上海理工大学光电信息与计算机工程学院

**《推荐系统》实验四**

****

|  |  |
| --- | --- |
| **学　　院** | 光电信息与计算机工程学院 |
| **专　　业** | 数据科学与大数据技术 |
| **学生姓名** | 罗震宇 |
| **学　　号** | 2135060321 |
| **指导教师** | 艾均 |
| **成 绩** |  |

**目录**

[实验四 SVD推荐系统算法实现 3](#_Toc152051504)

[1. 实验要求与环境 3](#_Toc152051505)

[1.1 实验要求 3](#_Toc152051506)

[1.2 实验环境 3](#_Toc152051507)

[2. 实验过程 3](#_Toc152051508)

[2.1 数据输入 3](#_Toc152051509)

[2.2 模型训练与评分预测 3](#_Toc152051510)

[2.4 评估与可视化 3](#_Toc152051511)

[2. 实验结果 5](#_Toc152051512)

[4. 个人感想 6](#_Toc152051513)

实验四 SVD推荐系统算法实现

## 实验要求与环境

### 1.1 实验要求

选择MovieLens small 数据集；

采用MovieLens数据，构造一个10x10的评分矩阵，其中每一个评分都是已知的，随机选取其中30%数据为测试集，在一个字典中记录它们，并在矩阵中移除这些测试集评分；

实现基于SVD的个性化推荐算法,预测已经选好的测试集；

计算并绘一张图，表示出该算法的MAE和RMSE误差。

### 1.2 实验环境

处理器 AMD Ryzen 7 5800H with Radeon Graphics

内存 16 GB

系统 Windows 11

IDE PyCharm Professional 2023.2.4

## 2. 实验过程

### 2.1 数据输入

选择MovieLens small 数据集，针对评分最多的前10个用户，构建一个10x10的评分矩阵，并返回该矩阵以及一个包含行索引、列索引和相应评分的三元组列表。对该矩阵进行SVD分解，并选择前3个奇异值进行矩阵分解。随机选取其中30%数据为测试集，在一个字典中记录它们，

### 2.2 模型训练与评分预测

对评分矩阵进行矩阵分解，通过梯度下降法不断迭代，更新模型参数，直到达到指定的迭代轮数或者误差小于0.01。

在每轮迭代中，计算预测评分与测试集中实际评分的均方根误差（RMSE）和平均绝对误差（MAE），并打印输出，同时将迭代过程中的误差信息存储在单独的数据框中。

### 2.4 评估与可视化

本实验采用平均绝对误差(MAE)和均方根误差(RMSE)作为评估标准。

Mean Absolute Error:

Root Mean Square Error:

使用Seaborn和Matplotlib库函数绘制MAE和RMSE随迭代轮数变化的曲线图。

## 实验结果

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **表1** | | | | | | | | |
| **Epoch** | MAE | RMSE | **Epoch** | MAE | RMSE | **Epoch** | MAE | RMSE |
| 1 | 3.12485 | 1.84220 | 6 | 0.08207 | 0.06349 | 11 | 0.03758 | 0.02886 |
| 2 | 0.38160 | 0.25048 | 7 | 0.06734 | 0.05209 | 12 | 0.03343 | 0.02579 |
| 3 | 0.22475 | 0.15528 | 8 | 0.05681 | 0.04380 | 13 | 0.03003 | 0.02329 |
| 4 | 0.14775 | 0.10737 | 9 | 0.04870 | 0.03753 | 14 | 0.02724 | 0.02121 |
| 5 | 0.10525 | 0.08025 | 10 | 0.04256 | 0.03268 | 15 | 0.02479 | 0.01946 |



图.1 Line plot of MAE＆RMSE vs. epoch

如图1所示，在前3轮迭代时MAE与RMSE急剧降低，训练效果较好；在第4轮开始下降趋势减弱。随着迭代次数的不断增加，误差呈减小趋势。

## 4. 个人感想

本次题目的目的是设计并实现一个基于协同过滤的推荐系统，利用用户的评分数据，为用户提供个性化的推荐服务，采用SVD模型的协同过滤（CF）个性化推荐算法。我对本次题目的完成情况给予了95分。在完成本次题目的过程中，有以下几个方面的收获：

在理论学习方面，这次实验明显比之前的实验要难度大一些，在理解实验原理的过程中要认真学习，明白奇异值分解在这个里面的用途。如果直接按照前面的思路来进行操作得到的图是错误的。

在编程实现方面，最开始没有很快的找到一个无空值的10\*10矩阵，因为当时的想法是直接在表格中按顺序找，后来经过老师指点，先找到评分过最多的前十部电影，然后找到评分过这几部电影的所有用户，对这几部电影的用户取交集，最后成功解决了这个问题。

在报告撰写方面，我遵循了上海理工大学论文的格式要求，详细地阐述了实验的背景、原理、过程和结果，较好地反映了本次实验的目的和意义。实验结果具有可重复性和科学性，可以作为推荐系统研究的一个参考案例。

整个过程中还遇到了很多问题，在进行SVD迭代时，前期没有想清楚究竟是要迭代什么，导致误差无法下降；并且在时间安排上较为不合理，到了最后来赶工的次数较多。希望自己下次能吸取教训。