数据

1）数据总体概述     参赛者可用的训练数据包括用户的基本属性user\_info.txt、银行流水记录bank\_detail.txt、用户浏览行为browse\_history.txt、信用卡账单记录bill\_detail.txt、放款时间loan\_time.txt，以及这些顾客是否发生逾期行为的记录overdue.txt。（注意：并非每一位用户都有非常完整的记录，如有些用户并没有信用卡账单记录，有些用户却没有银行流水记录。）     相应地，还有用于测试的用户的基本属性、银行流水、信用卡账单记录、浏览行为、放款时间等数据信息，以及待预测用户的id列表。     脱敏处理：(a) 隐藏了用户的id信息；(b) 将用户属性信息全部数字化；(c) 将时间戳和所有金额的值都做了函数变换。

2）数据详细描述

（1）用户的基本属性user\_info.txt。共6个字段，其中字段性别为0表示性别未知。

 用户id,性别,职业,教育程度,婚姻状态,户口类型

 6346,1,2,4,4,2

 2583,2,2,2,2,1

 9530,1,2,4,4,2

 6707,0,2,3,3,2

（2）银行流水记录bank\_detail.txt。共5个字段，其中，第2个字段，时间戳为0表示时间未知；第3个字段，交易类型有两个值，1表示支出、0表示收入；第5个字段，工资收入标记为1时，表示工资收入。

用户id,时间戳,交易类型,交易金额,工资收入标记

6951,5894316387,0,13.756664,0

6951,5894321388,1,13.756664,0

18418,5896951231,1,11.978812,0

18418,5897181971,1,12.751543,0

18418,5897293906,0,14.456463,1

（3）用户浏览行为browse\_history.txt。共4个字段。其中，第2个字段，时间戳为0表示时间未知。

用户id,时间戳,浏览行为数据,浏览子行为编号

34724,5926003545,172,1

34724,5926003545,163,4

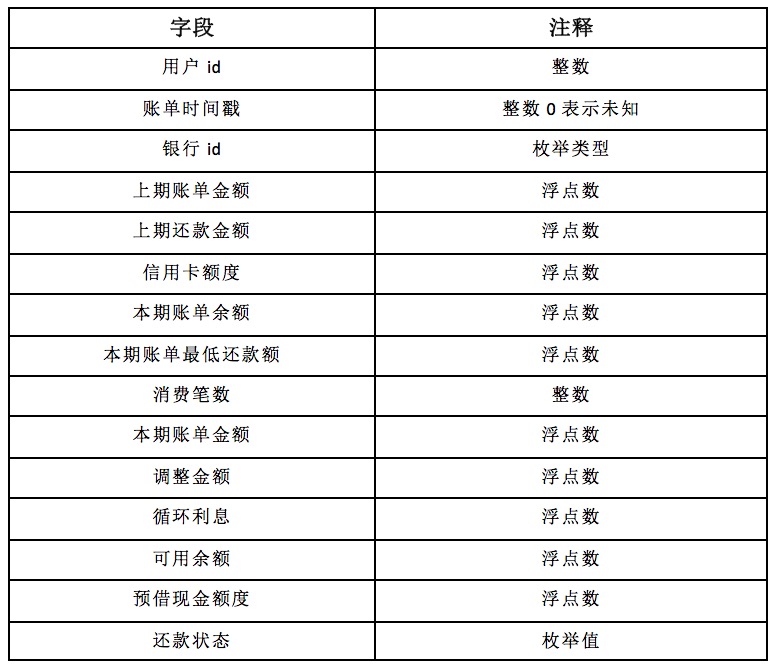
34724,5926003545,38,7

67215,5932800403,163,4

67215,5932800403,138,4

67215,5932800403,109,7

（4）信用卡账单记录bill\_detail.txt。共15个字段，其中，第2个字段，时间戳为0表示时间未知。为方便浏览，字段以表格的形式给出。



文件示例如下：     用户id,账单时间戳,银行id,上期账单金额,上期还款金额,信用卡额度,本期账单余额,本期账单最低还款额,消费笔数,本期账单金额,调整金额,循环利息,可用金额,预借现金额度,还款状态

 3147,5906744363,6,18.626118,18.661937,20.664418,18.905766,17.847133,1,0.000000,0.000000,0.000000,0.000000,19.971271,0     22717,5934018585,3,0.000000,0.000000,20.233635,18.574069,18.396785,0,0.000000,0.000000,0.000000,0.000000,0.000000,0

（5）放款时间信息loan\_time.txt。共2个字段，用户id和放款时间。

 用户id,放款时间

1,5914855887

2,5914855887

 3,5914855887

（6）顾客是否发生逾期行为的记录overdue.txt。共2个字段。样本标签为1，表示逾期30天以上；样本标签为0，表示逾期10天以内。注意：逾期10天~30天之内的用户，并不在此问题考虑的范围内。用于测试的用户，只提供id列表，文件名为testUsers.csv。

 用户id,样本标签

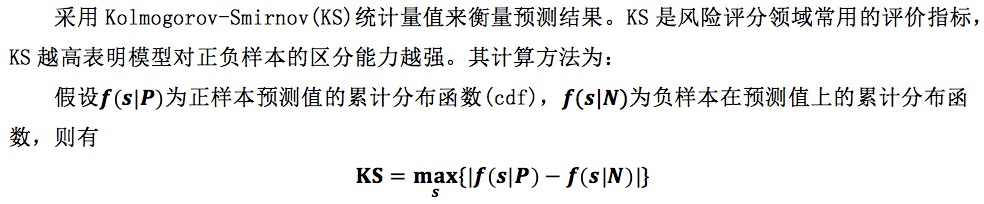
1,1

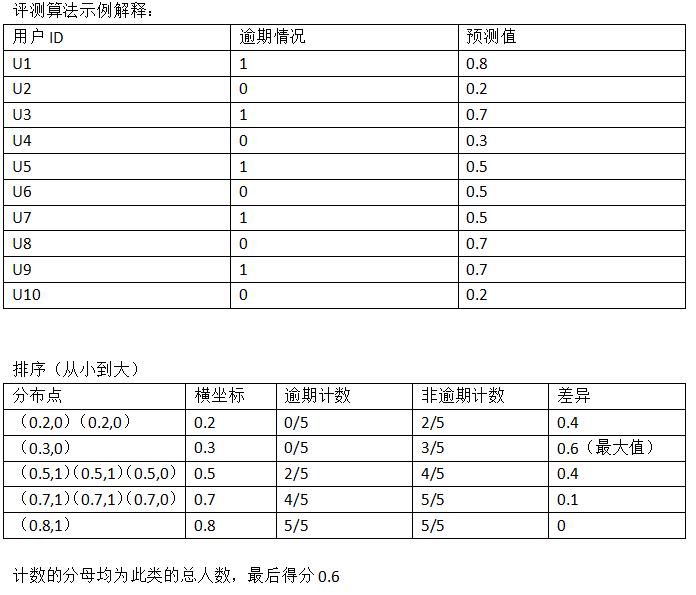
2,0

3,1

评分标准

评分算法：  
other  
评分标准:





提交规则

算法结果提交说明： 1.参赛者提交csv结果文件。 2.文件内容分为两列，第一类为“顾客id”，第二列为“逾期概率”，概率范围［0,1］格式如下。

顾客id,逾期概率

0,0.1

1,0.3

3,0.9