

区分線形回帰モデルを用いた非線形データの解析

Piecewise linear Regression Analysis of Nonlinear Data

NEC システムプラットフォーム研¹, JST さきがけ² ○石田 真彦¹, 岩崎 悠真^{1,2}

NEC Corp.¹, JST-PRESTO², ◯Masahiko Ishida¹, Yuma Iwasaki^{1,2}

E-mail: ishida@ah.jp.nec.com

新しい学術分野では、基本的な知見の蓄積が十分で無い状況で、材料の組成などデータ取得の条件を変えてかなければならない場合がある。実験の結果には、全く予期していない要因が隠れている可能性が高まるため、演繹アプローチを機能させることは容易ではない。そのような状況では、むしろデータ解析が主導する帰納的アプローチによって、意味のある情報の蓄積が効率よく進む場合があると考えている。

データ解析が主導する形で、課題に取り組むことができるかどうかは、得られるデータの種類、量、そして質に大きく依存する。そこで、我々はスピンの関与して生じる熱電変換材料¹⁾の探索にコンビナトリアル型の実験技術を用いて開発の効率化に取り組んできた²⁻⁵⁾。しかし扱うデータの量が増え、データ処理・解析を人の手で行うことが難しくなってくる。この点に関しては、機械学習やAIを活用した効率化を実現し、新しい材料の発見につなげることが出来始めている。

本講演では、我々が行っている解析技術として、多次元のデータ空間をいくつかに分けし、各区分における回帰を同時に最適化する、区分線形回帰と呼ばれるアルゴリズムを紹介する。データ空間全体を等価に扱う決定木や線形回帰とは異なり、部分的に表れているデータの相関を適切に抽出できる点が大きな特徴である。さらに、区分けも回帰もできるだけシンプルに表現されるよう、L0 正則化と呼ばれる非常に強い制限要素を課して最適化を行うことが出来る。後者は、目的変数の量に対して説明変数の量が極端に多いような場合、いわゆるスパースデータと呼ばれるデータ群から有意な相関を抽出する上で大きな利点となる。

本研究の一部は JST ERATO 齊藤スピン量子整流プロジェクト、ならびに JST さきがけ マテリアルズインフォの支援を受けて行われました。

1) K. Uchida et. al., “Thermoelectric generation based on spin Seebeck effects” Proceedings of the IEEE 104, (2016) 1946.

2) 石田真彦, “スピン流を用いた熱電変換素子開発の動向” 応用物理 87, (2018) 11.

3) Y. Iwasaki et. al., “Comparison of dissimilarity measures for cluster analysis of X-ray diffraction data from combinatorial libraries” npj Computational Mater. 3, (2017) 4.

4) Y. Iwasaki et. al., “Machine-learning guided discovery of a new thermoelectric material”, Sci. Rep. (2019).

5) R. Sawada et. al., “Boosting Material Modeling Using Game Tree Search”, Phys. Rev. Mater. 2, (2019) 103862.