

基于订单关联规则 改善性需求推荐

赵海臣

背景

截至2017-06-14，聚美优品首页千人千面已经使用了

- ★ 基于MF ALS的用户长期个性模型
- ★ 基于CF的用户短期兴趣模型
- ★ 基于点击的用户直接兴趣模型
- ★ 基于ECPM的商品实时热门模型

考虑到另外一种情况：用户购买了一台全新笔记本电脑，使用一段时间后会逐渐提出一些改善性需求，例如一个更好的鼠标、鼠标垫、游戏键盘、散热垫。基于这种后期的改善性需求，可以在首页推荐中适当地加入一些订单关联性的商品。

关联规则的实施

✎ 只考虑 $A \Rightarrow B$ 的单商品关联关系，挖掘出 $A \Rightarrow B$ 的强相关关系，然后基于用户近期订单，基于近期购买商品推荐对应的强相关商品。

- ★ 挖掘最近一年订单中的频繁-2项集，获得项集的支持度Support，并过滤掉支持度小于SupportThr的项集；
- ★ 计算频繁-2项集的置信度Confidence，并过滤掉置信度小于ConfidenceThr的项集；
- ★ 由上两步，获得了强相关的 $A \Rightarrow B$ 商品关系；
- ★ 收集用户近期的订单商品，基于订单商品A使用强相关的关联规则 $A \Rightarrow B$ 获得推荐列表B

几点考虑

✎ 由于关联规则的主要场景不适用于用户推商品，因此基于订单的推荐主体方针为“**少而精**”，追求高召回率，高准确率，并且应该限制此引擎的推荐数量：

- ✧ 关联规则的主要适用场景为购物车的商品推商品，因此使用于用户订单推商品的推荐方式必须要有强相关关系的商品，否则将有可能适得其反；
- ✧ SupportThr与ConfidenceThr阈值应该设置到较高的水准，获得更强的相关关系；
- ✧ 用户订单中的单个商品关联出的商品应该设置一个数量限制，防止关联规则推荐的滥用；

测试方法

❧ 离线测试方法：

- ★ 由于基于关联规则的推荐数量有限，平均一个用户 ≤ 10 个，并且用户与用户之间数量不一致，因此直接与top100热销商品进行召回率/精确度/销售额对比并不合理
- ★ 可以将用户订单关联规则推荐替换cf或者als推荐的top100末端商品，然后间接通过与原始的cf或者als推荐进行召回率/精确度/销售额的对比来评估订单关联规则推荐的效果。

● 问题：

- 订单推荐的用户-商品可能与cf/als重合
- 需要准确地将cf/als的末端相应数量的商品给挤出去

● 解决途径：

- 将订单推荐的商品分数加上100，一定大于所有cf/als的分数
- 将两个列表融合，相同用户-商品分数融合相加，再取Top100，就能保证订单推荐的商品一定在列表，并且将cf/als最后商品给挤出去。

测试指标参数

参数:

- ★ OrderConcernTimeAcross : Last 365 days
- ★ UserOrderTimeAcross : Last 15 days
- ★ supportThr : 5.0
- ★ confidenceThr : 1%
- ★ product => products Maximum Num : 5

结果指标:

- ★ TotalRecommendUsers : 1323688
- ★ ProductsPerUser : 8.978808450329685

cf结果(基于订单推荐用户)

*****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*****

* * * * *
[0.420196][0.009504][8271699.520726][0.392275][0.773953][1.000000][14.030577][21.573782][0.003583],[0.014379] *****

*****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*****

***** [0.084131][0.001768][1773225.772881][0.006231][0.012461][26.000000][37.000000] *****

*****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*****

***** [0.199087][0.003781][3453821.109138][0.006231][0.012461][20.000000][27.000000] *****

*****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*****

***** [0.243048][0.004507][3987458.368582][0.006231][0.012461][19.000000][27.000000] *****

混合结果(基于订单推荐用户)

*****推荐[_combinedTest]*****

***** 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*****

* * * * *
[0.427488][0.009668][8380350.450931][0.259221][0.746228][1.000000][14.386620][20.673350][0.021000],[0.077960] *****

*****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*****

***** [0.084131][0.001768][1773225.772881][0.006231][0.012461][26.000000][37.000000] *****

*****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*****

***** [0.199087][0.003781][3453821.109138][0.006231][0.012461][20.000000][27.000000] *****

*****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*****

***** [0.243048][0.004507][3987458.368582][0.006231][0.012461][19.000000][27.000000] *****

结论

✎ 基于订单关联规则的推荐能提升离线评测1%-2%的召回率/精确度/销售额；

与近期模型结果的融合方法

✎考虑到订单推荐的辅助性质与应用场景，用于用户推荐应该确保较优质的推荐结果，尽量精简，宁缺毋滥：

★ 增加关联规则的置信度、支持度阈值，提高关联强度：

- OrderConcernTimeAcross : Last 365 days
- UserOrderTimeAcross : Last 15 days
- supportThr : 10.0
- confidenceThr : 5%
- per product => Maximum Associated Products Num : 2

★ 将最后的结果按照分数排名分别插入原始推荐列表的排名10, 20, 30, ... 100，即最多插入10个订单关联的商品，分散在top100推荐商品排名第10, 20, 30, 40...,100位。

融合结果(基于订单推荐用户)

*****推荐[Combine_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 *****

* * * * *
[0.451660][0.010206][8416349.922334][0.381608][0.769038][1.000000][12.409220][18.654430][0.113200],[0.139690] *****

*****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 *****

***** [0.071872][0.001611][1701487.182503][0.006252][0.012503][27.000000][37.000000] *****

*****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 *****

***** [0.227726][0.004079][3277626.521610][0.006252][0.012503][21.000000][27.000000] *****

*****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 *****

***** [0.272240][0.004861][3713474.401564][0.006252][0.012503][19.000000][28.000000] *****

融合结果分段(基于订单推荐用户)

*****推荐[_Combine_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*****

***** [0.379421][0.008200][6994513.561887][0.736013][1.000000][1.000000][6.888010][6.888010][0.167270],[0.301170]*****

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*****

***** [0.036037][0.001008][709238.620147][0.751488][1.000000][1.000000][8.319644][8.319644][0.061416],[0.181910]*****

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*****

***** [0.016642][0.000464][319548.380083][0.756477][1.000000][1.000000][8.408103][8.408103][0.058205],[0.170411]*****

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*****

***** [0.010784][0.000297][211121.700119][0.756645][1.000000][1.000000][8.768631][8.768631][0.033191],[0.140929]*****

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*****

***** [0.008446][0.000228][175012.850124][0.756822][1.000000][1.000000][9.115847][9.115847][0.026316],[0.117909]*****

cf+点击(基于订单推荐用户)

*****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*****

* * * * *
[0.447799][0.010093][8320708.412263][0.392107][0.774629][1.000000][13.616908][20.855827][0.004330],[0.017102] *****

*****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*****

***** [0.071872][0.001611][1701487.182503][0.006252][0.012503][27.000000][37.000000] *****

*****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*****

***** [0.227726][0.004079][3277626.521610][0.006252][0.012503][21.000000][27.000000] *****

*****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*****

***** [0.272240][0.004861][3713474.401564][0.006252][0.012503][19.000000][28.000000] *****

cf+点击分段(基于订单推荐用户)

*****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*****

***** [0.378017][0.008156][6944026.741853][0.748794][1.000000][1.000000][7.438704][7.438704][0.083229],[0.229268]*****

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*****

***** [0.034490][0.000965][676419.570133][0.755911][1.000000][1.000000][8.035632][8.035632][0.077966],[0.211694]*****

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*****

***** [0.016168][0.000452][316643.840046][0.757669][1.000000][1.000000][8.318665][8.318665][0.063272],[0.177563]*****

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*****

***** [0.010415][0.000287][203682.620112][0.756322][1.000000][1.000000][8.758842][8.758842][0.034020],[0.141470]*****

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*****

***** [0.008288][0.000221][172059.500119][0.755038][1.000000][1.000000][9.128423][9.128423][0.026350],[0.117471]*****

分析

- ✎ 基于订单关联规则的推荐能提升离线评测1%-2%的召回率/精确度/销售额；
- ✎ 从分段结果看来，每个段的效果都有对应的提升，因此基于订单关联规则的结果可以较好地融合。

THE END

THANK YOU!