# user-based cf

基于用户相似度的协同过滤算法得到了用户相似度矩阵。主要可以分为对加购行为计算用户相似度和对浏览行为计算用户相似度。

1、基于加购行为

通过user找相似user（user-based CF），发现user行为太稀疏了，每个user一般只对4个以内的item有行为。即使是通过行为次数很多的user找相似user，对应的相似的user的行为次数也只有3~4个，所以没有看出特别有价值的规律

2、基于浏览行为

由于浏览行为相比于加购行为更为密集，因此比较容易找到有相似行为的用户，下面给出两个例子具体分析，分别给出一个user\_id，以及与其最相似的2个user\_id的浏览和购物历史：

案例一：

用户：dabb65e11aeb672d6f117a51bee1803d

行为时间 浏览 加购 商品名

"20150617" 1 0 "贝亲--奶瓶清洗剂400ml(新)"

"20150617" 1 0 "费雪牌--声光安抚海马-粉色"

"20150617" 1 0 "婴唯爱--天然玉米胚芽爽身粉100g"

"20150617" 1 1 "贝亲--浓缩型衣物柔软剂补充装500ml"

"20150617" 1 1 "晨光--可水洗水彩笔12色"

"20150617" 1 1 "贝亲--浓缩型衣物清洗剂补充装500ml"

"20150618" 1 0 "晨光--可水洗水彩笔12色"

"20151020" 1 1 "贝亲(喂养)--婴儿柔湿巾10\*6片/包"

"20151020" 1 0 "贝亲(喂养)--婴儿柔湿巾25\*4片/包"

"20151020" 1 1 "贝亲(喂养)--婴儿柔湿巾10\*6片/包"

"20151020" 1 0 "惠步舒(新)--TXH20370机能鞋TXH20370灰14cm双"

"20151020" 1 0 "基诺浦(新)--TXG2023机能鞋TXG2023红13.5cm双"

"20151020" 1 0 "贝亲(喂养)--婴儿柔湿巾25\*4片/包"

"20151020" 1 0 "惠步舒(新)--TXH20370机能鞋TXH20370灰14cm双"

"20151020" 1 0 "基诺浦(新)--TXG2023机能鞋TXG2023红13.5cm双"

用户：f1b57413264d60f8a3769bb5912630b4

行为时间 浏览 加购 商品名

"20150618" 0 0 "好奇--金装超值装初生号NB70片"

"20150618" 1 0 "爱他美--4段幼儿配方奶粉(12-24月)800g/桶"

"20150618" 0 0 "贝亲(喂养)--婴儿抗菌洗衣皂120g\*4块/包"

"20150707" 1 0 "贝亲(喂养)--婴儿抗菌洗衣皂120g\*4块/包"

"20150707" 1 1 "贝亲(喂养)--婴儿柔湿巾10\*6片/包"

"20150731" 0 0 "贝亲--防溢乳垫36片装(塑料袋装)"

"20150731" 0 0 "日康--婴儿专用剪刀"

"20150731" 0 0 "贝亲--婴儿柔湿巾80片3连包"

"20150731" 0 0 "贝亲--细轴棉棒180支"

"20150731" 0 0 "B&B--纤维洗涤剂(香草香)补充装1300ml"

"20151019" 0 0 "屁屁乐--专业护臀霜60G"

"20151019" 0 0 "贝亲--婴儿沐浴露200ml"

"20151019" 0 0 "贝亲--婴儿沐浴露200ml"

"20151019" 0 0 "屁屁乐--专业护臀霜60G"

"20151111" 1 0 "好奇--金装超值装初生号NB70片"

"20151111" 1 2 "花王--日本原装进口纸尿裤NB/90片(宝宝店专供)"

"20151111" 1 0 "贝亲--婴儿柔湿巾80片装"

"20151111" 0 0 "贝亲--婴儿柔湿巾80片3连包"

"20151111" 1 0 "好奇--金装超值装初生号NB70片"

"20151111" 1 2 "花王--日本原装进口纸尿裤NB/90片(宝宝店专供)"

"20151111" 0 0 "贝亲--婴儿柔湿巾80片3连包"

"20151111" 1 0 "贝亲--婴儿柔湿巾80片装"

用户：5852c9dc779f807f4578981cec2b2175

行为时间 浏览 加购 商品名

"20151007" 1 0 "歌瑞贝儿(新)--背袋熊立体袜/2双装GB130-212褐色12-24个月月包"

"20151007" 1 0 "贝亲—婴儿手口湿巾70片装(无酒精)"

"20151007" 1 0 "贝亲(喂养)--婴儿手口湿巾25\*4片/包"

"20151007" 0 0 "歌瑞贝儿(新)--时尚条纹袜/2双装GB141-013L混色24-36个月月包"

"20151007" 2 1 "贝亲(喂养)--婴儿柔湿巾10\*6片/包"

"20151007" 0 0 "歌瑞贝儿(新)--中性雪人毛巾袜GB154-029混色24-36个月包"

"20151007" 0 0 "歌瑞贝儿(新)--时尚条纹袜/2双装GB141-013L混色24-36个月月包"

"20151007" 1 0 "歌瑞贝儿(新)--男童条纹学步袜GB154-005混色6-12个月包"

"20151007" 1 0 "贝亲(喂养)--婴儿手口湿巾25\*4片/包"

"20151007" 1 0 "歌瑞贝儿(新)--背袋熊立体袜/2双装GB130-212褐色12-24个月月包"

"20151007" 1 0 "贝亲—婴儿手口湿巾70片装(无酒精)"

"20151007" 1 0 "歌瑞贝儿(新)--男童条纹学步袜GB154-005混色6-12个月包"

"20151007" 0 0 "歌瑞贝儿(新)--中性雪人毛巾袜GB154-029混色24-36个月包"

"20151007" 2 1 "贝亲(喂养)--婴儿柔湿巾10\*6片/包"

案例二

用户：151e9c1c6c2e2f94419988dd0c1ade97

行为时间 浏览 加购 商品名

"20150619" 0 0 "贝亲--自然实感宽口径玻璃奶瓶240ml(黄色)"

"20150619" 0 0 "贝亲(喂养)--新安抚奶嘴潜水艇(6-18月)1个"

"20150619" 0 0 "贝亲(喂养)--新安抚奶嘴小蓝鲸(3-6月)1个"

"20150619" 0 0 "津娃--迪娜娅浴盆L"

"20150623" 1 0 "贝亲--婴儿粉扑"

"20150623" 0 0 "贝亲--细轴棉棒180支"

"20150623" 0 0 "贝亲--耳孔清洁细轴棉棒"

"20150623" 1 0 "婴之侣--洗头帽"

"20150624" 1 0 "1500积分礼品--拼装益智积木玩具"

"20150624" 1 0 "歌瑞贝儿Great Baby(网)--单面提花皇冠短裤=Q131蓝90cm件"

"20150624" 1 0 "歌瑞凯儿Great Kid(网)--男童炫色图案背心（夏）GK15W-077蓝80件"

"20150624" 0 0 "歌瑞贝儿Great Baby(T)--40支双面布顽皮小恐龙对襟上衣GB133-2005VQ蓝80cmGB133-2005VQ蓝80CM件"

"20150624" 1 1 "歌瑞贝儿Great Baby(网)--莫代尔蜜蜂比尔可拆密档短裤=Q265蓝73cm件"

"20150624" 1 0 "歌瑞凯儿Great Kid(网)--男童炫色图案背心（夏）GK15W-077绿80件"

"20150624" 1 1 "歌瑞贝儿Great Baby(网)--莫代尔蜜蜂比尔背心=Q253蓝66cm件"

"20150624" 1 1 "歌瑞凯儿Great Kid(网)--男童炫色纯棉短裤（夏）GK15W-079蓝80件"

"20150624" 1 0 "歌瑞贝儿Great Baby(网)--圣麻提花单面布无骨缝短袖肩开套=Q58蓝80cm件"

"20150624" 1 0 "歌瑞凯儿Great Kid(网)--男童炫色纯棉五分裤（夏）GK15W-073灰80件"

"20150625" 1 0 "歌瑞贝儿Great Baby(网)--圣麻提花单面布无骨缝短袖套=Q61蓝110cm件"

"20150625" 1 0 "歌瑞贝儿Great Baby(网)--莫代尔蜜蜂比尔背心=Q247蓝66cm件"

"20150625" 1 0 "歌瑞贝儿Great Baby(网)--圣麻提花单面布无骨缝短袖肩开套=Q58蓝80cm件"

"20150626" 0 0 "贝亲--magmag鸭嘴式宝宝杯(绿色)"

"20150626" 0 0 "歌瑞贝儿Great Baby(网)--圣麻提花单面布无骨缝短袖套=Q61蓝110cm件"

"20150626" 1 1 "歌瑞贝儿Great Baby(网)--圣麻提花单面布无骨缝短袖肩开套=Q58蓝80cm件"

"20150626" 1 0 "津娃--儿童专用不锈钢餐碗(0岁以上)1个/盒"

"20150626" 0 0 "澳贝--好问爬行小蟹"

"20150626" 0 0 "歌瑞贝儿Great Baby(网)--单面提花皇冠短裤=Q131蓝90cm件"

"20150626" 0 0 "歌瑞凯儿Great Kid(网)--男童炫色纯棉短裤（夏）GK15W-079蓝120件"

用户：981b067e9e5233825deb51fa0ae3a407

行为时间 浏览 加购 商品名

"20150604" 1 0 "歌瑞贝儿Great Baby(网)--莫代尔蜜蜂比尔背心=Q253蓝66cm件"

"20150604" 1 0 "歌瑞凯儿Great Kid(网)--男童炫色图案背心（夏）GK15W-077绿80件"

"20150604" 1 0 "歌瑞贝儿Great Baby(网)--莫代尔蜜蜂比尔背心=Q247蓝66cm件"

"20150604" 1 1 "歌瑞贝儿Great Baby(网)--圣麻提花单面布无骨缝短袖肩开套=Q58蓝80cm件"

用户：e5aff3539a6b168d84345ddb6ef9186e

行为时间 浏览 加购 商品名

"20150707" 0 0 "歌瑞凯儿Great Kid(网)--女童炫色图案短袖T恤（夏）GK15W-068蓝90件"

"20150707" 1 0 "歌瑞贝儿Great Baby(网)--莫代尔蜜蜂比尔背心=Q247蓝66cm件"

"20150707" 0 0 "歌瑞凯儿Great Kid(网)--女童炫色图案短袖T恤（夏）GK15W-068黄90件"

"20150707" 1 0 "歌瑞凯儿Great Kid(网)--男童炫色纯棉五分裤（夏）GK15W-073灰80件"

"20150707" 2 1 "歌瑞凯儿Great Kid(网)--女童炫色梭织短裤（夏）GK15W-074咖啡80件"

"20150707" 1 1 "歌瑞贝儿Great Baby(网)--圣麻提花单面布无骨缝短袖肩开套=Q58蓝80cm件"

"20150707" 1 0 "歌瑞凯儿Great Kid(网)--男童炫色图案背心（夏）GK15W-077蓝80件"

"20150707" 0 0 "歌瑞凯儿Great Kid(网)--女童炫色纯棉背心两件装（夏）GK15W-088混色100件"

"20150707" 1 0 "歌瑞凯儿Great Kid(网)--女童炫色梭织短裤（夏）GK15W-074玫红90件"

"20150707" 1 1 "歌瑞凯儿Great Kid(网)--女童炫色图案短袖T恤（夏）GK15W-068蓝80件"

"20150707" 1 0 "歌瑞贝儿Great Baby(网)--莫代尔蜜蜂比尔背心=Q253蓝66cm件"

"20150707" 0 0 "歌瑞凯儿Great Kid(网)--女童炫色图案短袖T恤（夏）GK15W-068桃红90件"

"20150707" 2 2 "歌瑞凯儿Great Kid(网)--女童炫色图案短袖T恤（夏）GK15W-068桃红80件"

"20150707" 1 0 "歌瑞贝儿Great Baby(网)--莫代尔蜜蜂比尔可拆密档短裤=Q265蓝73cm件"

"20150707" 1 1 "歌瑞凯儿Great Kid(网)--女童炫色纯棉背心两件装（夏）GK15W-088混色80件"

"20150725" 0 0 "歌瑞凯儿Great Kid(网)--男童炫色条纹短袖T恤（夏）GK15W-069绿100件"

"20150725" 0 0 "歌瑞贝儿(新)--男生小熊三角内裤GB140-3002混色90CM包"

"20150725" 0 0 "歌瑞凯儿Great Kid(网)--女童炫色图案背心（夏）GK15W-090蓝100件"

"20150725" 0 0 "金盾婴宝--婴儿驱蚊手环(4条)"

"20150725" 0 0 "碧呵--艾草防蚊贴（36+6）片/盒"

"20150725" 0 0 "歌瑞贝儿(新)--男生小熊三角内裤GB140-3002混色90CM包"

"20150914" 0 0 "贝亲-标准口径塑料奶瓶(PP)240ML"

"20150914" 0 0 "贝亲—自然实感宽口径奶嘴(L)单个盒装"

"20150914" 0 0 "贝亲--标准奶嘴单个装L"

从案例一可以看出，这3个用户都是对“贝亲”牌的“婴儿柔湿巾”感兴趣的，但其关注的其他商品又互有不同。因此，我们可以将user1感兴趣的"贝亲--浓缩型衣物柔软剂补充装500ml"推荐给其他两个user，同理，也可以将user2的"贝亲--婴儿沐浴露200ml"等商品推荐给user1和user2。

案例二则是找出了对“歌瑞贝儿”牌的背心和短裤感兴趣的相似用户，也可以同理推荐其他类的商品。

通过案例分析可以发现，基于用户相似度的分析不仅可以发现有类似偏好的用户集合，还可以给用户集合内的用户交叉推荐感兴趣的商品，这样的商品不同于基于商品相似度的推荐，更有可能是不同类别的商品，因此更有新鲜感，同时又保证了用户对所推荐商品会比较感兴趣。

# prefixspan

该算法是一种序列模式挖掘算法，可以统计出经常出现的序列模式及其出现的次数。下面给出在乐友的加购数据上处理后得到的结果的案例分析：

序列模式 出现次数

[["贝亲--自然实感宽口径玻璃奶瓶240ml(绿色)"], ["贝亲—自然实感宽口径奶嘴(L)单个盒装"]] 10

[["启赋--幼儿配方奶粉(1-3岁)900g/听"], ["启赋--幼儿配方奶粉(1-3岁)900g/听"]] 58

[["花王--日本原装进口纸尿裤XL44片(宝宝店专供)"], ["花王--日本原装进口学步裤XL/38片(宝宝店专供)"]] 20

从第1行可以看出，奶瓶配奶嘴是经常一起出现的组合，可以在用户购买过奶瓶后推荐奶嘴，或者搭配促销出售。

从第2行可以看出，该幼儿配方奶粉的重复购买率很高，因此可以在用户购买过该品牌奶粉后继续重复推荐或者促销。

从第3行可以看出，纸尿裤和学步裤是一个很好的搭配推荐的组合。

通过案例分析可以发现，从该算法所挖掘出的序列模式中可以提取出很多适合搭配推荐的商品和容易重复购买的商品。