[BIOINFO 2023]

FragGNN: Enhancing Molecular Property Prediction through Fragment-based Hierarchical Graph Neural Networks

Dohyeon Kim[1], Dongmin Bang[2,3], Sangseon Lee[4], and Sun Kim[1,2,3,5,*]

[1]*Interdisciplinary Program in Artificial Intelligence, Seoul National University*
[2]*Interdisciplinary Program in Bioinformatics, Seoul National University*
[3]*AIGENDRUG Co., Ltd.*
[4]*Institute of Computer Technology, Seoul National University*
[5]*Department of Computer Science and Engineering, Seoul National University*
**Corresponding author: sunkim.bioinfo@snu.ac.kr*

The prediction of molecular properties, particularly physicochemical and bioactive attributes, plays a pivotal role in molecular design and has profound implications for drug design and discovery. Graph Neural Networks (GNNs) have emerged as a powerful tool, providing graph-specific representations. However, their performance can sometimes be hindered due to the lack of essential prior knowledge. To bridge this gap, we present FragGNN, a new GNN model specifically designed for molecular property prediction. Through a unique hierarchical approach, FragGNN effectively incorporates fragment information, anchored by fragment-based pooling. Three fragmentation methodologies, namely BRICS, Murcko, and Functional Group, were employed. This hierarchical approach enables bidirectional communication between atom and fragment levels, further enhanced by injecting information using molecular fingerprints. This structured, bi-directional information exchange strengthens the model's predictive capability. In a comparative study across various molecular property tasks, our method outperformed state-of-the-art models in 8 of 9 molecular property prediction benchmarks. FragGNN provides invaluable fragment-based embeddings, offering deeper insights into molecular structures. This groundbreaking aspect enables focused adjustments at the molecular fragment level. Through precise adjustments of these fragments, we can optimize and refine particular molecular characteristics.