

信贷项目特征提取

10215501443 袁芮琪

1 项目信息

- 项目背景：商户会存在逾期还款甚至不还款的现象。
- 项目目标：使用已有的商户信息、信用记录以及商铺流水信息，预测新申请的贷款是否会出现逾期30天未还款现象。

2 预测流程

贷款申请时，一项贷款包含贷款申请id和商户id。可以通过商户id查看他以往信用记录（table_datamining_history_loan），结合他所拥有的所有商铺的流水情况（table_tansaction）进行评估。信用记录可以反应该商户是否诚信，商铺的流水情况可以反应他所开的商铺是否值得，及是否会出现入不敷出的情况。根据两者所占的不同比重加权得出一个预测的概率值，因此，企业可通过定义这个概率值，低于该概率值的商户将不给予借贷。

3 项目所需特征

在此之前，我们需要掌握商户的行业和省份信息。因此，我们需要保留table_datamining_merchant表中的数据。

列名	comment
merchant_id	商户ID
industry_level	一级行业
province	省份

可以发现，待预测的贷款是以贷款申请id和商户贷款id为标识的，而非商铺id。一个商户可以管理多个商铺，所以在预测贷款是否为坏账时，需要结合该商户所管理的多个商铺的流水信息进行评估。为了预测该商户是否会出现逾期还款的现象，我将从两个维度进行衡量。一个是**商户是否有信用**，另一个是**商户是否有能力**。

3.1 商户是否有信用

- table_datamining_history_loan

查看历史的贷款申请记录，找出有关该商户的申请，从**商户是否实际结清日迟于理论结清日**和**是否有30天还款现象**两个方面进行刻画。这两方面并不冲突，因为虽然一个商户在理论结清日之前结清了所有贷款，但是可能在个别期数内存在逾期30天未还款的现象（例如，

```
apply_id,merchant_id,apply_date,lend_amount,lend_period,notional_settle_date,
real_settle_date,is_30days_overdue
24015336,037740656205727,2017-08-23 00:00:00,5000.0,6,2018-02-24
00:00:00,2017-12-21 00:00:00,1
```

该商户虽然在理论结清日前结清了，但在6个期数内，仍然存在逾期30天未还款的现象）

当然，对于同一商户，可能他有多笔贷款记录，因此，没笔贷款记录对于他的信用的影响程度也是不同的。所以，我用历史贷款的**贷款金额**来衡量该条记录对该商户的信用刻画的影响程度。如果一笔贷款的金额很大（大于某个值），那么这笔贷款重要程度要大于金额很小的贷款记录（一笔金额很小的良好贷款记录并不能抹去一笔金额很大的坏的贷款记录）。

因此，对于table_datamining_history_loan表，我选择以下特征

列名	comment
merchant_id	贷款人id
lend_amount	贷款金额
notional_settle_date	理论结清日
real_settle_date	实际结清日
is_30days_overdue	有无逾期30天还款

3.2 商户是否有能力

- table_transaction

在该表中，我只选取交易类型为付款成功的交易进行讨论。因此在预处理时需要根据type这一特征将类型为取消和退款的交易剔除。此外，一个商铺是否会入不敷出往往也取决于该商铺的客户的消费水平，如果客户们的消费水平较高的情况下，说明该商铺的收入来源是较稳定的，不会有很大的客户流失的现象。因此，可以用**信用卡和花呗支付金额占总支付金额的比重**来刻画该门店的客户画像。如果该比重很大（大于某一个值），那么说明在该商铺开设地段的客户整体的消费趋于一种“超前消费”的形势，该客户群体是较不稳定的，在生活出现变化时，可能该类客户就不会选择来该店面

消费。

因此，对于table_transaction表，我选择以下特征

列名	comment
type	交易类型 30-付款 10-取消 11-退款
merchant_id	商户id
paid_amount	消费者实际支付金额
alipay_huabei	花呗支付金额
bankcard_credit	信用卡支付金额