

题目：高水平科研合作模式预测

要求：建立回归模型，预测作者合作能否产生 CCF-A 的论文。

数据：

<http://lfs.aminer.cn/lab-datasets/aminerdataset/AMiner-Paper.rar>

<http://lfs.aminer.cn/lab-datasets/aminerdataset/AMiner-Author.zip>

<http://lfs.aminer.cn/lab-datasets/aminerdataset/AMiner-Coauthor.zip>

说明：

下载论文，作者，合作关系三者数据库后，根据 CCF-A 的列表对数据进行标注。如果论文（含期刊和会议论文）是 CCF-A 则标注为 1，否则为 0。

CCF 推荐论文列表：https://www.ccf.org.cn/Academic_Evaluation/。

合作关系能否产生 CCF-A 的判断依据：一个作者合作集合中，可能发表多篇论文，如果有一篇论文是数据与 CCF-A,则认为能产出 CCF-A，否则认为不能产出。

数据输入：可以是作者 ID，作者的历史论文列表，作者的历史合作关系等，根据实际情况确定。

实验划分方法：

将合作关系去重后，划分为 80%作为训练集,10%作为验证集，10%作为测试集

建议思路：

1.对合作关系打标签，判断是能产出 CCF-A

2.利用 GCN 网络对合作关系进行建模，分析作者在合作网络中的合作潜力。（建议设计

一个带有鉴别能否发表 CCF-A 能力的合作网络)

3.建立用户历史发表论文的个人画像 ,(建议设计历史画像的时候 , 融入 CCF-A 产出能力的鉴别信息)

4.建立可变用户数量的输入模型 (用户合作数量可以是 2-10)

5.先建立浅层次的网络模型 , 减少模型复杂度。

6.建立经典模型 (如 : SVM , GDBT 等分类模型) 作为对比实验。

报告内容 :

1. 含问题理解 , 设计思路 , 模型框架 , 实验设计 , 模型消融实验对比

2. 除了准确率分析以外 , 提交 CCF-A 与非 CCF-A 合作模式的可视化分析结果。