

下水処理施設における機器状態遷移の可視化 Visualizing State Cluster of Indicator Sequence in Swage Treatment Facilities

井上 千徳[†] 矢口 勇一[†] 佐藤 彰[‡] 菊池 稔[‡] 岡 隆一[†]
Yukinori Inoue Yuichi Yaguchi Akira Sato Minoru Kikuchi Ryuichi Oka

1. はじめに

下水処理施設において各装置が正常に動作しているかどうかを確認するため週に一度程度、人による点検が行われている。しかしながらこの方法による点検では突発的な装置の異常や変化に気づくことができない。そこで我々は各装置の動作状態を示すランプ盤を監視するためのモニタリングシステムを開発した。各装置の動作状態はランプ盤上のランプの色で判断することができる。ランプの色が緑色であればその装置は停止中であり、赤色ならば作動中、黄色であれば故障とみることができる。このモニタリングシステムにより遠隔地から常に各装置の動作状態をデータとして保存し確認することができるようになった。

下水処理施設には多くの装置が存在し汚水の流入量などその時々環境しだいで各装置の動作状況が変わる。装置そのものの故障はランプを見ることで確認することができるが装置の数も少なくないため施設全体の状態の兆候などをとらえることは難しい。本研究では、開発したモニタリングシステムを通して得られた各装置の動作状態のデータを基に施設全体の状態を主成分分析にかけたのちクラスタリングしそのクラスタの遷移を時系列で可視化する。

本研究に用いるデータは各装置のランプが各時点で何色なのかを示す離散的な値である。このデータの特徴量として本研究では主成分分析によるスコアを用いる。主成分分析はよく知られた統計的手法であり[1] 多次元のデータに対して適用することで次元削減をすることができる。本研究のデータは施設内のすべての装置のものであるので次元数が多いのでこれを要約してクラスタリングに用いる。また主成分スコアのユークリッド距離を用いてクラスタリングを行う。クラスタリングを用いて状態の可視化を行なっている研究はいくつかある[2][3]。

クラスタリングにはインクリメンタルクラスタリングを用いる。インクリメンタルクラスタリングは逐次的なクラスタリング手法である。はじめに最初の要素をひとつのクラスタとしそこに各要素を追加していく。追加された要素と既存のクラスタの中心との距離があらかじめ定めた閾値未満であれば同一のクラスタとし、閾値を超えた場合は新しいクラスタであるとする。これを時系列の古い方から順に適用していく。

これらの結果を表示するため各要素が所属するクラスタを色で分け時系列に沿って **bitmap** 形式で出力しクラスタの遷移を表示することによって施設全体の状態を可視化する。

2. 手法

2.1 データセット

我々が本研究で用いるデータセットは実際に観測した北会津にある下荒井地区の下水処理施設の装置の一年間の動作データである。二つのカメラを用いて 45 種類のランプ

を監視している。各ランプが示す色と装置の状態、本研究で用いる値との対応は表 1 のとおりである。

ランプの色	装置の状態	用いる値
赤	稼働中	1
緑	停止中	2
黄色	故障	3
無灯火 値なし	unknown	0

表 1 ランプの色と装置の状態と値の対応

各データは 1 分おきにカメラで撮影されクラウド上に保存される。

これらのデータを用いて行列を作成し主成分分析にかけられる。行列はランプの 45 種類の 1 分間隔で 1 年間撮影された 1,2,3,0 を要素としてもつ $45 \times 527040 (60 \times 24 \times 366)$ の行列である。

2.2 主成分分析とインクリメンタルクラスタリング

2.2.1 主成分分析

主成分分析を用いて主成分スコアを求め寄与率の高い変数を求め次元を削減する。与えられたポイント X

$$X = [x_1, x_2, \dots, x_n]^T$$

主成分 Y_1 は以下の式で求める

$$Y_1 = w_{11}x_1 + w_{12}x_2 + w_{13}x_3 + \dots + w_{1n}x_n$$

同様にして n 個目まで求めることができる

$$Y_n = w_{n1}x_1 + w_{n2}x_2 + w_{n3}x_3 + \dots + w_{nn}x_n$$

ここで

$$w_i^T w_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

$$\|w_i\|^2 = 1$$

2.2.2 インクリメンタルクラスタリング

主成分分析で寄与率の高い変数をいくつか選択しそれらを各要素の座標としてマッピングする。

まず初めの要素を第一のクラスタとする。そこから要素をひとつずつ追加していき逐次的にクラスタリングしていく。新たな要素と既に存在するクラスタの中心との距離が閾値以内であればそのクラスタに属するとし、閾値を超えていた場合は新しいクラスタを形成する。このとき新たな要素が追加されたクラスタは中心座標を計算しなおす。以下にアルゴリズムを示す。

- c_i : i 番目のクラスタの中心座標
- s_i : 要素 e_i が属するクラスタの番号
- N : 要素の総数
- Now : その時点でのクラスタの総数
- $Thres$: 閾値

```

Algorithm : Incremental clustering Flow
c1 <= e1
now = 1
s1 = now
for i = 2 to N do
  find <= 0
  dist ( MIN(DISTANCE(ei, c))
  si = i
  if dist < thres then
    addcluster
    find <= 1
    recalculate centroid coordinates
  end if
  if find is 0 then
    CreateNewCluster
    now ++
  end if
end for

```

3. 結果

表 2 は主成分分析の結果である。本実験では上位 3 位の主成分を用いてクラスタリングを試みた。その結果が図 1 である。クラスタは合わせて 10 個形成された。

これらの結果を基に所属クラスタを色で分け Bitmap で表示させた結果が図 2 である。この画像は横に分割みで 1 日分の要素、縦に 366 行連なっている。

主成分	寄与率	累積寄与率
PC1	0.6972	0.6972
PC2	0.1332	0.8305
PC3	0.01512	0.84560
PC4	0.01427	0.85987

表 2 主成分分析の結果(上位 4)

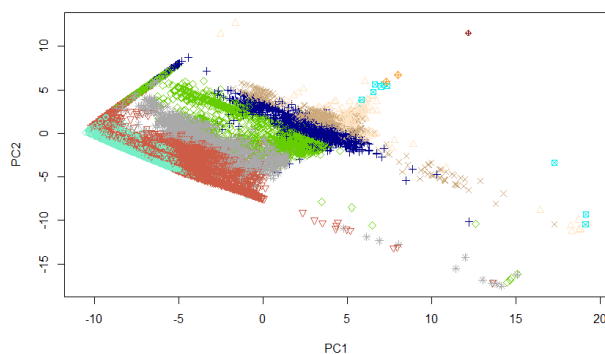


図 1 主成分スコアのクラスタリング結果

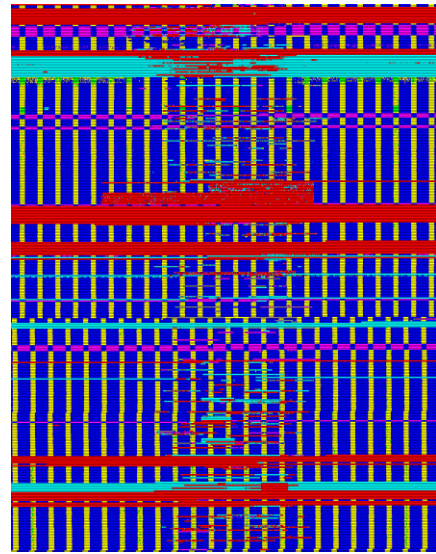


図 2 クラスタ遷移の bitmap 表示

4. まとめ

図 2 からクラスタの遷移には一定のパターンや周期があることが読み取れる。また一部の観測できなかった部分が多い場合についてはエラー値のように固まって出現しているのがわかる。

今後これらのクラスタに含まれる要素がどういった特徴をもっておりどういった要因で別のクラスタに移るのか調査していく必要がある。またインクリメンタルクラスタリングは自ら決めた閾値によって結果が変動するのでそのことも考量しつつ比較をしていく必要がある。

謝辞

研究をするにあたりアドバイスをいただいた同研究室の皆様と協力していただいたアクアクルー株式会社の皆様に心から感謝いたします。

参考文献

- [1] Ding., "K-means Clustering via Principal Component Analysis " Proceedings of the 21th International Conference on Machine Learning, Banff, Canada, 2004.
- [2] M. Kikuchi, M. Okamoto, T. Yamasaki, "Extraction of topic transition from document stream based on hierarchical clustering", DEWS2008 B3-3
- [3] N. Yamaguchi, "Visualizing State in Auto-Regressive Hidden Markov Models Using Generative Topographic Mapping", 2010

† 会津大学 University of Aizu

‡ アクアクルー株式会社