

2024 暑期大数据推荐系统课程

课程作业：基于隐式反馈的 Top-10 推荐列表预测

| | |
|-------|-------------------------|
| 分数: | 50%课程分数 |
| 截止日期: | 2024 年 7 月 28日 23:59:59 |

考察目的:

通过实践帮助学生理解推荐系统理论，包括数据处理、推荐算法以及算法评估。

数据集:

所使用的数据集为开源的真实推荐系统数据集，包含以下两个文件:

- 训练集 (training.txt) 包含了用于模型训练/计算的数据，内含 942 个用户，1412 个商品项目，44,234 条点击信息。文件内每一行包含三个字段，分别为: user_id, item_id, click
- 测试集 (test.txt) 包含了需要预测 top-10 推荐列表的 926 个用户，每个用户均在训练集中出现过。文件内每一行分别代表一个用户。

任务说明:

利用训练集 training.txt 内的信息对测试集内的用户进行 top-10 预测 (即，为每一个测试集内用户，从商品集合中预测 10 个最有可能被点击的商品。)。提示: 可以首先自行将给定的训练集进一步划分，如取 80%训练集数据为新的训练集和 20%数据为验证集，然后以新的训练集训练不同的推荐算法模型，并在验证集上验证推荐效果，最终选择在验证集上最优的推荐算法对测试集内用户进行 top-10 推荐预测。

编程语言与工具:

不限，可以使用任何语言。

提交材料与方式:

需要提交以下三项材料:

- 实验报告 (pdf) : 描述课程作业的完成过程 (如，实验设计细节，算法模型实现细节/思路)。以你的学号_report 进行命名，如: 20171001_report.pdf
- 实验代码 (类型不限) : 用于完成课程作业的代码。以你的学号 _code 进行命名，如 20171001_code.py。
如有多个代码文件以 zip 打包用以上命名方式提交。
- 实验结果 (txt): 生成的预测列表。包含 926 个用户的 top-10 推荐。文件格式应为如下:
UserID: Item1_ID,item2_ID,item3_ID,...,item10_ID。每一行包含一个用户以及推荐的 10 个项目，项目以逗号隔开。文件命名为学号_result.txt，如 20171001_result.txt

将三项材料打包后以自己的学号命名，如 20171001.zip，然后邮件发送到 rsbdcourse@gmail.com。邮件主题为 rsbd-by-你的学号-姓名，如 rsbd-by-20171001-张伟