

# 西门子订单数据分析过程

## 0. 问题

Q1：是否能够获取到商品的唯一标识

## 1. 原始数据处理

### 1.1 数据字段

#### 1.1.1 ExportOrderList201905101349（D1）

字段名	示例
订单编号	272676430516911307
买家会员名	hzzwmwg1111
买家支付宝账号	wg1111@126.com
买家应付货款	29.00
买家应付邮费	0.00
买家支付积分	0
总金额	29.00
返点积分	14
买家实际支付金额	29.00
买家实际支付积分	0
订单状态	交易成功
买家留言	（空）
收货人姓名	丁昊
收货地址	上海 上海市 虹口区 纪念路 480 号水利部太湖流域管理局 309 室(200434)
运送方式	快递
联系电话	（空）
联系手机	‘15902198734
订单创建时间	2019-02-03 17:21:06
订单付款时间	2019-02-03 17:21:13
宝贝标题	西门子闭路电视信号分配器 数字有线电视分支器一分三 天线分线器
宝贝种类	1
物流单号	No:75128315520361
物流公司	中通快递
订单备注	（空）
宝贝总数量	1
店铺 Id	（空）
店铺名称	siemens 雍宸专卖店
确认收货时间	2019-02-25 10:55:46
打款商家金额	29.00 元
是否村淘订单	否

#### 1.1.2 ExportOrderList201905101357（D2）

字段名	示例
订单编号	438762080881991151
买家会员名	oy 绅士
买家支付宝账号	oyss345572968@yahoo.cn
支付单号	2019050922001115891037261426
支付详情	支付方式：支付宝，支付单号：2019050922001115891037261426，金额：32.00；
买家应付货款	32.00
买家应付邮费	0.00
买家支付积分	0
总金额	32.00
返点积分	16
买家实际支付金额	32.00
买家实际支付积分	0
订单状态	交易成功
买家留言	null

收货人姓名	欧阳
收货地址	上海 上海市 静安区 北站街道上海市静安区新疆路 425 号 204 室 (200070)
运送方式	快递
联系电话	‘null
联系手机	‘18116309248
订单创建时间	2019-05-09 16:08:36
订单付款时间	2019-05-09 16:08:40
宝贝标题	西门子闭路电视信号分配器 数字有线电视分支器一分三 天线分线器
宝贝种类	1
物流单号	No:75146943595908
物流公司	中通快递
订单备注	null
宝贝总数量	1
店铺 Id	0
店铺名称	siemens 雍宸专卖店
订单关闭原因	订单未关闭
卖家服务费	0
买家服务费	0 元
发票抬头	(空)
是否手机订单	手机订单
分阶段订单信息	null
特权订金订单 id	null
是否上传合同照片	否
是否上传小票	否
是否代付	否
定金排名	(空)
修改后的 sku	null
修改后的收货地址	null
异常信息	null
天猫卡券抵扣	null
集分宝抵扣	null
是否是 020 交易	null
新零售交易类型	电商订单
新零售导购门店名称	null
新零售导购门店 id	null
新零售发货门店名称	null
新零售发货门店 id	null
退款金额	0.00
预约门店	null
是否村淘订单	否
确认收货时间	2019-05-10 09:51:46
打款商家金额	32.00 元
含应开票给个人的个人红包	(空)
主订单编号	(无)

1.2 数据提取

1) 结合 D1 与 D2，提取部分数据字段（如下表所示），组成数据 **D3**，并进行简单的统计分析

字段名	统计结果
订单编号	共 45486 个订单
买家会员名	共涉及 33927 个用户
总金额	无
订单状态	共五种状态：交易成功、交易关闭、等待买家付款、买家已付款，等待卖家发货、卖家已发货，等待买家确认
收货地址	无
订单创建时间	2014 年-2019 年
宝贝标题	共 5690 个标题
宝贝种类	共 35 种
宝贝总数量	无

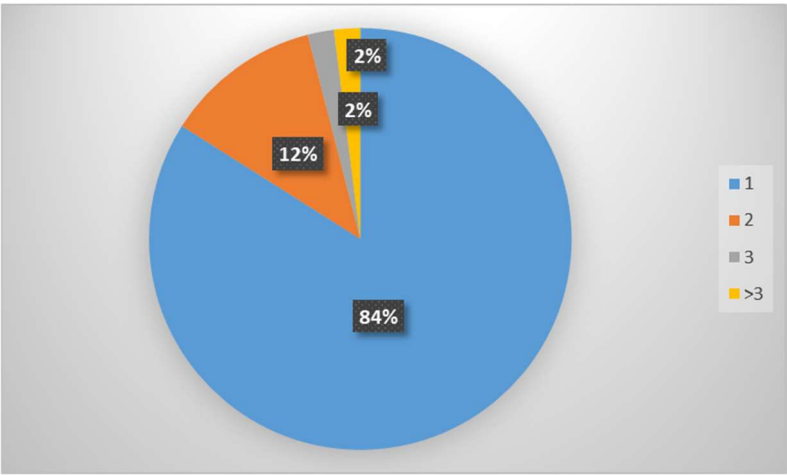
2) 根据 D3，筛选出订单状态除“交易关闭”之外的订单，组成数据 D4，并进行简单的统计分析

字段名	统计结果
订单编号	共 38319 个订单
买家会员名	共涉及 31202 个用户
总金额	无
订单状态	共四种状态：交易成功、等待买家付款、 买家已付款，等待卖家发货、卖家已发货，等待买家确认
收货地址	无
订单创建时间	2014 年-2019 年
宝贝标题	共 5005 个标题
宝贝种类	共 33 种
宝贝总数量	无

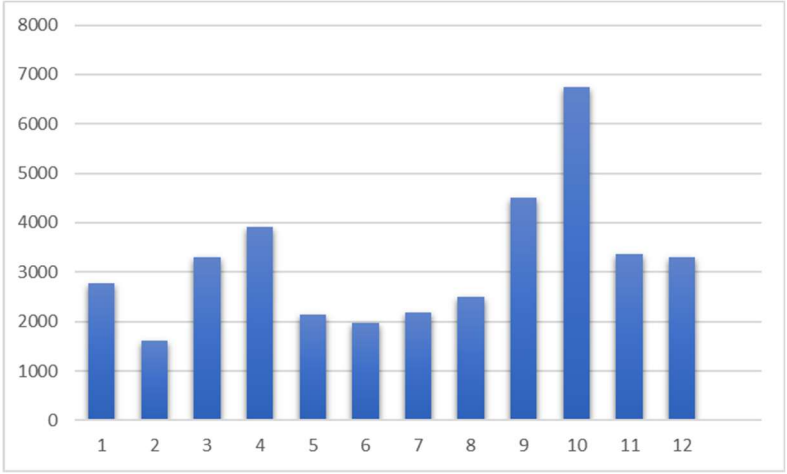
## 2. 数据分析

根据 D4，进一步进行数据统计分析

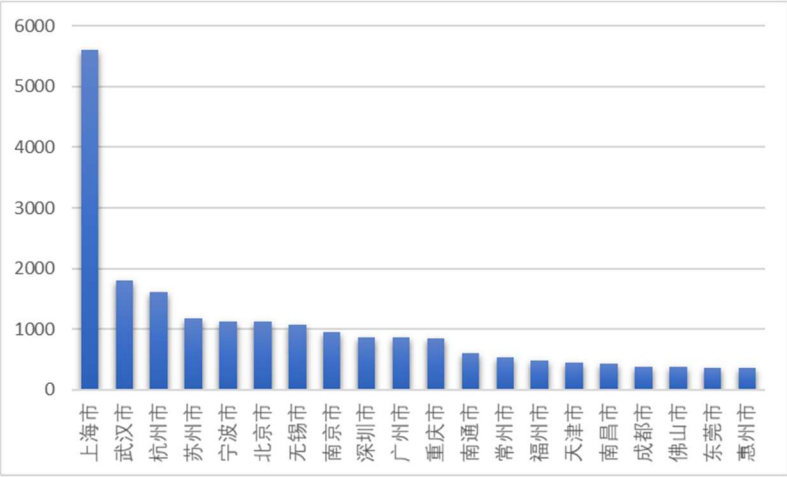
1) 统计每个用户购买次数，如下图所示，其中有 26318（84%）个用户只购买过一次，4884 个用户购买过两次以上



2) 统计不同月份的销售情况，如下图所示，可以看出销售旺季有 3 月、4 月、9 月、10 月、11 月、12 月



3) 统计不同城市的销售情况，如下图所示，可以看出上海、江苏、浙江、北京等城市销售量较多



4) 统计每个用户随时间的购买情况，并筛选出购买次数达到两次以上（含两次）的用户，组成数据

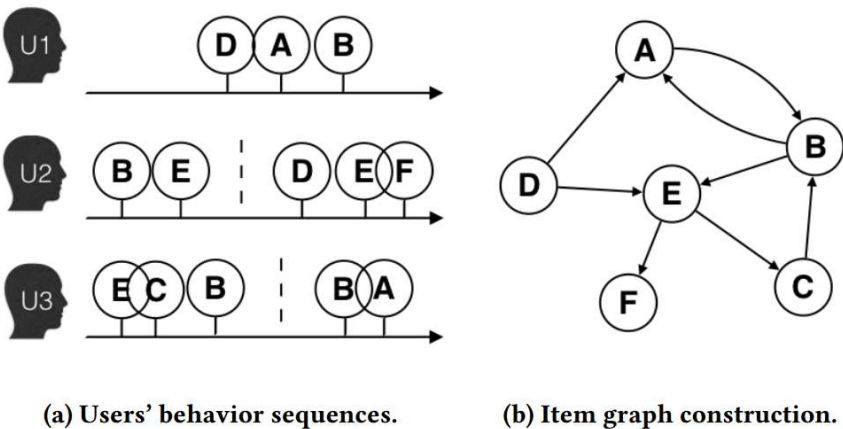
D5

字段名	统计结果
订单编号	共 12001 个订单
买家会员名	共涉及 4884 个用户
宝贝标题	共 2376 个标题
宝贝种类	共 31 种

3. 构建商品网络

3.1 构建方法

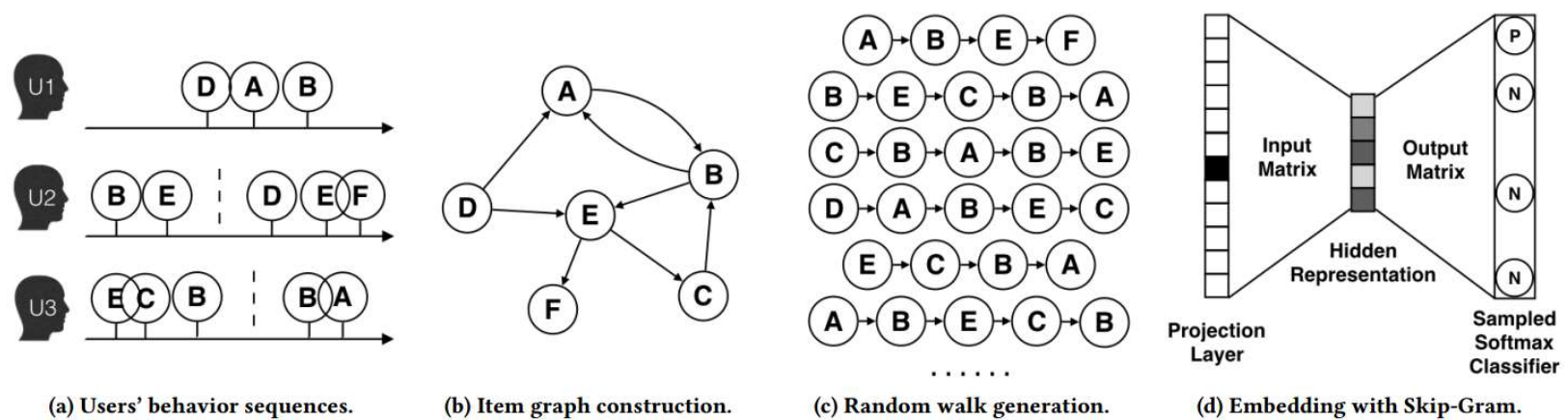
- 1) 如下图所示，首先根据不同用户的购买情况，生成用户随时间购买商品的序列（图（a））
- 2) 根据用户购买商品序列，生成对应的商品网络，比如用户 U1 按时间排序分别购买了商品 D->A->B, 则相应形成两条边，D->A, A->B（图（b））
- 3) 在本次应用中，考虑两种商品网络：
  - ① 节点为商品种类（TypeNet）
  - ② 节点为商品标题（NameNet）



3.2 DeepWalk

- 1) DeepWalk:

DeepWalk 是经典的 Graph Embedding 方法，2014 年提出，它的主要思想是在由物品组成的图结构上进行随机游走，产生大量物品序列，然后将这些物品序列作为训练样本输入 word2vec 进行训练，最后得到物品的 Embedding
- 2) 下面简单介绍一下 DeepWalk 实现过程，如下图所示，整个 DeepWalk 的算法流程可以分为四步：
  - ① 获取原始的用户行为序列（图（a））；
  - ② 基于这些用户行为序列构建物品相关图（图（b））；
  - ③ 采用随机游走的方式随机选择起始点，重新产生物品序列（图（c））；
  - ④ 最后将这些物品序列输入 word2vec 模型，生成最终的物品 Embedding 向量（图（d））。



### 3.3 实验结果

#### 3.3.1 TypeNet

① 节点数：31

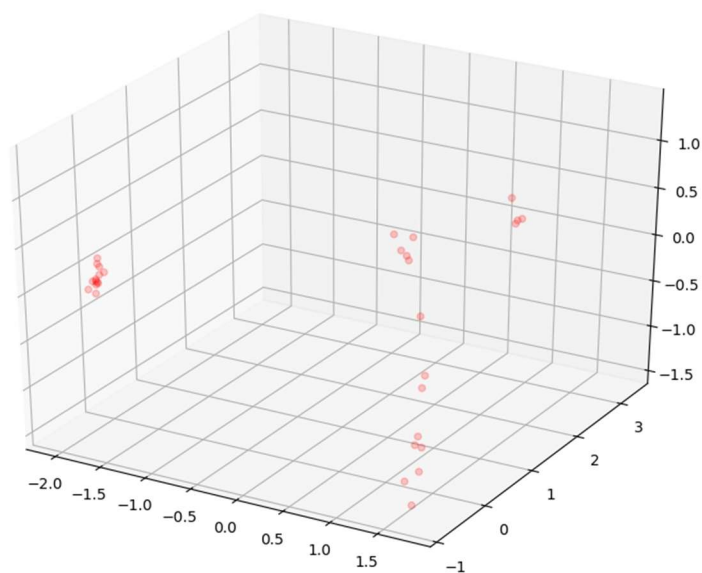
② 边数：199

1) DeepWalk 训练结果

```
1 -0.5310618 0.012471064 0.43551406 0.0071696276 -0.4182303 0.3698531 0.29231927 0.136051 -0.20419927 0.29894325
0.2961722 -0.40628257 0.11683022 -0.2433214 -0.22001658 -0.25407752 -0.2768643 0.11413346 -0.18499948 0.24833266 -
0.3093475 -0.32241055 0.35169527 0.35360438 -0.21931976 0.23014773 0.28058147 -0.58088666 -0.436246 -0.36429593
0.20434734 -0.05442135 -0.32711637 0.66186684 -0.15076835 -0.064011745 0.2509828 0.504061 0.6601866 0.44931963 -
0.25407088 -0.14481016 -0.055507198 -0.14369258 -0.67776316 -0.85663414 -0.5144873 0.094147645 0.29257363 0.0388294
-0.036610406 -0.39137092 0.19193947 -0.07914912 0.11129486 -0.43385264 0.43779477 0.045504775 -0.8697689 -0.4773096
0.14722769 0.06480824 -0.23501119 0.45101008
```

注：1 代表节点（宝贝种类），其后为 64 维的向量

2) 结果的三位展示图



#### 3.3.2 NameNet

① 节点数：2376

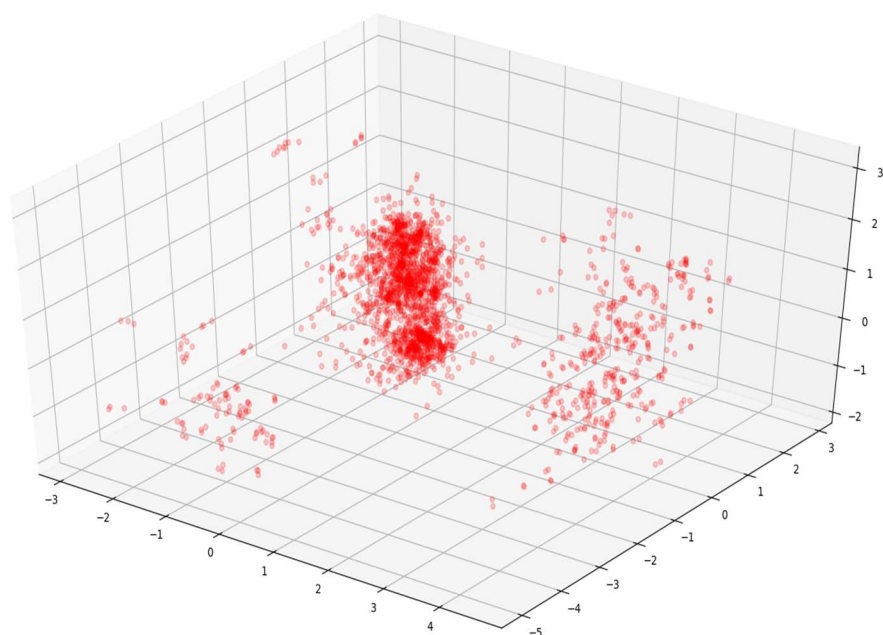
② 边数：3383

1) DeepWalk 训练结果

```
5 -0.14725839 0.95603466 -0.7432987 0.81592643 -1.1906476 0.002429132 -0.07700288 0.5907321 -0.54880166 0.5329487 -
0.5130559 0.92919916 0.06593084 1.2975826 -0.15412864 -1.1687489 2.2389393 -0.89899015 0.27200055 0.6618499 -
0.17494829 1.8289541 0.49643874 0.11597226 -0.39521137 1.116125 0.4495142 0.39502355 1.0659947 -0.6147371 1.5983468
0.8554322 -1.4558047 1.1803766 -0.50739145 1.3787873 0.5184076 2.5642583 -0.03773351 -0.18089643 -0.28835502
0.44952378 1.6980665 0.5570727 -0.27076483 0.52707726 2.0978897 0.3843645 1.1711103 -0.007653008 0.12323899
1.8586572 0.51664495 -1.4997358 -0.6950756 0.53869414 -1.1252033 -0.08559704 -1.7853224 0.29119915 0.10595711
0.8920225 -0.54930097 -0.91032135
```

注：5 代表节点（宝贝标题），其后为 64 维的向量

2) 结果的三位展示图



4. 未完待续...