# 半導体シリサイドのバンドギャップエネルギーのベイズ回帰分析

## Bayesian-Regression Analysis of Band-Gap Energies of Semiconducting Silicides

九工大院情報工　○前田 佳均

Graduate School of Information Eng., Kyutech, ○Yoshihito Maeda

E-mail: maeda@cse.kyutech.ac.jp

【はじめに】マテリアル・インフォマティックス（MI）は，測定や計算で得られた多数の物性・特性データから機械学習用の統計モデルを構築し，未知の物質・材料を効率よく探索する新たなデータ駆動型材料工学である [1]．われわれは，16 種類の半導体シリサイドのバンドギャップエネルギー（$E_g$）のデータから， Elastic Net（EN）正則化回帰と一般化線形回帰（GLM）によって，電気陰性度，原子番号，融点，擬ポテンシャル半径，元素グループ番号を記述子とした線形回帰モデルを構築した[1]．本研究では，ベイズ統計による線形回帰を用いて，半導体シリサイドの $E_g$ についてベイズ線形回帰モデル(Bayesian linear regression model: BLM)を検討した．

【解析方法】ベイズ線形回帰には CRAN ホームページ[2]から library(MCMCpack)を入手して，データ解析に用いた．目的変数 $E_g$ の記述子には，データが完備している格子定数($a$, $b$, $c$)，平均電子密度($N_e$)，最近接原子間距離($d$)を用いた．

【結　果】記述子の縮約のために EN 正則化回帰[1]を行った．最適値 $\alpha$ =0.2 のときに記述子は，{$a$(回帰係数:0.162), $b$(0.129), $N_e$(−0.018)}に縮約できた．次にマルコフ連鎖モンテカルロ法（MCMC）によるランダムサンプリング（メトロポリス法）[3]を行い，収束判定した後， Fig. 1 に示した $a$ と $b$ の定常分布を得た．ベイズの原理から，これらが記述子 $a$ と $b$ の最尤分布に一致する．23 個のモデルをベイズ因子の大小で判定した結果, EN 正則化回帰で縮約した記述子セット{$a$, $b$, $N_e$}を用いた線形回帰（交差項も含めて）が最尤の BLM であることを明らかにした．Fig. 2 に BLM と GLM の回帰結果を比較した. BLM では回帰偏差が小さく, GLM よりも当てはまりがかなり良いことが分かる．

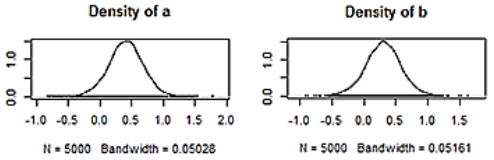


Fig.1 Most likelihood distributions of descriptors {$a$, $b$} obtained from MCMC sampling (N: iterant number for random) sampling).

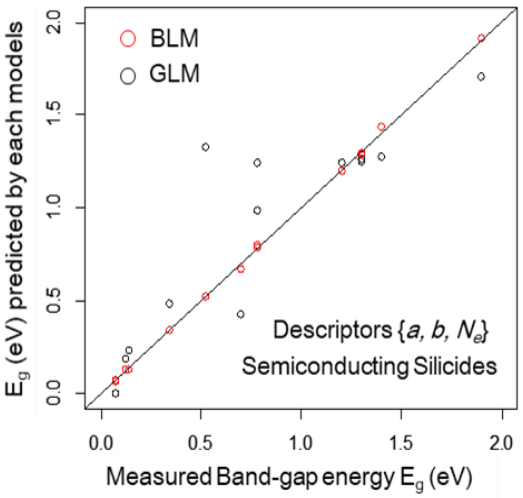


Fig.2 Results of linear regression using BLM and GLM for $E_g$ of 16 semiconducting silicides.

【参考文献】

[1] 淵 雅也，有馬幹尋，前田佳均：2017 春季応物講演会 16p-B5-12.

[2] CRAN Homepage, https://cran.r-project.org/web/packages/MCMCpack/index.html

[3] 久保拓弥：「データ解析のための統計モデリング入門」第 8 章，2012，岩波書店.