

機械学習による高分子フラーレン太陽電池の分子設計

Molecular design of polymer-fullerene solar cells based on machine learning

阪大院工¹, JST さきがけ² ○長澤慎司¹, 佐伯昭紀^{1,2}Osaka Univ.¹, JST-PRESTO², Shinji Nagasawa¹, Akinori Saeki^{1,2}

E-mail: saeki@chem.eng.osaka-u.ac.jp

近年の急速な人工知能 (AI) の発展と共に、材料開発におけるマテリアルズ・インフォマティクス (MI) が注目を集めている。一方で、有機薄膜太陽電池の高分子の分子設計は、未だ研究者の勘・経験・こだわりを軸に、量子化学計算 (主に DFT) を参考にして構造を絞る事が主流である。この状況を打破するため、本研究では有機薄膜太陽電池 (OPV) 評価と MI の概念を融合させ、材料の化学構造と変換効率の関係を見出し、ひいては高効率材料を提案することを目指した^[1]。データセットとして、約 1000 個の高分子フラーレン太陽電池の素子性能と物性値 (J_{sc} , V_{oc} , FF, PCE, HOMO、重量平均分子量: M_w 、バンドギャップ: E_g) を既報論文から抽出し、さらに高分子の化学構造は MACCS key (分子構造の指紋キー: 1~166 の値が与えられる)、PubChem key (881 ビット)、あるいは ECFP6 key (1064 ビット) に変換した。機械学習のアルゴリズムとして、人工ニューラルネットワーク (ANN) 法、ランダムフォレスト (RF) 法を用いた。ともに、入力には M_w , E_g , HOMO、および MACCS key などの Fingerprint、目的変数は変換効率 PCE とした。

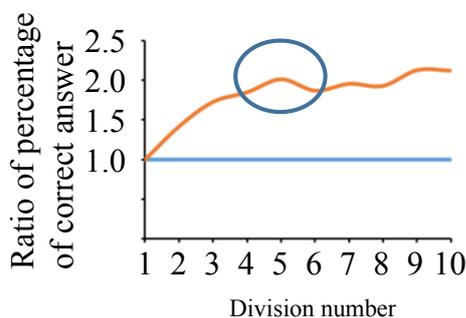


Figure 1. Percentage of correct answer as a function of the number of group division.

Table 1. Results of test data using random forest computing. The row and column are the input and output group (A: highest PCE – D: lowest PCE), respectively.

		Output			
		A	B	C	D
Input	A	10	2	3	0
	B	4	6	4	1
	C	3	4	7	3
	D	1	4	3	7

ANN 法により予測した出力データは入力データに対して最大で約 0.35 という低い相関係数を示し、変換効率を直接予測するモデルを作成する事は出来なかった。一方で RF 法では ECFP6 Key を用いることで、最大 0.62 の相関係数を示す事が出来た。しかし、変換効率を予測するには不十分であったため、変換効率に応じてグループ分けを行い、性能予測を検討した。例えば、4 分割では変換効率が高い順に A, B, C, D としてラベル付けを行い学習させた。結果的に 4 分割では、約 50% の確率 (62 個中 30 個) でラベルを正解でき、単純確率の 25% (=1/4) の 2 倍の正答率となった (Table 1)。同様に 2~10 分割で RF を構築したところ、いかなる分割数でも単純分割を上回ることが出来た (Figure 1)。単純な MACCS key では、高分子骨格の選別は可能であったが、アルキル鎖の違いまでは判別できなかった。しかし、ECFP6 Key で構築した RF モデルでは、最適なアルキル鎖の提案も可能である。

[1] S. Nagasawa, A. Saeki, *et al.* Manuscript in preparation.