

TIAの新展開とデータ駆動科学

New Development of TIA and Data Driven Science

概要 outline

データ駆動科学とは、実験/計測/計算データの背後にある潜在的構造に関して、データが対象とする学問に依存しない普遍的な学問体系である。本ポスターでは、データ駆動科学の基本テクノロジーを紹介するとともに、幾つかの応用事例を紹介する。

Data-driven science is a general framework for the extraction of hidden structures embedded in data. In this poster, we introduce the basic technologies of data-driven science and these applications.

データ駆動科学：TIAの新展開との親和性

Data-Driven Science : affinity to TIA

研究開発(RD)

- AIの重要な研究領域(AI for Science)
- マテリアルズインフォマティクス(MI)
- 計測/計算/データ解析(量子ビームとスパコン)の融合

産官学協創

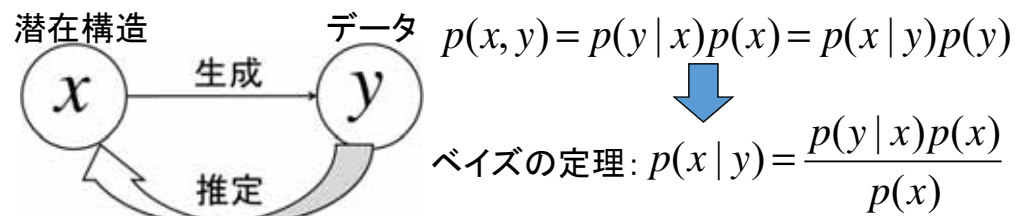
- 産官学の枠を超えた共通学理の提供
- ITによるRD組織のフラット化のロールモデル
- 基礎科学と産業応用をシームレスにつなぐ

教育人材育成

- 産業構造の変革に耐える人材の育成
- 研究開発の雇用のミスマッチ対策/人材再教育の戦略

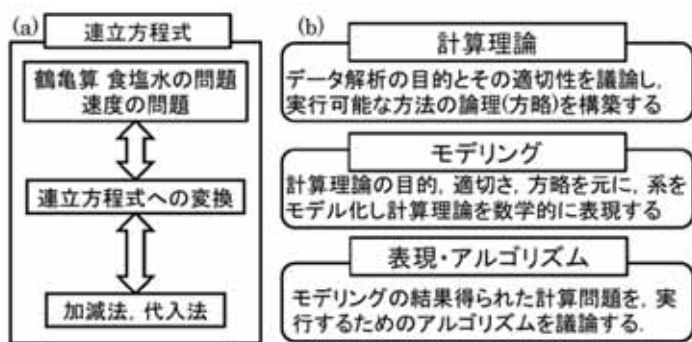
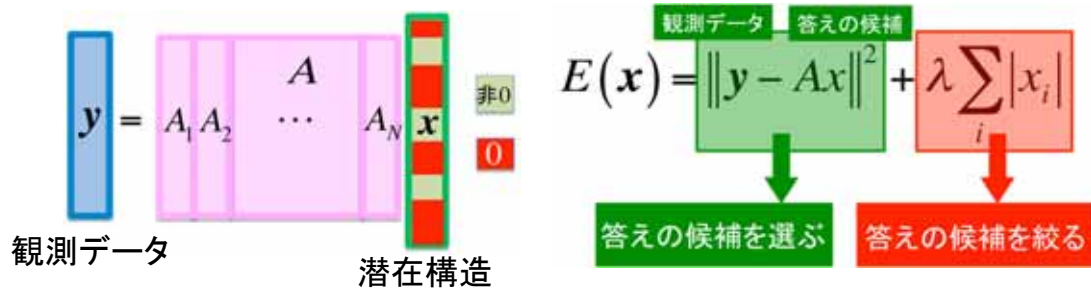
ベイズ推定

- データの生成プロセスを確率的に定式化
- ベイズの定理に基づき、因果律を遡ることで、データの背後に潜む構造を確率的に推定する枠組み



スパースモデリング

- 高次元データの説明変数が次元数よりも少ない(スパース(疎)である)と仮定し、
- 説明変数の個数なるべく小さくなることと、データへの適合とを同時に要請することにより、
- 人手に頼らない自動的な説明変数の選択を可能にする枠組み

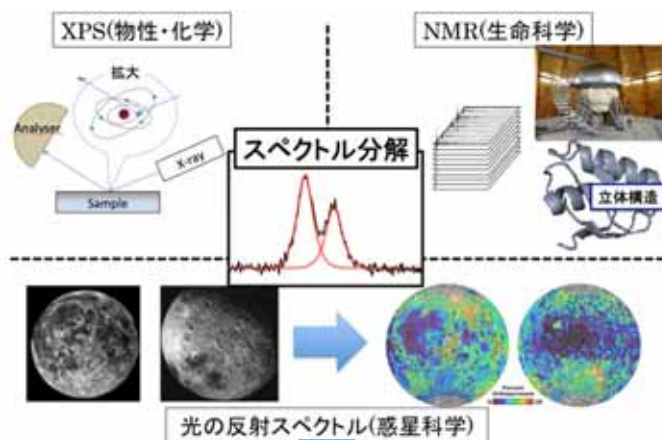


(a)連立方程式とその応用 (b)データ駆動科学の三つのレベル (Igarashi et al., 2016)

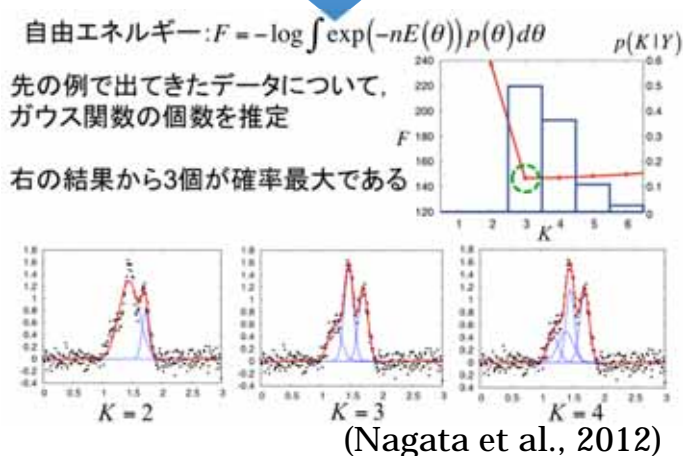
データ駆動科学の応用事例

Application of Data-driven science

ベイズ推定によるスペクトル分解

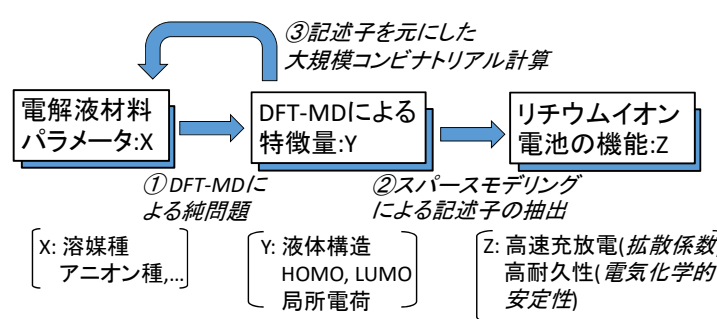


ベイズ推定によるモデル選択 交換モンテカルロ法による大域解の自動探索



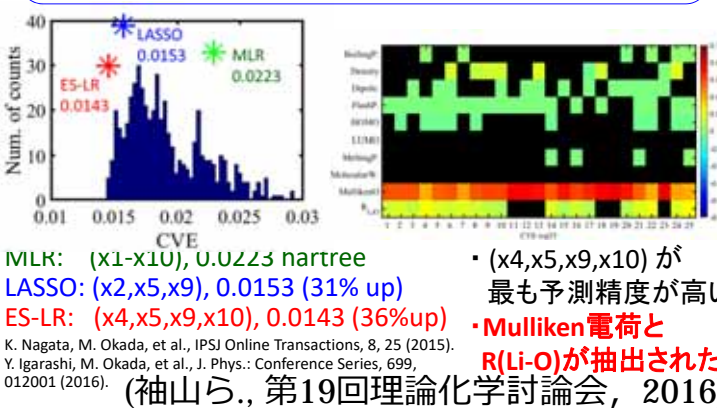
リチウムイオン電池電解液の新材料探索

電解液材料における3つのステップ



線形回帰による全状態探索 (ES-LR)

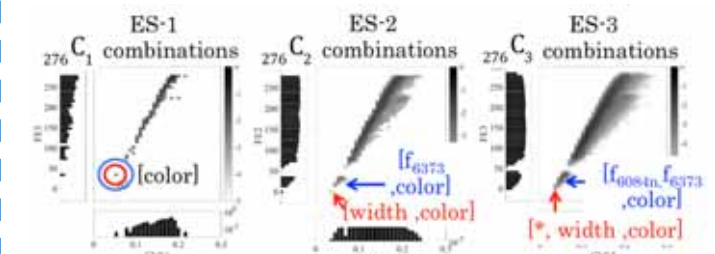
MLR (記述子Xの全てを使って線形回帰)
LASSO (刈り込んだXのみを使って線形回帰)
ES-LR ((x1), (x1,x2), ..., (x1-x10)の全ての組合せで線形回帰)(2^N-1通り) 最も高精度な組合せが分かる



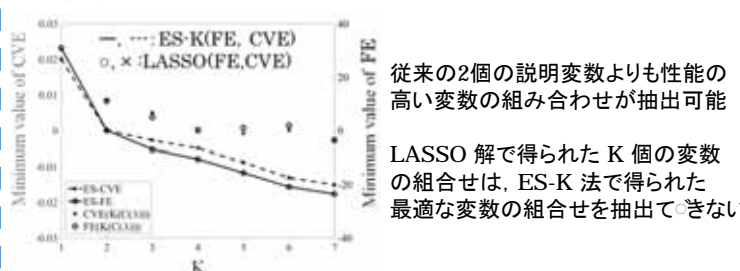
全状態探索による線形回帰のスパース変数選択

K-スパース全状態探索(ES-K)法
最適な説明変数の組合せが K スパースと仮定して、その説明変数の組合せ ($N C_K$ 通り, N=変数の個数)を網羅的に探索する枠組み。

実データ解析の結果(ES-1.2.3/LASSO):
交差誤差(CVE), ベイズ自由エネルギー(FE)の2次元状態密度



従来より説明変数として用いられている、widthとcolorをもつ解の組合せが、他の変数の組合せに比べてFE, CVEともに小さくなる傾向にあり、この2変数を中心にして上位のモデルがクラスターを形成されている



すべての近似アルゴリズムをインディケータ生成器として考え、既存及び今後の提案されるスパース変数選択の近似アルゴリズムの性能を、ES-K法がもつめる状態密度の上で評価する。

スパース変数選択の系統的プロトコル
(五十嵐ら., 信学技報, 2016)