

機械学習による 3元系酸化物半導体の光電流応答に関する解析

小西由也・佐山和弘
産業技術総合研究所 太陽光発電研究センター 機能性材料チーム

研究の目的

- 人工光合成技術(半導体光触媒・半導体光電極)
- 水素製造
 - 有用化学品製造
- 環境浄化用の可視光応答型光触媒
色素増感太陽電池・ペロブスカイト型太陽電池
- 材料としての
3元系酸化物
半導体の探索
- 材料開発のための探索
- 膨大な候補を探索
 - 探索に長時間
 - 多量のデータ
 - 実験のコスト大
- 自動合成高速スクリーニング装置による探索
(High Throughput Screening 実験)
- 得られるデータの機械学習による解析で
さらに探索を促進できないか?

実験



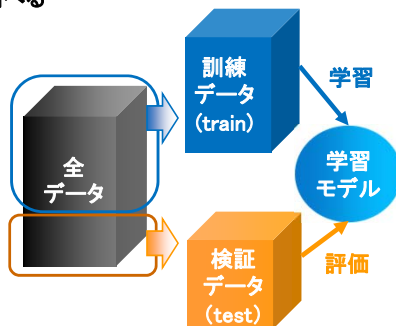
半導体膜ライブラリの自動合成装置

- 有機金属熱分解法(MOD法)により含有する元素の種類・比率が異なる半導体光電極薄膜ライブラリを自動塗布でFe系・Cu系を中心に合成
 - 半導体評価システムによって光電流応答を自動評価
- 光電流応答がある組成を探索
同じ条件で得られた実験データ
⇒ 機械学習での解析に適する

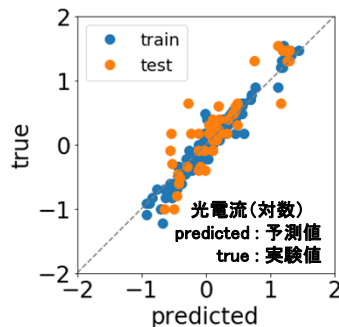
機械学習による解析

得られたデータセットに対して機械学習により規則性(パターン)を調べる
→ 光電流応答のあったもの 前処理後の226データ

- 30種類の元素(Al, B, Ba, Bi, Ca, Ce, Co, Cr, Cs, Cu, Dy, Fe, In, La, Mg, Mn, Nb, Ni, Pb, Sb, Si, Sm, Sn, Sr, Ta, Ti, V, Y, Zn, Zr)の中から1~3種類を含む酸化物半導体薄膜ライブラリ
- 原料の元素含有比率と半導体の型(p型・n型)を説明変数(試料薄膜の構造の情報ではなく原料における元素比率)
- 光電流値の対数を目的変数 -1.22 (0.06 μ A) ~ 1.54 (35 μ A)
- PythonによるScikit-learnライブラリを使用
- 学習モデルはRandom Forest (回帰木のアンサンブルモデル)
- ホールドアウト法でデータを分割して学習モデルを評価
- 5分割交差検証法によるGrid SearchでHyperparameterを決定
 - 回帰木の深さ: 14
 - 回帰木の数: 20



学習モデルの評価方法
「未知のデータをどの程度まで予測するか」(汎化性能)



Random Forest による
学習モデルの評価

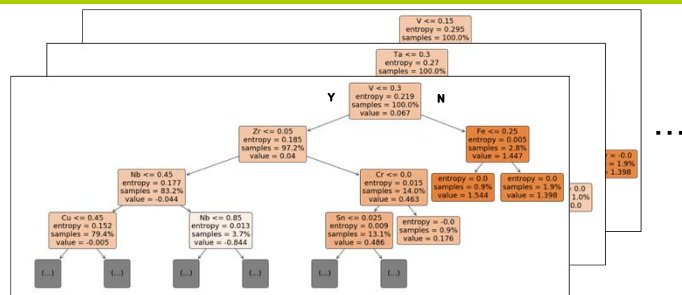
考察

- 訓練データから学習モデルを構築
→ 決定係数(R^2 値): 0.93
- 検証データの傾向を再現
→ 決定係数(R^2 値): 0.77

説明変数	重要度
Cuの比率	0.252
Vの比率	0.144
n型・p型	0.110
Zrの比率	0.101
Biの比率	0.071
Feの比率	0.070

光電流応答のある元素の組み合わせには規則性がある。

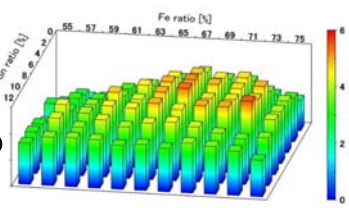
学習モデルにおける説明変数の重要度からは光電流応答へのCuの比率とVの比率が寄与が大きく、Feの比率の寄与はあまり大きくない。



回帰木の例(Random Forestではこれらを平均)

今後の展開

- 学習結果の解釈(既知の情報?)
- 未実験の組み合わせの予測
- 組成比の詳細な実験データへの適用
(→ 光電流マッピングの解析)
- 各元素の特徴量を説明変数に加える
(価電子数・電気陰性度・イオン半径...)
- 他の機械学習モデルとの比較
- 深層学習(ニューラルネットワーク)
- 実験計画への寄与(ベイズ最適化)



Fe-Sn-Zr-O系の光電流マッピング

結論と参考文献

結論

- 検証データの傾向を再現したことから光電流応答のあった元素の組み合わせには規則性があることが分かった。
- Cuの比率とVの比率が光電流応答には重要であることが分かった。それに対してFeの比率の寄与はあまり大きくなかった。

参考文献

- Takeo Arai, Yoshinari Konishi, Yasukazu Iwasaki, Hideki Sugihara, Kazuhiro Sayama, J. Comb. Chem. 9, 574-581 (2007).
- Leo Breiman, Mach. Learn. 45, 5-32 (2001).