有两个用法，1是用训练集来预测测试集计算rmse，2是用训练集获得推荐结果。

用法1

spark-submit --class org.apache.spark.mllib.cf.UserbasedApp --driver-memory 12G Userbased.jar test recommendation/data/ua.base recommendation/data/ua.test

参数说明

a test表示计算rmse

b recommendation/data/ua.base训练集

c recommendation/data/ua.test测试集

用法2

spark-submit --class org.apache.spark.mllib.cf.UserbasedApp --driver-memory 12G Userbased.jar rec recommendation/data/ua.base recommendation/output/userbased

参数说明

a rec符号表示给用户推荐

b recommendation/data/ua.base 用户评分历史记录数据

c recommendation/output/userbased推荐结果存放目录

用户数量大，userbased计算用户相似度时花较长时间

我在几千万的用户数据集测试过，计算相似度就失败了。

解决该问题需要对spark原理更深入了解。（计算相似度用的是mllib中的一个函数）

但无论如何，用户数量大，计算用户的相似度的时间是需要很长时间的。

因此，对于userbased的算法，适合用于user数量少的推荐应用

鉴于大多数应用场景，用户数量比物品数量大很多，所以userbased比itembased的速度慢很多。这也是为什么后者较前者应用比较广泛。