# 作业8

# 数据处理思路

## 前期算法选择工作

目前基于知识的推荐系统有如下几种：基于协同的推荐系统、基于内容的推荐系统和基于约束的推荐系统，对于当前的数据集而言，用户-电影之间的评分关系明显，电影的详细资料较为充分，用户的详细资料不够充分，并且对数据没有显著约束条件，所以基于协同的推荐系统和基于内容的推荐系统是最适用于当前数据集的。

基于协同的推荐系统又分为物品-物品协同过滤和用户-用户协同过滤两种，无法直观的通过估算来评判推荐算法的质量，故需要编写相应算法的代码进行测试。

此外，还可以混合多种推荐算法搭建混合推荐系统，单一混合、并行混合与管道混合都需要分别进行算法设计，通过测试来判断推荐算法质量。

## 算法编写思路

目前通过粗略估计可能取得较好推荐效果的单一算法有物品-物品协同推荐、用户-用户协同推进和基于内容的协同推荐。

1. 数据预处理：

* 1. 对表格中的数据进行读取和格式处理。

1. 准备工作：

* 1. 缺省值处理：利用算法对数据中的缺省值进行处理。

* 1. 计算相似度矩阵，物品-物品协同推荐相似度矩阵、用户-用户相似度矩阵和物品-物品内容相似度矩阵。

1. 推荐评分计算：

* 1. 单一策略：使用三个算法对每个用户对所有电影的评分（推荐指数）进行计算，并储存为矩阵。

* 1. 混合策略：

* + 1. 并行混合：对三个算法的结果进行并行混合得到结果矩阵。

* + 1. 管道混合：对三个算法进行搭配，进行混合。

1. 推荐准确率计算：

* 1. 基于用户对电影的已有评分进行推荐的准确率计算，分为4个浮动范围分别衡量。

1. 最优算法选择：

* 1. 给予4个浮动范围不同的权重，计算综合准确率并比较得出相对最优推荐算法。