

日期: /

Pre-Training GNNs

深度学习模型的两大挑战:

① 标签数据的稀缺性易导致模型过拟合

② 外推能力较差,常基于数据集中的虚假相关性进行预测

idea: 在应用模型前先向其中注入领域知识 (pre-training)

朴素的策略甚至有负面

Q: 什么样的预训练策略是有效的? 效果

A: 同时对节点嵌入和图嵌入进行预训练, 以使网络捕获局部及全局信息。

Self-supervised (No need for external labels)	Node-level	Graph-level
	Attribute prediction	Supervised Attribute Prediction
Attribute prediction	Attribute Masking	
	Context Prediction	Structural Similarity Prediction

1. Attribute Masking

- 对节点属性进行随机掩码
- 使用 embeddings 预测掩码属性的身份

日期: /

2. Context Prediction

- 对每个图提取一个中心节点
- 提取中心节点的 ^{k-hop} 邻居 和 上下文图 与邻居直接相连的子图
- 对邻居和上下文图进行嵌入
- 最大化真实邻域-上下文对嵌入的点积, 同时最小化
虚假~

鼓励上下文接近的节点取得接近的嵌入

(分布式假设)