

# 第二次作业

---

## 简述

---

任意选择一种课上讲过的network embedding方法实现（Deep random walk或者Node2Vec或者LINE），在Cora, Citeseer和Amazon\_Photo数据集上计算nodes embedding，将计算得到的nodes embedding作为特征训练分类器对结点的标签进行预测，并且和直接基于特征的分类器做对比。

## 数据集介绍

---

对于每一个数据集，包含以下内容：

edges: 每一行表示一条边，包含两个整数，分别是源结点的ID(ID编号从0开始)，目标结点的ID。

labels: 每一行的第一个整数表示结点的ID，第二个整数表示标签的ID(标签ID也从0开始)。

attributes: 每一行代表一个结点的特征向量，例如第*i*行表示ID为*i*的结点特征。

train\_nodes/test\_nodes/val\_nodes: 训练结点的ID/测试结点的ID/验证结点的ID。

## 作业说明

---

需要完成的内容：

- 对每个数据集，任选一种network embedding的方法实现，计算nodes embedding并保存。
- 利用保存的nodes embedding在训练结点上训练分类器（可以是线性的也可以是非线性的MLP），预测测试结点的标签，验证结点可以用于early stop的判断，也可以不使用验证结点。
- 直接利用结点的特征向量在训练结点上训练分类器，预测测试结点的标签。

需要提交的内容：

- 代码：network embedding的代码实现以及利用计算得到的nodes embedding训练分类器的代码，可以不提交直接利用结点特征向量训练分类器的代码。
- 报告：需要给出每个数据集nodes embedding作分类和直接训练分类器作分类的结果，并简单分析效果的差异。

作业以邮件形式发送到:[2001213110@stu.pku.edu.cn](mailto:2001213110@stu.pku.edu.cn)，邮件主题和报告名均为:学号+姓名+第二次作业报告。作业截止时间为:2021年11月27日晚24:00。

