推荐系统后端程序

功能一：猜您喜欢（为您推荐）

1. 数据分析探索---------------->ExploringDataAnalysis.ipynb
2. 全量数据召回 通过surprise svd 召回 ---------------->movieidprocess.py
3. 对用户侧数据user.csv 和电影侧数据做特征
   1. embedding 特征 ---------------->Embeddingourdata.scala
   2. 其他特征----------------> FeatureEngForourdata.scala
4. 对3.2产生的train test all的数据集进行处理,进行训练----------------> WideNDeep.py （flag=train）
5. 对全量召回集进行处理。产生召回集特征---------------->recallprocess.py
6. WideNDeep对全量数据进行预测---------------->WideNDeep.py（flag=predict）
7. 送给WideNDeep的特征有：

'userid', 'movieid','userPositiveType1', 'userPositiveType2', 'userPositiveType3',

'userPositiveType4', 'userPositiveType5', 'userRatedMovie1',

'userPositivefeature1', 'userPositivefeature2', 'userPositivefeature3',

'userPositivefeature4', 'userPositivefeature5', 'userPositivearea1',

'userPositivearea2', 'userPositivearea3', 'userPositivearea4',

'userPositivearea5', 'userRatingCount', 'userRatingStddev',

'userAvgRating', 'movieactor1', 'movieactor2', 'movieactor3',

'movieactor4', 'moviedirecotr1', 'movieRatingCount', 'movieAvgRating',

'movieRatingStddev'

功能二：喜欢这部电影的人也喜欢 （用户之间的关系）

算电影和电影之间相似度。只不过是通过user来表达。

功能四：与这部电影相似的电影。

功能三：双路召回。双路排序。