**GenerativeDGI-DNE：**

**基于生成式虚拟药物的异构蛋白质-药物网络建模框架**

**医24英才班居凡博 指导老师：胡广**

**研究背景与意义**

**当**前药物靶点预测面临两大核心挑战：

- 蛋白质互作网络（PPI）与药物-靶点相互作用（DTI）数据的异构性导致难以统一建模；

- 罕见靶点样本不足（如孤儿受体仅占已知靶点的5.3%）。现有的药物靶点互作数据不足以建立泛化性能强大的模型。

**G**enerativeDGI-DNE提出创新解决方案：

-跨模态特征融合：整合网络拓扑（拉普拉斯向量）、三维结构（接触图）、序列语义（ESM3）和化学特征（指纹）四维数据，突破传统单模态建模局限

-动态表示学习：通过门控注意力机制实现不同模态特征的动态权重分配，解决异构图数据的不均衡表征问题。

-合成数据增强：采用DRUGGEN生成6000种虚拟化合物，在某种意义上对其进行蒸馏，缓解数据稀疏性问题

**方法论设计**

2.1 多模态输入特征

|  |  |  |
| --- | --- | --- |
| **特征类型** | **蛋白质节点处理** | **药物节点处理** |
| 拉普拉斯向量① | 128维谱嵌入 | 128维谱嵌入 |
| 残基接触图 | edge index:R | 空矩阵 |
| ESM3嵌入 | 1536维残基表示 | 空矩阵 |
| Morgan指纹 | 全零占位矩阵 | 1024morgan指纹向量（radius=2） |

①（LEdropout，防止产生拓扑依赖，逼迫机器学习蛋白质和小分子的化学特征）

2.2 模型架构 超参数：LEdropout，p,q。

1. GAT蛋白质特征融合路径：

- 图注意力层：

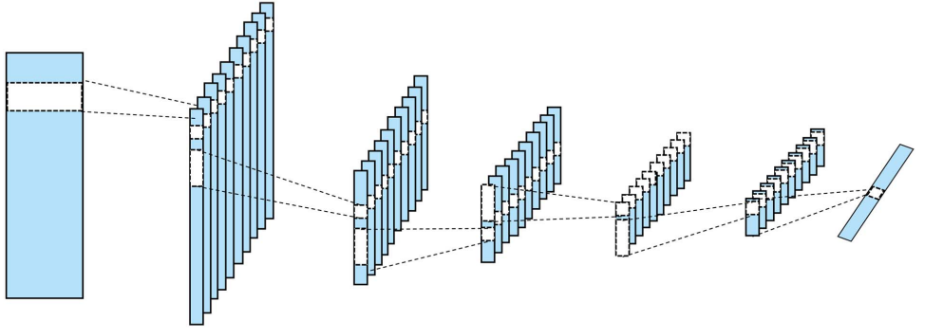
单层8头GAT聚合接触图与ESM3特征，学习残基级重要性权重：



- 多尺度卷积编码：

采用TextCNN提取局部结构模式，捕获β折叠等二级结构特征

TextCNN参数：



2. 对比学习模块：

- DNE方法：使用Node2Vec随机游走取样，将p和q作为超参数设置三组：



- 损失函数Contrasive Loss：交叉熵损失函数





3. 数据准备方案

3.1 基准数据集

- HuRI PPI网络：采用2024年更新版，包含53,151对互作（Y2H-SCORES>0.55）

- DGIdb 4.2：筛选19,724个高置信度（Confidence≥0.5）药物-靶点对