本周报告  
  
这个星期，我按计划完成了电子书的数据清洗工作。我将数据导入了python的数据分析库中，利用pandas库进行数据清洗和预处理，以供下一步的推荐算法作准备。通过对数据的初步了解，我发现电子书内容类型涵盖了小说、人文社科、流行读物等多个领域，这个结果有利于我们推荐系统的多样性和用户群体广泛性。  
  
接着，我对电子书数据进行了标签化，将每本书的主题进行了切分词语处理，去除无效词语和停用词。我使用jieba分词库和nltk库，采用了tf-idf算法和word2vec算法对电子书进行标签化和向量化。这样我们就得到了每本书的主题和特征向量，可以为下一步的推荐算法做好准备。  
  
本周，我还完成了flask服务器的搭建工作。我创建了主要的api端口，测试了调试了接口的响应和返回数据。我还添加了一些中间件来解决一些常见的请求问题，例如支持跨域请求、对请求进行认证和限流等。现在我们的服务器已经可以接收请求并返回数据，我们可以在此基础上构建推荐算法的接口。  
  
这个星期，由于我忙于准备调剂复试，做的工作不多。但我会尽快补上这些工作的进度。下个星期，我将主要进行推荐算法的编写。首先，我会选择一种适合我们数据特点的算法进行实现，例如基于内容的推荐、协同过滤算法等。然后，我会根据我们服务器的api结构，实现相应的接口，将算法结果以json形式传递给客户端。  
  
总体来说，工作进展在这个星期有所放缓，但我们已经解决了数据清洗和标签化问题，服务器也已经搭建好并通过测试。下个星期我们将主要进行推荐算法的编写工作。虽然进度可能会受到调剂复试的影响，但我们会尽力做好工作，争取早日实现一个稳定的推荐系统。