

次元削減技術を用いた視覚的テンソルデータ分析

藤田啓二郎・坂本尚久（神戸大学）・藤原孝紀（Linköping University）・野中丈士・塚本俊之（理研R-CCS）



はじめに

HPCシステムの発展

- システム及びプログラムの大規模化・複雑化
- 信頼性への要求

環境ログデータ

- センシング技術の向上
- 蓄積されるログデータの膨大化
- 有効活用への要求

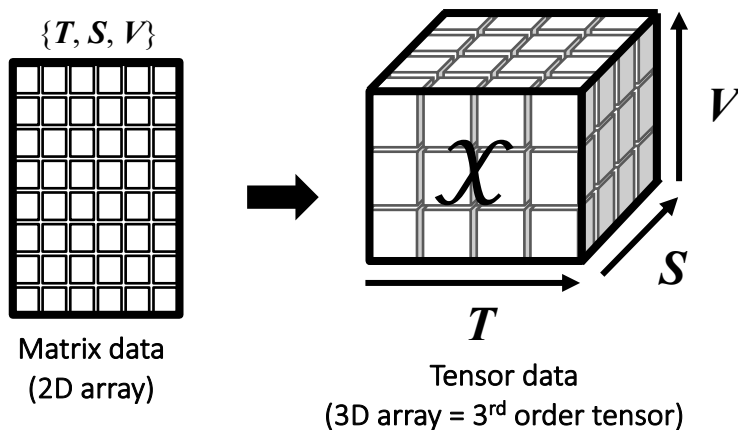
目的

多次元時系列データを時間・空間・計測値の3軸からなる3次テンソル（3次元配列）データとして表現し、複数の次元削減ステップを適用することで、特徴的な時空間領域を特定し、システムの状態や動作の理解を支援する。

テンソルデータ表現

3次テンソルデータ

- 多変量時系列データを時間（ T ）、空間（ S ）、変数（ V ）の3つの軸を持つテンソルデータとして表現

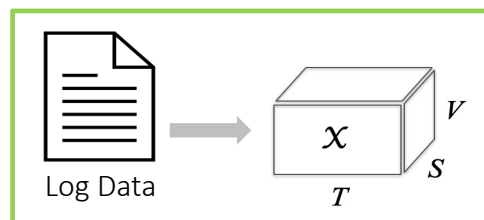


方法

次元削減技術を用いた視覚的テンソルデータ分析

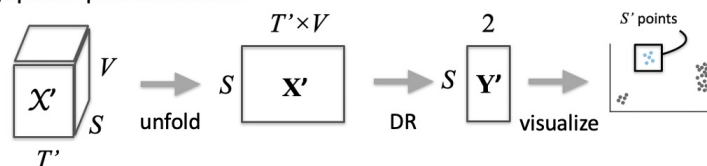
- 多変量時系列データを3次テンソルデータとして表現し、それぞれの軸ごとに次元削減を段階的に適用し、特徴的な時空間領域を対話的に選択する

1. HPC環境ログデータを3次テンソルデータとして表現



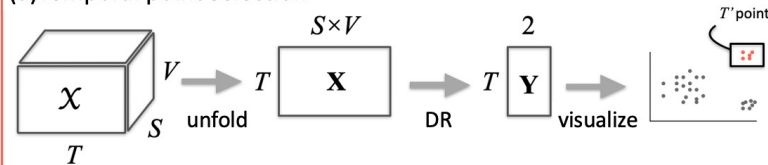
2. 次元削減とクラスタ選択により注目する時間領域の選択

(b) Spatial point selection



3. 次元削減とクラスタ選択により注目する空間領域の選択

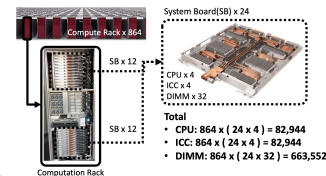
(a) Temporal point selection



結果

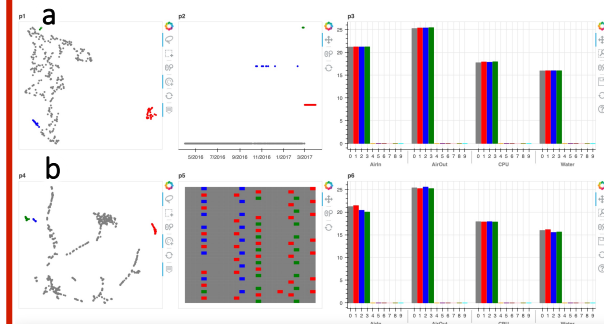
データ

- 「京」コンピュータ環境ログデータ
 - ✓ 期間：2016/04～2017/03
 - ✓ それぞれの計算ラックで測定されたデータを1日ごとに平均化
 - ✓ 時間点：361日
 - ✓ 空間点：864個の計算ラック
 - ✓ 変数：4種類の温度データ



変数

AirIn: 空冷システムの吸気温度
AirOut: 空冷システムの排気温度
CPU: 計算ラック中のCPU温度
Water: 水冷システムの冷却水温度



視覚的分析システム

- a: 時間領域選択
- b: 空間領域選択

左: 次元削減プロット
中央: 時間・空間分布
右: 各クラスタの持つ測定値の平均値プロット

分析事例

- 選択した時間・空間点群の分布に注目
 - ✓ 連続した時間点がクラスタとして現れた
 - ✓ 同じ列に配置されている計算ラックがクラスタとして現れた

ジョブの充填率・計算ノードにかかる負荷によって、空冷システムの吸排気温度が変化すると推定

考察

- 特徴的な時空間を選択することができた
- データの粒度を調整する必要
- クラスタの持つ情報・クラスタ間の差異をより工夫して可視化する必要