

实验总体要求

数据来源为豆瓣电影的评分记录。请根据训练数据中的用户评分信息，判断用户偏好，并为测试数据中 user-item 对进行评分。

必考核内容:个性化推荐技术(方法自选)。

可选考核内容:标签处理(NLP 技术)、社交推荐(社交网络分析技术)。

基于评分预测的 **RMSE** 指标进行评估。

严禁抄袭代码，一经查实本实验作 0 分处理。

RATINGS FILE DESCRIPTION

训练集中所有的打分信息都以如下格式保存在"training.dat"文件中：

```
User ID, Movie ID, Rating (0-5), Timestamp[, Tag 1, Tag 2, ...]
```

- User ID 用户ID
- Movie ID 电影ID
- Rating 用户给电影的打分
- Timestamp 用户打分的时间，精确到月
- Tag 电影的标签，部分电影存在标签缺失的现象

例如: 1000001, 1293510, 3, 2005-06, black humor

表明用户 ID 为“1000001”给电影“1293510”打了 3 分，时间为“2005-06”，同时留下了“black humor”的标签。

USERS FILE DESCRIPTION

用户之间的社交关系都以如下格式保存在"relation.txt"文件中：

```
A:B,C,D 表示A关注了BCD
```

社交关系均为单向的

TESTING FILE DESCRIPTION

测试集中的用户与训练集一致，相比训练集而言抹去了打分信息，以如下格式保存在"testing.dat"中：

```
User ID, Movie ID, Timestamp[, Tag 1, Tag 2, ...]
```

你需要提交的是针对测试集中每一条记录，给定用户的打分，提交模版见文件“exp3_example_submission.txt”，提交方式将在课程群内进行通知，提交时请务必确保文件格式正确。

以上所有文件可以通过如下链接下载：

链接：<https://rec.ustc.edu.cn/share/531d2aa0-46b0-11eb-9692-7b2b71171220>密码：0h7q

本说明文档将根据实验进行不断更新。更新时将通过课程主页、课程 QQ 群及课上等渠道进行通知，敬请关注。