

## 研究成果報告書

研究テーマ (和文)	機械学習を活用したスペクトルデータの直接的解析に基づく分子特性予測法の開発		
研究テーマ (英文)	Machine Learning Approach for the Prediction of Molecular Properties Based on Direct Analysis of Spectra Data		
研究期間	2019年 ~ 2022年	研究機関名	国立研究開発法人産業技術総合研究所
研究代表者	氏名	(漢字)	矢田 陽
		(カタカナ)	ヤダ アキラ
		(英文)	Akira Yada
	所属機関・職名	国立研究開発法人産業技術総合研究所 触媒化学融合研究センター 主任研究員	
共同研究者 (1名をこえる場合は、別紙追加用紙へ)	氏名	(漢字)	石原 司
		(カタカナ)	イシハラ ツカサ
		(英文)	Tsukasa Ishihara
	所属機関・職名	国立研究開発法人産業技術総合研究所 バイオメディカル研究部門 主任研究員	

概要 (600字~800字程度にまとめてください。)

機械学習による分子特性予測において、高精度な予測モデルを構築するためには、分子特性と関連する分子構造や特徴の適切な分子記述子の準備が必須である。これまで研究者が様々な分子記述子を考案してきたが、それらが分子の電子的要因や立体的要因、エネルギー状態等を必ずしも適切に表現できているとは限らなかった。そこで本研究では、直接的に分子の特徴を表現可能な分光スペクトルを記述子として活用し、分子特性の予測モデルを構築することを目指して研究を行なった。

本研究では、触媒反応の触媒1成分である配位性分子の違いによる触媒活性を、分光スペクトルを用いて予測することを試みた。まず、赤外線吸収スペクトルを入力とした機械学習に取り組んだ。配位性分子の赤外線吸収スペクトルを量子化学計算によって計算し、それを機械学習の入力として触媒活性予測を試みた。しかしながら、高精度な予測モデルの構築には至らなかった。赤外線吸収スペクトルは分子の運動状態遷移を表現するもので、触媒活性に重要な構造を表現できていないことが原因と考えた。そこで次に、電子線エネルギー損失スペクトルに現れる吸収端近傍微細構造(ELNES)に着目した。ELNESは元素の配位環境や化学結合に関する情報を表現でき、触媒活性予測に適していると期待した。ELNES計算の第一人者である東京大学 溝口照康教授の協力のもと、配位性分子の元素N端スペクトルの計算を行い、そのスペクトルを機械学習の入力として予測モデルの構築を試みた。機械学習法としてニューラルネットワークを用いたところ、高精度な活性予測モデルが構築できることを明らかにした。本成果は、配位環境を機械学習させることが、触媒活性の予測において重要であることを示唆するものである。今後、本研究成果の論文発表に向けた結果の取りまとめとともに、本手法の適用範囲の拡大に向けた研究を推進する予定である。

一方で、配位性分子の赤外線吸収スペクトルは、分子が有する官能基に由来する複数のピークの重ね合わせによって成り立っている。そのため、触媒活性の重要なスペクトルピークを同定する必要があるが、それには複数の重ね合わせたスペクトルを分離する必要があることを認識した。そのため、機械学習を活用したピーク分離技術の開発にも取り組んだ。現在、本研究成果の論文発表に向けて研究を推進している。

発表文献（この研究を発表した雑誌・図書について記入してください。）					
雑誌	論文課題				
	著者名		雑誌名		
	ページ	～	発行年	巻号	
雑誌	論文課題				
	著者名		雑誌名		
	ページ	～	発行年	巻号	
雑誌	論文課題				
	著者名		雑誌名		
	ページ	～	発行年	巻号	
図書	書名				
	著者名				
	出版社		発行年	総ページ	
図書	書名				
	著者名				
	出版社		発行年	総ページ	

英文抄録（100語～200語程度にまとめてください。）

In order to predict molecular properties with high accuracy by machine learning, it is essential to prepare appropriate molecular descriptors of the molecular structure and features related to their properties. Although various molecular descriptors have been proposed by researchers so far, they are not necessarily descriptors that appropriately describe the molecules' electronic and steric factors and energy states. This study investigated the development of new methodologies for building predictive models of molecular properties using spectral data as input (descriptor) for machine learning. Thus, the prediction of catalytic activity of catalytic reactions using spectra calculated by quantum chemistry calculations as input for machine learning was investigated. Various spectra were calculated by quantum chemistry calculation and applied as input for machine learning to the predictive model. It was found that the energy-loss near-edge structure (ELNES) appearing in electron energy-loss spectroscopy (EELS) of the ligand molecules, which is one component of catalytic reactions, is promising that predicts the catalytic activity with high accuracy.

On the other hand, the infrared absorption spectrum of a coordinating molecule consists of the superposition of multiple peaks derived from molecular functionalities. Therefore, we recognized the need to identify the key spectral peaks of catalytic activity, which required the separation of multiple overlaid spectra. Consequently, we also worked on developing peak separation technology using machine learning. The result will be reported in due course.