目录

[1、数据集划分 2](#_Toc516467711)

[2、特征工程 3](#_Toc516467712)

[2.1、用户线下相关的特征 3](#_Toc516467713)

[2.2、商家相关的特征 5](#_Toc516467714)

[2.3、用户-商家交互特征 8](#_Toc516467715)

[2.4、优惠券相关的特征 11](#_Toc516467716)

[2.5、用户-优惠卷特征 12](#_Toc516467717)

[2.6、其它特征 13](#_Toc516467718)

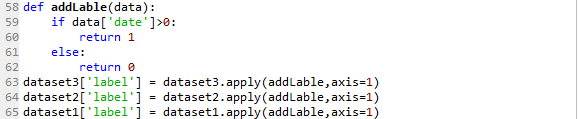
[3、特征合并 15](#_Toc516467719)

# 1、数据集划分

使用滑窗的方法得到多份训练数据集，特征区间越小，得到的训练数据集越多。

|  |  |  |
| --- | --- | --- |
|  | 预测区间（提取label） | 特征区间（提取feature） |
| 测试集 | 20160701-20160731 | 20160315-20160630 |
| 训练集1 | 20160515-20160615 | 20160201-20160514 |
| 训练集2 | 20160414-20160514 | 20160101-20160413 |





# 2、特征工程

从不同的角度出发，提取相关特征

## 2.1、用户线下相关的特征

用户领取优惠券次数 user\_couoon\_total\_received 1

用户获得优惠券但没有消费的次数 user\_coupon\_received\_notuse 2

用户获得优惠券并核销次数 user\_coupon\_received\_use 3

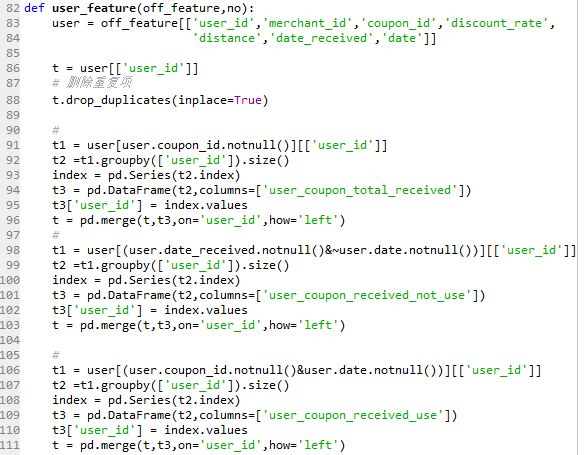
用户领取优惠券后进行核销率 user\_coupon\_use\_rate 4

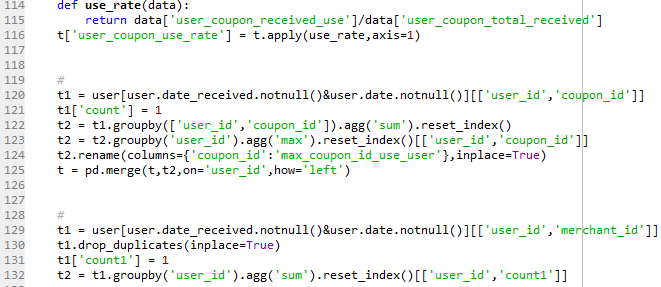
用户核销过优惠券的不同商家数量，及其占所有不同商家的比重 try\_fresh\_merchant 5

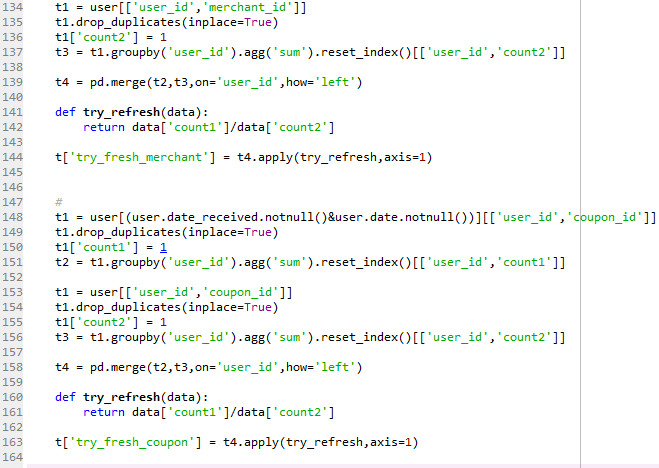
用户核销过的不同优惠券数量，及其占所有不同优惠券的比重 try\_fresh\_coupon 6

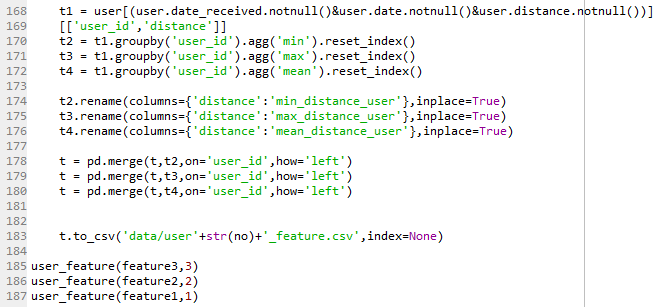
用户核销优惠券中的平均/最大/最小用户-商家距离 mean\_distance\_user/max\_distance\_user/min\_distance\_user 7/8/9

用户核销最多优惠券id max\_coupon\_use 10









## 2.2、商家相关的特征

商家优惠券被领取次数 coupon\_total 11

商家优惠券被领取后核销次数 coupon\_use 12

商家优惠券被领取后核销率 coupon\_use\_rate 13

商家被核销最多优惠券id max\_coupon\_id\_use\_merchant 14

核销商家优惠券的不同用户数量，及其占领取不同的用户比重 have\_fresh\_user 15

商家优惠券平均每个用户核销多少张 use\_per\_user 16

商家被核销过的不同优惠券数量 use\_diff\_coupon\_num 17

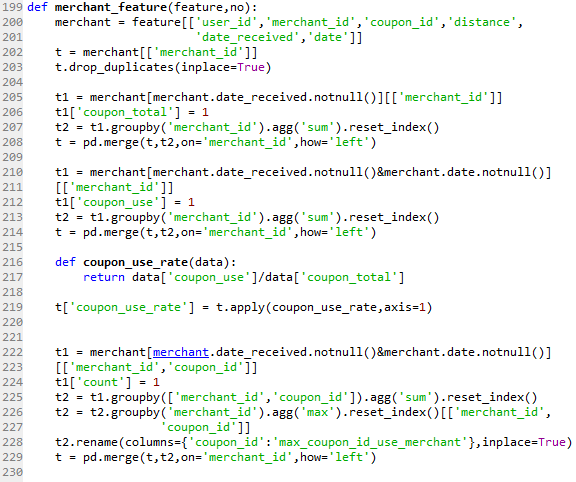
商家被核销过的不同优惠券数量占所有领取过的不同优惠券数量的比重 usr\_rate\_merchant 18

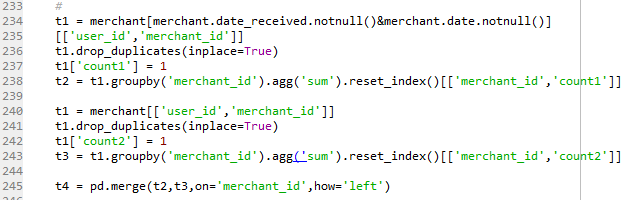
商家平均每种优惠券核销多少张 use\_per\_coupon 22

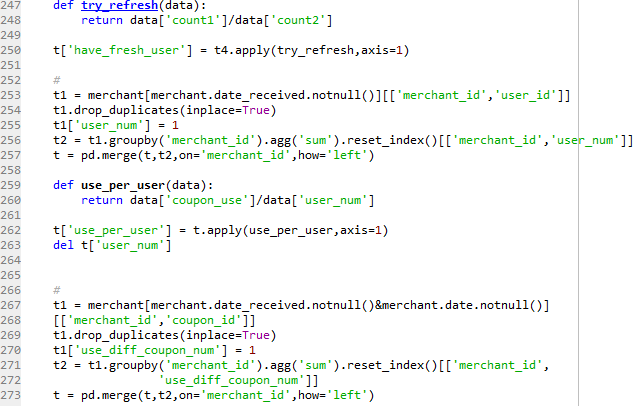
商家被核销优惠券的平均时间率 time\_gap\_per\_coupon 23

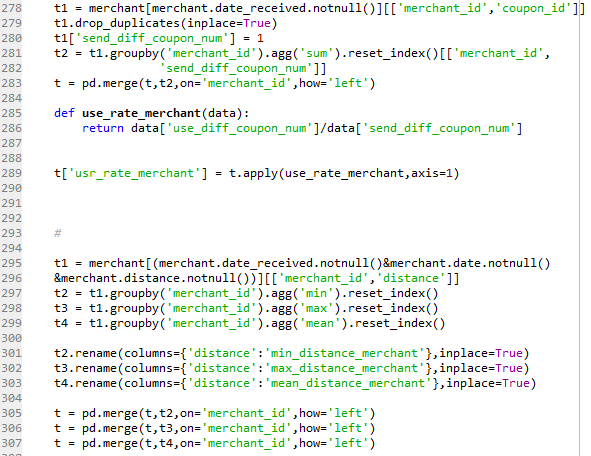
商家被核销优惠券中的平均/最小/最大用户-商家距离

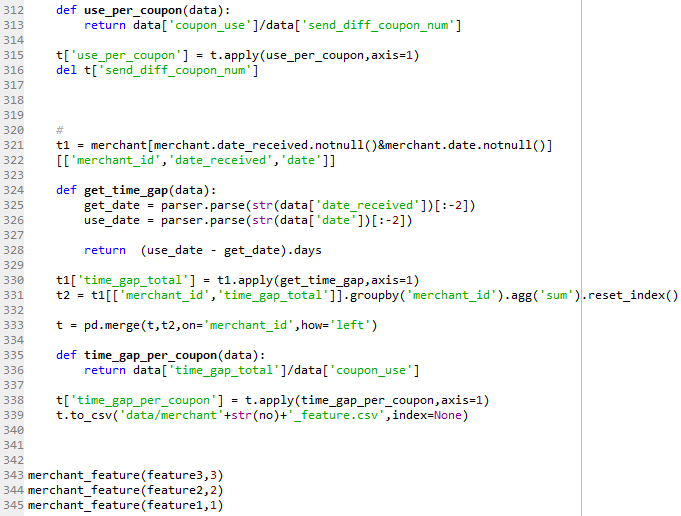
max\_distance\_merchant/mean\_distance\_merchant/min\_distance\_merchant 19/20/21











## 2.3、用户-商家交互特征

用户领取商家的优惠券次数 coupon\_receive\_times 24

用户领取商家的优惠券后不核销次数 coupon\_receive\_not\_use\_user\_merchant 25

用户领取商家的优惠券后核销次数

coupon\_receive\_use\_user\_merchant 26

用户领取商家的优惠券后核销率

coupon\_receive\_use\_user\_merchant\_rate 27

用户对每个商家的不核销次数占用户总的不核销次数的比重

user\_coupon\_received\_not\_use\_per\_merchant\_rate 28

用户对每个商家的优惠券核销次数占用户总的核销次数的比重

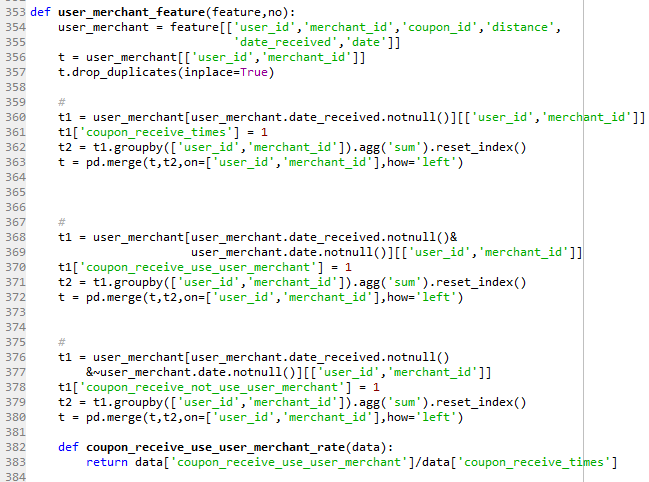
user\_coupon\_received\_use\_per\_merchant\_rate 29

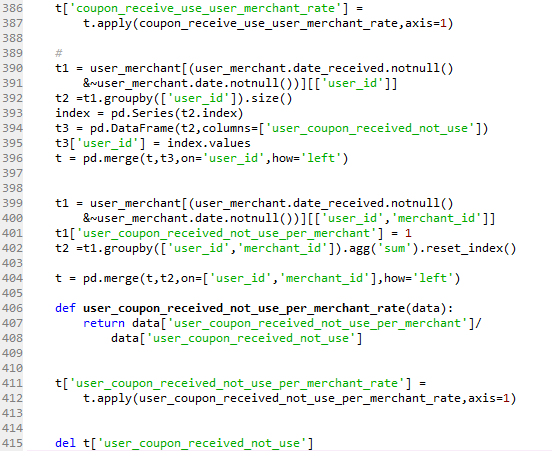
用户对每个商家的不核销次数占商家总的不核销次数的比重

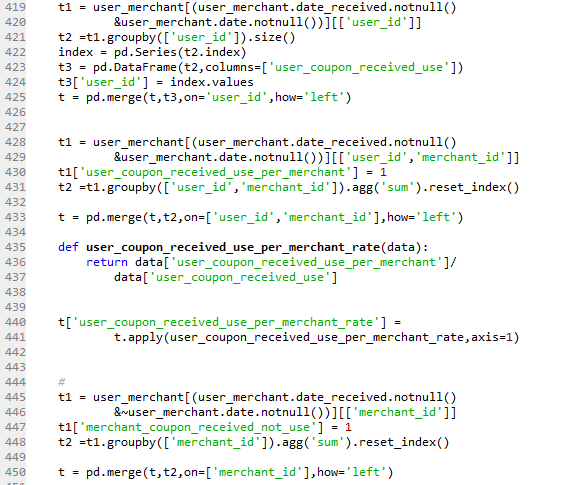
merchant\_coupon\_received\_not\_use\_per\_merchant\_rate 30

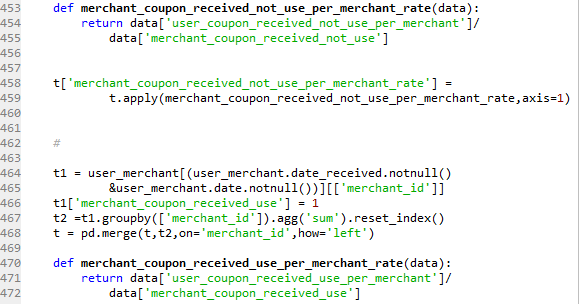
用户对每个商家的优惠券核销次数占商家总的核销次数的比重

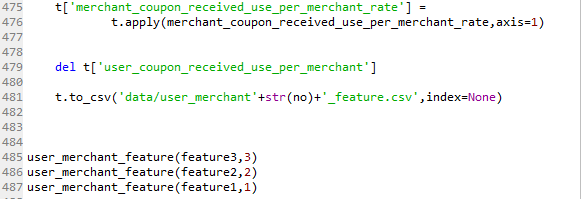
merchant\_coupon\_received\_use\_per\_merchant\_rate 31











## 2.4、优惠券相关的特征

优惠券类型(直接优惠为0, 满减为1) is\_full\_cut 32

优惠券折率 discount\_rate\_fair 33

历史出现次数 coupon\_count 34

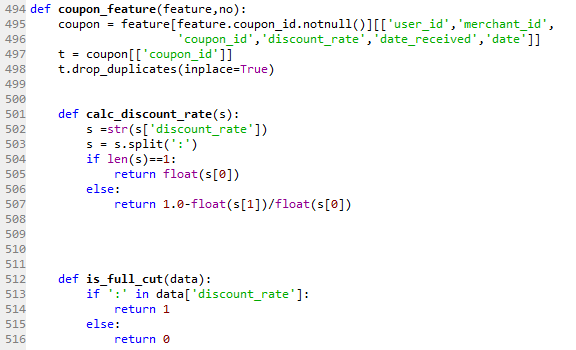
历史核销次数 coupon\_use\_count 35

历史核销率 coupon\_use\_rate\_coupon 36

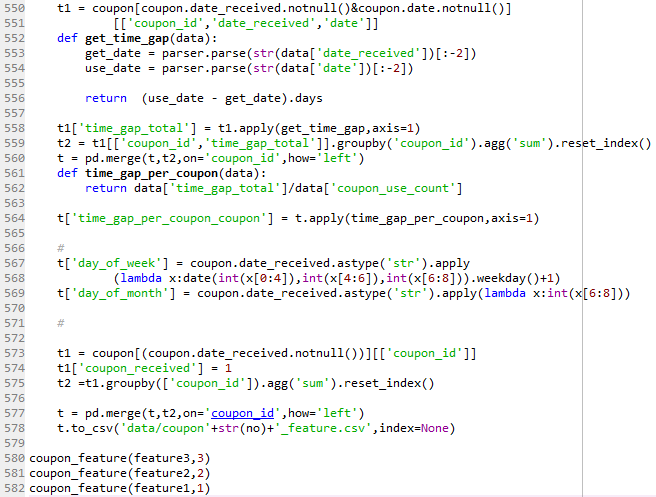
历史核销时间率 time\_gap\_per\_coupon\_coupon 37

领取优惠券是一周的第几天 day\_of\_week 38

领取优惠券是一月的第几天 day\_of\_month 39







## 2.5、用户-优惠卷特征

历史上用户领取该优惠券次数 user\_coupon\_received 40

历史上用户消费该优惠券次数 user\_coupon\_use 41

历史上用户对该优惠券的核销率 user\_coupon\_use\_rate 42



## 2.6、其它特征

用户领取的所有优惠券数目

this\_month\_user\_receive\_all\_coupon\_count 43

用户领取的特定优惠券数目

this\_month\_user\_receive\_same\_coupon\_count 44

用户此次之后/前领取的特定优惠券数目

this\_month\_user\_receive\_same\_coupon\_lastone 45

this\_month\_user\_receive\_same\_coupon\_firstone 46

用户当天领取的优惠券数目

this\_day\_user\_receive\_all\_coupon\_count 47

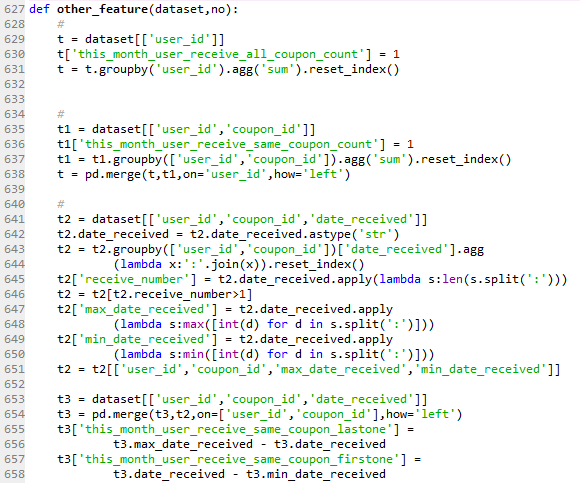
用户当天领取的特定优惠券数目

this\_day\_user\_receive\_same\_coupon\_count 48

用户领取的优惠券前/后时间间隔

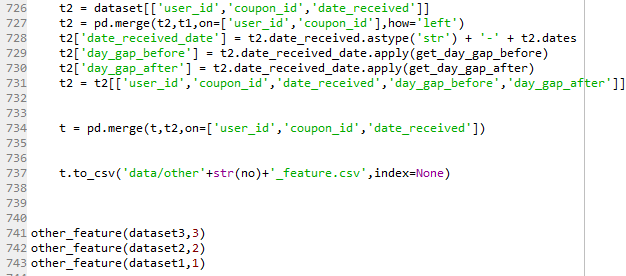
day\_gap\_before 49

day\_gap\_after 50









# 3、特征合并

最后将提取到的各种特征与标签相连接

标签列在55列

