

智能营销大作业指南

一、实验目的

本次大作业的内容是基于图学习的客群分类任务，主要考察同学们搭建客群挖掘算法的能力。具体来说，请基于以下的图学习算法（三选一）预测图上每个节点的所属的客群标签。每个算法都有提供开源代码供大家学习，请适当参考链接中相应的开源解决方案。

二、数据集

CiteSeer 是一个经典的图数据集，数据集的统计信息如下表所示。数据集下载链接 <https://github.com/kimiyoung/planetoid>。

节点数量	边数量	特征数量	客群标签数量
3, 327	9, 104	3, 703	6

三、图学习算法

1. 图卷积神经网络（Graph Convolutional Network, GCN）

- 算法描述：GCN 通过在图上执行卷积操作来学习节点的表示向量，融合节点自身特征和邻居节点特征，适用于节点分类和图分类任务。
- 参考论文：《Semi-Supervised Classification with Graph Convolutional Networks》（ICRL 2017）
- 参考代码：<https://github.com/tkipf/gcn>

2. 图注意力网络（Graph Attention Network, GAT）

- 算法描述：GAT 利用注意力机制实现节点级别的特征聚合，允许节点在聚合邻居信息时具有不同的重要性，适用于节点分类、链接预测和图生成等任务。
- 参考论文：《Graph Attention Networks》（ICLR 2018）
- 参考代码：<https://github.com/PetarV-/GAT>

3. 图聚合卷积（GraphSAGE）

- 算法描述：GraphSAGE 是一种采样和聚合的图卷积算法，通过采样节点的邻居子图并聚合邻居节点的特征，以迭代方式更新节点表示，从而在保留图结构的同时有效地学习节点的低维度表示，适用于节点表示学习和图分类任务。
- 参考论文：《Inductive Representation Learning on Large Graphs》（NeurIPS 2017）

- 参考代码: <https://github.com/williamleif/GraphSAGE>

四、评分标准

1. 数据集预处理
2. 算法复现代码
3. 预测指标结果: 评测指标包括 Accuracy、Micro-averaged F1 scores (Micro-F1)
 - $\text{Accuracy} = \text{分类正确节点数} / \text{总节点数}$
 - Micro-F1: 指标讲解见链接
https://blog.csdn.net/qq_43190189/article/details/105778058

五、提交内容

1. 实验报告
2. 实验代码

注意: 请在学在浙大平台上提交, 请勿上传数据集。