[BIOINFO 2023]

Improving Out-of-Distribution Generalization in Molecule Graphs with Hierarchical Semantic Environments

Yinhua Piao[1], Sangseon Lee[2], Yijingxiu Lu[1] and Sun Kim[1,3,4,5*]

[1]*Department of Computer Science & Engineering, Seoul National University,*
[2]*Institute of Computer Technology, Seoul National University,*
[3]*Interdisciplinary Program in Bioinformatics, Seoul National University,*
[4]*AIGENDRUG Co., Ltd., Seoul 08826, Korea*
[5]*MOGAM Institute of Biomedical Research, Yongin-si 16924, Korea,,*
**Corresponding author: sunkim.bioinfo@snu.ac.kr*

Out-of-distribution generalization in the graph domain is challenging due to complex distribution shifts and a lack of environmental contexts. Previous works try to generate ***flat*** environmental contexts to help learn from different distribution shifts. However, these ***flat*** environments have shown limitations in capturing complex data changes. To address these limitations, we propose a hierarchical approach to generate semantic environments for each graph.

Specifically, we first explicitly extract variant subgraphs from the input graph to generate proxy predictions on local environments. Stochastic attention mechanisms are employed to re-extract the subgraphs for regenerating global environments in a hierarchical manner. To ensure the meaningfulness of these generated hierarchical environments, we introduce a new learning objective that encourages our model to learn diversity of environments within the same hierarchy while maintaining consistency across different hierarchies. In addition, this approach enables our model to consider the relationships between environments and facilitates robust graph invariant learning.

We adopted DrugOOD benchmark (Assay, Scaffold, and Size splits) that containing various realistic graph distribution shifts to evaluate the out-of-distribution performance of our model. Extensive experiments on real-world graph data have demonstrated the effectiveness of our framework.