

## スパースモデリングとデータ駆動科学

### Sparse modeling and data driven science

東大新領域 岡田 真人

Univ. of Tokyo Masato Okada

E-mail: okada@edu.k.u-tokyo.ac.jp

本講演では我々のグループの研究を具体例として、今年度 H29 年度に終了する文科省科研費新学術領域「スパースモデリングの深化と高次元データ駆動科学の創成(略称:疎性モデリング)」を紹介する[1]. データ駆動科学とはデータ科学の一分野であり、計測データや数値計算データの背後にある潜在的構造の抽出に関して、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. 同じアルゴリズムがスケールや対象を超えて、有用であることが多いという経験的事実を背景として、その理由を問い、背後にある普遍性から、データ解析自体を対にする学問的枠組みである[2]. 本領域は先端的データ解析手法の活用による科学的方法論の革新を目的としたものであり、各計画研究においては、データ解析による限界打破を目指さない実験・観測研究、あるいは、実験・観測研究の目的を忘れた手法開発に陥らないように領域を運営した. その結果、David Marr が指摘した三つのレベルが、データ駆動科学を創成するためにも重要な着眼点となることを確信した. データ駆動科学においては、データ解析の目的である計算理論のレベルと、データ解析手法である表現・アルゴリズムのレベルの間に、モデリングのレベルが存在することに気づいた. これらの三つのレベルを新たに“データ駆動科学の三つのレベル”と名づけ、データ駆動科学の学理の原点に位置付けた[2].

データ駆動科学の情報数理基盤は、ベイズ推論とスパースモデリングである. ベイズ推論は、データの生成過程である因果律を確率モデルで記述し、ベイズの定理を用いて、因果の逆をたどる数学的枠組みである. 量子力学と電磁気学からの第一原理的なアプローチが可能な物理学や化学などの理科 I 分野では、ベイズ推論を用いるのがもっとも系統的な枠組みである. データ駆動科学におけるベイズ推論の一例として、ベイズのスペクトル分解を紹介する[3]. 一方、生物学や地学などの理科 II 分野では、かならずしも第一原理的なアプローチが可能ではない場合は多く存在する. この場合に威力を発揮するのがスパースモデリング(SpM)である. 国内では、SpM の近似の一手法にすぎない L1 最適化が主流であるが、世界レベルでは、説明変数の部分集合全体を探索する全状態探索アルゴリズムの動向も無視できない状況である. データ駆動的視点から、我々は説明変数の部分集合全体を記述する方法論の必然性を痛感し、ES-DoS(Exhaustive Search with Density of State)法を提案した[4].

[1] <http://sparse-modeling.jp/> [2] Y. Igarashi, K. Nagata, T. Kuwatani, T. Omor, Y. Nakanishi-Ohno and M. Okada: J. Phys. Conf. Series, 699, pp.012001-1–012001-13, 2016 [3] K. Nagata, S. Sugita and M. Okada: Neural Networks, 28, pp.82-89, 2012. [4] Y. Igarashi, H. Ichikawa, Y. Nakanishi-Ohno, H. Takenaka, D. Kawabata, S. Eifuku, R. Tamura, K. Nagata and M. Okada: J. Phys. Conf. Series, in press.