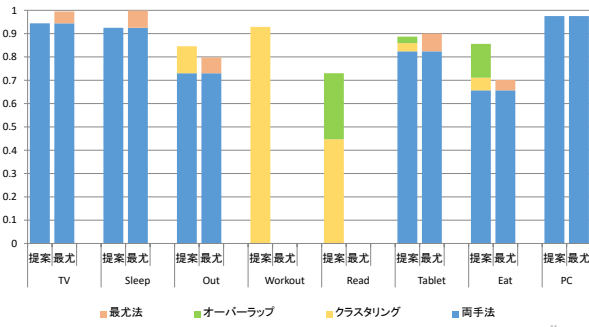


図 7 推定成功の内訳

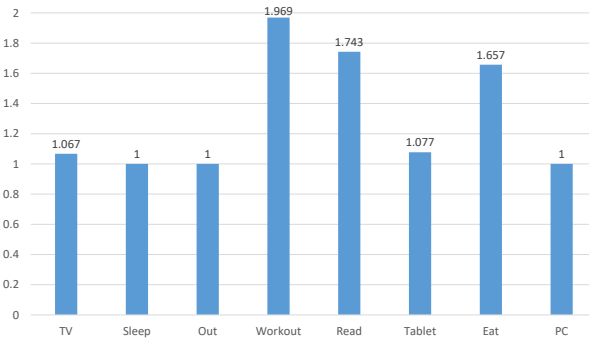


図 8 平均オーバーラップ数

シーケンスとの距離は 14.875 となった。提案手法が 3.125 短い結果となり、提案手法によって推定された行動シーケンスが、より正解行動シーケンスと近いと言える。また、提案手法による完全間違いの行動数は 2、最尤法は 6 となった。

6 ま と め

本研究では、環境センサのみからの情報による学習器から尤度ベクトルを取得し、行動ラベル付き尤度ベクトルのクラスタリングを行い、行動クラスタと行動ラベル不明の尤度ベクトルとの距離を測ることで、尤度の低い行動シーケンスを考慮した推定手法を提案した。提案手法の実装後、行動シーケンス推定に関する 3 つの評価実験とオーバーラップについての考察を行った。正解行動におけるヒット率に関して、最尤法の絶対誤差を 5%とした場合、最尤法でのヒット率が 10%を下回った 7 つの行動のうち、提案手法では 6 行動を推定できた。各行動ラベルの個別評価について、各行動ラベルのヒット率は、全体的に提案手法のヒット率が高く、特「Workout」「Read」に関しては、それぞれ 0.929, 0.73 の差があることが分かった。各行動ラベルの平均オーバーラップ数は、実験に使用したデータでは、重要な行動は矛盾関係の解消処理によってオーバーラップを 1 にしたが、ヒット率が大幅に下がることはなかった。また、一部の行動はある程度オーバーラップしていたが、許容範囲であると判断した。正解行動シーケンスとの距離に関して、提案手法による推定行動シーケンスと正解行動シーケンスとの距離は 11.75、最尤法による推定行動シーケンスと正解行動シーケンスとの距離は 14.875 であり、提案手法による行動シーケンスの方が正解行動シーケンスに近い結果となった。

今後の課題として、行動の遷移検知精度の向上、本研究で用いた学習器作成アルゴリズム、クラスタリング手法の改善やパラメータの自動推定などが挙げられる。また、行動シーケンスを複数集め、行動のパターンマイニングも行いたいと考えている。

謝 辞

本研究の一部は、JSPS 科研費 (18H03242, 18H03342, 16H02908, 17K12684), JST ACT-I の助成を受けたものである。ここに記して謝意を表す。

文 献

- [1] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB*, Vol. 1215, pp. 487–499, 1994.
- [2] Roy De Maesschalck, Delphine Jouan-Rimbaud, and Désiré L Massart. The mahalanobis distance. *Chemometrics and intelligent laboratory systems*, Vol. 50, No. 1, pp. 1–18, 2000.
- [3] Richard Dubes and Anil KJain. Validity studies in clustering methodologies. *Pattern Recognition*, Vol. 11, pp. 235–254, 1979.
- [4] Jiawei Han, Jian Pei, Behzad Mortazavi-Asl, Qiming Chen, Umeshwar Dayal, and Mei-Chun Hsu. Freespan: frequent pattern-projected sequential pattern mining. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 355–359. ACM, 2000.
- [5] Subhas Chandra Mukhopadhyay. Wearable sensors for human activity monitoring: A review. *IEEE sensors journal*, Vol. 15, No. 3, pp. 1321–1330, 2015.
- [6] Garib N Murshudov, Alexei A Vagin, and Eleanor J Dodson. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D*, Vol. 53, No. 3, pp. 240–255, 1997.
- [7] Pelleg, Dan, Moore, Andrew W, et al. X-means: Extending k-means with efficient estimation of the number of clusters. In *Icml*, Vol. 1, pp. 727–734, 2000.
- [8] Korbinian Strimmer and Arndt Von Haeseler. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Molecular Biology and Evolution*, Vol. 13, No. 7, pp. 964–969, 1996.
- [9] Tim Van Kasteren, Athanasios Noulas, Gwenn Englebienne, and Ben Kröse. Accurate activity recognition in a home setting. In *Proceedings of the 10th international conference on Ubiquitous computing*, pp. 1–9. ACM, 2008.
- [10] Ian H Witten, Eibe Frank, Mark A Hall, and Christopher J Pal. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.
- [11] 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, 2016.
- [12] 杉浦健人, 石川佳治, 佐々木勇和. 確率的データストリームにおけるパターン問合せ結果のグループ化 (データ工学). 電子情報通信学会技術研究報告= IEICE technical report: 信学技報, Vol. 114, No. 173, pp. 113–118, 2014.
- [13] 大内一成, 土井美和子. スマートフォンを用いた生活行動認識技術. *東芝レビュー*, Vol. 68, No. 6, pp. 40–43, 2013.
- [14] 大和淳司, 大谷淳, 石井健一郎. 隠れマルコフモデルを用いた動画像からの人物の行動認識. 電子情報通信学会論文誌 D, Vol. 76, No. 12, pp. 2556–2563, 1993.
- [15] 鉦龍, 佐伯幸郎, 中村匡秀. 屋内環境センシングデータを用いた独居者の行動生活の検知. 電子情報通信学会技術研究報告

= IEICE technical report: 信学技報, Vol. 117, No. 75, pp. 41–46, 2017.