第一章 基础理论

1、推荐系统定义：采集用户历史行为信息，结合具体推荐模型帮助用户选择商品或提供建议的过程。

2、个性化推荐模型：

①数据采集和预处理：收集用户属性、项目属性和用户对项目的行为信息。预处理包括清理、减噪。

②推荐算法：将收集并处理好的数据通过推荐算法为用户产生推荐。

③产生推荐：通过推荐算法计算得到目标用户的最近邻集合，将最近邻评价过的项目推荐给目标用户；利用模型对未知项目进行预测，将预测评分最高的项目推送给目标用户。

3、协同过滤推荐技术：

①基于内存的协同过滤：根据用户或者项目的相似度，选出与目标用户最相似的若干用户的评分，来对未评分的项目进行评分评测。

②基于模型的协同过滤：通过分析用户和项目的内部规律，预测用户对项目的偏好。

典型代表：概率矩阵分解技术。

4、打分记录：



每一行是用户i评价过的电影集合。

每一列是评价过电影j的用户集合。

每一个元素是用户i对电影j的评分，通常取1~5整数，数据越大用户对该项目越满意。

5、基于近邻的协同过滤算法：

①余弦相似度：



定义向量a和向量b为中的第u行和第v行，将两个向量的夹角余弦值定义为用户u和v的相似度，sim（u,v）的值越接近1，说明两者相似度越高。

②修正余弦相似度：

相对于传统的余弦相似度考虑了用户的评分偏好。



③评分预测：



给定目标用户u，可以找出与其最相似的前n个用户记为。由上式可以对任一用户未进行评分的项目进行预测，将所有预测结果降序排列，从中选出前n个推荐给用户。

6、TOP-N预测准确率

①召回率：

②精确率：