

## B 题：推荐书籍

随着信息技术和互联网的发展，人们逐渐从信息匮乏的时代走入了信息过载的时代。此时，无论是信息消费者还是信息生产者都遇到了很大的挑战：对于信息消费者，从大量信息中找到自己感兴趣的信息是一件非常困难的事情；对于信息生产者，让自己生产的信息脱颖而出，受到广大用户的关注，也是一件非常困难的事情。

推荐，就是解决这一矛盾的重要工具，在互联网的产品和应用中被广泛采用，包括大家经常使用的相关搜索、话题推荐、电子商务的各种产品推荐、社交网络上的交友推荐等。

我们获得了一个著名网上书店的用户行为信息，包括对于书籍的评分数据，书籍的标签信息以及用户的社交关系，请你根据数据完成以下问题。

- 1.分析影响用户对书籍评分的因素；
- 2.建立一个模型，预测 predict.txt 附件中的用户对未看过书籍的评分；
- 3.针对 predict.txt 附件中的用户，给每个用户推荐 3 本没看过的书籍。

数据说明：

数据一共包括 5 部分。

predict.txt 为预测集合，共 2 列，从左到右依次是用户 ID、书籍 ID，参赛者需要预测出第三列的值，即该用户对书籍的评分。

user\_book\_score.txt 中显示了用户评分数据，共 3 列，从左到右依次是用户 ID、书籍 ID、该用户对书籍的评分（评分分值范围 1-5 分，分数越高，可视为用户越喜欢这本书籍）。

book\_tag.txt 为书籍的标签数据，共 2 列，从左到右依次是书籍 ID、该书籍对应的标签；标签之间以逗号隔开。书籍的受欢迎程度是有差别的，所以书籍的标签数也是有差别的。

user\_social.txt 为用户的社交关系数据，共 2 列，从左到右依次是用户 ID、用户关注的好友的 ID。这里的好友关系是单向的。

user\_read\_history.txt 为用户看过的书籍数据，共 2 列，从左到右依次是用户 ID、该用户看过的书籍的 ID。