

作业三——客户星级和信用等级评估

团队成员分工

姓名	主要工作
郑伟鑫	star_level, credit_level部分的数据处理和训练模型并预测，文档编写
王文渊	star_level模型训练
谢瀚杵	star_level部分的数据处理
徐晨阳	credit_level部分的数据处理

数据处理的思路和处理过程

star_level部分

- 由于star_level与交易数据与客户存款有关，本小组将每个用户的每个类型的交易总额提取出来并进行预测
- 使用到的数据处理过程由以下代码给出

```
create table
sum_etc ENGINE=Memory as
select uid, sum(tran_amt_fen) as etc
from dm_v_tr_etc_mx
group by uid
```

```
create table
sum_grwy ENGINE=Memory as
select uid, sum(tran_amt) as grwy
from dm_v_tr_grwy_mx
group by uid
```

```
create table
sum_sa ENGINE=Memory as
select uid, sum(tran_amt) as sa
from dm_v_tr_sa_mx
group by uid
```

```
create table
sum_sbyb ENGINE=Memory as
select uid, sum(tran_amt_fen) as sbyb
from dm_v_tr_sbyb_mx
group by uid
```

```
create table
sum_sdrq ENGINE=Memory as
select uid, sum(tran_amt_fen) as sdrq
from dm_v_tr_sdrq_mx
group by uid
```

```
create table
sum_sjyh ENGINE=Memory as
select uid, sum(tran_amt) as sjyh
from dm_v_tr_sjyh_mx
group by uid
```

```
create table
sum_asset ENGINE=Memory as
select uid, max(all_bal) as asset
from pri_cust_asset_info
group by uid
```

```
select a.uid as uid, asset, etc as etc, grwy as grwy, sa as sa,
sbyb as sbyb, sdrq as sdrq, sjyh as sjyh, star_level
from pri_star_info a
left join sum_asset b on a.uid=b.uid
left join sum_etc c on a.uid=c.uid
left join sum_grwy d on a.uid=d.uid
left join sum_sa e on a.uid=e.uid
left join sum_sbyb f on a.uid=f.uid
left join sum_sdrq g on a.uid=g.uid
left join sum_sjyh h on a.uid=h.uid
```

credit_level部分

- 由于credit_level与交易数据与客户贷款贷记业务有关，本小组将表pri_cust_liab_info中各项数据提取用于模型训练
- 主要的数据处理逻辑如下述代码

```
select
    c.uid uid,
    l.all_bal all_bal,
    l.bad_bal bad_bal,
    l.due_intr due_intr,
    l.norm_bal norm_bal,
    l.delay_bal delay_bal,
    c.credit_level credit_level
from pri_credit_info c
left join pri_cust_liab_info l on c.uid=l.uid
where c.credit_level!='-1'
```

模型训练

- 本次实验，小组基于python和sklearn框架，使用KNN模型进行预测，具体的实现细节可见源码
 - 我们将此前获得的数据，分为训练集，测试集和预测集
 - 训练集与测试集的比例为7:3
- 最终训练出的模型的准确率如下图所示
 - star_level部分：

```
C:\Users\zwx-\Desktop\Courses\SE\codePractice\python\nothing\dm\Scripts\python.exe "C:/Users/zwx-/Desktop/to code/dm/ml.py"
预测分数为: 0.7866896403187594
```

- credit_level部分：

```
C:\Users\zwx-\Desktop\Courses\SE\codePractice\python\nothing\dm\Scripts\python.exe "C:/Users/zwx-/Desktop/to code/dm/ml.py"
预测分数为: 0.8530407454375034
```

- 预测结果分别保存在 `predict_star_level.txt`, `predict_credit_level.txt` 中

代码仓库地址

<https://github.com/zwx-zwx/sklearn-demo>