

LASSO の適用事例：材料物性の予測及び関係構造

DAM Hieu Chi

国立大学法人 北陸先端科学技術大学院大学
〒923-1211 石川県能美市旭台 1-1

E-mail: dam@jaist.ac.jp

キーワード：マテリアルズ・インフォマティクス，機械学習

マテリアルサイエンスの重要な課題の一つは物質のデータから効率的に学習することである。そのためには、各物質のデータを識別できるように物質を記述する属性・記述子を適切に設計しないといけないだけでなく、その属性・記述子を基にした物質同士の定量的に比較する方式も提案する必要がある。物質の構造を記述するには全原子の座標を記述すればすべての情報が含まれるが、物性値まで記述に入れようとすると、膨大な属性・記述子が必要になり、物質空間を表現するためには非常な高次元のデータが必要となる。さらに、これらの属性・記述子を基にすると、研究者の直感的行為である思索や試行錯誤が追いつかなくなる。

一方、物質は一般的な物理法則に従うだけでなく、それぞれの物質群はそれぞれ共通な特色を持ち、共通な法則に従う。それらの共通な特色と法則はマテリアルサイエンティストが探すものである。ここで、特色や法則の存在は属性・記述子同士がお互いに完全に独立でないことを意味する。数学的なイメージでは高次元空間における各物質のデータが低次元超局面に乗っていることになる。そのため、物質の高次元データを次元削減できることが暗黙的に法則を見つけることと等価である。

近年、高次元データから現象の本質を低次元で抽出するスパースモデリングという技術がトレンドとなっている。そのなか、Tibshirani らが提案した変数選択アプローチの LASSO 法が有名である。LASSO 法は l_1 ノルム [1] の正則化を用いた回帰解析手法で、スパース性の学習・重要な記述子の識別選択という問題への貢献度が大きい。LASSO 法を用いた各記述子の網羅的な関係性を構築することによって、データの背後にある本質的な構造を抽出ことができ、それを用いて因果関係を議論することに近づくことができ、マテリアルサイエンスにおいてはデータ駆動型の材料設計への道が開かれる。本講演ではこの観点で LASSO 法の適用事例を紹介したい。

[1] R. Tibshirani, Journal of the Royal Statistical Society, Series B Vol. 58, Issue 1 267-288 (1996)