Movielens数据集介绍：

links.csv:

文件里面的内容是帮助你如何通过网站id在对应网站上找到对应的电影链接的。数据格式如下：

movieId, imdbId, tmdbId

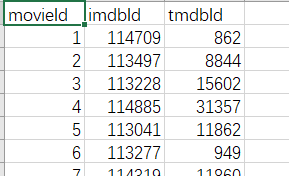
movieId:表示这部电影在movielens上的id，可通过链接https://movielens.org/movies/(mId) 来得到。

imdbId:表示这部电影在imdb上的id，可以通过链接http://www.imdb.com/title/(imdbId)/

来得到。

tmdbId:表示电影在themoviedb上的id，可通过链接http://www.imdb.com/title/(tmdbId)/

来得到。



movies.csv:

movieId, title, genres

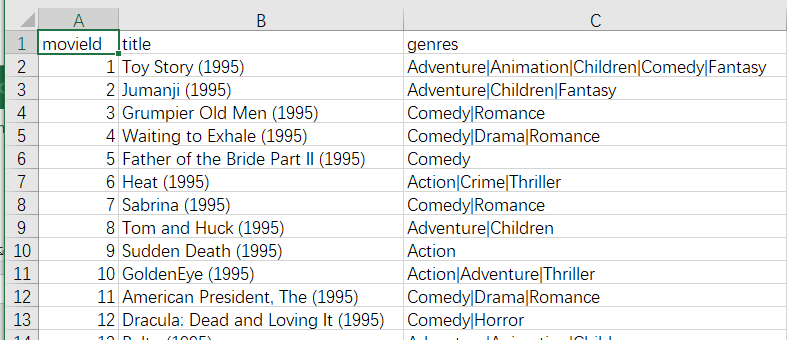
文件里包含了一部电影的id和标题，以及该电影的类别。数据格式如下：

movieId, title, genres

movieId:每部电影的id

title:电影的标题

genres:电影的类别（19个类别）



\* Action

\* Adventure

\* Animation

\* Children's

\* Comedy

\* Crime

\* Documentary

\* Drama

\* Fantasy

\* Film-Noir

\* Horror

\* Musical

\* Mystery

\* Romance

\* Sci-Fi

\* Thriller

\* War

\* Western

\* (no genres listed)

ratings.csv:

文件里面的内容包含了每一个用户对于每一部电影的评分。数据格式如下：

userId, movieId, rating, timestamp

userId: 每个用户的id

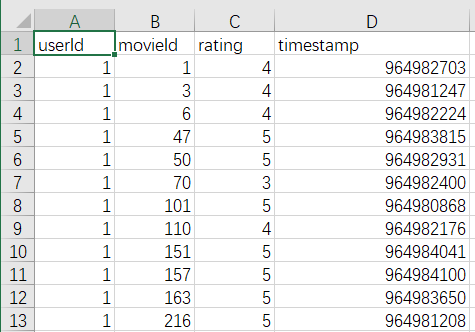
movieId: 每部电影的id

rating: 用户评分，是5星制，按半颗星的规模递增(0.5 stars - 5 stars)

timestamp: 自1970年1月1日零点后到用户提交评价的时间的秒数

数据排序的顺序按照userId，movieId排列的。

预测：ratings



tags.csv:

文件里面的内容包含了每一个用户对于每一个电影的分类。数据格式如下：

userId, movieId, tag, timestamp

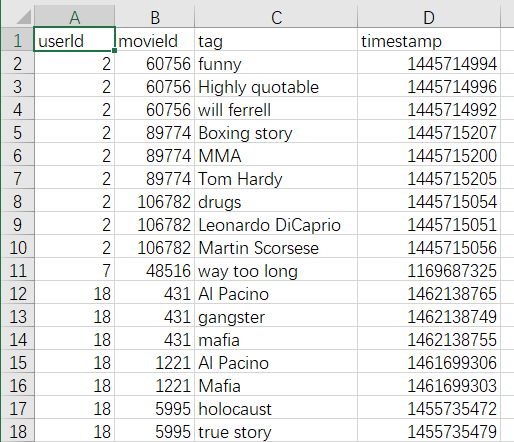
userId: 每个用户的id

movieId: 每部电影的id

tag: 用户对电影的标签化评价

timestamp: 自1970年1月1日零点后到用户提交评价的时间的秒数

数据排序的顺序按照userId，movieId排列的。



最新：2019.12 MovieLens 25M：

2.5千万评级，1百万标签，62000个电影，162000用户，1.5千万相关性标签基因数据，1129个标签。

genome-scores.csv, genome-tags.csv,

Yelp Dataset Challenge是Yelp官方公布的挑战赛，只要你的研究项目用到这个数据集，并且研究文章得到Yelp的青睐，就有机会获取5000美元以上的奖金。Yelp是美国著名商户点评网站，创立于2004年，囊括各地餐馆、购物中心、酒店、旅游等领域的商户，用户可以在Yelp网站中给商户打分，提交评论，交流购物体验等。在Yelp 中搜索一个餐厅或者旅馆，能看到它的简要介绍以及网友的点论，点评者还会给出多少星级的评价，通常点评者都是亲身体验过该商户服务的消费者，评论大多形象细致。

PURS目标：提高预测率