

# 多次元時系列データの時系列可視化分析

岡見 直樹 237X012X

## 1. Introduction

近年、様々なデータが時系列データとして保存されている。時系列データは、時間の経過に伴って収集されるデータで気象学、株式市場分析、医療診断、センサーデータなど、多くの分野で使われている。

時系列データには、複数の変数を持つ多次元時系列データとして保存されるデータがある。これらのデータは多くの変数によって構成されており、相互に値が関係していることが多い。

多次元時系列データを分析する際、複数の変数の影響を考慮する必要があるため、主成分分析や判別分析などが使われることが多い。これらを分析する際、よく、時間点ごとに変数の影響を考慮した低次元空間に投射され、散布図などで表されることが多い。

多次元時系列データにおいて時間点とそれぞれの変数の影響を考慮した分析ができることが重要である。

本システムでは、多次元時系列データを複数の変数の影響を考慮しながら、時系列の情報を得ることを目的とする。

## 2. Method

本手法では、多次元時系列データを TULCA によってラベルを考慮したクラスタを分ける。そして散布図を用いて TULCA を用いて低次元空間に投射したデータをプロットする。

散布図から選択した範囲の時間点を対話的にハイライトすることでクラスタが出来ている部分の時間的かわりが可視化できる。

## 3. Result

本システムの多次元時系列データを複数の変数の影響を考慮しながら、時系列の情報を得ることができるか確認する。

### 3.1 使用データ

スーパーコンピュータ「京」のログデータを使用した.864 台の計算ラックがあり、4 種類の測定値が計測されている。測定値は、空冷システムの冷却風の吸気温度、排気温度、ラック中の CPU 温度の平均、水冷システムの冷却水の温度の 4 種類の温度データである。本実験では、2014 年 4 月から 2017 年 3 月までのデータを利用した。

### 3.2 分析事例

実際に TULCA を用いて 2014 年から 2017 年を年度ごとにラベルを分け、それぞれの年度ごとの分散を大きくした散布図からクラスタを選択する。Fig.1 では、図の通り、選択を行い、下の時系列プロットからほぼすべてのデータが選択されていることが分かる。しかし、一部クラスタから外れている部分が見て取れる。

一部外れている部分は散布図を見ると、3 つに分かれていることが分かる。ここでは、それぞ

れを Fig.2 で見てみると選択したそれぞれ 3 つのクラスタがそれぞれ連続した違う時間が集まっていることが分かる。

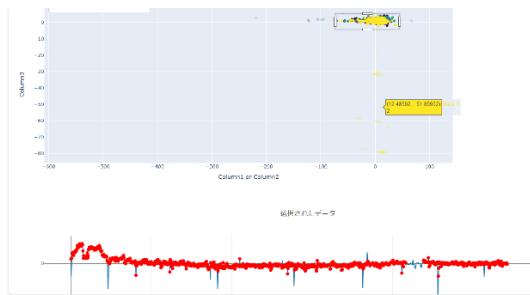


Fig.1: Selection Cluster

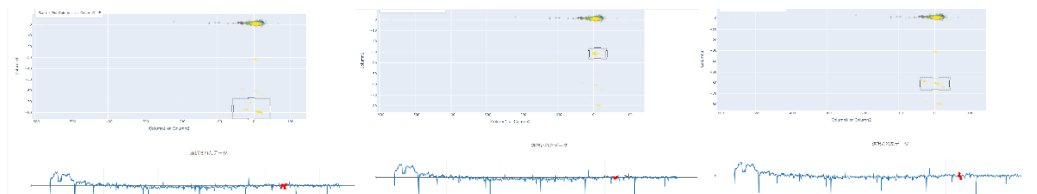


Fig.2: Detailed Selection Clusters

#### 4. Discussion

実験結果より、散布図プロットからクラスタを選択し、時系列プロットすることで散布図でのクラスタの時間軸での意味を分析できていると考えられる。

#### 5. Conclusion

本システムの多次元時系列データを複数の変数の影響を考慮しながら、時系列の情報を得ることができるか確認できた。

このシステムの改善点として時間での情報を得ることができるようになったが変数での情報を見ることができないため、分析できるようにすることでより詳細な分析ができると考えられる。