

# 人行動分類のための類型パターンに基づく最近傍法

須賀 佑太朗<sup>1</sup> 安藤 晋<sup>2</sup> 関 庸一<sup>2</sup>

**概要:** 近年, Kinect などの安価な撮影機器や画像認識ライブラリ等の普及により, 人行動認識に有用な多変量時系列データを比較的容易に観測できるようになった. しかし, これらの時系列は欠測期間を多く含み, また実用的には十分な観測が得られていない時点で出力を求められる場合が起こりうるため, 従来の時系列分類手法では高い精度を実現することは難しい. 本研究では, クラスに特徴的な類型パターンを利用し, それらが明確に観測された場合のみ予測を出力する最近傍法を提案する. さらに, 欠測期間に対応するアンサンブル学習の枠組みを実装する. 提案手法を行動データに適用し, 既存手法との比較において, より短い時間で, かつ精度の高い分類が実現できることを示した.

**キーワード:** アンサンブル学習, 最近傍法, 階層的クラスタリング, 人行動分類, 多変量時系列分類

## Nearest Neighbor Classification based on Exemplary Pattern for Classifying Peoples' Behavior

**Abstract:** The advent of affordable monitoring devices such as Kinect and image recognition libraries have made it easier to observe multivariate time series data useful for recognizing people's behavior. However, such time series data are prone to durations of missing values and a classifier may be asked of an output at an arbitrary timing without a sufficient amount of observation which makes it difficult to achieve accurate predictions by existing time series classification techniques. In this paper, we propose a nearest neighbor classification model which exploits exemplary extracted from the cluster hierarchy of the time series. In addition, we implement an ensemble framework which outputs predictions only when a decisive pattern is observed, in order to address the missing values. Our empirical study using a real-world behavior dataset showed that the proposed method achieves better precision in shorter time compared to the baseline methods.

**Keywords:** Ensemble Learning, Nearest Neighbor Classification, Hierarchical Clustering, People' Behavior Classification, Multivariate Time Series Classification

### 1. はじめに

近年, Kinect などの安価な撮影機器や画像認識ライブラリ等が普及している. これにより, 行動分類のためデータ収集基盤が整いつつある. 動画像から人の各部位検出・追跡する技術は PIA (People Image Analysis) と呼ばれ, サポートを提供し生活の質を向上させ, サポートを提供するアプリケーションを実現する上で, PIA によって認識した人体の状態から行動を予測する問題は重要な役割を持つ [2].

しかしながら, 画像から抽出された特徴に基づいて行動分類を行う上で 2 つの大きな問題がある. このような特徴量の時系列はハードウェアの技術的限界によって欠測期間

を多く含む. 時系列データの中には多くの欠測期間を含むものがあり, 分類問題でこれらのデータを特徴量として使用した場合, 欠測期間により大きな影響を受けるため高い精度を実現することが難しい. また観測に成功する特徴量は事前に分からないため分類に有用な特徴を選択することも難しい. このため, 従来の時系列分類手法では高い精度を実現することは難しい. また, 実用的には, 行動の完了する時点は明示的でないため, 任意の時点で出力を求められた場合, 十分な観測が得られているとは限らない. このような条件は既存の分類手法では扱うことが難しい.

近年, 時系列分類問題において時系列に含まれる部分系列パターンを用いた事例ベース手法が有望なアプローチとして注目されている. 例えば Shapelet と呼ばれる部分系列からの距離に基づいて決定木その他の分類器を学習する手

<sup>1</sup> 群馬大学理工学府工学研究科情報工学専攻

<sup>2</sup> 群馬大学理工学研究科工学研究科情報工学専攻

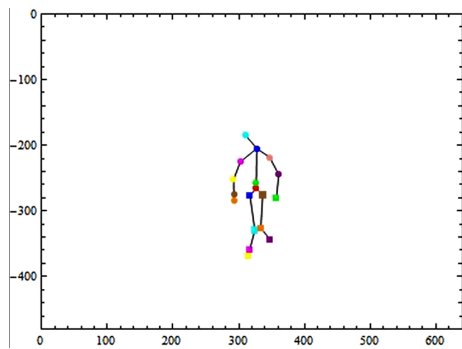


図 1 作成したスケルトン

法が多数提案されている [7][8].

本研究では、十分な観測が得られた場合にのみ予測を出力するため、抽出した類型パターンが近傍となる入力に対してのみ予測を出力する最近傍法を提案する。さらに、観測に成功した特徴量毎に予測を出力し、その結果を統合するアンサンブル学習の枠組みを提案する。提案手法をUHC-NN(Uniform Hierarchical Clustering-Nearest Neighbor)法と呼ぶ。

UHC-NN法は多変量時系列中に観測に成功した類型パターンが見つからない場合は予測を保留するため欠測に対応することができる。Kinectで観測した行動データを用いた実験により提案手法を評価し、既存手法よりも短時間で精度の高い分類を実現できることを示した。

## 2. データ概要

### 2.1 Kinect 観測データ

本研究では、群馬大学工学部の講義におけるティーチングアシスタント(以下TA)の行動データを収集し、分析を行った。行動データの収集にはマイクロソフト社のKinect for Windows™[5]を利用した。Kinectは映像及び距離センサーにより、RGB画像及び深度画像を取得できる[4]。さらに付属のライブラリ(Kinect for Windows Developer Toolkit v1.6.0)ではRandom Forestを用いて20種の関節(頭、中央肩、両肩、両手首、両手、両肘、両膝、両足、両足首、両臀部、中央臀部、背骨)を認識する[6]。それぞれの関節について $x, y$ 座標(640×480画像)、計測時刻、トラッキング状態を本研究で使用した。

トラッキング状態は、各関節の認識結果を表し、認識(Tracked)、推測(Inferred)の2つの値を取る。Inferredの場合は座標の値が不正確であるため、本研究ではこのときの座標の値を欠測として扱った。以下では全関節のデータを指してスケルトンと呼ぶ。図1に認識したスケルトンの例を示す。各関節は異なる色または形で表され、適宜直線によって接続する。実験に使用した関節データの詳細を表1に示す。

関節の座標は大きなばらつきを含むため、0.1秒間隔で平均を取り平滑化を行った。

表 1 関節データ概要

番号	データ数(平滑化後)	Tracked 率	行動
(1,-1)	19,038	0.67	座る・歩く
(2,-1)	19,038	0.73	歩く
(3,-1)	13,397	0.64	歩く
(4,-1)	14,720	0.71	歩く
(1,1)	29,840	0.79	教える
(2,1)	16,880	0.78	教える
(3,1)	23,032	0.69	教える
(4,1)	22,720	0.73	教える

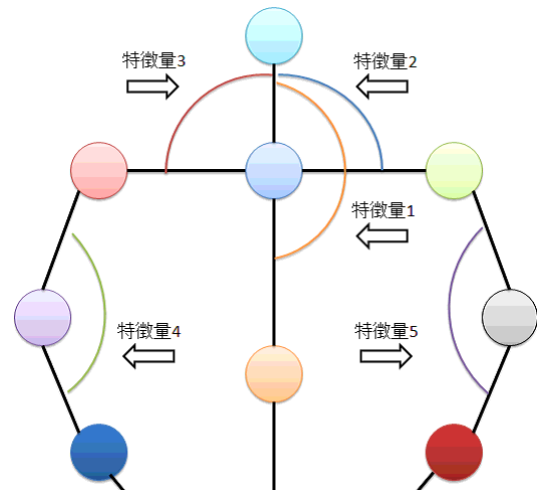


図 2 分類に用いる姿勢特徴量

### 2.2 姿勢特徴量の求め方

本研究では、関節間の角度が人の姿勢、さらには行動を表す特徴として有用であると考え、図2に示す5つの角度を用いる。ある角が3つの関節 $A, B, C$ としたとき、 $\vec{AB}, \vec{BC}$ のベクトル間から求められるコサイン距離を特徴量として用いる。

$$\text{コサイン距離} = 1 - \frac{\vec{AB}\vec{BC}}{|\vec{AB}||\vec{BC}|} \quad (1)$$

観測された特徴量3の時系列の例を図3に示す。ただし、負の値は関連する間接座標の欠測を表す。さらに、各特徴量時系列から滑走窓により長さ $L$ の部分系列集合に分割し、それぞれが一つの行動事例の特徴ベクトルを表すものとする、すなわち、行動事例は $L/10$ 秒分の5つの部分系列によって表される。

## 3. 従来研究

### 3.1 Local Shapelet

Local Shapeletは、時系列分類において有用な情報を持つ部分系列であり、ラベル付き時系列集合から距離に基づく評価に従って抽出される。マッチングに基づく分類に用いられる他、データの特徴を視覚的に判断しやすいというメリットがある。Local Shapeletは、部分時系列、閾値、ターゲットクラスからの3つからなるタプルであり、以下の

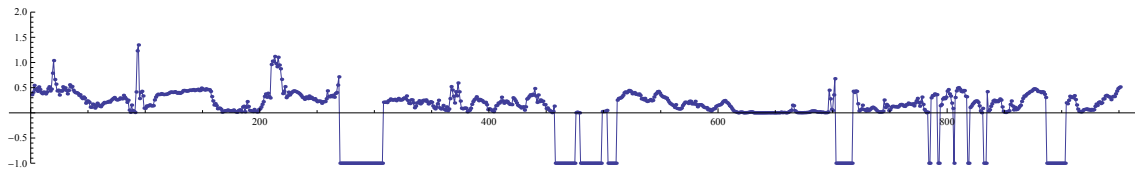


図 3 特徴量の時系列データの例

手順によって訓練データから抽出される。

- (1) Local Shapelet と全ての訓練データとの距離を計算する。
- (2) 計算された距離の中から、ターゲットクラス以外の訓練データとの距離のみを選択する。
- (3) 選択された距離を確率変数の実現値の集合とし、この集合の平均から、標準偏差の  $k = \sqrt{1/(1-x)}$  倍離れた値を Local Shapelet の閾値とする。

Local Shapelet の部分時系列  $s$  と時系列の距離は最小一致距離 (Best Match Distance) によって求められる。ただし、5 節に示す実験ではユークリッド距離を用いる。

テストデータ  $t$  を分類する場合は各 Local Shapelet への距離を求め、距離が閾値以下となる Local Shapelet のターゲットクラスを予測とする。該当する Local Shapelet が複数ある場合は最も近傍にある Local Shapelet を選択する。

$x$  は Local Shapelet の部分時系列と異なるクラスのデータとの距離が  $k\sigma$  以下になる確率を表すパラメータであり、距離の分布に関するチェビシェフの不等式によって閾値と関連づけられる [1]。

## 4. 提案手法

### 4.1 類型パターンに基づく最近傍法

本節では類型パターンの抽出方法および類型パターンを用いた最近傍法について述べる。類型パターンの抽出は一変量時系列の部分系列集合を入力とし、それぞれの部分系列に対応するラベルを出力する。また、中間データとして訓練データの階層クラスタ構造を単連結クラスタリング [3] により生成する。

訓練データ、テストデータは単一特徴量の部分時系列とする。訓練データの集合を  $X$ 、そのラベルの集合を  $Y$  とする。ここで扱う問題は 2 クラス分類問題とみなし、ラベルの値は  $\{1, -1\}$  とする。さらに、テストデータに対する予測の値は保留を表す 0 加え、 $\{1, 0, -1\}$  からとるものとする。テストデータが欠測を含む場合は NA を出力するものとする。

- (1) 部分系列の集合を入力とし、単連結クラスタリングにより階層木構造を生成する。
- (2) 部分木のうち葉の数が  $m$  以上でなおかつ全ての葉の

### Algorithm 1 UHC:類型パターンの抽出

```

    入力: 訓練データ (部分系列集合)  $X = \{x_i\}_{i=1}^N$ 
        訓練ラベル  $Y = \{y_i\}_{i=1}^N$ 
         $y_i \in \{1, 0, -1\}$ 
         $m$  (最小一様クラスタサイズ)
    出力: 一様クラスタラベル  $Z = \{z_i\}_{i=1}^N$ 
         $z_i \in \{1, 0, -1\}$ 
    関数: HC: 階層的クラスタリング (単結合)
     $t$ : (単連結木)  $Y = \{y_i\}_{i=1}^N$ 
     $t \leftarrow HC(X)$ 
     $T \leftarrow t$  に含まれるラベル  $y_i$  が一様な部分木を列挙
     $T' \leftarrow T$  の中で葉の数が  $m$  以上ある要素を列挙
    for  $x \in X$  do
        if  $x_i \in T$  の葉の集合 then
             $z_i = y_i$ 
        else
             $z_i = 0$ 
        end if
    end for
    return  $Z$ 
    
```

ラベルが同じであるものを抽出する。これを一様類型パターン (UHC) と呼ぶ。

- (3) 各訓練データについて UHC ラベルを次のように定める。
    - (a) あるデータが UHC の要素でなければ UHC ラベル値を 0 とする。
    - (b) さもなくば、元のラベルを UHC ラベル値とする。
- 上記の手順の疑似コードを Algorithm1 に示す。

UHC ラベルを用いた最近傍法はテストデータに対して最も距離の近いの訓練データを求め、その UHC ラベルを予測クラスとして出力する。UHC 要素は二つのクラスの分布が重複する領域には存在せず、それを用いることで最近傍法がノイズやクラスの重複を受け難くなる。

### 4.2 最近傍法予測のアンサンブル

新たに観測される行動は  $p$  個の特徴量部分系列によって表される。よって UHC ラベルを用いた最近傍法を特徴量毎に実行することで  $p$  個の予測が得られる。ただし、予測値  $\{-1, 0, 1, NA\}$  のいずれかの値をとる。この予測ベクトルを以下の手続きで一つの予測値に集約する。

- (1) ラベル  $\{1, -1\}$  のうち多数派を選択
- (2) NA を除いた全体のラベル数との割合が確信度閾値以上ならばそのラベルを出力、未満または多数派が存在しない場合は 0 を出力。

**Algorithm 2** アンサンブルによる予測

入力:  $p$  変量時系列の部分系列集合  $\{X_i\}_{i=1}^N$   
 部分系列のラベル集合  $\{Y_j\}_{j=1}^p$   
 $r$ (確信度閾値)  
 出力: 予測ラベル  $Y' = \{y_i\}_{i=1}^N$   
 $Y'_j = \{y'_{ij}\}_{i=1}^N, y'_{ij} \in \{1, -1, 0, NA\}$   
 関数: UHC(訓練データ, ラベル, 最小一様クラスタサイズ  $m$ )  
 1NN(訓練データ, ラベル, テストデータ)  
**for**  $j = 1$  to  $p$  **do**  
      $Z_j = UHC(X_j, Y, m)$   
      $Y_j = 1NN(X_j, Z_j, X'_j)$   
**end for**  
**for**  $i = 1$  to  $N$  **do**  
      $Z'_i = Z \in \{1, -1\}$   
      $y'_i = Z_i : r_{Z_i} \geq r, 0 : otherwise$   
     ただし,  $r_{Z_i}$  は,  $\{y_{ij}\}_{j=1}^p$  に含まれる  $Z$  の個数から  $NA$  を除いた中で  $Z_i$  の占める割合の大きい方  
**end for**  
**return**  $Z$

上記手順の疑似コードを Algorithm2 に示す。提案手法では観測に成功した特徴量に類型パターンが近傍となる部分系列が含まれる場合のみ予測を出力するため、欠測がある場合にも予測が可能となる。

**5. 数値実験**

**5.1 実験設定**

本研究では8つの動画から抽出した姿勢特徴量を用いて評価実験を行った。TAの行動は「歩く」と「教える」の2種類がある。5変量時系列から、滑走窓により抽出した部分時系列を8つの部分集合に分割し4×2交差検定を実施した。訓練データとテストデータに含まれる事例数は表2のようになる。最小クラスタサイズ  $m = 5$  の時、確信度閾値  $r = \{0.2, 0.33, 0.4\}$ , 最小クラスタサイズ  $m = 10$  の時、確信度閾値  $r = \{0.2, 0.33, 0.4\}$  として実験を行った。部分時系列の長さを1秒相当とした。実験は、Intel<sup>R</sup> Core i7<sup>TM</sup> 2.30Hz CPU (OS Windows) の計算機を使用し、Mathematica 環境で実施した。

**5.2 ベースライン手法**

比較として、同様の設定で提案手法、Local Shapelet および4つの基本的な正解率の分類結果を評価する。Local Shapelet は、 $x = \{0.7, 0.8\}$  の場合について実験を行った。SMO(Sequential Minimal Optimisation),  $kNN(k = 1, 3, 5)$ , J48(決定木) による分類は Weka \*1 上で実行した。これらの手法の出力は保留を含まず予測ラベル値のみであるため、部分系列の長さを3,5,7秒相当とした場合についても評価を行った。

**5.3 視覚的分析**

本実験で得られた類型パターンを視覚的に分析した結果

\*1 <http://www.cs.waikato.ac.nz/ml/weka/downloading.html>

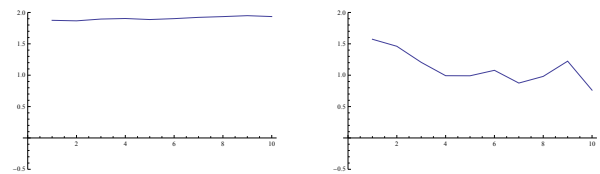


図4 類型パターンの例

表2 訓練データとテストデータの事例数

	Fold1	Fold2	Fold3	Fold4
訓練データ	1228	888	900	926
テストデータ	1227	888	900	926

から、2つの行動の特徴として解釈可能な例を図4(a)(b)に示す。(a)は行動「歩く」の時に現れた特徴量1の典型的な類型パターンの一例であり、数値は高いままで直線的である。これは人が歩いている時は、頭は大きく動かないので安定した値を保ったままであるということを表している。一方、(b)は行動「教える」の時に現れた特徴量1の典型的な類型パターンの一例であり、数値は上下に動き値はそれほど高くない。これは人が教えている時は頭を傾けたりすることによって不安定な値を取るということを表している。

**5.4 結果**

提案手法と Local Shapelet の分類結果をそれぞれ表3と4に示す。ただし、最右の列には保留率を除いた場合の正答率を正答率\*と表記する。基本的な分類手法の分類結果を表5に示す。

表3より、提案手法では保留を許したことで正答率は1秒以内で約90%を実現し、既存の分類手法よりも短い時間でかつ精度の高い行動分類を実現していることが分かる。パラメータ  $m$  または  $r$  を大きくした場合、いずれも保留率が増加し、正答率は減少するという傾向がある。この傾向を実用的に利用するならば、許容されるある正答率を上回れば予測を行うという制約を付けることができる応用として期待できる。例えば、表3を踏まえると、 $m = 10$  の時、許容正答率が0.88ならば  $r = 0.2$  を選択すべきだが、許容正答率が0.89ならば  $r = 0.4$  を選択すべきであると考えられる。また許容される観測時間が7秒であるとする表5から最大の正答率は86.83%であるためそれ以上の正答率であれば実用的には有効と考えられる。

表4より、Local Shapelet も同様に  $x$  の値または  $r$  の値を大きくすると保留率は増加し、正答率は減少するという傾向が見られる。提案手法と Local Shapelet の比較では、全体的には保留率と正答率\*については一方が増大するともう一方も同時に増大する傾向が見られる。表3, 4の結果からは、パラメータ設定ごとに両者が異なるため2つの手法を直接的に比較することは難しい。ただし、保留率が0.5以下の場合に限ると提案手法の正答率\*が高いといえる。

表 3 提案手法の実行結果

$m$	$r$	誤答率	保留率	正答率	正答率*
5	0.20	0.037 ± 0.011	0.439 ± 0.102	0.525 ± 0.110	0.934 ± 0.030
5	0.33	0.041 ± 0.017	0.617 ± 0.112	0.343 ± 0.098	0.881 ± 0.029
5	0.40	0.011 ± 0.007	0.693 ± 0.111	0.296 ± 0.112	0.965 ± 0.074
10	0.20	0.041 ± 0.021	0.638 ± 0.136	0.322 ± 0.116	0.886 ± 0.025
10	0.33	0.033 ± 0.019	0.690 ± 0.110	0.278 ± 0.092	0.883 ± 0.023
10	0.40	0.016 ± 0.008	0.814 ± 0.085	0.171 ± 0.173	0.898 ± 0.031

表 4 Local Shapelet の実行結果

$x$	$r$	誤答率	保留率	正答率	正答率*
0.7	0.20	0.111 ± 0.050	0.327 ± 0.060	0.562 ± 0.093	0.835 ± 0.076
0.7	0.33	0.020 ± 0.013	0.749 ± 0.071	0.231 ± 0.069	0.921 ± 0.045
0.7	0.40	0.020 ± 0.013	0.749 ± 0.071	0.231 ± 0.069	0.921 ± 0.045
0.8	0.20	0.077 ± 0.031	0.486 ± 0.073	0.437 ± 0.093	0.851 ± 0.068
0.8	0.33	0.006 ± 0.004	0.879 ± 0.036	0.115 ± 0.036	0.952 ± 0.030
0.8	0.40	0.006 ± 0.004	0.879 ± 0.036	0.115 ± 0.036	0.952 ± 0.030

表 5 各種分類手法の正答率 (単位:%)

時間 (秒)	SMO	$kNN(k=1)$	$kNN(k=3)$	$kNN(k=5)$	J48
1	77.17	52.23	50.54	51.15	78.62
3	80.60	48.24	52.11	57.20	82.08
5	82.77	54.11	56.21	57.65	84.67
7	84.79	57.56	59.86	61.30	86.83

Early Prediction on Time Series: A Nearest Neighbor Approach. In Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI'09), Hiroaki Kitano (Ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1297-1302.

- [8] Xing, Z., Pei, J., Yu, P. S. and Wang, K.: Extracting Interpretable Features for Early Classification on Time Series, Proc. of the 11th SIAM International Conference on Data Mining, pp. 247-258(2011).

## 6. おわりに

本研究では、欠測期間による大きな影響、分類に有用な特徴量の選択方法といった問題に対処するために類型パターンを利用した最近傍法及び予測のアンサンブルを提案した。既存手法との比較では、同程度以上の正答率を出すことに成功した。一般的な分類手法と比較した場合、保留を許すことによって、正答率を大幅に上げることに成功した。

今後の課題として、各パラメータの設定基準、早い段階での分類によって実現できる応用、保留を許す場合の結果の比較方法などが上げられる。

## 参考文献

- [1] 辻本貴昭 上原邦明 Local Shapelet を用いた時系列分類に最適な距離尺度の選択 情報処理学会研究報告.BIO, パイオ情報学 2012-BIO-32(27), 1-6, 2012-11-29.
- [2] 藤吉弘亘「人間を理解するための ICT 技術-人間を対象としたセンシング・情報処理からその応用まで-」『電子情報通信学会誌』 Vol.95 pp.416-421. 電子情報通信学会.2012.
- [3] 元田浩・津本周作・山口高平・沼尾正行 (2006): データマイニングの基礎, オーム社.
- [4] Kinect のトラッキング原理「部位認識に基づく 3D 姿勢推定, <http://derivecv.tumblr.com/post/2106495200>
- [5] Kinect for Windows, <http://www.microsoft.com/en-us/kinectforwindows/>
- [6] J.Shotton, A.Fitzgibbon, M.Cook, T.Sharp, M.Finocchio, R.Moore, A.Kipman, and A.Blake, "Real-time human pose recognition in parts from single depth images." CVPR, vol.2, pp.1297-1304, 2011.
- [7] Zhengzheng Xing, Jian Pei, and Philip S. Yu. 2009.