

3P107

L1 型正則化線形回帰モデルを用いたシミュレーションデータからの二元合金の物性予測

○鈴木 大輔¹, 川崎 隆史¹, 杉山 歩¹, 水上 卓², Dam Hieu Chi¹, Ho Tu Bao¹
(北陸先端大・知識科学¹, 北陸先端大・マテリアル²)

Prediction of binary alloy property from the simulation data by using L1 type regularization
linear regression model

Daisuke Suzuki¹, Takahumi Kawasaki¹, Ayumu Sugiyama¹, Taku Mizukami², Dam Hieu Chi¹, Ho Tu Bao¹
(JAIST Knowledge Sci. ¹, JAIST Materials Sci. ²)

【序】

情報科学技術の進展とともに発展してきた計算科学は幅広い分野で成果を上げ、量子化学計算による材料設計、分子動力学シミュレーションによるタンパク質や高分子ポリマーの解析など産業面においても必要不可欠なものとなっている。

計算機シミュレーションは実際の物質を用いた実験と比べ、多くのサンプルを多彩な条件下で簡便に実行することが出来るとの期待から時間と設備の節約になるものと考えられてきたが、近年の大規模シミュレーションでは数万コアのリソースを必要とされ、金銭的負担は看過出来ない程となっている。

一方、目的の物性と関連のあるパラメータの探索を試み、その結果を用いて予測を行う手法の研究もされており、これらの研究はいわゆる逆問題的設計法として期待が高まっている。Yousefら[1]は CaAg, BaPb といった二元合金の融点を、構成元素の価電子数、原子番号、電気陰性度、沸点など計 16 個の基本物性のみを用いたデータマイニングにより予測し、また、予測結果に寄与する物性を示している。このような研究は、帰納的知識発見法である機械学習やデータマイニング手法の発展と共に盛んになってきているが、まだ実用的な手法として確立されたものはない。

本研究では、基礎物性データと第一原理シミュレーションから得られる物性データから成る物性データベースを構築し、本データベースから重回帰分析により物性間の相関グラフ構造を明示化する方法を確立した。本手法の概要及び実験結果を以下の通り報告する。

【方法】

本研究は以下の手順で行う。

1. 第一原理計算による合金物性データベース構築
2. 予測に使用する物性特徴空間設計
3. 重回帰分析による物性値（融点）予測モデル構築
4. 並列した予測モデル構築からの全物性間の関係性を行列化及びグラフ化

本研究で行った全ての物性計算は、密度汎関数論 (DFT) に基づいた第一原理計算であり、Dmol3 によって実行した。構造最適化及びパラメータ決定は密度勾配近似 (GGA)、汎関数に PBE、基底関数に DNP を用いて計算した。尚、内殻電子処理は相対性全電子とした。

本計算によって求められた二元合金の結合距離、Mulliken 電荷など、19 種の物性値と二元合

金の構成元素の基礎物性データから物性データベースを作成し、融点予測に使用する特徴空間とした。

予測モデルとして様々な既存のモデルがあるが、本研究では予測対象とパラメータ間の関係性の解釈が容易であり、グラフ化にも適用しやすいという理由から線形回帰モデルによる予測を行う。本研究では下に示す代表的 L1 型正則化線形回帰手法である Least absolute shrinkage and selection operator (Lasso) を採用した。

$$\text{Lasso} : \min \left((y - X\beta)^T (y - X\beta) + \gamma \sum_{j=1}^p |\beta_j| \right)$$

y : 実際値 X : パラメータ β : 回帰係数

p : パラメータ数 γ : チューニングパラメータ

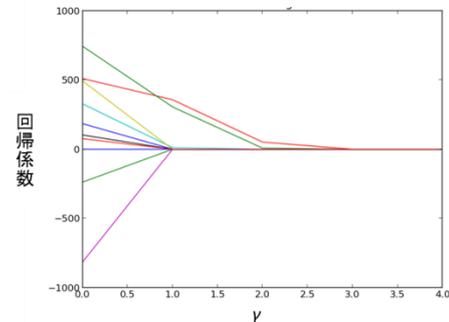


図1. Lassoの正則化効果

Lasso は正則化効果により、図1の通り、チューニングパラメータ γ の値が決定すると一意的にパラメータが決定されモデルが求まり、また、 γ の値が大きいほどモデルのパラメータ数が減少する特徴を持つ。従って、Lasso を用いることによりスパースな線形モデルを得る事が可能となる。また、汎用性のある予測モデルを得るために、データをモデル構築用とモデル評価用に分割し適用する交差検定を行った。この Lasso と交差検定の融合手法を用いて、順次 γ の値を変えてモデルを構築し、予測誤差が最小となるモデルを最良の予測モデルと決定した。

本手法による物性-パラメータ間の関係性の予測モデル構築を、予測対象物性毎を入れ替え並列に実行することにより、二元合金の全物性間の関係性を全て行列化した。また、この関係性行列を以って二元合金物性の関係のグラフィカルモデルを作成した。

【結果】

結果は図2の通りとなった。プロットが対角に近い程、融点の実際値と予測値の差が小さいことを意味する。

本手法を用いて二元合金の融点予測を行ったところ、平均二乗誤差 101 K での予測を行うことが出来た。従って、本手法を用いることにより、予測対象物性-パラメータ間の関係性を高い精度で表現することが可能であると言える。

融点予測の結果の詳細及び二元合金物性間の関係のグラフ化の結果については当日報告する。

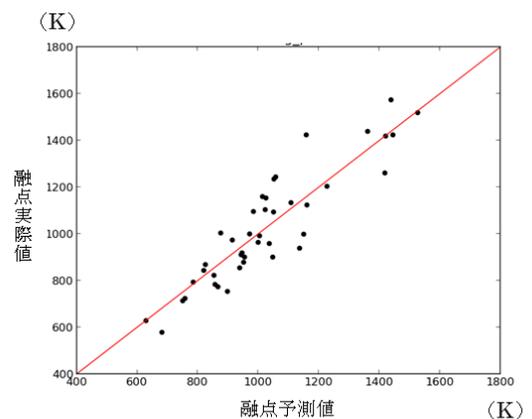


図2.二元合金融点の予測値と実際値

【参考】

- [1] Y.Saad, D.Gao, S.Bobbitt, J.R.Chelikowsky, and W.Andreoni, Phy. Rev. B 85 104104 (2012)
- [2] R. Tibshirani, J. R. Statist. Soc. B 58, 267 (1996)
- [3] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, Annals of Statistics 32, 409 (2004)
- [4] N. Meinshausen and P. Buhlmann, Ann. Statist. 34 (2006)