

活动首页

答题与数据

提交记录

排行榜

招商银行FinTech精英训练营线上复赛

[返回全部题目列表](#)

现金流预测

招行做为金融科技的先行者，存有大量的用户收支数据。根据用户收支数据对用户的现金流进行预测，可以更好地理解用户的收支习惯，更好地为用户提供金融服务。

本次大赛为参赛选手提供了用户90天内的收支数据和天级别的资产信息，希望参赛选手通过分析建模，解决如下问题：根据用户最近60天的收支数据和资产信息，预测用户7天后的账户余额。

训练数据集(train.txt) 提供了用于建模的数据。验证数据集(test\_input.txt)有3000个用户在某个日期以前60天内的收支数据和天级别的资产信息，请对用户7天后的账户余额进行预测。这两个个文件都拥有2种数据：资产信息数据和收支数据。

资产信息的格式为“用户id,bal,日期,账户余额”（bal是余额的简称）。账户包括用户所有的借记卡/信用卡/投资账户等，账户余额包括借记卡里面的存款和投资账户的持仓金额，信用卡等负债账户里面的应还金额或可用额度不会算在账户余额里面的。例如验证数据集中第一个用户是ID为5的用户，他在第15天到第74天内的账户余额信息如下：

```
5,bal,15,50462.78
...
5,bal,74,1335.40
（第15天的账户余额是50462.78元，第74天的账户余额是1335.40元）
```

收支是指进出银行卡的交易，投资账户的收益和亏损不会出现在收支数据里，但会体现在账户余额上。收支数据的格式有两种，分别是7列和8列。绝大多数收支数据有8列，为“用户id,trx,卡id,交易日期,2G摘要码,交易商户名称(交易备注),交易分类,交易金额”（trx是交易的简称）。5号用户在第15天到第74天内的收支数据如下：

```
5,trx,26706,24,奖金,代发奖金,收入-奖金,2981.00
...
5,trx,38272,69,转账汇款,孔某某1933转账,支出-转账给他人,-2000.00
（第24天有一笔2981.00元的钱从“奖金”渠道进账到26706的卡中，改交易的备注为“代发奖金”，交易的一级分类为“收入”，二级分类为“奖金”；第69天有一笔2000元的钱从38272的卡中支出，支出渠道为“转账汇款”，交易备注为“孔某某1933转账”，交易的一级分类为“支出”，二级分类为“转账给他人”）
```

极少量没有交易商户名称或交易备注的收支数据，有7列，格式为“用户id,trx,卡id,交易日期,2G摘要码,交易分类,交易金额”。2G摘要码为“null”的交易是信用卡交易，2G摘要码非“null”的交易是借记卡交易。

参赛选手需要提交一份预测结果test\_output.txt,格式同训练集和验证集的资产信息的格式保持一致，即“用户id,bal,日期,账户余额”，第三列“日期”的值为7天后的日期，第四列余额数字保留小数点后2位。例如：

```
5,bal,81,3513.54
(验证集中有5号用户第15天到第74天的数据，预测结果需要有一条5号用户在第81天的资产信息；第74天过7天后是第81天，所以第三列取值“81”，第四列是5号用户在第81天的账户余额)
每个用户的预测结果为一行，所有数据按用户id从小到大排序。
预测结果的打分方式采用平均绝对误差MAE。
```

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| = \frac{1}{n} \sum_{i=1}^n |e_i|$$

$f_i$  :预测值


$y_i$  :真实值

$e_i = |f_i - y_i|$  绝对误差

^













扫一扫下载牛客APP

注：

1) 为了给参赛选手提供真实完整的数据，我们尽可能多地提供信息，可能不是所有的字段都对预测任务有帮助，是否采用由参赛选手自行判断。特别解释下一些字段的含义：

2G摘要码可以理解为某些交易渠道，可用于辅助判断交易分类，该字段可能为空；

交易商户名称(交易备注)字段也可能为空（为空的收支数据只有7列）；

交易分类的一级分类有三种：收入、支出、本人资金往来。如果用户的交易对象是我们能识别到户主是用户本人的银行卡/股票基金等账户时，一级分类为本人资金往来，例如还招行信用卡、购买基金、网商银行转账等交易的一级分类是本人资金往来，60个分类详见附录《收支分类表》；

交易金额带负号表示资金流出；

2) 数据采用CSV格式，使用逗号分隔，交易商户名称(交易备注)字段如果含有逗号，整个字段会用双引号包住，例如："实物[牛肉饺子,羊肉饺子,猪肉大]"

建议使用专门处理csv格式的模块处理数据，就不用关注这些细枝末节。

3)采用UTF-8无BOM编码的txt文件提交，每列用逗号分隔。

本次比赛提供招行用户的真实消费行为、资产信息；为保护用户隐私和数据安全，所有数据均会进行必要的脱敏。

特别说明：

- 1) 数据使用：本赛题数据仅允许用于本次竞赛相关活动，禁止参赛者用做他用。
- 2) 外部数据：本赛题仅允许使用“竞赛数据”中提供的数据，不允许使用任何外部数据。

附录

收支分类表

分类ID	一级分类	二级分类
101	支出	服饰鞋包
102	支出	日常饮食
103	支出	外出餐饮
104	支出	零食烟酒
105	支出	房贷
106	支出	房租
107	支出	水电物业
108	支出	话费宽带
109	支出	家装维护
110	支出	家私家电
111	支出	公共交通
112	支出	私家车
113	支出	医疗保健
114	支出	家居用品
115	支出	宠物
116	支出	网络购物
117	支出	超市便利
118	支出	商场百货
119	支出	奶粉辅食
120	支出	尿裤湿巾
121	支出	玩具
122	支出	孕婴用品
123	支出	旅行
124	支出	商务差旅
125	支出	美容护肤
126	支出	运动健身
127	支出	书籍教育
128	支出	休闲爱好
129	支出	电子数码
130	支出	保险
131	支出	红包
132	支出	转账给他人
133	支出	现金取出
134	支出	利息手续费
135	支出	其他还款
136	支出	其他支出
201	收入	工资福利
202	收入	奖金
203	收入	兼职
204	收入	分红
205	收入	报销

- 206 收入 社保公积金
- 207 收入 现金存入
- 208 收入 他人转入
- 209 收入 退款
- 210 收入 借款
- 211 收入 他人还款
- 212 收入 收礼金
- 213 收入 其他收入
- 301 本人资金往来 基金
- 302 本人资金往来 理财产品
- 303 本人资金往来 定期
- 304 本人资金往来 外汇
- 305 本人资金往来 贵金属
- 306 本人资金往来 股票
- 307 本人资金往来 第三方理财
- 308 本人资金往来 转账给自己
- 309 本人资金往来 还招行信用卡
- 310 本人资金往来 分期
- 311 本人资金往来 其他理财

附件说明

- 1. train.zip 里面有train.txt，用于建模
  - 2. test\_input.txt是验证集
  - 3.提交文件
    - 3.1.文件格式txt。
    - 3.2.文件每条数据用逗号隔开。
    - 3.3.每行的格式为：“用户ID， bal， 日期， 账户余额”。 第三列“日期”的值为7天后的日期， 第四列余额数字保留小数点后2位。
    - 3.4.必须按照给定的验证数据集进行预测， 不能有遗漏的数据或多出的数据。
- 郑重声明 我们发现有的同学共享答案或共享数据，触发了我们的报警，这是不被允许的。请各位同学遵守复赛规则，一定要下载自己的数据，运行自己的算法，按要求提交自己的结果。复赛试题有一定难度，但相信对于优秀的你们来说，研究一段时间后一定能交出一份满意的答卷。

文件	文件格式
train.zip	zip(70.07M)  <a href="#">高速下载</a> <a href="#">普通下载</a>
test_input.txt	txt(22.67M)  <a href="#">高速下载</a> <a href="#">普通下载</a>

上传预测结果数据，格式仅限txt文件，大小不得超过10M

运行

牛客网，程序员必备求职神器

QQ群：169195721

微 信：www\_nowcoder\_com [关注](#)

微 博：牛客网 [关注](#)



扫一扫，把题目装进口袋

