

## 回帰モデルにおける記述子の選択理由

物性特徴量の回帰予測モデルにおいて記述子選択は予測性能を決める重要な問題である。本論文では記述子間距離を利用した階層的クラスタリングでデータから同種・異種グループを定義し、樹形図による表示で同種グループ内の記述子の重要性を示す手法を開発した。これにより最良回帰予測モデルの記述子選択理由を解釈することが可能になった。また、最良記述子組み合わせだけでなく次善記述子組み合わせも系統的に行えるようになった。

従来の理論駆動による物性物理は膨大な研究の積み重ねがあり強力な理論体系である。しかし、原理的に方程式は作れても計算量が多すぎて適用に問題がある物理量も多い。例えば、キュリー温度は自動車用モーターにも使われる磁石の重要な特性の一つであるが理論駆動による物性量予測は未だに実用的ではないためデータ駆動により予測値を補完することが重要である。このような理由により高速で有用なデータからの予測手法を物性物理に適用した研究が近年増えてきている。

理論的な物理研究においてはモデルをつくりそのモデルに従い物性値を計算するためモデルそれ自体が物理的意味を与える。(モデルの解法で特定の相互作用のみを強調して物理的意味を考えることもある。)機械学習の場合は記述子組み合わせがモデルに相当する。最良の記述子組み合わせの回帰モデルを得る最も正確な手法が(記述子組み合わせ)全探索であるが、先駆者である岡田真人教授(東大新学術領域)らのグループにおける近年の一連の全探索研究を参照すると、最良性能回帰モデルとほぼ同じ回帰性能を示す回帰モデルは一般的には多数存在することが分かる。その為全探索それ自体からは記述子の選択指針や重要性は明らかにはならない。一方、記述子重要性を評価できる回帰手法としては、例えば、random forest 回帰があるが、未知物質探索など予測のための回帰モデル作成では「次元の呪い」を避けるために回帰性能を落とさない範囲内でより少ない記述子選択によるスパースモデリングの枠内で重要性評価を行うことが重要である。

最近、物質材料研究機構、情報統合型物質・材料研究拠点のメンバーを中心とする研究グループは、記述子間距離を用いた階層クラスタリング手法によりデータから同種・異種記述子を定義した上で全探索回帰モデルリストに対して関連性解析を行うことで同種記述子グループ内の重要性を系統的に示す手法(subgroup relevance analysis)を開発した。この成果は、日本物理学会が発行する英文誌 Journal of the Physical Society of Japan (JPSJ)の 2018 年 11 月号に掲載された。

図 1 は遷移金属希土類二元金属のキュリー温度に対して subgroup relevance analysis を適用した結果であり、縦数字は group R,T,S 内の記述子 subgroup の重要性を示している。これによると group R(機械学習によりクラスタリングされた理論駆動モデルにおける希土類金属相当グループ)の最重要記述子は  $Z_R$  であり group S(構造相当)の最重要記述子は  $C_R$  であることが分かる。group T(遷移金属相当)に関しては最重要記述子は一つではなく  $J_T-r_T$  までの複数記述子から成る subgroup である。この知見から、重要性とは別に評価される最良回帰モデルのスパースな記述子選択が説明可能であることは興味深い。また、記述子間距離から作成した樹形図に表示していることから、最良組み合わせだけでなく次善組み合わせに対しても系統的に記述子選択を行えるようしており有用性も高い。

ある記述子組み合わせのカーネル回帰モデルはデータ間距離を用いて作成される。それに対して、本研究で記述子組み合わせの異なる回帰モデル集合が記述子間距離で整理できた点もまた興味深

い。物性物理では元素組み合わせとそれらの原子配置という異種構成要素からなる物質の特徴量を扱う事が多い為に手法の有効範囲は広いと思われる。また、大域探索手法としての発展も考えられるため、今後の展開を期待したい。

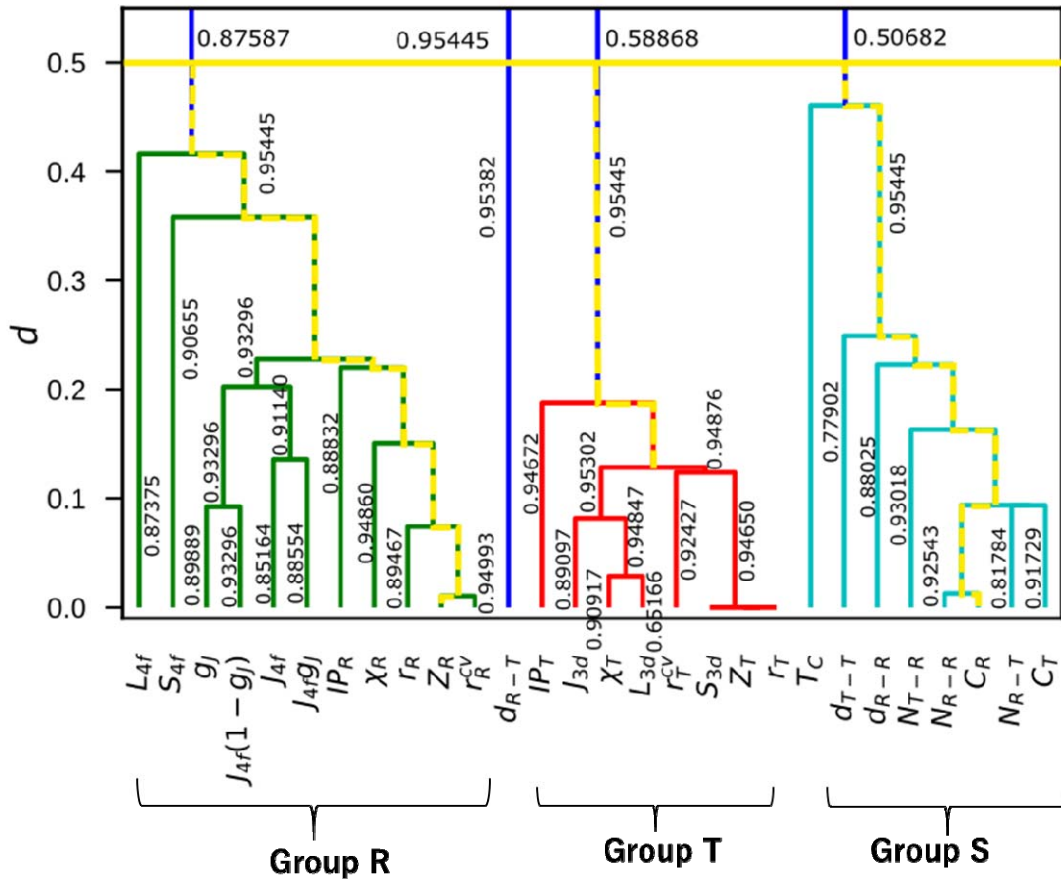


図 1：遷移金属希土類二元金属のキュリー温度のカーネルモデル全探索結果を樹形図による部分グループ関連性解析に適用した例。縦軸は記述子間距離( $d$ )、横軸は記述子、黄色点線が最重要記述子を示す。

### 原論文

[Important Descriptors and Descriptor Groups of Curie Temperatures of Rare-earth Transition-metal Binary Alloys](#)

Hieu Chi Dam, Viet Cuong Nguyen, Tien Lam Pham, Anh Tuan Nguyen, Kiyoyuki Terakura, Takashi Miyake, Hiori Kino, *J. Phys. Soc. Jpn.* **87**, 113801 (2018).

問合せ先：木野日織（物質・材料研究機構）