

实验四：基于预训练的节点分类

一、任务描述

本次任务为一个节点分类任务，类别数为4。该数据集共包含34282个节点，每个节点为一段文本。除节点属性信息外，我们还提供了网络的结构信息(edges.csv)，对于graph数据而言，通常节点的属性信息和网络的结构信息都会对节点分类的最终结果产生影响，但具体影响需要自行探索。本次测试集不提供标签，同学们要上传预测文件pred.csv，由助教进行最终的准确度计算。最终提交内容包括一份详细的实验报告及预测文件pred.csv。

二、数据集

- train.csv → 包含id, name, text, label
- val.csv → 包含id, name, text, label
- test.csv → 没有label，提交预测文件pred.csv(格式详见submit_example.csv)
- edges.csv → 图的结构信息，每行包含一条边连接的两个节点node1,node2
- submit_example.csv → 需要提交的预测文件的示例，包含id, label(预测的类别标签)

三、方法

- Word2vec/Fasttext/Glove/BERT/XLNet等。

四、评价指标

- val_acc
- test_acc → 提交pred.csv文件，由助教计算最终的test_acc值
- 最终的评分包括两个方面：1. 实验报告 2. test_acc
- pred.csv的文件形式严格按照submit_example.csv进行提交，文件形式不对一律按不合格处理。