

## 深層学習を利用した化学的知識獲得に関する考察

(東大院総文<sup>1)</sup>) ○貞金 輝久<sup>1</sup>・横川 大輔<sup>1</sup>

Insights into the process of acquiring chemical knowledge with deep learning (<sup>1</sup>*Graduate School of Arts and Sciences, The University of Tokyo*) ○Teruhisa Sadakane,<sup>1</sup> Daisuke Yokogawa<sup>1</sup>

Deep learning has seen significant advancements in chemistry, enabling a wide range of applications. These achievements largely rely on the use of large-scale datasets. However, in chemistry, the high technical demands of experiments often result in the prevalence of small-scale datasets, leading to a critical issue where model performance significantly declines in such environments. This study aims to address this issue by incrementally increasing the number of training samples to analyze how models learn the influence of substituents and functional groups on molecular properties. By doing so, we seek to elucidate the relationship between improved prediction accuracy and the process of acquiring chemical knowledge.

In this study, the water/octanol partition coefficient (LogP) was used as a molecular property to analyze the performance of deep learning in low-data environments and the process of acquiring chemical knowledge. Specifically, LogP prediction was targeted, and the number of samples was incrementally increased to evaluate prediction accuracy and the interpretability of features learned by the model. A Graph Convolutional Network (GCN) was employed, and Integrated Gradients were applied to visualize chemical knowledge. As a result, it was confirmed that as the number of samples increased, prediction accuracy improved, and the model more accurately captured the chemical characteristics of molecular structures.

**Keywords :** *Explainable AI; Feature Attribution; Solubility; Substituent; Graph Neural Network*

深層学習は、化学分野においても幅広い応用が進展しており、その成果は大規模データセットの活用が大きく依存している。しかし、化学分野では実験の技術的要求が高いため、小規模データセットが多くを占めており、このような環境ではモデルの性能が著しく低下するという課題がある。本研究では、この課題を解決する手がかりとして、学習サンプル数を段階的に増加させながら、分子構造の中で置換基や原子団が物性値に与える影響をモデルがどのように学習しているかを解析した。これにより、予測精度の向上と化学的知識の獲得過程の関係性を明らかにすることを目的とした。

本研究では、化学物質の物性値として水/オクタノール分配係数 (LogP) を用い、少量データ環境における深層学習の性能と化学的知識の獲得過程を解析した。具体的には、LogP 予測を対象に、サンプル数を段階的に増加させながら予測精度と学習モデルによる特徴量の解釈性を評価した。モデルには Graph Convolutional Network (GCN) を使用し、Integrated Gradients<sup>1</sup> を適用して化学的知識の可視化を行った。その結果、サンプル数が増加するにつれ、予測精度が向上するとともに、学習モデルが分子構造の化学的特徴をより正確に捉えることが確認された。

1) Interpretable Attribution Assignment for Octanol–Water Partition Coefficient. D. Yokogawa, K. Suda, *J. Phys. Chem. B.* **2023**, 127, 7004.