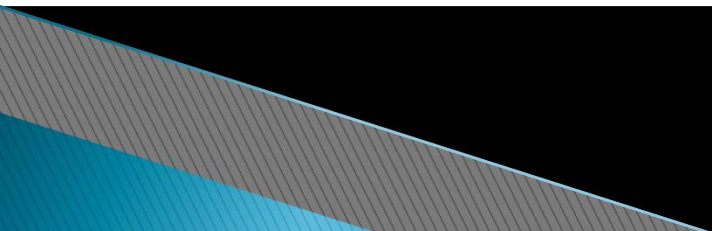


聚美优品 item-based CF协同过滤推荐

赵海臣

核心特点

- ▶ 7天短期兴趣模型(与ALS长期个性模型互补)
- ▶ 基于用户session浏览数据建模
- ▶ 使用库存数据减低运算规模
- ▶ 融合商品点击模型
- ▶ 相似度矩阵模型与用户得分模型独立分开可控



背景

- ▶ 由于SPARK ALS Matrix Factorization使用的数据为长期数据，因此导致推荐结果比较符合用户的长期个性，不能及时捕捉到用户短期的兴趣关注点，所以通过短期CF来捕捉用户的短期兴趣关注点，形成“长期稳定兴趣”+“短期动态关注点”的联合推荐结果。

基于物品的cf推荐

- ▶ 基于物品的推荐，总体流程：
 - item-item相似度矩阵
 - 为每件物品预先计算好最为相近的其它物品，形成一张item-item相似度矩阵表；
 - 目标用户-物品行为表
 - 提取出用户发生过行为的物品列表；
 - 计算每个item推荐分数
 - 根据用户发生过关系的物品，基于item-item相似度矩阵计算所有物品的分数。
 - 推荐列表
 - 基于上一步的所有商品分数，提取分数最高的商品进行推荐。

建立相似度矩阵的数据选择

- ▶ 由于每个用户的单次session行为的关注性较强，因此对用户的单次session记录进行处理能达到相对于用户生涯浏览记录区分更高的精细度
- ▶ 历史数据的考察时长
 - 在基于用户生涯的标准行为记录中，区分粒度为用户级别，较长用户行为历史时长将模糊用户的专注点，导致单个商品相似度精确度的下降。
 - 在基于session的记录中，区分粒度为Session级别，单次session时间为短时段、有明确关注点的记录，因此历史数据的考察时长并不会显著降低相似度矩阵的精准度。

item-item相似度矩阵

- 目标：
 - itemA – itemB的相似度矩阵
- 使用cosine-similarity作为相似度距离标准：

$$w_{uv} = \cos A = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \cdot \|\vec{v}\|} = \frac{u_1 v_1 + u_2 v_2 + \dots + u_n v_n}{\sqrt{u_1^2 + u_2^2 + \dots + u_n^2} \cdot \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}} = \frac{|N(u) \cap N(v)|}{\sqrt{N(u) \times N(v)}}$$

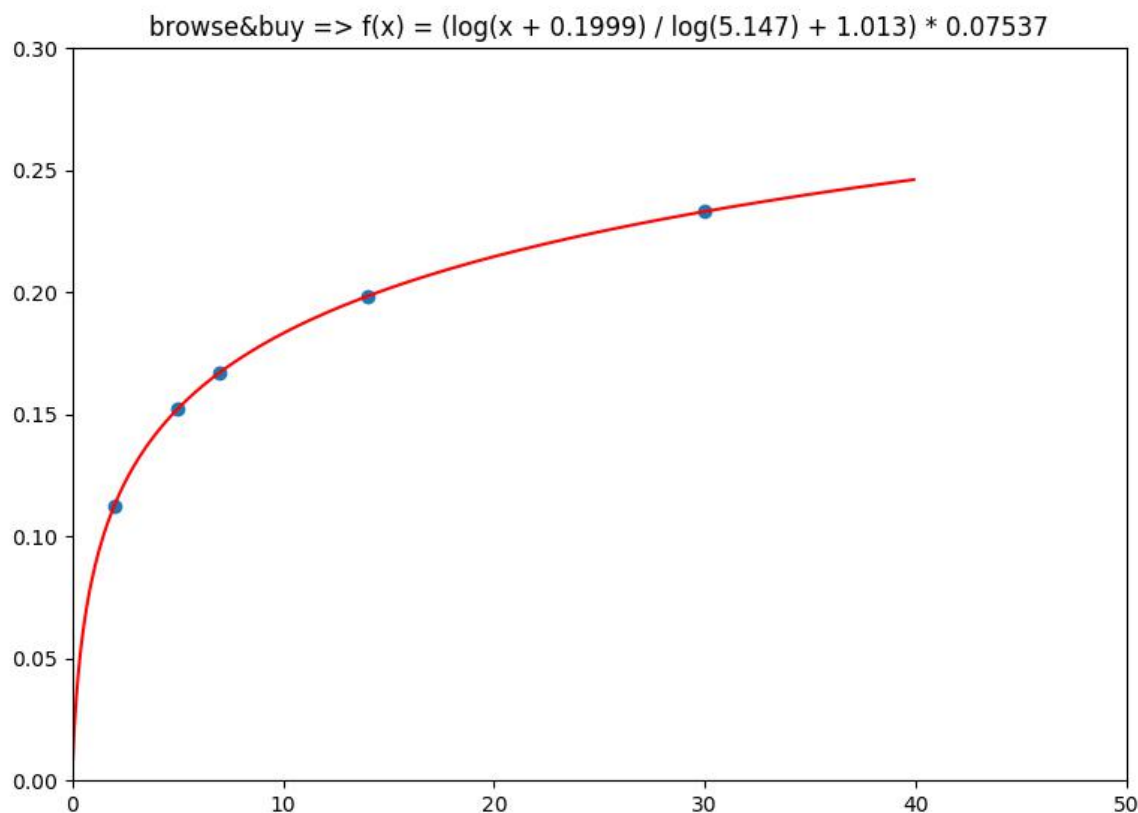
- 库存过滤非可售itemB
 - 只计算itemA-可售itemB的距离，itemB若不被推出，无计算价值。

user-item打分模型融合点击模型

- ▶ 用户的历史浏览与最近购买行为关系分析：
 - 用户最近2天所购买的商品在过去30天浏览过的比例：
 - 23.326%
 - 用户最近2天所购买的商品在过去14天浏览过的比例：
 - 19.805%
 - 用户最近2天所购买的商品在过去7天浏览过的比例：
 - 16.7079%
 - 用户最近2天所购买的商品在过去5天浏览过的比例：
 - 15.2455%
 - 用户最近2天所购买的商品在过去2天浏览过的比例：
 - 11.2537%

user-item打分模型融合点击模型

- ▶ 用户的历史浏览与最近购买行为关系拟合：



user-item打分融合点击模型

- ▶ 所以，用户的历史浏览点击商品是用户高购买倾向的商品，应该融合进入CF推荐模型
- ▶ 用户点击的融合：
 - 在item-item相似度矩阵中考虑itemA-itemA自身的相似度，设为参数 $\omega = 1.0$ ，而itemA-其它商品itemB的相似度大部分分布于0.5以下，因此自身点击可以显著增加商品自身的推荐权值。
 - 参数 ω 是点击模型的权重，直接影响最终推荐结果中浏览点击商品的比例：
 - 若 ω 设为极大值，例如10000，则点击过的商品将全部排在最前面，且按照点击次数排序
 - 若 ω 设为极小值，例如0，则表示点击模型关闭

user-item打分模型融合点击模型

- ▶ 根据用户发生过关系的物品，基于item-item相似度矩阵计算所有物品的分数
 - 用户每次点击商品itemA将对所有itemB加权值，itemB的权值分根据itemA-itemB的分数确定

$$Score_{user-itemB} = \sum_{user-itemA} clickTimes_{itemA} \times similarity_{itemA-itemB}$$

user-item的数据选择

- ▶ 不同于建立item-item相似度矩阵，基于session的相似度矩阵时间越长越精确不同，由于cf推荐的目标是考察用户短期的兴趣关注点，因此选择7~15天比较合适。
- ▶ 由于用户不是每天都会登陆，因此过短时间的用户群体覆盖率过低，导致很多短期未登陆用户将无法有效获得推荐。
 - 改进方法：可以采用用户最近的行为记录而非硬性的时长限制。

复购限制

- ▶ 对用户最近购买过的商品进行过滤，不进行推荐
 - *此举降低推荐效果，用户的复购行为比想象中的频繁，在“用户重复行为分析”中发现，大约占总销量的2~3%的商品为过去15天内购买过的商品。

关闭复购、打开点击模型

```
> *****推荐[_CF_test]*****
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.464736][0.011798][10171038.024047][0.401056][0.797499][1.000000][14.696020][25.050250][0.002170],[0.009610] *****
> *****mainpageTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性 *****
> ***** [0.072363][0.001946][2060895.144848][0.006931][0.013862][23.000000][34.000000] *****
> *****allTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性 *****
> ***** [0.214456][0.005135][4140545.829504][0.006931][0.013862][17.000000][24.000000] *****
> *****sevenDaysTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性 *****
> ***** [0.260687][0.006340][5264806.509224][0.006931][0.013862][23.000000][27.000000] *****
> *****推荐[_CF_test] *****
> lowRank = 0, highRank = 19
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.365921][0.008804][7881010.809784][0.740831][1.000000][1.000000][7.308400][7.308400][0.078640],[0.244710] *****
> *****推荐[_CF_test] *****
> lowRank = 20, highRank = 39
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.041584][0.001255][982768.221946][0.732919][1.000000][1.000000][8.770514][8.770514][0.046633],[0.142190] *****
> *****推荐[_CF_test] *****
> lowRank = 40, highRank = 59
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.024639][0.000748][575786.940949][0.727514][1.000000][1.000000][9.947216][9.947216][0.025642],[0.089667] *****
> *****推荐[_CF_test] *****
> lowRank = 60, highRank = 79
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [1.000000][1.000000][10.590359][10.590359][0.016132],[0.061836] *****
> *****
> lowRank = 80, highRank = 99
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.013731][0.000425][11.131995][11.131995][0.010271],[0.045225] *****
```

关闭复购、关闭点击模型

```
> *****推荐[_CF_test]*****
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比*****
> ***** [0.268069][0.007221][6653193.190034][0.351991][0.775852][1.000000][14.103850][24.733070][0.003700],[0.014270] *****
> *****mainpageTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性*****
> ***** [0.072905][0.001980][2121193.405135][0.006931][0.013862][23.000000][34.000000] *****
> *****allTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性*****
> ***** [0.215273][0.005204][4219494.219826][0.006931][0.013862][17.000000][24.000000] *****
> *****sevenDaysTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性*****
> ***** [0.260993][0.006421][5375994.749382][0.006931][0.013862][23.000000][27.000000] *****
> *****推荐[_CF_test]*****
> lowRank = 0, highRank = 19
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比*****
> ***** [0.117749][0.003044][3051123.733992][0.710167][1.000000][1.000000][6.510140][6.510140][0.130440],[0.331750] *****
> *****推荐[_CF_test]*****
> lowRank = 20, highRank = 39
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比*****
> ***** [0.057856][0.001583][1441932.612397][0.720434][1.000000][1.000000][8.866061][8.866061][0.051974],[0.147520] *****
> *****推荐[_CF_test]*****
> lowRank = 40, highRank = 59
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比*****
> ***** [0.038918][0.001081][922736.631422][0.720965][1.000000][1.000000][10.000470][10.000470][0.027814],[0.090242] *****
> *****推荐[_CF_test]*****
> lowRank = 60, highRank = 79
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比*****
> ***** [1.000000][1.000000][10.672468][10.672468][0.017403],[0.060030] *****
> *****
> lowRank = 80, highRank = 99
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比*****
> ***** [0.023675][0.000675][11.214649][11.214649][0.010492],[0.042268] *****
```

打开复购、关闭点击模型

```
> *****推荐[_CF_test]*****
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比*****
> ***** [0.226419][0.006200][5808857.549158][0.352934][0.776327][1.000000][14.066850][24.644600][0.003910],[0.014600] *****
> *****mainpageTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性*****
> ***** [0.072130][0.001955][2069747.055031][0.006931][0.013862][23.000000][34.000000] *****
> *****allTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性*****
> ***** [0.215444][0.005182][4176945.289788][0.006931][0.013862][17.000000][24.000000] *****
> *****sevenDaysTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性*****
> ***** [0.262500][0.006399][5335683.369158][0.006931][0.013862][23.000000][27.000000] *****
> *****推荐[_CF_test] *****
> lowRank = 0, highRank = 19
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比*****
> ***** [0.100356][0.002612][2689184.853209][0.711911][1.000000][1.000000][6.491490][6.491490][0.139480],[0.337870] *****
> *****推荐[_CF_test] *****
> lowRank = 20, highRank = 39
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比*****
> ***** [0.047685][0.001346][1222997.232196][0.719674][1.000000][1.000000][8.867711][8.867711][0.052064],[0.148400] *****
> *****推荐[_CF_test] *****
> lowRank = 40, highRank = 59
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比*****
> ***** [0.033173][0.000933][797028.851646][0.719000][1.000000][1.000000][9.999110][9.999110][0.028262],[0.091506] *****
> *****推荐[_CF_test] *****
> lowRank = 60, highRank = 79
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比*****
> ***** [1.000000][1.000000][10.641824][10.641824][0.017162],[0.061526] *****
> *****
> lowRank = 80, highRank = 99
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比*****
> ***** [0.019996][0.000584][11.192337][11.192337][0.010651],[0.043126] *****
```

打开复购、打开点击模型

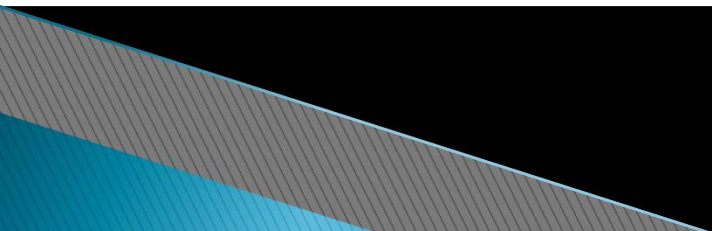
```
> *****推荐[_CF_test]*****
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.340061][0.009420][8998315.875058][0.406664][0.799973][1.000000][14.751870][25.154450][0.002510],[0.009880] *****
> *****mainpageTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性 *****
> ***** [0.074133][0.002032][2184767.995360][0.006931][0.013862][23.000000][34.000000] *****
> *****allTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性 *****
> ***** [0.217669][0.005298][4301676.820091][0.006931][0.013862][17.000000][24.000000] *****
> *****sevenDaysTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性 *****
> ***** [0.263419][0.006521][5461465.739362][0.006931][0.013862][23.000000][27.000000] *****
> *****推荐[_CF_test] *****
> lowRank = 0, highRank = 19
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.254989][0.006770][6796474.110820][0.745637][1.000000][1.000000][7.307880][7.307880][0.078400],[0.246210] *****
> *****推荐[_CF_test] *****
> lowRank = 20, highRank = 39
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.034295][0.001080][915304.931759][0.736247][1.000000][1.000000][8.820047][8.820047][0.042243],[0.138880] *****
> *****推荐[_CF_test] *****
> lowRank = 40, highRank = 59
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.021623][0.000673][548414.971198][0.729217][1.000000][1.000000][9.977438][9.977438][0.024653],[0.086490] *****
> *****推荐[_CF_test] *****
> lowRank = 60, highRank = 79
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [1.000000][1.000000][10.602026][10.602026][0.016243],[0.061290] *****
> *****
> lowRank = 80, highRank = 99
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.012850][0.000394][11.164313][11.164313][0.010032],[0.043109] *****
```


分析

- 将用户的兴趣特征与商品的特征映射到同一个空间平面
- 用户点击一个商品，意味着意识中对此商品所代表的特征空间中的一个具备特定特征的点感兴趣，此点在平面上突起
- 推荐系统的目标是挖掘出该特征点附近的其它点(相近商品)
- 在用户的特征空间中，点击过的特征点形成了一个突尖，同时衬托起一片临近区域，特征点点击次数越多，用户对该特征点的confidence(implicit feedback)越高，突尖越高，临近区域也越高，用户感兴趣的probability越高，但突尖仍然是具有最高probability的点
- 所以用户的频繁点击商品仍然是用户最可能点击的点

分析

- 若用户心目中的理想特征点一直未找到，而找到了临近的特征点且频繁点击，这样理想点将被临近的特征点叠加衬托到最高，从而获得用户心目中最佳特征(推荐)商品



基于用户最近15天全体建相似度

```
> *****推荐[_CF_test]*****
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.568941][0.015374][1.4054515.148734][0.398517][0.793326][1.000000][14.095460][23.761890][0.002100],[0.010400] *****
> *****mainpageTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性 *****
> ***** [0.071250][0.001940][2094037.034526][0.006512][0.013024][24.000000][33.000000] *****
> *****allTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性 *****
> ***** [0.214987][0.005593][4858879.231318][0.006512][0.013024][15.000000][22.000000] *****
> *****sevenDaysTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性 *****
> ***** [0.243916][0.006315][5669468.810301][0.006512][0.013024][21.000000][26.000000] *****
> *****推荐[_CF_test] *****
> lowRank = 0, highRank = 19
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.469711][0.012154][11631238.555044][0.739973][1.000000][1.000000][6.937080][6.937080][0.085170],[0.265670] *****
> *****推荐[_CF_test] *****
> lowRank = 20, highRank = 39
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.044218][0.001438][1109485.551772][0.736494][1.000000][1.000000][8.647199][8.647199][0.043793],[0.144209] *****
> *****推荐[_CF_test] *****
> lowRank = 40, highRank = 59
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.024121][0.000790][589421.650809][0.730819][1.000000][1.000000][9.717527][9.717527][0.025622],[0.092857] *****
> *****推荐[_CF_test] *****
> lowRank = 60, highRank = 79
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [1.000000][1.000000][10.279148][10.279148][0.015902],[0.063036] *****
> *****
> lowRank = 80, highRank = 99
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.013575][0.000435][10.743469][10.743469][0.010391],[0.045085] *****
```

基于用户最近15按天建相似度

```
> *****推荐[_CF_test]*****
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.568941][0.015374][1.4054515.148734][0.398517][0.793326][1.000000][14.095460][23.761890][0.002100],[0.010400] *****
> *****mainpageTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性 *****
> ***** [0.071250][0.001940][2094037.034526][0.006512][0.013024][24.000000][33.000000] *****
> *****allTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性 *****
> ***** [0.214987][0.005593][4858879.231318][0.006512][0.013024][15.000000][22.000000] *****
> *****sevenDaysTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性 *****
> ***** [0.243916][0.006315][5669468.810301][0.006512][0.013024][21.000000][26.000000] *****
> *****推荐[_CF_test] *****
> lowRank = 0, highRank = 19
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.469711][0.012154][11631238.555044][0.739973][1.000000][1.000000][6.937080][6.937080][0.085170],[0.265670] *****
> *****推荐[_CF_test] *****
> lowRank = 20, highRank = 39
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.044218][0.001438][1109485.551772][0.736494][1.000000][1.000000][8.647199][8.647199][0.043793],[0.144209] *****
> *****推荐[_CF_test] *****
> lowRank = 40, highRank = 59
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.024121][0.000790][589421.650809][0.730819][1.000000][1.000000][9.717527][9.717527][0.025622],[0.092857] *****
> *****推荐[_CF_test] *****
> lowRank = 60, highRank = 79
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [1.000000][1.000000][10.279148][10.279148][0.015902],[0.063036] *****
> *****
> lowRank = 80, highRank = 99
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.013575][0.000435][10.743469][10.743469][0.010391],[0.045085] *****
```

基于用户最近30按天建相似度

```
> *****推荐[_CF_test]*****
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.568532][0.016000][14535347.168958][0.396059][0.780078][1.000000][12.939580][20.196960][0.003570],[0.016700] *****
> *****mainpageTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性 *****
> ***** [0.071763][0.001992][2163573.834842][0.006512][0.013024][24.000000][33.000000] *****
> *****allTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性 *****
> ***** [0.216395][0.005765][5006544.911476][0.006512][0.013024][15.000000][22.000000] *****
> *****sevenDaysTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性 *****
> ***** [0.244459][0.006489][5815772.670499][0.006512][0.013024][21.000000][26.000000] *****
> *****推荐[_CF_test] *****
> lowRank = 0, highRank = 19
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.479849][0.012968][12223354.765542][0.748945][1.000000][1.000000][7.029730][7.029730][0.079980],[0.248320] *****
> *****推荐[_CF_test] *****
> lowRank = 20, highRank = 39
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.043379][0.001518][1169664.871816][0.754142][1.000000][1.000000][7.782609][7.782609][0.062423],[0.210411] *****
> *****推荐[_CF_test] *****
> lowRank = 40, highRank = 59
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.020266][0.000695][520772.300747][0.757759][1.000000][1.000000][8.251805][8.251805][0.043823],[0.165590] *****
> *****推荐[_CF_test] *****
> lowRank = 60, highRank = 79
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [1.000000][1.000000][8.732129][8.732129][0.030262],[0.127140] *****
> *****
> *****
> lowRank = 80, highRank = 99
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比 *****
> ***** [0.010563][0.000339][9.109191][9.109191][0.021842],[0.099360] *****
```

基于用户最近15按天建相似度 15天预测

```
> *****推荐[_CF_test]*****
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比*****
> ***** [0.552019][0.015085][13698608.629501][0.421783][0.816810][1.000000][13.467540][21.056820][0.002200],[0.011890] *****
> *****mainpageTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性*****
> ***** [0.071252][0.001922][2089078.304771][0.006512][0.013024][24.000000][33.000000] *****
> *****allTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性*****
> ***** [0.215495][0.005520][4815391.841528][0.006512][0.013024][15.000000][22.000000] *****
> *****sevenDaysTOP召回*****
> ***** 召回,准确率,销售额,top50覆盖率,top100覆盖率,top50多样性,top100多样性*****
> ***** [0.243976][0.006205][5554533.660745][0.006512][0.013024][21.000000][26.000000] *****
> *****推荐[_CF_test] *****
> lowRank = 0, highRank = 19
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比*****
> ***** [0.451867][0.011847][11243030.265518][0.744316][1.000000][1.000000][7.196440][7.196440][0.067760],[0.224580] *****
> *****推荐[_CF_test] *****
> lowRank = 20, highRank = 39
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比*****
> ***** [0.048017][0.001576][1204687.171776][0.756743][1.000000][1.000000][8.119732][8.119732][0.048291],[0.174883] *****
> *****推荐[_CF_test] *****
> lowRank = 40, highRank = 59
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比*****
> ***** [0.023776][0.000768][597472.951067][0.755671][1.000000][1.000000][8.610341][8.610341][0.034522],[0.134437] *****
> *****推荐[_CF_test] *****
> lowRank = 60, highRank = 79
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比*****
> ***** [1.000000][1.000000][9.026873][9.026873][0.022033],[0.101302] *****
> *****
> lowRank = 80, highRank = 99
> ***** 召回,准确率,销售额,top10覆盖率,top50覆盖率,top100覆盖率,top50多样性,top100多样性,2个以下百分比,4个以下百分比*****
> ***** [0.012153][0.000374][1000000.000000][0.000000][0.000000][1.000000][9.373113][9.373113][0.016363],[0.080704] *****
```

分析

- ▶ 按天建相似度，能提高预测召回/准确度
- ▶ 将建相似度的时间由15天延长到30天，能提高预测召回/准确度 => 建相似度时间延长，能提高相似度矩阵的效果
- ▶ 将用户行为预测时间由7天延长到15天，将降低预测召回/准确度 => 将用户行为预测时间延长，将模糊用户的近期行为，降低用户的预测效果

The end