

Adaptive Resonance Theory (ART2) Clustering on SX-ACE

NEC 横山晴道, NEC 荒木拓也

目的 ビッグデータの効率的な高速クラスタリング
いままで解析できなかったサイズのデータを解析し、大気化学の新たな知見につなげる

内容

- 高速分散処理フレームワークFelissのSX-ACEでの開発環境整備.
- APIを利用したMatrix Library や Machine Learning Algorithm Libraryの動作を検証.
- ART2 Clusteringアルゴリズムを実装、およびSX-ACEで実行.

結果

- 大気中PM2.5粒子の質量スペクトルデータ (15GB)
- SX-ACEのメモリを用いて測定した全データを解析可能となった
 - いままでは全体の5%しか解析できなかった
- 高速化パフォーマンス
 - スモールデータセットで、Matlabソフトウェアと比較して1.5倍高速化

利用した計算機 SX-ACE

- ノード時間 21時間
- 使用メモリ 15GB
- ベクトル化率
 - メイン処理 (行列ベクトル積) : 99.80、全体 : 99.6
- 並列化 1 並列
 - 今後、複数コア/ノードに拡張予定