

大規模データが拓く学理と技術

—核融合データの統合的取り扱いと、データ解析技法の高度化

高密度プラズマ物理 大舘暁

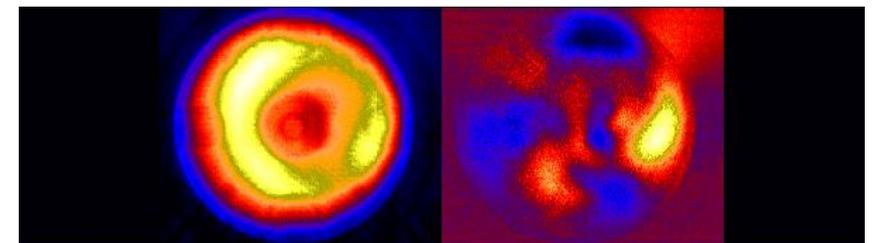
- 極めて複雑なデータであっても、その構造を抽出することができれば、その物理的な意味を考える糸口になる。
- 核融合プラズマ実験の大規模データには、データサイエンスの手法など新たなアプローチで解析を行う価値がある。

$m=3$ のモードと $m=1$ の逆回転するモードの相互作用。少数の項による説明できる。



2次元乱流データ PANTA (Kyushu Univ.)

Full Image Fluctuating component

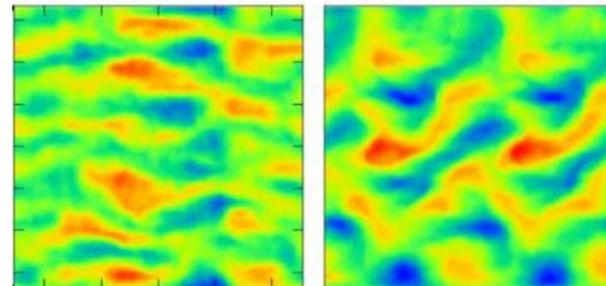


$m = 1$

$m = 2$

$m = 3$

大規模データが拓く学理と技術



ユニットの目標

複雑かつ大規模なデータから人間が理解可能な構造を抽出する

- ・ プラズマ乱流データの解析
- ・ トモグラフィ解析
- ・ 深層学習とディスラプション等イベント予測
- ・ 高次元シミュレーションデータからの構造抽出
- ・ データ同化シミュレーション etc.

乱流・輸送など物理現象の深い理解

計測・制御の高度化

核融合炉での運転手法の確立

研究を構成する学理・環境

国内のプラズマ実験の統一されたデータ提供環境

ビッグデータ・マネジメント
データ・マイニング

機械学習

数理統計学

可視化の技法

時系列データ解析

超分野の大規模データ利活用

データ解析手法の学理の確立

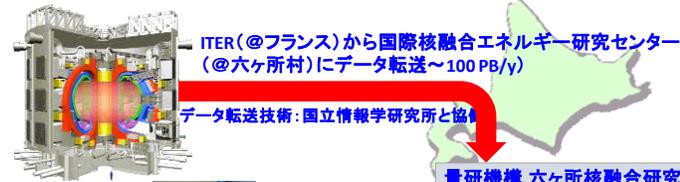
産業界/ICTへの貢献

基盤となる技術

- ✓ データ・セマンティック
- ✓ 機械学習(解析)計算サーバ
- ✓ 分散処理システム技術
- ✓ データアクセス方式の統合
- ✓ ITER/JT-60SAデータの利用
- ✓ 高速ネットワークデータ転送

- ✓ 核融合のオープンサイエンス化とLHD等デジタル資産の永続化
- ✓ 可視化・解析ツールのコミュニティ・ベース開発と共有
- ✓ 研究データ管理 (RDM)
- ✓ 高精度時間同期・実時間処理
- ✓ 遠隔実験技術・仮想化

コミュニティ及び一般社会への還元



ITER (@フランス) から国際核融合エネルギー研究センター (@六ヶ所村) にデータ転送 ~100 PB/y

データ転送技術: 国立情報学研究所と協働

量研機構 六ヶ所核融合研究所
核融合研六ヶ所研究センター
ITERデータ利活用を最大化

国内の様々な核融合実験及びITERのシームレスなデータアクセス
データ解析・可視化手法の共同開発や共有・再利用

核融合クラウド

核融合データの統一フォーマット
オープンデータ化

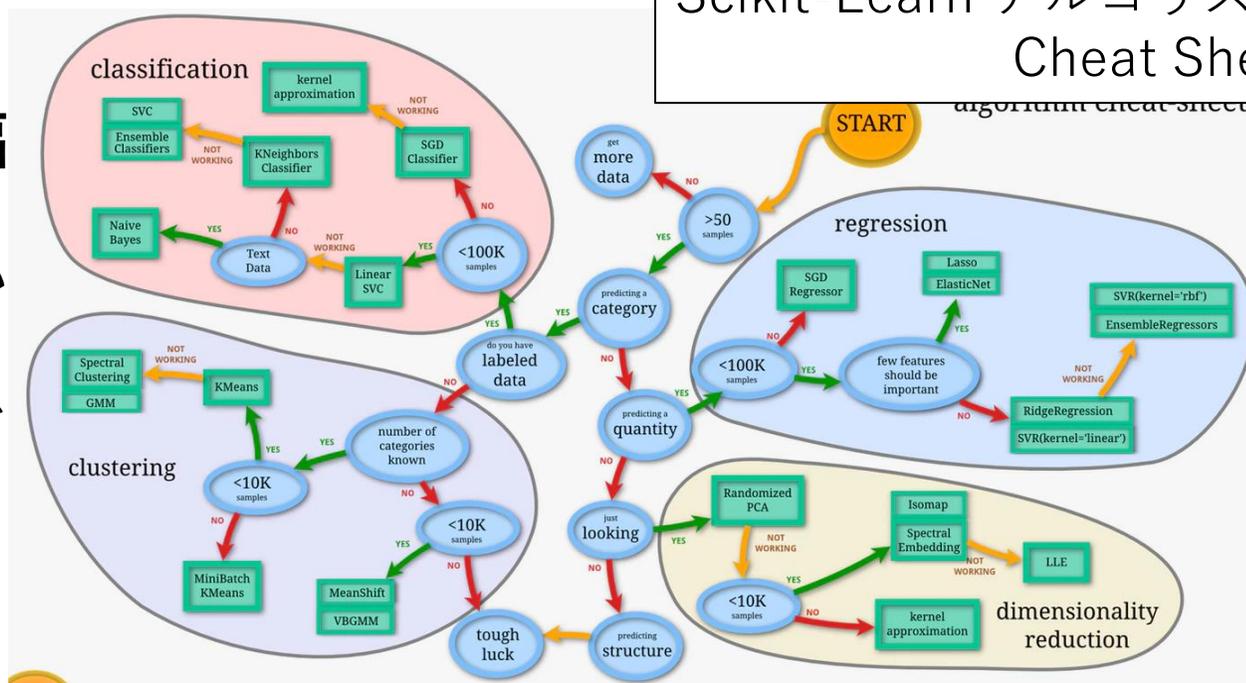
データ科学への貢献と
同研究者の参入促進

ユニットの意義

- 大規模データの背後にある構造を理解可能な形として可視化することは、**現象や対象に依存しない普遍的な学問**である。この手法を研究対象とする。

- データサイエンスの手法は非常に幅広く、全く同様の研究を行っている研究者は（特に核融合研内には）いない。しかしユニット内における、研究についての密接な議論・協力を通じて研究レベルの向上を狙うことができる。

Scikit-Learn アルゴリズム
Cheat Sheet



- データサイエンスを支える技術、データマネージメント、データフォーマットの共通化、認証、高速転送、可視化などはそれ自体、学術的研究である。プラットフォームと連携することで、**核融合研が蓄積してきた大規模なデータ**を研究対象にすることができる。

外部との連携

- **プラズマ・核融合学会（プラズマ・核融合学会）専門委員会**

- 「データサイエンスとの連携によるプラズマ・核融合データ利活用の推進」
- 横山（「データ駆動」研究に関するコミュニティ内外の「情報共有・発信の場」）・大舘（データ標準化・データアクセスの向上）で運営
- 国内の主要大学の研究者 + アドバイザーとして情報・システム研究機構 上野 玄太先生・吉田亮先生

- **画像計測研究会 ⇒ プラズマインフォマティクス研究会**

（九大・稲垣先生代表）

- 1995年より研究会を開いてきており、多くのプラズマを専門としないプラズマを専門としない研究者との議論も重ねてきた。
- QST石井さん、NII山中先生など客員教員の先生方との連携
- 核融合プラウドの提案（長壁さん・横山さん・渡邊さん・中西さん）