

光電子収量分光 (PYS) の閾値の自動判定

Determination of threshold of photoelectron yield spectroscopy (PYS)

物材機構¹ ○柳生進二郎¹・吉武道子¹・知京 豊裕¹

NIMS¹, °Shinjiro Yagyū¹, Michiko Yoshitake¹, Toyohiro Chkyow¹

E-mail: YAGYU.Shinjiro@nims.go.jp

1. はじめに

有機半導体や電極材料開発において、電子・ホールが移動する価電子帯や HOMO レベルが重要である。この値は、光電子分光収量分光法 (PYS) の光電子放出閾値から求められる。この閾値の一般的な解析方法は、Fowler の光電効果の理論をもとに、PYS 強度の平方根をとり、その傾きとバックグラウンドとの交点から閾値を求める。現状では、この解析方法を自動化したものではなく、解析者が一つ一つのスペクトルを見て、閾値を求めている。そのために、解析者の経験などから値に違いが生じることがあるとともに、今後の Big data 化のためには、解析の自動化が求められる。従来からある Fowler の式による最小二乗回帰 (Fitting) と Fowler の式よりパラメータを変えて計算したデータによる機械学習の 2 つの方法で、実際の測定データ (Au) を用いて閾値の自動解析の検討を行った。

2. データセット

教師あり機械学習には、千のオーダー以上のデータセット (特微量と教師データ) が必要である。そのような大量の実験データセットを持っていないため、PYS の解析式を用い、そのパラメータ (閾値 (教師データ)、温度、規格化、バックグラウンドの 4 つ) を変え、ノイズ (ポアソンノイズ、ガウシアンノイズ)、を重畳したスペクトルを特微量データとして作成した。なお、特微量は 45 個 (4.0eV から 6.2eV まで 0.05eV 刻みの強度の値)、教師データは 1 個 (閾値) のデータセットを 6300 作成した。

3. 結果・検討

作成した計算データを用いて機械学習の識別結果を表に示す。決定木を用いたアンサンブル識別機である (Gradient boosting) では、98% 以上の高い識別性能を示した。Fitting 及び機械学習による残差プロット (予測結果と教師データとの差) から、機械学習では、初期値の推定が不要な点、ノイズ耐性の点で Fitting より優れていると考えられる。実際の Au のデータについて熟練した解析者が解析した結果に対して、機械学習では近い値を予測したのに対し、Fitting では大きな信号成分の影響を受け高い値を予測した。今後、自動解析の精度や信頼性を高めていくためには、より多くの実験データを用いて機械学習の検証を行い、モデルの改良などを行う必要がある。

Model	Accuracy (Test Data)	Prediction [eV] (Au(exp.) Data)	Residual [eV] Au(exp.): 4.47eV
Linear Regression	0.713	3.97	0.49
SVM (kernel=rbf)	0.981	4.5	-0.03
Decision Tree	0.944	4.5	-0.03
Gradient Boosting	0.989	4.5	-0.03
Fitting		4.97	-0.5

表 モデルによる識別率 (Accuracy) と実際のデータと予測の差