

# Compression-based Dissimilarity Measure(CDM)を用いた 人感センサ情報の類似判定

菊地 誠<sup>†</sup>阿部 洋丈<sup>†</sup>岡部 正幸<sup>‡</sup>梅村 恭司<sup>†</sup><sup>†</sup>豊橋技術科学大学 情報工学系<sup>‡</sup>豊橋技術科学大学 情報メディア基盤センター

## 1 はじめに

人感センサを用いて人間の行動を判定する場合、その情報は膨大であり、可変長であることやノイズが含まれることなどから、取得した人感センサ情報をそのまま用いて判定することは困難である。人感センサの情報から人間の行動を判定し異常を検知する研究に、藤原ら<sup>2)</sup>の異常検知システムがある。藤原らのシステムでは確率モデルに基づいて解析しているが、本研究では文字列処理やデータ圧縮技術を用いて解析する。

本研究では Keogh らが提案する、Compression-based Dissimilarity Measure(CDM)<sup>1)</sup>を用いて、人感センサ情報を類似判定する。CDMとは、文字列化した時系列情報の圧縮を用いて、多種多様な時系列情報を類似判定するための尺度である。本稿では実際に観測した人感センサ情報を文字列化し、既存のデータ圧縮方式である zip, lha, rar を用いて、CDMで人感センサ情報を類似判定した結果を報告する。

## 2 Compression-based Dissimilarity Measure(CDM)

Keogh らが提案する CDM は以下の式により求められる。

$$CDM_{\alpha}(x, y) = \frac{C_{\alpha}(x, y)}{C_{\alpha}(x) + C_{\alpha}(y)} \quad (1)$$

$x, y$  : 文字列

$C(x), C(y)$  : 文字列  $x, y$  のファイルをデータ圧縮したファイルのバイト数

$C(x, y)$  : 文字列  $x, y$  を結合したファイルをデータ圧縮したファイルのバイト数

$\alpha$  : データ圧縮方式

複数の文字列に対して CDM を算出し、最も小さな値になる文字列同士が最も関係があると判定する。

Categorization of Infrared Detector Information Using Compression-based Dissimilarity Measure(CDM)

<sup>†</sup>Makoto KIKUCHI, Hirotake ABE, Kyoji UMEMURA

Department of Information and Computer Sciences

Toyohashi University of Technology

<sup>‡</sup>Masayuki OKABE

Information and Media Center, Toyohashi University of Technology

同じ条件から生じた 2 つの時系列情報を結合して圧縮すると、圧縮のために必要な辞書などの情報が共有できるため、別々に圧縮した場合よりも圧縮率が向上することが期待できる。一方、別の条件から生じた 2 つの時系列情報では共有ができないため、結合しても圧縮率が向上しないと考えられる。CDMはこの考えを利用したものである。

## 3 人感センサ情報の類似判定実験

### 3.1 使用するセンサ

本実験では人感センサとして焦電型センサを用いる。焦電型センサはセンシング範囲内で物が動いた時にオンとなり、今回の実験ではオンを 1, オフを 0 とする二値センサとして使用する。センサは図 1 のような方向で配置し、計 6 個のセンサを使用する。これらのセンサは我妻ら<sup>3)</sup>のセンサ方向制御の研究で用いられるものである。一人の人間が入口から入り席に着席する行動を 1 つのイベント ( $e_1, e_2, e_3, e_4, e_5, e_6$ ) とし、イベントごとに取得されたセンサデータをもとに類似判定実験した。

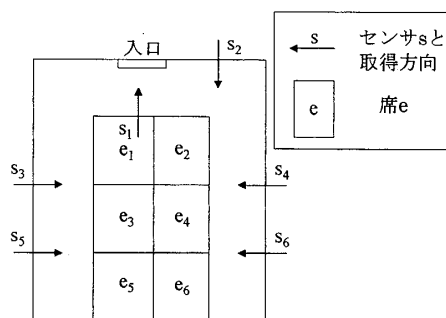


図 1 センサ配置

### 3.2 センサデータの時間間隔

焦電型センサによるセンシングの時間間隔は、通常の人間の行動を対象としているため 0.2 秒となっている。このためセンサの出力を一定間隔でサンプリングして得た実時間間隔データは、センサがオフの状態が非常に長くなる。ここで、我々は同じ状態があまりに長く続くデータでは、CDMが正常に動作しないのではないかと考えている。そこで、実時間間隔データを可変時間間隔データに変換した。以下の場合のみ値を取り出し、それ以外の値は破棄する。

- ・ センサ状態が変化した場合
- ・ 一定時間センサ状態に変化がない場合

### 3.3 センサデータの文字列化

CDM を算出するにあたり、センサデータを文字列化する必要がある。センサ 6 個の場合、図 2 のようにセンシングされた時間ごと縦に並ぶセンサデータを取り出し、区切り用文字を 1 文字置いて 1 行の文字列とする。

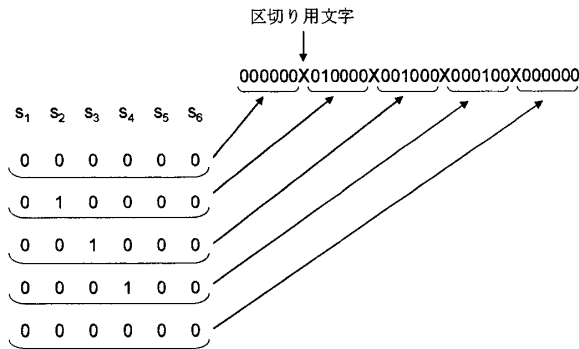


図 2 センサデータの文字列化

### 3.4 CDM を用いた類似判定

各席に座るイベント 6 つに対しそれぞれ 30 件のセンサデータにおいて、10 件を正解データ、20 件をテストデータとし、文字列化されたセンサデータに対し zip, lha, rar を用いてデータ圧縮し、CDM を算出する。正解データとテストデータが無作為に選択し、CDM を平均した結果が正解データに対し最も小さくなる場合に正解と判定する。センサ 6 個 (s<sub>1</sub>,s<sub>2</sub>,s<sub>3</sub>,s<sub>4</sub>,s<sub>5</sub>,s<sub>6</sub>) を用いた場合において、類似判定した結果を図 3 に示す。

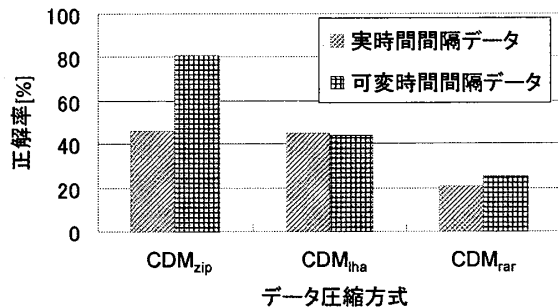


図 3 類似判定結果

図 3 から、まずデータ圧縮方式の違いにより正解率が異なっていることがわかる。データ圧縮アルゴリズムの違いが、CDM の結果に大きく影響を与えることがわかった。今回の目的では、rar は zip や lha よりも圧縮率が高いデー

タ圧縮方式ではあるが、CDM を正常に動作させるためのデータ圧縮アルゴリズムとしては適さない。また、各データ圧縮方式において実時間間隔データと可変時間間隔データを用いた場合を比較してみると、lha や rar ではほぼ変わらない結果となっているが、zip においては良い結果となることがわかる。これは zip のデータ圧縮アルゴリズムにおいては、実時間間隔データではセンサがオフの状態が長すぎるために CDM が正常に動作していなかったと考えられる。我々の可変時間間隔データへの変換によりオフの状態を縮め、CDM をより正常に動作させることができると思う。

## 4 おわりに

人感センサの情報を分析し人間の行動を判定するために、CDM を用いた類似判定の手法を示した。今回は既存のデータ圧縮方式である zip, lha, rar を用いることで、どの程度の正解率が得られるかを実験で確認した。その際に、zip においては実時間間隔データを用いる場合よりも可変時間間隔データを用いるほうが、正解率は良くなることがわかった。今後、その原因を分析したい。

今回用いた既存のデータ圧縮方式では、まだ十分な正解率が得られていない。今後は正解率上げるため、我々が分析対象としている人感センサの情報を、うまくデータ圧縮する方法を検討することが課題となる。

**謝辞** この研究は、戦略的情報通信研究開発推進制度 (SCOPE) の課題「インターユビキタスネットワーク情報基盤の研究」の成果である。また、平成 20 年度科学研究補助金課題番号 (19500120) の研究成果を使用した。

### 参考文献

- 1) Eamonn Keogh, Stefano Lonardi, Chotirat Ann Ratanamahatana, Towards Parameter-Free Data Mining: Proceedings of the tenth ACM SIGKDD, pp. 206-215, 2004
- 2) 藤原大輔, 菊地誠, 阿部洋丈, 岡部正幸, 梅村恭司. 人間の動作に対するアノマリ型異常検知システムの実装: 第 50 回プログラミングシンポジウム, pp. 181-184, 2009
- 3) 我妻裕樹, 阿部洋丈, 岡部正幸, 梅村恭司. 情報量の最大化に基づく指向性センサの方向制御: 第 50 回プログラミングシンポジウム, pp. 177-180, 2009