

Annual Meeting of the Japan Society of Vacuum and Surface Science 2021

有機物・高分子材料 TOF-SIMS スペクトルの 機械学習による予測システム開発

○家持 圭佑¹, 長谷川 慶², 青柳 里果^{1,2*}¹成蹊大学大学院理工学研究科, ²成蹊大学理工学部

Development of prediction system on TOF-SIMS spectra of organic and polymer materials by machine learning

○Keisuke Kamochi¹, Kei Hasegawa¹, Satoka Aoyagi^{1*}¹Faculty of Science and Technology, Seikei University, ²Materials and Life Science, Seikei University

1. はじめに

飛行時間形二次イオン質量分析 (TOF-SIMS) は、高い分解能で2次元および3次元分子分布が得られることから、様々分野への応用が期待されているが、二次イオンの発生機構およびフラグメント化が複雑であることから、スペクトルの解釈が難しいという課題がある。多くの分光法や質量分析分野ではスペクトルのデータベースの活用が進んでいるが、TOF-SIMS では測定条件によるスペクトルの変化が大きいため、特定のピークに注目した同定が難しく、データベース化が難しい。そこで、本研究ではデータベースに変わるシステムとして、機械学習による未知試料スペクトルの予測システムの構築 [1, 2] を目指している。あらゆる試料に関する TOF-SIMS データの予測が可能な学習システムは実現が難しいため、予測対象試料範囲を絞る必要があるが、本研究では有機物・高分子試料を対象とし、試料範囲は絞らずに本研究室で所蔵するデータを用いて、基礎検討を実施した。機械学習手法としては、予測結果を得る過程で重要視された特徴量 (二次イオンピーク) に関する情報が得られる Random Forest を用いた。

2. 実験方法

TOF-SIMS データの機械学習による予測に関する国際ラウンドロビン (VAMAS TW2 A26) [2] 用に作成したピークリストを用いて、本研究室で所蔵する TOF-SIMS スペクトルについて、各二次イオン強度を数値化した。二次イオンピーク 4230 個について、各スペクトルの総二次イオン強度で規格化した強度を特徴

量とし、有機物の化学構造 33 種類を試料ラベルとして、機械学習用のデータフォーマットを作成した。TOF-SIMS スペクトル 297 個をデータセットとして、Python, Scikit-learn の Random Forest を用いて5回交差検定を実施した。Table.1 にラベルの例を示す。設定した構造が含まれる場合は 1、含まれない場合は 0 とした。

Table.1. Result of Random Forest on TOF-SIMS spectra of high-molecular and organic materials

Sample	C(=O)OH	COH	...	C ₆ H ₆
Material A	0	1	...	1
Material B	0	1	...	0
...	0	1	...	1
Material X	0	1	...	0

3. 結果と考察

高分子・有機物材料試料データ 297 個について、33 種の化学構造を試料ラベルとして Random Forest による予測を実施した結果、単純な構造によるラベルは誤予測が多かった。未知試料の予測により、効果的な試料ラベルに必要な条件および各試料データの測定条件に基づいた記述子や試料ラベル設定に関する知見が得られた。

文 献

- 1) 石倉航, 高橋一真, 山岸崇之, 青木弾, 福島和彦, 志賀元紀, 青柳里果, J. Surf. Anal. 25(2) 103-114 (2018).
- 2) Satoka Aoyagi, Yukio Fujiwara, Akio Takano, et al., Anal. Chem. 93(9), 4191-4197, 2021

*E-mail: aoyagi@st.seikei.ac.jp