Task: Predict the rating scores of the pairs (u, i) in the Test.txt file.

Dataset：

(1) Train.txt, which is used for training your models.

(2) Test.txt, which is used for test.

(3) ItemAttribute.txt, which is used for training your models (optional).

(4) ResultForm.txt, which is the form of your result file.

The formats of datasets are explained in the DataFormatExplanation.txt.

Note that if you can use ItemAttribute.txt appropriately and improve the performance of the algorithms, **additional points (up to 10)** can be added to your final course score.

In this project, you need to report the predicted rating scores of the unknown pairs (u, i) in the Test.txt file. You can use any algorithms you have learned from the course or from other resources (such as MOOC).

One group (consisting of at most three students) needs to write a report about this project. The report should include but not limited to the following contents:

1. Basic statistics of the dataset (e.g., number of users, number of ratings, number of items, etc.);

2. Details of the algorithms;

3. Experimental results of the recommendation algorithms (RMSE, training time, space consumption);

4. Theoretical analysis or/and experimental analysis of the algorithms.

**The deadline is June. 8, 2023. （2023年6月8日晚上24点）**

**Send this homework to bigdatacomputing@163.com**

All your submitted assignments must be entirely your own groups. Any student found cheating or performing plagiarism will receive a final score of zero for this course.

作业内容：预测Test.txt文件中的对(u, i)的评分。

数据集：

(1) Train.txt，用于训练你的模型。

(2) Test.txt，用于测试。

(3) ItemAttribute.txt，可以用来训练你的模型（可选）。

(4) ResultForm.txt，你的结果文件的格式。

数据集的格式在DataFormatExplanation.txt中有详细说明。

请注意，如果你能够合理使用ItemAttribute.txt并提高算法的性能，你的最终课程成绩可以额外增加最多10分。

在这个项目中，你需要报告Test.txt文件中未知对(u, i)的预测评分。你可以使用你从课程或其他资源（比如MOOC）中学到的任何算法。

每组（最多三个学生）需要为这个项目写一份报告。报告应包含但不限于以下内容：

数据集的基本统计（例如，用户数量，评分数量，项目数量等）；

算法的详细信息；

推荐算法的实验结果（RMSE，训练时间，空间消耗）；

算法的理论分析或/和实验分析。

截止日期是2023年6月8日。

请将此作业发送至bigdatacomputing@163.com

你提交的所有作业必须完全是你自己的小组完成。任何被发现作弊或剽窃的学生将在这门课程中获得零分。

DataFormatExplanation.txt文件中的数据格式解释：

1、train.txt

这个文件是训练数据集。每个用户的数据开始于一个特别的行，这行包含两个部分：用户ID和该用户评价的项目数量。在这一行之后，就是用户评价的各个项目以及对应的评分。每个项目及其评分占据一行，项目ID和评分之间用空格分隔。

例如：

用户ID | 评价的项目数量

项目ID 评分

...

2. test.txt

这个文件是测试数据集。其格式与train.txt类似，但是没有评分。每个用户的数据开始于一个特别的行，这行包含两个部分：用户ID和该用户评价的项目数量。在这一行之后，就是用户需要评价的各个项目，每个项目占据一行。

例如：

用户ID | 评价的项目数量

项目ID

...

3. itemAttribute.txt

这个文件包含每个项目的属性信息。每行代表一个项目，包含项目ID和两个属性（属性1和属性2）。如果一个项目不属于任何一个属性，那么属性值就标为'None'。

例如：

项目ID | 属性1 | 属性2