

CellCLAT: Preserving Topology and Trimming Redundancy in Self-Supervised Cellular Contrastive Learning

基于拓扑保持和冗余修剪的胞腔对比学习

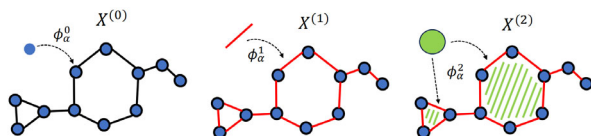
秦滨*, 戢启瑞*, 李江梦*, 王宇鹏, 吴学淞, 曹建文, 徐帆江

KDD 2025

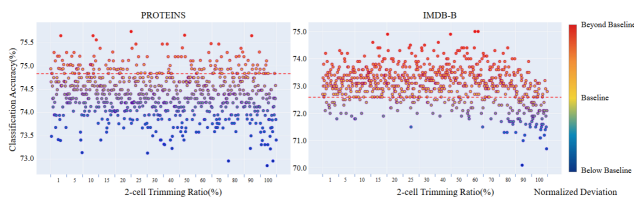
✉ qinbin@ios.ac.cn

研究背景

图到胞腔复形的粘贴映射



2维胞腔冗余的动机实验



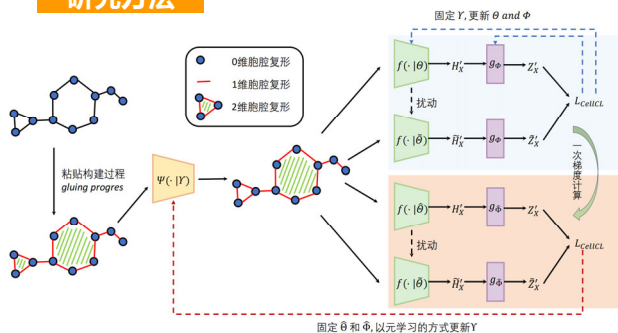
研究现状

拓扑深度学习将图转向组合拓扑空间（如单纯复形、胞腔复形），这些空间基于消息传递的神经网络具有3-WL测试的表达能力，能捕捉图的高阶拓扑信息。

存在问题

- 传统图增强技术会破坏胞腔复形固有的拓扑结构约束（如附着映射连续性和闭包有限性）
- 胞腔表示中存在内在语义冗余，部分高阶胞腔结构损害目标任务的效果

研究方法



解决方案

- 基于参数扰动的拓扑保持增强方法：**针对传统增强破坏拓扑结构的问题，该方法将受控高斯噪声直接注入到胞腔复形编码器的参数中，而非扰动原始胞腔结构本身
- 基于元学习的复形修剪调度器：**为去除任务无关的拓扑冗余，研究团队设计了一个可学习的2维胞腔修剪器。该修剪器以胞腔特征为输入，预测其是否应被掩码（即修剪）

理论视角

- 表达能力：**证明CellCLAT编码器的表达能力超过了WL测试，至少达到3-WL测试的能力
- 修剪操作的有效性：**从因果的角度证明了2维胞腔为混杂因子，将胞腔裁剪视为干预操作的具体实现

实验结果

无监督设置

Dataset	NCI1	PROTEINS	MUTAG	NCI109	IMDB-B	IMDB-M	A.R. ↓
node2vec	54.9 ± 1.6	57.5 ± 3.6	72.6 ± 10.0	-	-	-	11.0
sub2vec	52.8 ± 1.5	53.0 ± 5.6	61.1 ± 15.8	-	55.3 ± 1.5	-	11.8
graph2vec	73.2 ± 1.8	73.3 ± 2.0	83.2 ± 9.3	-	71.1 ± 0.5	-	9.3
InfoGraph	76.2 ± 1.0	74.4 ± 0.3	89.0 ± 1.1	76.2 ± 1.3	73.0 ± 0.9	48.1 ± 0.3	5.7
GraphCL	77.9 ± 0.4	74.4 ± 0.5	86.8 ± 1.3	78.1 ± 0.4	71.2 ± 0.4	48.9 ± 0.3	6.2
ADGCL	73.9 ± 0.8	73.3 ± 0.5	88.7 ± 1.9	72.4 ± 0.4	70.2 ± 0.7	48.1 ± 0.4	8.3
JOAO	78.1 ± 0.5	74.6 ± 0.4	87.4 ± 1.0	77.2 ± 0.6	70.2 ± 0.1	48.9 ± 1.2	6.7
JOAOv2	78.4 ± 0.5	74.1 ± 1.1	87.7 ± 0.8	78.2 ± 0.8	70.8 ± 0.3	49.2 ± 0.9	5.3
RGCL	78.1 ± 1.0	75.0 ± 0.4	87.7 ± 1.0	77.7 ± 0.3	71.9 ± 0.9	49.3 ± 0.4	4.2
SimGRACE	79.1 ± 0.4	75.3 ± 0.1	89.0 ± 1.3	78.4 ± 0.4	71.3 ± 0.8	49.1 ± 0.8	2.7
HTML	78.2 ± 0.7	75.0 ± 0.3	88.9 ± 0.8	77.9 ± 0.2	71.7 ± 0.4	48.9 ± 0.6	4.0
CellCLAT	79.4 ± 0.2	75.7 ± 0.1	89.7 ± 0.3	78.9 ± 0.4	73.4 ± 0.1	50.6 ± 0.2	1.0

半监督设置

Dataset	NCI1	PROTEINS	NCI109	IMDB-B	IMDB-M	A.R. ↓
InfoGraph	68.9 ± 0.8	69.5 ± 0.9	67.8 ± 0.7	68.5 ± 0.6	43.8 ± 0.2	6.4
GraphCL	69.6 ± 0.3	69.7 ± 0.1	69.2 ± 0.3	67.8 ± 1.3	43.9 ± 0.3	4.4
ADGCL	65.6 ± 0.6	68.5 ± 2.0	65.7 ± 0.8	67.7 ± 0.9	41.9 ± 0.9	8.8
JOAO	69.1 ± 0.1	70.9 ± 2.0	69.2 ± 0.3	67.6 ± 0.8	43.9 ± 1.0	5.0
JOAOv2	69.1 ± 0.4	70.1 ± 1.4	69.2 ± 0.2	68.3 ± 1.0	43.6 ± 1.1	5.2
RGCL	70.2 ± 0.7	71.2 ± 0.9	69.1 ± 0.3	68.9 ± 0.3	44.5 ± 0.9	2.4
SimGRACE	69.3 ± 0.3	71.3 ± 0.7	69.0 ± 0.2	68.6 ± 0.7	44.2 ± 0.6	3.4
HTML	69.4 ± 0.1	69.1 ± 1.5	68.9 ± 0.1	68.4 ± 0.3	43.2 ± 0.3	6.4
CellCLAT	70.4 ± 0.5	71.8 ± 0.1	69.5 ± 0.3	68.5 ± 0.4	44.0 ± 0.3	1.8

- 在TU数据集上无监督和半监督设置的实验证明了CellCLAT的有效性