

ノンパラメトリック混合隠れマルコフモデルの日常生活行動分類への応用

Nonparametric Bayesian HMM and its Application to Daily Activity Classification

菅野 隼[†] 松重 龍之介[†] 岡留 剛[†]
 Jun Kanno Ryunosuke Matsushige Takeshi Okadome

1. はじめに

センサデータから日常生活行動を分類する場合には、通常、教師あり学習を用いて多クラス分類器を作成する。教師あり学習による行動認識の研究は数多く存在し、それらは、データの時系列性を加味せずに、単一行動の全時間区間における加速度の平均値やゼロクロスといった多次元の量の特徴として分類器を作成するものと、単一行動時間区間を短い時間区間に分け各区間の特徴量を用いて、データの時系列性を反映させるように、例えば隠れマルコフモデル (HMM) により分類器を構築するものとに大別される。

一般に、認識課題におけるデータの時系列性を反映されるために HMM が用いられているが、出力がガウス分布など単峰性の分布を仮定する識別学習を用いた HMM では、例えば、個人差などに起因する出力の多峰性により精度向上に限界が生じる。

これに対し、Tied-Mixture Hidden Markov Model (TM-HMM) [3] は、出力に混合分布を導入することによりこの問題を回避している。しかし、TM-HMM における学習では、多峰性分布の尤度最大化によるため計算が発散することがしばしば起こる。この発散は、学習のデータ数が少ない場合に特に顕著である。また、TM-HMM では、出力分布の混合数をあらかじめ定めておく必要があるが、それは一般には明らかではない。

本研究は、各クラス 10 から 100 個程度の比較的小数の学習データで高い精度を出す分類器の構築を目指し、TM-HMM を拡張して、そこに、潜在変数およびパラメータの事前分布を導入したベイズモデルを提案する。とりわけ提案手法は、パラメータに対する事前分布を仮定することにより計算の発散問題を回避し、潜在変数にディリクレ過程事前分布を導入して、事後確率の最大化に基づき HMM の状態数および混合分布の要素数を観測データから自動的に決定する。HMM の状態数および混合分布の要素数は、日常生活行動の種類ごとに異なる可能性があり、これを捉えることは認識率の向上に寄与すると考える。事後確率の最大化は、誤識別率の最小化と等価 [4] であり、その意味でも、センサデータからの行動分類器構築において有用なモデルであると考えられる。

本研究では、日常生活行動を、(1)「歩く」や「走る」

といった周期的な身体動作を伴う行動 (以下、周期的な運動) と、(2)「椅子に座る」や「立ち上がる」・「横たわる」といった周期的な動作を含まない単発的な行動 (以下、非周期的な運動) とに大別し、それぞれ異なる特徴量を用いて学習データから上記のベイズ化された TM-HMM により周期的・非周期的な運動のそれぞれに対する分類器を作成する。新たに来たデータに対しては、まずランダムフォレストにより周期・非周期のどちらに属するかを推定し、その推定結果に応じた上記分類器で行動を分類する。

本研究では、非周期的な運動を伴う行動と、周期的な運動を伴う行動の両者を含む Daily and Sports Activities の加速度データセットを使用して、比較的小数の学習データによる分類実験を行った。その実験において、本提案手法はベースラインとする SVM(Support Vector Machine) や KLR(Kernel Logistic Regression) より高く、また TM-HMM と同等以上の識別率を示した。

なお、本稿は、情報処理学会論文誌 コピキタスコンピュティングシステム特集 (VI) に、採録が決定している論文のダイジェストである。

2. 関連研究

2.1 行動認識

数多くの先行研究があるが、ここでは代表的なものと、近年の成果を挙げる。Bao [1] は、「歩く」や「走る」・「自転車に乗る」・「掃除をする」・「食事をする」などの 20 種類の行動を分類している。20 名の実験参加者の手首や肘・腰・足首・膝の 5ヶ所に、サンプリングレートが 76.25Hz の 2 軸加速度センサーを取り付け、行動データを収集している。文献 [1] で用いられている特徴量は、平均やエネルギー・エントロピー・相関である。本研究では、加速度データから特徴量を抽出する際に、これらの特徴量を参考にしている。

文献 [12] では、HASC 2010 corpus を利用して、一般的な特徴量から新たな特徴量を加えて行動推定を行った。3 軸加速度センサーを実験参加者の腰に取り付けて、「静止」、「歩く」、「座る」、「スキップ」、「階段を上る」、「階段を下る」の 6 種類の行動を対象として識別を行っている。特徴量には、平均、分散、周波数領域毎のエネルギー (3.91~6.25Hz, 7.03~12.50Hz, 13.28~25.00Hz, 25.78~50.00Hz)、零点交差率、加速度の積分値が用い

[†] 関西学院大学, Kwansai Gakuin University

られている。分類器には、決定木が用いられている。

次に、本研究でも用いる階層ディリクレ過程を行動認識の手法に用いている研究について述べる。文献 [7] [13] は、行動の認識に階層ディリクレ過程隠れマルコフモデル (HDP-HMM) を用いて、隠れマルコフモデルの状態数を観測データから自動で決定している。文献 [7] は、センサデータから異常な行動を検知するモデルとして、HDP-HMM と Fisher カーネルを One-Class Support Vector Machine (OCSVM) モデルに組み込んだモデルを組み合わせた異常行動検知モデルを提案している。文献 [13] は、骨格関節位置に基づく特徴を用いて、連続深度画像において生じる人間の動作を分類する研究を行っている。分類モデルとして、HDP-HMM を使用している。

2.2 ディリクレ過程混合モデル

分類モデルの一つに混合分布モデル [10] がある。混合分布モデルは、各クラスのデータがある分布に従うと仮定し、各クラスの分布を混合したモデルである。例えば、各クラスのデータがガウス分布に従うと仮定し、各クラスのガウス分布を混合した混合ガウスモデルなどがある。さらには、潜在変数の事前分布にディリクレ過程 [5] を導入することにより、事後確率の最大化に基づき混合分布の要素数を観測データから自動的に決定する、ディリクレ過程混合モデル (Dirichlet Process Mixture model; DPM) が提案されている。以下で、DPM に関する研究を簡単にまとめる。

文献 [14] は、ネスト型無限混合ガウスモデルを提案している。これは、各クラスのデータが混合ガウス分布に従うと仮定し、各クラスの混合ガウス分布を混合させたモデルである。文献 [16] は、DPM に基づく拡張結合混合モデル (Extended Tied-Mixture model; ETM) を提案している。これは、各クラスのデータが混合分布に従うと仮定し、各クラスの混合分布を混合させたモデルである。上述のネスト型無限混合ガウスモデルが各クラスごとに独立な混合分布を用いているのに対し、DPM-ETM はすべてのクラスで共通な混合分布を用いている。すなわち、ネスト型無限混合ガウスモデルは、DPM-ETM の特殊な場合に相当する。

ネスト型無限混合ガウスモデルや DPM-ETM は、各クラスのデータが単一のガウス分布に従うと仮定する Semi-Supervised Gaussian Mixture Model (SSGMM) と比較して、データに対する分布の仮定が柔軟なモデルである。しかし、両モデルは時系列データに対応していない。本研究では、加速度の時系列データを分類するため、DPM に基づく時系列分類モデルを構築する。

3. 隠れマルコフモデルの拡張

提案手法のもととなる、隠れマルコフモデルの混合モデルについて詳述する。まず HMM を拡張したモデルである I-HMM について述べる。次に教師あり学習の Tied-Mixture Hidden Markov Model を概観する。

3.1 Infinite Hidden Markov Model

Infinite Hidden Markov Model (I-HMM) [9] は、観測データから HMM の潜在変数の状態数を自動的に決定するモデルである。通常の HMM を拡張したモデルであり、状態遷移確率分布が階層ディリクレ過程 (HDP) [15] に従っているとし、HDP を用いて潜在変数の状態を推定する。潜在変数の状態の確率の計算は、中華料理店過程 (Chinese restaurant process: CRP) に基づいており、学習毎に潜在変数の状態の確率と状態数を更新していく。本研究では I-HMM を採用する。

3.2 Tied-Mixture Hidden Markov Model

Tied-Mixture Hidden Markov Model (TM-HMM) [3] は、HMM の出力確率分布に混合ガウス分布を仮定したモデルである。図 1 に、TM-HMM のグラフィカルモデルを示す。

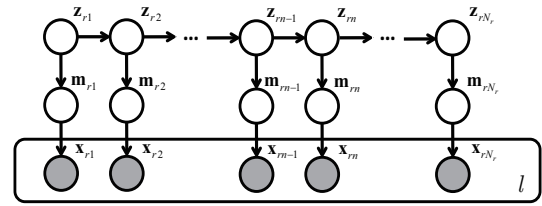


図 1: TM-HMM のグラフィカルモデル。 l はラベルつきデータの個数、 \mathbf{x}_{rn} は r 番目の時系列データの時刻 n における D 次元の特徴ベクトル、 \mathbf{z}_{rn} は r 番目の時系列データの時刻 n の状態、 \mathbf{m}_{rn} は r 番目の時系列データの時刻 n の混合分布の要素、 N_r を r 番目の時系列データの長さを表す。

以下、 $\{(\mathbf{X}_i, y_i)\}_{i=1}^l$ をラベルつきデータ、 l をラベルつきデータの個数、 N_r を r 番目の時系列データの長さ、 Y をクラス数、 D を次元数とする。すなわち、入力 \mathbf{X}_i に対するラベルが $y_i \in \{1, \dots, Y\}$ である。また、 $x_{r,n,1}$ のような表記に対しては x_{rn1} のように「,」を略記して示す。

ここで、 $\mathbf{X}_r = \{\mathbf{x}_{r1}, \dots, \mathbf{x}_{rN_r}\}$ は r 番目の時系列データ、 $\mathbf{x}_{rn} = (x_{rn1}, \dots, x_{rnD})^T$ は \mathbf{X}_r の時刻 n における D 次元の特徴ベクトルである。

まず、状態を表す潜在変数 $\{\mathbf{Z}_r\}_{r=1}^l$ を導入する。ここで、 $\mathbf{Z}_r = \{\mathbf{z}_{r1}, \dots, \mathbf{z}_{rN_r}\}$ 、 r 番目の時系列データの時刻 n の状態 $\mathbf{z}_{rn} = (z_{rn1}, \dots, z_{rnUy})^T$ 、 $z_{rni} \in \{0, 1\}$ 、

$\sum_{i=1}^{U^y} z_{rni} = 1$, U^y はクラス y の HMM の状態数である。次に、混合分布の要素を表す潜在変数 $\{\mathbf{M}_r\}_{r=1}^l$ を導入する。ここで、 $\mathbf{M}_r = \{\mathbf{m}_{r1}, \dots, \mathbf{m}_{rN_r}\}$, r 番目の時系列データの時刻 n の混合分布の要素 $\mathbf{m}_{rn} = (m_{rn1}, \dots, m_{rnK})^\top$, $m_{rnk} \in \{0, 1\}$, $\sum_{k=1}^K m_{rnk} = 1$, K は混合分布の要素数である。

初期状態 \mathbf{z}_{r1} の周辺分布は以下で書き表せる。

$$p(\mathbf{z}_{r1} | y_r) = \prod_{i=1}^{U^{y_r}} \pi_i^{y_r z_{r1i}},$$

ここで、 $\{\boldsymbol{\pi}_y\}_{y=1}^Y$, $\boldsymbol{\pi}_y = (\pi_1^y, \dots, \pi_{U^y}^y)^\top$, $\pi_i^y \equiv p(z_{r1i} = 1 | y)$, $0 \leq \pi_i^y \leq 1$, $\sum_{i=1}^{U^y} \pi_i^y = 1$ である。

状態 \mathbf{z}_{rn} の確率分布は直前の状態 $\mathbf{z}_{r,n-1}$ に依存する。条件付き確率分布 $p(\mathbf{z}_{rn} | \mathbf{z}_{r,n-1})$ は遷移確率分布と呼ばれ、以下で書き表せる。

$$p(\mathbf{z}_{rn} | \mathbf{z}_{r,n-1}, y_r) = \prod_{j=1}^{U^{y_r}} \prod_{i=1}^{U^{y_r}} a_{ij}^{y_r z_{r,n-1,i} z_{rnj}},$$

ここで、 $\{\mathbf{A}_y\}_{y=1}^Y$, \mathbf{A}_y は (i, j) 成分が a_{ij}^y で与えられる $U^y \times U^y$ の行列, $a_{ij}^y \equiv p(z_{rnj} = 1 | z_{r,n-1,i} = 1, y)$, $0 \leq a_{ij}^y \leq 1$, $\sum_j a_{ij}^y = 1$, $\mathbf{a}_i^y = (a_{i1}, \dots, a_{iU^y})^\top$ である。

混合分布の要素 \mathbf{m}_{rn} の確率分布はその時刻の状態 \mathbf{z}_{rn} に依存する。条件付き確率分布 $p(\mathbf{m}_{rn} | \mathbf{z}_{rn})$ は以下で書き表せる。

$$p(\mathbf{m}_{rn} | \mathbf{z}_{rn}, y_r) = \prod_{k=1}^K \prod_{j=1}^{U^{y_r}} c_{jk}^{y_r z_{rnj} m_{rnk}},$$

ここで、 $\{\mathbf{C}_y\}_{y=1}^Y$, \mathbf{C}_y は (j, k) 成分が c_{jk}^y で与えられる $U^y \times K$ の行列, $c_{jk}^y \equiv p(m_{rnk} = 1 | z_{rnj} = 1, y)$, $0 \leq c_{jk}^y \leq 1$, $\sum_k c_{jk}^y = 1$, $\mathbf{c}_j^y = (c_{j1}, \dots, c_{jK})^\top$ である。

観測変数の条件付き確率分布 $p(\mathbf{x}_{rn} | \mathbf{z}_{rn})$ は出力確率分布と呼ばれ、以下で書き表せる。

$$\begin{aligned} p(\mathbf{x}_{rn} | \mathbf{z}_{rn}, y_r) &= \sum_{\mathbf{m}_{rn}} p(\mathbf{x}_{rn}, \mathbf{m}_{rn} | \mathbf{z}_{rn}, y_r) \\ &= \sum_{\mathbf{m}_{rn}} p(\mathbf{x}_{rn} | \mathbf{m}_{rn}, \mathbf{z}_{rn}, y_r) p(\mathbf{m}_{rn} | \mathbf{z}_{rn}, y_r), \end{aligned}$$

ここで、 \mathbf{m}_{rn} で条件付けたときに、 \mathbf{x}_{rn} と \mathbf{z}_{rn}, y_r は条件付き独立である。したがって、出力確率分布は以下で書き表せる。

$$p(\mathbf{x}_{rn} | \mathbf{z}_{rn}, y_r) = \sum_{\mathbf{m}_{rn}} p(\mathbf{x}_{rn} | \mathbf{m}_{rn}) p(\mathbf{m}_{rn} | \mathbf{z}_{rn}, y_r).$$

混合分布の要素 \mathbf{m}_{rn} が与えられたもとで観測データ \mathbf{x}_{rn} が得られる確率はガウス分布に従うとする。すなわち、

$$p(\mathbf{x}_{rn} | \mathbf{m}_{rn}, \boldsymbol{\phi}) = \prod_{k=1}^K \mathcal{N}(\mathbf{x}_{rn} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{m_{rnk}},$$

ここで、 $\boldsymbol{\phi} = \{\phi_k\}_{k=1}^K$, $\phi_k = \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}$ はガウス分布の平均と分散である。

観測変数と潜在変数の同時分布は以下で書き表せる。

$$p(\mathbf{X}_r, \mathbf{Z}_r, \mathbf{M}_r | \boldsymbol{\Theta}_{y_r}) =$$

$$p(\mathbf{z}_{r1} | y_r) \left[\prod_{n=2}^{N_r} p(\mathbf{z}_{rn} | \mathbf{z}_{r,n-1}, y_r) \right] \prod_{n=1}^{N_r} p(\mathbf{x}_{rn} | \mathbf{z}_{rn}, y_r),$$

ここで、 $\boldsymbol{\Theta}_{y_r} = \{\boldsymbol{\pi}_{y_r}, \mathbf{A}_{y_r}, \mathbf{C}_{y_r}, \boldsymbol{\phi}\}$ である。

4. 提案手法

本研究で使用する分類器は、TM-HMM を拡張したベイズモデルである。以下では、これを BTM-HMM と呼ぶ。本研究では、潜在変数およびパラメータの事前分布を導入し、事後確率の最大化に基づき HMM の状態数および混合分布の要素数を観測データから自動的に決定する。本章では、時系列データを周期的な運動と非周期的な運動に分ける推定手法と、提案モデルについて詳述し、パラメータの推定方法について述べる。

4.1 行動の周期・非周期判定

日常生活行動は、「歩く」や「走る」といった一定の時間間隔で同じ状態が繰り返される行動と、「座る」や「立つ」のような一定の時間間隔で同じ状態が表れない行動に大別される。行動分類の特徴として、周期的な運動に適した特徴は、非周期的な運動には適さないことも起こりうる。本研究では、データに応じて、周期的な運動と非周期的な運動に判別することでこの問題に対処する。すなわち、あらかじめ周期的な運動に属する行動の学習データで BTM-HMM により学習した分類器と、非周期的な運動に属する行動の学習データで学習した分類器を作成する。新たに来たデータに対しては、周期・非周期どちらかを推定し、適切な時間窓幅の設定と特徴抽出を行い、周期・非周期の推定結果に応じた分類器で分類を行う。テストデータに対する周期的か非周期的かの判別には Random Forests(RF)[18] を用いて 2 値分類を行う。RF に使用する特徴量については第 5 章で詳しく述べる。

4.2 提案モデル：BTM-HMM

まず、初期確率 $\boldsymbol{\pi}_y$ および遷移確率 \mathbf{a}_i^y の事前分布として、ディリクレ分布を導入する。

$$\begin{aligned} \boldsymbol{\pi}_y &\sim \text{Dir}(\boldsymbol{\pi}_y | \alpha\beta_1^y, \dots, \alpha\beta_{U^y}^y), \\ \mathbf{a}_i^y &\sim \text{Dir}(\mathbf{a}_i^y | \alpha\beta_1^y, \dots, \alpha\beta_{U^y}^y), \end{aligned}$$

ここで、 $\boldsymbol{\beta} = \{\boldsymbol{\beta}_y\}_{y=1}^Y$, $\boldsymbol{\beta}_y = (\beta_1^y, \dots, \beta_{U^y}^y)$ である。 $\boldsymbol{\beta}_y$ はディリクレ分布に従うとする。

$$\boldsymbol{\beta}_y \sim \text{Dir}(\boldsymbol{\beta}_y | \frac{\gamma}{U^y}, \dots, \frac{\gamma}{U^y}).$$

ここで、 $U^y \rightarrow \infty$ の極限を考えると、これはディリクレ過程 [5] となる。すなわち、初期確率 π_y および遷移確率 \mathbf{a}_i^y はディリクレ過程 $\text{DP}(\alpha, G_i^y)$ に従う。ここで、 α はディリクレ過程の集中度パラメータである。ただし、各時刻で状態を共有する必要があるため、基底分布 G_i^y 自体が、もう一つのディリクレ過程からのサンプル $G_i^y \sim \text{DP}(\gamma, G_0^y)$ であるとする。これは、可能な遷移先とその事前確率を各状態間で共有することを意味する。ディリクレ過程が階層化されているため、これは階層ディリクレ過程 [15] と呼ばれる。ここで、 $G_0^y = \sum_{k=1}^{\infty} \beta_k^y \delta_{\phi_k}$ 、 $G_i^y = \sum_{j=1}^{\infty} a_{ij}^y \delta_{\phi_j}$ である。

次に、混合分布の要素を選択する確率 \mathbf{c}_j^y の事前分布として、ディリクレ分布を導入する。すなわち、

$$\mathbf{c}_j^y \sim \text{Dir}(\mathbf{c}_j^y \mid \sigma \lambda_1^y, \dots, \sigma \lambda_K^y),$$

ここで、 $\boldsymbol{\lambda} = \{\lambda_y\}_{y=1}^Y$ 、 $\lambda_y = (\lambda_1^y, \dots, \lambda_K^y)$ である。 λ_y は以下のようにディリクレ分布に従うとする。

$$\lambda_y \sim \text{Dir}(\lambda_y \mid \frac{\eta}{K}, \dots, \frac{\eta}{K}).$$

ここで、 $K \rightarrow \infty$ の極限を考え、混合分布の要素の選択確率 \mathbf{c}_j^y はディリクレ過程 $\text{DP}(\sigma, H_j^y)$ に従うとする。ここで、 σ はディリクレ過程の集中度パラメータである。ただし、各状態で混合分布の要素を共有する必要があるため、分布 H_j^y 自体がもう一つのディリクレ過程からのサンプル $H_j^y \sim \text{DP}(\eta, H_0)$ であるとする。すなわち、混合分布の要素とその事前確率を各状態間で共有する。ここで、 $H_0 = \sum_{k=1}^{\infty} \lambda_k^y \delta_{\phi_k}$ 、 $H_j^y = \sum_{k=1}^{\infty} c_{jk}^y \delta_{\phi_k}$ である。このもとで、BTM-HMM は以下で書き表せる。

$$\begin{aligned} \beta_y \mid \gamma &\sim \text{GEM}(\gamma), \\ \pi_y \mid \alpha, \beta_y &\sim \text{DP}(\alpha, \beta_y), \\ \mathbf{a}_i^y \mid \alpha, \beta_y &\sim \text{DP}(\alpha, \beta_y), \\ \lambda_y \mid \eta &\sim \text{GEM}(\eta), \\ \mathbf{c}_j^y \mid \sigma, \lambda_y &\sim \text{DP}(\sigma, \lambda_y), \\ \phi_k \mid H &\sim H, \\ \mathbf{z}_{r1} \mid \pi_y &\sim \pi_y, \\ \mathbf{z}_{rn} \mid \mathbf{z}_{r,n-1}, (\mathbf{a}_i^y)_{i=1}^{\infty} &\sim \mathbf{a}_{\mathbf{z}_{r,n-1}}^y, \\ \mathbf{m}_{rn} \mid \mathbf{z}_{rn}, (\mathbf{c}_j^y)_{j=1}^{\infty} &\sim \mathbf{c}_{\mathbf{z}_{rn}}^y, \\ \mathbf{x}_{rn} \mid \mathbf{m}_{rn}, (\phi_k)_{k=1}^{\infty} &\sim F(\phi_{\mathbf{m}_{rn}}), \end{aligned}$$

ここで、GEM は棒折り過程で、 H は基底分布、 F は出力分布である。

4.3 BTM-HMM のパラメータ推定

本研究では、ギブスサンプリングにより各パラメータを求める。以下、 $\mathbf{X} = \{\mathbf{X}_r\}_{r=1}^l$ 、 $\mathbf{y} = \{y_r\}_{r=1}^l$ 、 $\mathbf{Z} = \{\mathbf{Z}_r\}_{r=1}^l$ 、 $\mathbf{M} = \{\mathbf{M}_r\}_{r=1}^l$ 、 $\phi = \{\phi_k\}_{k=1}^K$ とする。

・状態 \mathbf{Z} の推定

状態 \mathbf{Z} の推定には、beam sampling [6] を用いる。まず、補助変数 u_{rn} を導入する。 u_{rn} は状態 $\mathbf{z}_{r,n-1}$ と \mathbf{z}_{rn} に基づき一様分布からサンプリングされる。すなわち、

$$u_{rn} \sim \text{Uniform}(0, a_{\mathbf{z}_{r,n-1}\mathbf{z}_{rn}}^{y_r}).$$

この u_{rn} の確率密度関数は以下で書き表せる。

$$p(u_{rn} \mid \mathbf{z}_{r,n-1}, \mathbf{z}_{rn}, y_r) = \frac{\delta(0 < u_{rn} < a_{\mathbf{z}_{r,n-1}\mathbf{z}_{rn}}^{y_r})}{p(\mathbf{z}_{rn} \mid \mathbf{z}_{r,n-1}, y_r)},$$

ただし、 $\delta(C)$ は、条件 C が成立するときに 1、そうでないときに 0 をとる関数である。

次に、前向きメッセージを伝播し、バックトラックの際に状態 \mathbf{z}_{rn} をサンプリングする。前向きメッセージは以下で書き表せる。

$$\begin{aligned} \hat{\alpha}(\mathbf{z}_{rn} \mid y_r) &= p(\mathbf{z}_{rn} \mid \mathbf{x}_{r1:r,n}, u_{r1:r,n}, y_r) \\ &\propto p(\mathbf{z}_{rn}, u_{rn}, \mathbf{x}_{rn} \mid \mathbf{x}_{r1:r,n-1}, u_{r1:r,n-1}, y_r) \\ &= p(\mathbf{x}_{rn} \mid \mathbf{z}_{rn}, y_r) \sum_{\mathbf{z}_{r,n-1}: u_{rn} < a_{\mathbf{z}_{r,n-1}\mathbf{z}_{rn}}^{y_r}} \hat{\alpha}(\mathbf{z}_{r,n-1} \mid y_r). \end{aligned}$$

すなわち、 $a_{\mathbf{z}_{r,n-1}\mathbf{z}_{rn}}^{y_r} > u_{rn}$ を満たす状態遷移のみが非零の確率を持つ。前向きメッセージを伝播する際に、状態 $\mathbf{z}_{r,n+1}$ の U^y 個の各々の値に対する最大化に対応する \mathbf{z}_{rn} の値を記録しておく。一旦メッセージが鎖の終端まで伝播し、 \mathbf{z}_{rN_r} の最も確からしい状態が求めれば、鎖を後戻りしてたどっていくことにより、 \mathbf{z}_{rn} が得られる。 $r = 1, \dots, l$ について状態系列 \mathbf{Z}_r を求める。

・混合分布の要素 \mathbf{M} のサンプリング

混合分布の要素 \mathbf{M} をサンプリングするために、上述の beam sampling を応用する。まず、補助変数 v_{rn} を導入する。 v_{rn} は、状態 \mathbf{z}_{rn} と混合分布の要素 \mathbf{m}_{rn} に基づき一様分布からサンプリングされる。すなわち、

$$v_{rn} \sim \text{Uniform}(0, c_{\mathbf{z}_{rn}\mathbf{m}_{rn}}^{y_r}).$$

この v_{rn} の確率密度関数は以下で書き表せる。

$$p(v_{rn} \mid \mathbf{z}_{rn}, \mathbf{m}_{rn}, y_r) = \frac{\delta(0 < v_{rn} < c_{\mathbf{z}_{rn}\mathbf{m}_{rn}}^{y_r})}{p(\mathbf{m}_{rn} \mid \mathbf{z}_{rn}, y_r)}.$$

混合分布の要素 \mathbf{m}_{rn} の事後分布は以下で書き表せる。

$$\begin{aligned} p(\mathbf{m}_{rn} \mid \mathbf{m}_{r,\setminus n}, \mathbf{M}_{\setminus r}, \mathbf{X}, \mathbf{y}, \mathbf{Z}, \Theta) \\ &\propto p(\mathbf{m}_{rn}, v_{rn}, \mathbf{x}_{rn} \mid \mathbf{z}_{rn}, y_r) \\ &= p(\mathbf{x}_{rn} \mid \mathbf{m}_{rn}) \delta(v_{rn} < c_{\mathbf{z}_{rn}\mathbf{m}_{rn}}^{y_r}). \end{aligned}$$

すなわち、 $c_{\mathbf{z}_{rn}\mathbf{m}_{rn}}^{y_r} > v_{rn}$ を満たす混合分布の要素のみが非零の確率を持つ。ここで、 $\mathbf{m}_{r,\setminus n}$ は $\mathbf{m}_{r1}, \dots, \mathbf{m}_{rN_r}$ から \mathbf{m}_{rn} を除いたもの、 $\mathbf{M}_{\setminus r}$ は $\mathbf{M}_1, \dots, \mathbf{M}_{l+1}$ から

\mathbf{M}_r を除いたもの、 $\mathbf{x}_{r,\setminus n}$ は $\mathbf{x}_{r1}, \dots, \mathbf{x}_{rN_r}$ から \mathbf{x}_{rn} を除いたものを表す。 $r = 1, \dots, l$ および $n = 1, \dots, N_r$ について、混合分布の要素 \mathbf{m}_{rn} をサンプリングする。

・パラメータ \mathbf{A} の更新

初期確率 π_y および遷移確率 \mathbf{a}_y と、状態数 U^y を更新する。

初期確率 π_y および遷移確率 \mathbf{a}_i^y の事前分布にディリクレ分布を仮定したため、事後分布は以下のディリクレ分布で書き表せる。

$$\pi_y \sim \text{Dir}(\pi_y | \alpha\beta_1^y + n_1^y, \dots, \alpha\beta_{U^y}^y + n_{U^y}^y, \alpha \sum_{k=U^y+1}^{\infty} \beta_k^y),$$

$$\mathbf{a}_i^y \sim \text{Dir}(\mathbf{a}_i^y | \alpha\beta_1^y + n_{i1}^y, \dots, \alpha\beta_{U^y}^y + n_{iU^y}^y, \alpha \sum_{k=U^y+1}^{\infty} \beta_k^y), H(\phi_k) = \mathcal{N}(\boldsymbol{\mu}_k | \mathbf{m}_0, (\beta_0 \boldsymbol{\Lambda}_k)^{-1}) \mathcal{W}(\boldsymbol{\Lambda}_k | \mathbf{W}_0, \nu_0).$$

ただし、 n_i^y は、クラス y の HMM において、初期状態 i が選ばれた回数で、 n_{ij}^y は、クラス y の HMM において、状態 i から状態 j に遷移した回数である。新たな状態が出現する確率は、 $\alpha \sum_{k=U^y+1}^{\infty} \beta_k^y$ である。

β_y をサンプリングするために、補助変数 h_{ij} を導入する [6]。

$$p(h_{ij} | \mathbf{Z}, \beta_y, \alpha) \propto S(n_{ij}^y, h_{ij}) (\alpha \beta_j^y)^h$$

ここで、 $S(\cdot, \cdot)$ は第 1 種スターリング数である。 h_{ij} を用いて、 β_y の事後分布は以下のディリクレ分布で表せる。

$$\beta_y \sim \text{Dir}(\beta_y | h_{\cdot 1}, \dots, h_{\cdot U^y}, \gamma),$$

ここで、 $h_{\cdot j}^y = \sum_{i=1}^{U^y} h_{ij}$ である。詳細については、[15] を参照されたい。

・パラメータ \mathbf{C} の更新

混合分布の要素を選択する確率 \mathbf{c}_j^y と混合要素数 K を更新する。

混合分布の要素を選択する確率 \mathbf{c}_j^y の事後分布は以下のディリクレ分布で書き表せる。

$$\mathbf{c}_j^y \sim \text{Dir}(\mathbf{c}_j^y | \sigma\lambda_1^y + m_{j1}^y, \dots, \sigma\lambda_K^y + m_{jK}^y, \sigma \sum_{i=K+1}^{\infty} \lambda_i^y),$$

ただし m_{jk}^y は、クラス y の HMM において、状態 i のときに混合分布の要素 k が選択された回数である。新たな混合分布の要素を選択する確率は、 $\sigma \sum_{i=K+1}^{\infty} \lambda_i^y$ である。

λ_y をサンプリングするために、さらに補助変数 k^* を導入する [17]。

$$k^* \sim 1 + \text{Poisson}(-\sigma \log v^*)$$

ここで、 $v^* = \min(v_{1n}, \dots, v_{rn})$ で、 $\text{Poisson}(\zeta)$ はパラメータ ζ のポアソン分布である。従って、 λ_y の事後分

布は以下のディリクレ分布で書き表せる。

$$\lambda_y \sim \text{Dir}(\lambda_y | \frac{\eta}{K} + m_{\cdot 1}^y, \dots, \frac{\eta}{K} + m_{\cdot K}^y, k^*),$$

ここで、 $m_{\cdot k}^y = \sum_{j=1}^{U^y} m_{jk}^y$ である。詳細については、[17] を参照されたい。

・パラメータ ϕ の更新

混合分布の要素 \mathbf{M} の値が決まれば、事前分布 $H(\phi_k)$ およびクラス k に所属するデータに対する尤度 $p(\{\mathbf{x}_{rn} | \mathbf{x}_{rn} \in k\} | \phi_k)$ を用いて、ベイズの定理より事後分布が得られる。事前分布には、ガウス分布の共役事前分布であるガウス-ウィシャート分布を用いる。

事前分布に共役事前分布を用いることで、事後分布もガウス-ウィシャート分布となる。

$$\begin{aligned} p(\phi_k | \{\mathbf{x}_{rn} | \mathbf{x}_{rn} \in k\}) &= \frac{p(\{\mathbf{x}_{rn} | \mathbf{x}_{rn} \in k\} | \phi_k) H(\phi_k)}{p(\{\mathbf{x}_{rn} | \mathbf{x}_{rn} \in k\} | \phi_k)} \\ &= \frac{H(\phi_k) \prod_{\mathbf{x}_{rn} \in k} p(\mathbf{x}_{rn} | \phi_k)}{\int H(\phi_k) \prod_{\mathbf{x}_{rn} \in k} p(\mathbf{x}_{rn} | \phi_k) d\phi_k} \\ &= \mathcal{N}(\boldsymbol{\mu}_k | \mathbf{m}_k, (\beta_k \boldsymbol{\Lambda}_k)^{-1}) \mathcal{W}(\boldsymbol{\Lambda}_k | \mathbf{W}_k, \nu_k), \end{aligned}$$

ただし、以下のように定義した。

$$\begin{aligned} \bar{\mathbf{x}}_k &= \frac{1}{N_k} \sum_{\mathbf{x}_{rn} \in k} \mathbf{x}_{rn}, \\ \mathbf{S}_k &= \frac{1}{N_k} \sum_{\mathbf{x}_{rn} \in k} (\mathbf{x}_{rn} - \bar{\mathbf{x}}_k)(\mathbf{x}_{rn} - \bar{\mathbf{x}}_k)^T, \\ \beta_k &= \beta_0 + N_k, \\ \mathbf{m}_k &= \frac{1}{\beta_k} (\beta_0 \mathbf{m}_0 + N_k \bar{\mathbf{x}}_k), \\ \mathbf{W}_k^{-1} &= \mathbf{W}_0^{-1} + N_k \mathbf{S}_k + \frac{\beta_0 N_k}{\beta_0 + N_k} (\bar{\mathbf{x}}_k - \mathbf{m}_0)(\bar{\mathbf{x}}_k - \mathbf{m}_0)^T, \\ \nu_k &= \nu_0 + N_k. \end{aligned}$$

N_k は k 番目の混合要素に属するデータの数である。 $k = 1, \dots, K$ についてパラメータ ϕ_k を更新する。新たな混合分布の要素が選択されたときには、1つの観測データのみを用いてその新たな混合分布の要素のパラメータを決定する。なお、超パラメータ \mathbf{W}_0 は対角行列であり、予備実験により各対角要素を 0.28 とした。また、 ν_0 は自由度で次元数より大きくななければならないという制約があるので、特徴量の次元数に 1 を足した値とした。

5. 特徴抽出

本研究では、運動の 3 軸加速度データを使用して行動認識を行うため、3 軸加速度データから抽出する特徴量について詳述する。

5.1 周期・非周期推定に用いる特徴量

まず、加速度の3軸方向の中から振幅が最も大きい軸に着目し、ここではそれをx軸とする。予備実験により推定精度が高かった以下の特徴を用いる。まず、x軸の加速度のパワースペクトルの最大値から前後の幅1, 2, 3点を最大値からの近傍とみなして平均と25%値, 中央値, 75%値を求める。また、x軸の加速度データの1つ前の時間の値との差分と25%値, 75%値を求める。さらに同様にx軸の加速度の値の25%値, 75%値を求める。以上合計10個を特徴量とし、RFに入力として与えることでテストデータが周期的・非周期的かの推定を行う。なお、RFの木の本数は、予備実験により最も高い精度が得られた500と設定する。

5.2 行動分類のための特徴量

はじめに周期的な運動に属する行動の分類に用いる特徴量について述べる。

加速度データからの教師あり学習に関する先行研究 [1] や [8]・[12] を参考にして、本研究では、周期的な運動の加速度データに対して、以下で述べる16の特徴量を用いる。ある時間幅の窓を設定し、窓長の50%のオーバーラップでずらしながら時間窓毎に16の特徴量を抽出する。本研究では、各時系列データに自己相関関数を適用して周期を求め、その2倍の値を窓幅に設定した。

3軸加速度センサーから得られるデータは、デバイスの保持姿勢によって値が変化するので、重力成分を推定し、除去する必要がある [8]。3軸加速度センサーには常に1Gの重力がかかるので、一定の時間幅 w_G でのXYZ軸の平均ベクトルを重力ベクトルと推定する。すなわち、時刻 t での重力ベクトル $v_G(t)$ は、3軸加速度ベクトル $v(t)$ を用いて以下で書き表せる。

$$v_G(t) = \frac{\sum_{i=t-w_G}^t v(i)}{w_G}.$$

3軸加速度ベクトルから重力ベクトルを減算し、正規化された運動加速度ベクトルを得る。時刻 t での正規化された運動加速度ベクトル $v_n(t)$ は以下で書き表せる。

$$v_n(t) = v(t) - v_G(t).$$

まず、正規化された加速度ベクトルから特徴量として、以下のベクトル長 $F_1(t)$ と重力ベクトルとの内積値 $F_2(t)$ ・重力ベクトルとの外積値 $F_3(t)$ を算出する。

$$\begin{aligned} F_1(t) &= \|v_n(t)\|, \\ F_2(t) &= v_n(t) \cdot v_G(t), \\ F_3(t) &= v_n(t) \times v_G(t). \end{aligned}$$

文献 [8] では、 $F_1(t)$, $F_2(t)$, $F_3(t)$ それぞれについて平均値, 最大値, 最小値, 分散値の4種類の統計量を計

算しているが、これらの統計量は外れ値の影響を受ける可能性がある。そのため、本研究では、25%値, 中央値, 75%値の3種類の統計量を用いた。また、「歩く」や「走る」などの行動は、周期的に繰り返される動作であるので、周波数軸方向における特徴量を考慮する。本研究では、エネルギーと周波数領域エントロピーの2種類の特徴量を用いる [11]。フーリエ変換をする際には、切り出した時系列データにハニング窓関数をかける。エネルギーは、各データに関してFFTを行い、得られた周波数成分の絶対値の合計で表される。すなわち、

$$E_n = \frac{1}{N} \sum_{i=2}^n |F_i|^2.$$

本研究では文献 [12] を参考にして、周波数帯域ごとのエネルギー (1~3Hz, 4~6Hz, 7~12Hz, 13~25Hz, 26~50Hz) も特徴量に加えた。周波数領域エントロピーは、FFTの全成分の総和で各成分を正規化し、確率分布 p を求め、そのエントロピーで表される。すなわち、

$$p(i) = \frac{|F_i|^2}{\sum_{i=2}^n |F_i|^2},$$

$$FDE = - \sum_{i=2}^n p(i) \log p(i).$$

得られた16次元の特徴量に対して、白色化(ホワイトニング)を行う [4]。すなわち、各データ点 \mathbf{x}_n に対して、

$$\mathbf{y}_n = \mathbf{L}^{-1/2} \mathbf{U}^T (\mathbf{x}_n - \bar{\mathbf{x}}),$$

で与えられる変換値を定義する。ここで、 \mathbf{L} は、データ集合の共分散行列の固有値 λ_i を持つ対角行列で、 \mathbf{U} は、 i 番目の列が固有ベクトル \mathbf{u}_i で与えられる直交行列、 $\bar{\mathbf{x}}$ はサンプル平均である。これにより、特徴空間の各軸が無相関化される。

次に非周期的な運動に属する行動の分類に用いる特徴量について述べる。これは、先行研究 [2] を参考に、各軸の加速度データについての最小値と最大値・平均・分散・歪度・尖度・フーリエ変換の最大値の合計21種類の特徴量として選択した。なお、予備実験により推定精度が最も高い値を窓幅とし800msとする。21次元の特徴量に対して、周期的な運動の特徴量と同様に白色化を行う。

6. 評価

窓をかけずに運動の全区間に対する上記特徴量を用いたKLRとSVMをベースラインとして分類性能を比較する。

評価の方法として、複数の分類器による教師あり学習による評価を行う。それぞれの評価において、全データ

の 1 割をテストデータとし、残りの 9 割を学習データとして交差確認を行う。実際に分類器の学習に使用するデータは、学習データとして分けられた各クラスデータから 5%、10%、30% の割合で抽出されたデータとする。なお、提案手法と同様に KLR と SVM、TM-HMM の学習では、周期的な運動に属する行動と、非周期的な運動に属する行動に対してはそれぞれ異なる特徴量を用いて別々の分類器を作成した。

6.1 実験条件

本研究では、Daily and Sports Activities の 3 軸加速度データから、「座る」などの非周期的な運動 5 種類と「時速 4km の速さでランニングマシンを平坦な道で走る」などの周期的な運動 14 種類を用いる [2]。ここで、1 クラスがデータ数を 480 とし、合計 9120 個のデータを用いた。

各分類器におけるそれぞれのパラメータは、予備実験を通して次のように設定した。まず、SVM の正則化係数は $C = 1, C^* = 1$ とした。また、カーネル関数にガウスカーネルを使用し、その精度パラメータを 0.1 とした。KLR も SVM と同様に、カーネル関数にガウスカーネルを使用し、その精度パラメータを 0.1 とした。BTM-HMM については、 $\alpha = 1, \gamma = 1, \sigma = 1, \eta = 1$ とした。また、TM-HMM の潜在変数の状態数や混合数は、予備実験により最も高い精度が得られた値を使用し、状態数と混合数ともに 6 と設定している。

6.2 実験結果

Daily and Sports Activities データでは、学習に用いるデータの割合が各クラスの全学習データの 5%、10%、30% のとき、各クラスの学習データの個数は、それぞれ 22 個、43 個、130 個である。各テストデータを、RF によって周期的・非周期的かを推定し、推定結果に応じた特徴量を用いてクラス分類したときの 19 種類の運動に対する各分類器の F 値の平均を示す。図 2 に、周期的な運動と非周期的な運動が混在している総合的な F 値の平均を示す。

提案手法の F 値は、学習データ数が各クラスデータの 5% のとき 74.9%、10% のとき 77.5%、30% のとき 82.0% であり、全学習データを使用したときは 84.4% である。図 2 に示したように、提案手法が最も高い F 値を示した。ただし、5% と 10% のときは、TM-HMM の学習が収束しなかったため図に結果は示されていない。分散分析の結果、手法とデータ数の主効果は統計的に有意である ($p < 0.001$)。また、30% と 100% のデータ数の結果を見ると、提案手法と TM-HMM では、いずれのデータ数においても提案手法が高い F 値の平均を示した。ただし、テューキーの HSD による F 値の差の

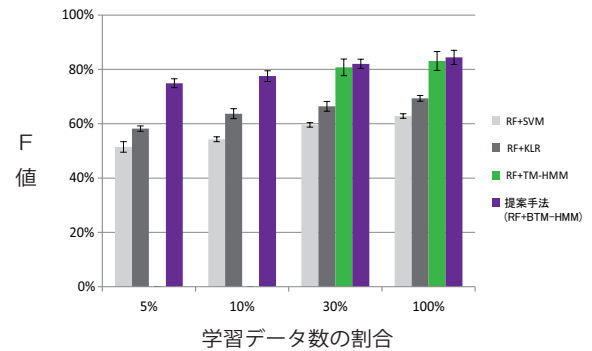


図 2: Daily and Sports Activities データに対する各分類器の F 値の平均と標準偏差 (ヒゲ) を表す。5% と 10% に対する TM-HMM の結果は欠損。

多重比較では、提案手法と TM-HMM の F 値の差は統計的に有意ではない。

7. 議論

TM-HMM における学習では、多峰性分布の尤度最大化によるため計算が発散することがしばしば起こる。この発散は、学習のデータ数が少ない場合に特に顕著であり、実際に Daily and Sports Activities の学習データのうち周期的なものに対してパラメータの学習を行ったとき、1 つのクラスの学習データ数が 130 個よりも下回ると、乱数を 1000 回振り直してもすべて計算が発散した。また、TM-HMM では、出力分布の混合数をあらかじめ定めておく必要があるが、それは一般には明らかではなく、学習するデータのクラス数が増えるほど出力分布の混合数を、例えば予備実験などで決定することが困難になる。

これに対し、提案した BTM-HMM では、発散が起らず、少数の学習データでモデルを学習することができる。TM-HMM にベイズモデルを導入し、事後確率最大化によってパラメータの推定を行っているためと考えられる。また提案手法では、HMM の状態数および出力分布の混合数をデータから事後確率最大により自動的に決定している。実験の結果より、従来の TM-HMM よりも高い F 値を示したことから BTM-HMM は TM-HMM の問題点を克服したモデルといえる。

比較的少数の学習データより、時系列性を考慮して学習したときの BTM-HMM の F 値は、時系列性を加味しない分類器である SVM や KLR と比較しても高い。また、少数の学習データにおける F 値を全学習データを用いたときの F 値と比較しても、BTM-HMM の F 値に大きな低下は見られない。このことから、BTM-HMM は比較的少数の学習データによる学習において、高い識

別率を示すことがいえる。

8. 結論

本研究では、センサデータから日常生活行動を分類するための、教師あり学習に基づく時系列分類法として、Tied-Mixture Hidden Markov Model (TM-HMM) をベイズモデルに拡張した手法、BTM-HMM を提案した。BTM-HMM は、潜在変数およびパラメータの事前分布を導入し、事後確率最大化により HMM の状態数や混合分布の要素数を観測データから自動的に決定することが可能になる。BTM-HMM と、TM-HMM、時系列を加味しない既存の教師あり学習モデルである SVM と KLR に、周期的な運動に対する加速度のデータセット、非周期的な運動に対する加速度のデータセットの両者の運動を含むデータセットに対して分類性能の比較検討を行った。本提案手法は、比較的少数のデータに対しても、結果が安定しており、高い精度の認識率を示した。

参考文献

- [1] Bao, L. and S. S. Intille (2004). Activity recognition from user-annotated acceleration data. *Proceedings of the Second International Conference on Pervasive Computing*, 3001, 1-17.
- [2] Barshan, B. and M. C. Yuksek (2014). Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units. *The Computer Journal*, 57, 11, 1649-1667.
- [3] Bellegarda, J. R. and D. Nahamoo (1990). Tied mixture continuous parameter modeling for speech recognition. *IEEE Transactions on Acoustics Speech and Signal Processing*, 38, 12, 2033-2045.
- [4] Bishop, C. M. (2012). *パターン認識と機械学習*, 丸善出版.
- [5] Ferguson, T. S. (1973). A bayesian analysis of some nonparametric problems. *Annals of Statistics*, 1, 1, 209-230.
- [6] Gael, J. V. et al. (2008). Beam sampling for the infinite hidden Markov model. *Proceedings of the 25th International Conference on Machine Learning*, 1088-1095.
- [7] Hu, D. H. et al. (2009). Abnormal activity recognition based on HDP-HMM models. *Proceedings of IJCAI*, 1715-1720.
- [8] 池谷直紀・菊池匡晃・長 健太・服部正典 (2008). 3 軸加速度センサを用いた移動状況推定方式. *電子情報通信学会研究報告, ユビキタス・センサネットワーク (USN)*, 108, 138, 75-80.
- [9] Matthew J. B. et al. (2014). The Infinite hidden markov model. *Advances in Neural Information Processing Systems 14*, 577-584.
- [10] Nigam, K. et al. (2000). Text classification from labeled and unlabeled documents using EM. *Machine Learning*, 39, 2-3, 103-134.
- [11] 小川兼人・伊藤雄一・安部登樹・岸野文郎 (2009). 実物体によるモーショクエリを用いた 3 次元形状モデル検索. *情報処理学会シンポジウム論文集*, 4, 9-16.
- [12] 小川延宏・梶 克彦・河口信夫 (2011). 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 (DICOMO 2011) シンポジウム論文集*, 76-82.
- [13] Raman, N. and Maybank, S. J. (2015). Action classification using a discriminative multilevel HDP-HMM. *Neurocomputing 154*, 149-161.
- [14] Sasaki, Y., K. Yoshii, and S. Kagami (2013). A nested infinite Gaussian mixture model for identifying known and unknown audio events. *Proceedings of International Workshop on Image Analysis for Multimedia Interactive Services*, 1-4.
- [15] Teh, Y. W. et al. (2006). Hierarchical Dirichlet process. *Journal of the American Statistical Association*, 101, 476, 1566-1581.
- [16] 上田修功・山田武士・桑田修平 (2007). ディリクレ混合過程モデルに基づく半教師有り学習. *電子情報通信学会技術研究報告, パターン認識・メディア理解 (PRMU)*, 107, 115, 87-92.
- [17] Walker, S.G.(2007). Sampling the dirichlet mixture modelwith slices. *Communications in Statistics - Simulationand Computation*, 36, 45.
- [18] 波部 斉 (2010). ランダムフォレスト. *情報処理学会研究報告, Vol. 2012-CVIM-182, No. 31*.