（1）item\_feature:商品粒度相关特征

|  |  |  |  |
| --- | --- | --- | --- |
| 字段 | 类型 | 含义 | 示例 |
| date | String | 日期 | 20150912 |
| item\_id | String | 商品ID | 132 |
| cate\_id | String | 叶子类目ID | 18 |
| cate\_level\_id | String | 大类目ID | 12 |
| brand\_id | String | 品牌ID | 203 |
| supplier\_id | String | 供应商ID | 1976 |
| pv\_ipv | Double | 浏览次数 | 2 |
| pv\_uv | Double | 流量UV | 2 |
| cart\_ipv | Double | 被加购次数 | 0 |
| cart\_uv | Double | 加购人次 | 0 |
| collect\_uv | Double | 收藏夹人次 | 0 |
| num\_gmv | Double | 拍下笔数 | 0 |
| amt\_gmv | Double | 拍下金额 | 0 |
| qty\_gmv | Double | 拍下件数 | 0 |
| unum\_gmv | Double | 拍下UV | 0 |
| amt\_alipay | Double | 成交金额 | 0 |
| num\_alipay | Double | 成交笔数 | 0 |
| qty\_alipay | Double | 成交件数 | 0 |
| unum\_alipay | Double | 成交人次 | 0 |
| ztc\_pv\_ipv | Double | 直通车引导浏览次数 | 0 |
| tbk\_pv\_ipv | Double | 淘宝客引导浏览次数 | 0 |
| ss\_pv\_ipv | Double | 搜索引导浏览次数 | 0 |
| jhs\_pv\_ipv | Double | 聚划算引导浏览次数 | 0 |
| ztc\_pv\_uv | Double | 直通车引导浏览人次 | 0 |
| tbk\_pv\_uv | Double | 淘宝客引导浏览人次 | 0 |
| ss\_pv\_uv | Double | 搜索引导浏览人次 | 0 |
| jhs\_pv\_uv | Double | 聚划算引导浏览人次 | 0 |
| num\_alipay\_njhs | Double | 非聚划算支付笔数 | 0 |
| amt\_alipay\_njhs | Double | 非聚划算支付金额 | 0 |
| qty\_alipay\_njhs | Double | 非聚划算支付件数 | 0 |
| unum\_alipay\_njhs | Double | 非聚划算支付人次 | 0 |

（2）item\_store\_feature: 商品和分仓区域粒度相关特征

|  |  |  |  |
| --- | --- | --- | --- |
| 字段 | 类型 | 含义 | 示例 |
| date | String | 日期 | 20150912 |
| item\_id | String | 商品ID | 132 |
| store\_code | String | 仓库CODE | 1 |
| cate\_id | String | 叶子类目ID | 18 |
| cate\_level\_id | String | 大类目ID | 12 |
| brand\_id | String | 品牌ID | 203 |
| supplier\_id | String | 供应商ID | 1976 |
| pv\_ipv | Double | 浏览次数 | 2 |
| pv\_uv | Double | 流量UV | 2 |
| cart\_ipv | Double | 被加购次数 | 0 |
| cart\_uv | Double | 加购人次 | 0 |
| collect\_uv | Double | 收藏夹人次 | 0 |
| num\_gmv | Double | 拍下笔数 | 0 |
| amt\_gmv | Double | 拍下金额 | 0 |
| qty\_gmv | Double | 拍下件数 | 0 |
| unum\_gmv | Double | 拍下UV | 0 |
| amt\_alipay | Double | 成交金额 | 0 |
| num\_alipay | Double | 成交笔数 | 0 |
| qty\_alipay | Double | 成交件数 | 0 |
| unum\_alipay | Double | 成交人次 | 0 |
| ztc\_pv\_ipv | Double | 直通车引导浏览次数 | 0 |
| tbk\_pv\_ipv | Double | 淘宝客引导浏览次数 | 0 |
| ss\_pv\_ipv | Double | 搜索引导浏览次数 | 0 |
| jhs\_pv\_ipv | Double | 聚划算引导浏览次数 | 0 |
| ztc\_pv\_uv | Double | 直通车引导浏览人次 | 0 |
| tbk\_pv\_uv | Double | 淘宝客引导浏览人次 | 0 |
| ss\_pv\_uv | Double | 搜索引导浏览人次 | 0 |
| jhs\_pv\_uv | Double | 聚划算引导浏览人次 | 0 |
| num\_alipay\_njhs | Double | 非聚划算支付笔数 | 0 |
| amt\_alipay\_njhs | Double | 非聚划算支付金额 | 0 |
| qty\_alipay\_njhs | Double | 非聚划算支付件数 | 0 |
| unum\_alipay\_njhs | Double | 非聚划算支付人次 | 0 |

我们还提供每个商品在全国和分仓区域的补少、补多的成本，可以用来计算总成本

(3) config: 每个商品在全国和分仓区域的补少、补多的成本

|  |  |  |  |
| --- | --- | --- | --- |
| 字段 | 类型 | 含义 | 示例 |
| item\_id | String | 商品ID | 333442 |
| store\_code | String | 仓库CODE, 注意如果是全国成本，这一列是all | 1 |
| a\_b | String | 商品补少补多cost，用"\_"联接起来。前一个数是补少的成本，后一个是补多的成本 | 10.44\_20.88 |

target: 每个商品在全国和分仓区域的目标库存

|  |  |  |  |
| --- | --- | --- | --- |
| 字段 | 类型 | 含义 | 示例 |
| item\_id | String | 商品ID | 333442 |
| store\_code | String | 仓库CODE（如果是全国的target，这列是all） | 1 |
| target | Double | 未来两周的全国或分仓区域目标库存（如果store\_code为all，就是全国的target,否则就是分仓的target） | 30.0 |

数据：

20141010到20151227

目标

20151228-20160110

# Off-line 1.0

Use last 15~28 days for training, last 1~14 for testing.

## Train

[1,416]->[417,430]

## Test

[15,430]->[431,444]

# Off-line 2.0

consider the influence of 1212, select days just after 1111 for training, days after 1212 for testing.

## Train

[1,399]->[400,413]

## Test

[32,430]->[431,444]

# Off-line 3.0

consider the influence of 1212, select last 7 days for predict, last 8~14 for validation

## Train

[1,430]->[ 431, 437]

## Test

[8,437]->[438,444]

双11：398

双12：429

# Off-line 4.0

训练：通过预测4天与4天的差值训练回归模型

测试：提取7天的特征预测后7天

## Train

[430,433]->[ 434, 437]

## Test

[431,437]->[438,444]

# On-line

## Train

[1,430]->[ 431, 444]

## Test

成本可能也是某种重要特征？放入考虑

中值滤波掉脉冲信号，把常见的脉冲信号日期对应的值提取出来单独作为特征。

# 框架：多模型加权融合。

## 周期性模型：

表示按年的周期性，

输出：今年与去年的比值、周期性的显著性

## 相似度模型

对于那些15年上架的商品，寻找其他趋势相近的14年就上架了的商品来辅助预测

## 趋势模型

用中位数滤波后的曲线进行预测，包括直接回归预测14天总销量和单天多步预测再求和。

分别对每个item训练一个模型

分别对每个category训练同一模型

分别对每个销售量级的item训练一个模型

用所有数据训练一个模型。

重新定义损失函数：不是通过拟合每一天，而是计算总的销量，再把误差沿着每天传播回去。

## 特征模型

查找商品销量单天爆发增长的时间节点（包括双11和12），提取出来作为特征（可以考虑除以本身销量中位数或者所有商品当天销量来归一化）

提取商品趋势特征：向热or向凉

直接将目标归一化到[0,1]，将特征标准化。每个样本根据缩放程度赋予相应的权重（同时也作为一个特征），样本

统计各个维度(item, brand, supplier)的cost分布，重点研究前80%cost的来源（前400个item）

有用的直接特征：

15,17,21…

最有用的信息：

1、前k天的波形走势：提取n阶多项式拟合的参数和置信系数

2、移动平滑14天（可加权），看趋势

3、取item、category、supplier间的rank后作为特征，及其交叉组合

4、时间上的rank

5、纵轴取对数，再做预测（降低双十一十二的值）

6、考虑协方差、相关系数等

除了预测值，还给出预测的精度