

次元削減を使った 視覚的テンソルデータ解析



藤田 啓二郎・坂本 尚久 (神戸大学大学院システム情報学研究科)

はじめに

スーパーコンピュータシステムの運用

- 科学的・社会的課題を解決するためには、スーパーコンピュータ (スパコン) は欠かすことのできない基盤環境となっている。
- スパコンの高性能化が進む一方で、システムの長期安定運用・省電力性への要求も高まっている。
- センシング技術の発展により、CPUやメモリ温度、使用電力量などのログデータがリアルタイムにモニタリングされストレージに蓄積されている。

課題

- 安定稼働に加え、エネルギー効率高く、より効率よく計算を行うためには、複雑なシステムの挙動を正確に理解しなくてはならない。

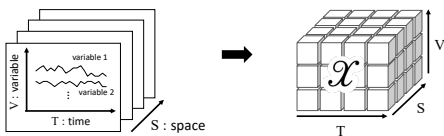
目的

ログデータを多変量時系列データ (テンソルデータ) として表現し、そこに内在する特徴構造を抽出し、その構造を効率よく理解するための視覚分析法を提案する。

方法

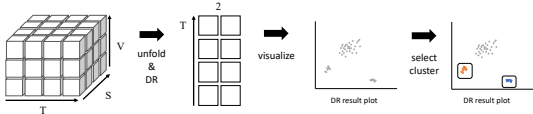
ログデータ

$\mathcal{X} \in \mathbb{R}^{N_t \times N_s \times N_v}$
 N_t : 時間点の数
 N_s : 空間点の数
 N_v : 測定値の種類数

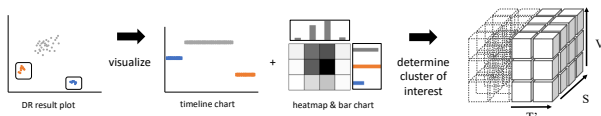


手順

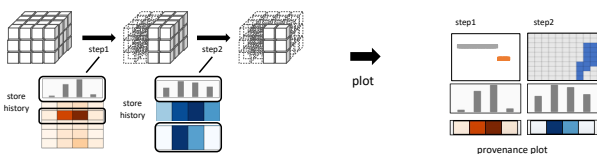
1. 多段階次元削減による特徴構造の抽出



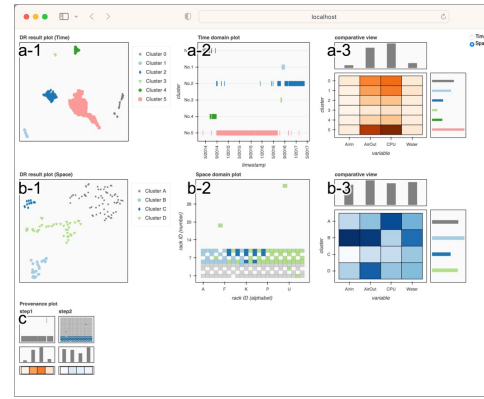
2. 選択する複数の部分特徴の比較可視化



3. 分析操作の履歴可視化



視覚的テンソルデータ解析システム



a. 時間領域プロット

- 1 次元削減プロット
- 2 タイムラインプロット
- 3 測定値プロット

b. 空間領域プロット

- 1 次元削減プロット
- 2 空間分布プロット
- 3 測定値プロット

c. 操作履歴プロット

システム実装

- Python
- Pandas
- Bokeh
- Scikit-learn
- UMAP-learn

結果

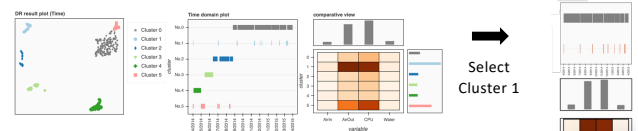
ログデータ

- 京コンピュータの環境ログデータ
- 2014/04~2017/03 の3年間
- 1086日
- 5分ごとに測定したデータを日平均化
- 864計算ラック
- 温度データ (4種類)
 - 空冷システム (AirIn, AirOut)
 - CPU温度 (CPU)
 - 水冷システム (Water)

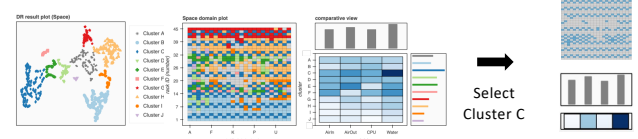
分析事例

- 1年分(2014/4~2015/3)のデータを対象にして年次の挙動を分析

Step1: 時間軸に注目して次元削減



Step2: 空間軸に注目して次元削減



- Step1の操作では…
 - 大規模ジョブ実行期間では?
 - 負荷高により排熱増と推察
- Step2の操作では…
 - ディスクラックに隣合う計算ラックでは?
 - CPU温度と制御チップの排熱が冷却水温に影響?

→ 抽出された部分構造は大規模ジョブ実行とラック配置に影響を受けている。

まとめ

- 先入観なく次元削減結果をもとにクラスタを選択することで特徴的な挙動をとっていた時空間領域を抽出し解釈することが可能である。
- 結論づけるには更なる検証が必要である。