

论文标题

摘要

这里是摘要

关键字： 聚类分析

目录

一、问题提出与重述	2
1.1 问题的提出	2
1.2 问题的重述	2
二、问题分析	2
2.1 问题的分析	2
三、模型的假设	3
3.1 模型的假设	3
四、名词解释与符号说明	4
4.1 名词解释	4
4.2 符号说明	4
五、模型建立与求解	5
5.1 问题一的建模与解答	5
5.1.1 数据清洗	5
5.1.2 数据统计	5
5.1.3 支付行为模型	6
5.1.4 支付特征分析	6
5.1.5 聚类统计分析	6
5.2 问题二的建模与解答	6
5.2.1 第三方支付平台的收入	6
5.2.2 第三方支付平台的支出	8
5.2.3 第三方支付平台的盈利模型	9
5.2.4 问题二模型的求解	9
5.2.5 问题二模型的分析	9
5.3 问题三的建模与解答	9
5.3.1 问题三模型的建立	11
5.3.2 问题三模型的求解与分析	11
六、模型的评价	11
6.1 模型的优点	11

6.2 模型的缺点	11
6.3 模型的改进	11
七、模型的推广与应用	11
参考文献	11
附录 A 遗传算法-python 源程序	12
附录 B 程序-X 代码	12

一、问题提出与重述

1.1 问题的提出

随着互联网的整体升级与手机行业的迅速发展，移动支付也逐渐得到普及，受到用户与商家的青睐。移动支付即通过手机而非现金或银行卡完成支付，具有移动性，及时性，定制化，集成性等特点，发展前景巨大，应用领域丰富。其中公交移动支付可以解决现金支付，刷卡支付等带来的不便，也可减少人工成本。

在出行领域相比于打车与共享单车，公交地铁的体量更大，交易量也十分可观，应用到移动支付行业会带来大量利润，所以通过分析市民乘车的支付行为建立盈利模型来分析估计支付平台的商业盈利十分必要。

1.2 问题的重述

1. 通过附件 1，2 给出的某市部分公交支付的信息与数据说明，分析该市乘车人的支付行为特征。

2. 参考附件 3 的信息，建立第三方支付平台的商业盈利数学模型，定量分析第三方平台的收支和盈利情况。

3. 通过附件 1，2 给出的该市四分之一的公交与地铁安装移动支付设备后试营运期间数据，估计该市实现全部公交移动支付后第三方平台的盈利情况。

二、问题分析

2.1 问题的分析

问题一分析：

为对用户的支付行为特征进行分析，我们分别用统计与聚类分析两个方法提取用户的支付行为特征并进行分析。分析附件一二给出的公交支付信息，发现其中有缺失数据与错误数据，需要先进行数据清洗与数据填充，得到有效数据。从而，我们需要先建立

统计模型，通过建立哈希表对数据进行整理，分别从不同时间，不同个体等角度对用户的支付行为进行统计分析，得到统计数据，绘制出相应图表，根据图表与数据提取用户支付行为特征并尝试分析其原因。建立聚类分析模型，采用 k-means 方法对用户群体进行聚类，画出不同用户群体的月均出行次数与一天内各时间段的曲线图，分析图像得出不同用户的支付行为特征。

问题二分析：

附件三给出了第三方支付平台的盈利模式，了解到第三方支付平台有手续费，广告费，沉淀资金的利息收入，服务费等多种盈利模式。其中每个盈利模式都有自己的方案，通过对每个盈利模式的方案进行分析，可以分别得到每种收益的表达式。支出方面分为机器成本，维护成本与推广费用几个方面，通过对第三方平台在各个环节产生的支出进行分析可得出每一项支出的表达式。平台盈利即为平台收入减平台支出。构建好平台盈利模型，通过定量分析的方法更改影响模型的因素，绘制盈利变化的曲线，分析收支与盈利情况。

问题三分析：

公交移动支付覆盖率从该市四分之一扩展到全部，会对移动支付流量产生影响，进而影响第三方支付的盈利。针对客流量，可以通过一个城市的实例来找到不同线路客流量的比例。以南京市地铁为例，分析其中热门 1/4 线路与全部线路的客流量，得到热门 1/4 线路客流量占比，再通过模型一得到的有关移动支付份额数据，就可以对移动支付全覆盖之后该市移动支付的人数与交易额进行估算。将移动支付交易额带入第二问建立的模型中得到第三方平台盈利数据。根据模型一得出的支付行为特征，我们发现移动支付市场份额会随移动支付试运营的时间发生变化。通过模型一数据，可建立移动支付市场份额随时间增长的变化曲线，应用于模型二得出的盈利模型中，可建立移动支付覆盖率提升之后，平台盈利随时间变化的曲线，即可估计与分析该市在移动支付全覆盖后第三方支付平台的盈利情况。

三、模型的假设

3.1 模型的假设

- (1) 假设表中显示的支付时间均为有效时间，即乘客在支付时认为公交处于运营状态。
- (2) 假设同一个人只能拥有一个 ID。
- (3) 假设移动支付试运营发生在该市客流量最大的 1/4 线路中。
- (4) 假设南京地铁的部分运营数据可以反映该市公共交通的运营数据。
- (5) 假设该市乘坐公交以及地铁的均价为 2 元。
- (6) 假设第三方支付平台资金周转能力良好。

四、符号说明

表 1 符号说明

符号	表示含义	单位
$SumCharge$	所有天数的公共交通支付笔数数目的总和	RMB
$SumCharge^{MobilePay}$	所有天数移动支付公共交通的成交量	
$DailyCharge_{(i,j)}$	第 i 月星期 j 的公共交通支付数目	
$DailyCharge_{(i,j)}^{MobilePay}$	第 i 月星期 j 当天移动支付公共交通的笔数	
$CustomerSum(i)$	乘客 i 公共交通支付总量	
$CustomerSum(i)^{mobilePay}$	乘客 i 公共交通移动支付的数量	
$CustomerValidSum(i)$	乘客 i 有效公共交通支付总量	
S_{sum}	销售总额	
Pay_{sum}	用户支付笔数	
$FreeRate$	第三方平台对用户的手续费率	
$ServiceRate$	第三方平台对商户的服务费率	RMB/CPM
$adFree$	第三方平台向商户收取的广告费	
\overline{remain}	用户人均余额	RMB
$NumberOfPeople$	使用该第三方支付平台的人数	RMB
$SedMoney(i)$	第 i 月的存入银行的沉淀资金	
$SedMoney^j(i)$	第 i 月的沉淀资金经过 j 月以后的总额	RMB
$r = IncreaseRate$	沉淀资金的增长率	

五、模型建立与求解

5.1 问题一的建模与解答

5.1.1 数据清洗

附件一，二给出了某市部分公交支付的信息与数据，分别取了 2，5，8，11 月份中连续的 7 天的支付信息，包括支付 ID、上一次乘车交易时间、本次乘车时间、付款方式，当月地铁与公交乘车次数与乘车总次数。每天大概有一百多万条乘车数据，而其中也有许多无效数据与错误数据，需要我们在统计前对数据进行清洗，现对数据进行以下三个方面的数据填充以及清洗：

1. 对于 PAYTYPE 值为 Null 的数据：将部分 PAYTYPE 项中的 Null 值替换为 0。附件一中的 PAYTYPE 项表示的是支付方式，其中 0 代表公交移动支付，1 代表公交卡支付，而通过附件 2 提供的信息可知 Null 为没刷卡，所以将 Null 将 NULL 算为非公交卡支付。通过观测 2 月、5 月、8 月和 11 月的公交数据可以发现，除二月以外的其他月中，支付类型只为 0，1 和 NULL，又因为 NULL 表示非公交卡支付，所以这里我们可以把 NULL 归为移动支付。对于二月的数据，我们可以把支付类型为 NULL 的数据，按比例分配给移动支付方式和其它支付方式。

2. 对于 UPTIME 没有记录的数据：UPTIME 中的 0001-1-1 表示没有刷卡记录，对这部分数据进行填充：

(1) 若该乘客相同的月份的相邻两次记录中，前一次的 UPTIME 丢失，而后一次的乘车记录给出了 LASTTIME，则我们使用后一次记录的 LastTime 填充上一次记录的 UPTIME。

(2) 若找不到可以补充的 LASTTIME，则按照以下步骤填充：

Step1: 将乘客 i 所有有效乘车时间提取出来,构成数组 $Uptime_i[k](k=0,1...CustomerValidSum(i))$ 。

Step2: 在 $0 \sim CustomerValidSum(i)$ 范围内随机生成数字 n 。

Step3: 将 n 当作数组下标在 $Uptime_i$ 数组中取 $Uptime_i[n]$ 作为填充数据。

Step4: 对于多个 Uptime 缺失的，重复 Step2~Step4 步骤

3. 清理极少乘坐公交的人：对于 $CustomerSum(i) \leq 2$ 的乘客信息，由于他们乘坐公交次数很少，支付行为特征难以分析，或者不是该市市民，对于这种乘客的信息进行清理，删掉这条 ID 相关的信息。

5.1.2 数据统计

由于支付数据量很大，我们先遍历一边所有 excel 表格，根据每个人的 ID 构建哈希表储存他们的信息，根据有效数据整理出如下数据：

(1) 每天的移动支付笔数： $\overline{times}_{(i,j)}^{mobilePay}$

- (2) 每天的公交支付笔数: $\overline{times}_{(i,j)}$
- (3) 一天中每小时内公交支付笔数: $HourCharge_i$
- (4) 一天中每小时内公交移动支付笔数 $HourCharge_i^{mobilePay}$

5.1.3 支付行为模型

1. 日均公交移动支付与总支付笔数小时分布图

统计平均一天中每小时内的公交支付总笔与移动支付的总笔数，绘制柱状图，其中移动支付笔数包含在总笔数之中。

2. 公交移动支付与总支付笔数按日期分布图

统计每天的公交支付笔数与公交移动支付笔数，按日期增长顺序绘制折线图。

3. 公交支付笔数星期内分布图

将四周的公交支付笔数分别按一星期七天画出四条折线图

4. 移动支付占比随乘车次数分布图

计算乘车次数为 i 的乘客的移动支付笔数总和与总乘车次数和，之后求出乘车次数为 i 的移动支付笔数占总乘车次数比例，绘制移动支付占比随乘车次数增长的折线图。

5. 公交移动支付星期内分布图

将四周的公交移动支付笔数分别按一星期七天画出四条折线图

6. 公交移动支付占比图

将四周每周的移动支付占比绘制饼图（4 张）

5.1.4 支付特征分析

...

5.1.5 聚类统计分析

数据统计模型可以获得用户整体的支付行为特征，若想对用户群体进行分类，得到用户整体中具有各自支付行为特征的群体，就需要建立聚类分析模型，将用户群体进行划分。我们采用 K-means 的聚类算法，对用户群体的支付行为特征进行聚类分析。

K-means 算法：K-Means 算法是无监督的聚类算法，应用广泛。K-Means 算法的思想为对于给定的样本集，按照样本之间的距离大小，将样本集划分为 K 个簇。让簇内的点尽量紧密的连在一起，而让簇间的距离尽量的大。应用于本题，我们要将用户整体划分为几个不同的群体，其中每个群体内部的支付行为特征相似，不同群体的支付行为特征不同。K-means 算法流程如下：

Step1: 从数据集中随机选取 K 个样本作为初始聚类的中心 $C = c_1, c_2, \dots, c_k$

Step2: 针对数据集中每个样本 x_i , 计算它到 K 个聚类中心的距离并将其分到距离最小的聚类中心所对应的类中

Step3: 针对每个类别 c_i , 重新计算它的聚类中心 $c_i = \frac{1}{|c_i|} \sum_{x \in c_i} x$ (即属于该类的所有样本的质心)

Step4: 重复 Step2 和 Step3 直到聚类中心的位置不再变化

K-means 模型构建:

本题应用于 K-means 模型中的特征, 输入。。。

由构建的 K-means 算法得到核数为

图

基于 K-means 的支付行为特征分析:

5.2 问题二的建模与解答

5.2.1 第三方支付平台的收入

由附件三分析, 第三方支付平台的收入主要由手续费, 服务费, 广告费, 沉淀资金的利息几个部分组成。现从以下四个方面构建收入模型:

1. 手续费 T_1 : 即第三方向用户收取手续费和向银行支付的手续费之差, 无论是, 线上支付的支付宝, 还是线下支付的拉卡拉, 手续费都是第三方支付平台的传统盈利方式。附件三中对手续费的区间描述为 $0.08\% - 1.25\%$, 根据微信, 支付宝等主流第三方支付平台的手续费率, 在模型中取手续费率 $FreeRate = 0.1\%$ 。可得手续费 t_1 的表达式为:

$$T_1 = S_{sum} * FreeRate \quad (1)$$

2. 服务费 T_2 : 这里所指的服务费是指第三方支付平台为其客户提出支付解决方案, 提供支付系统以及各种增值服务。这也应该是第三方支付平台最核的盈利模式。支付宝服务费为 0.6% , 在模型中取服务费率 $ServiceRate = 0.6\%$, 可得服务费 T_2 表达式为:

$$T_2 = S_{sum} * ServiceRate \quad (2)$$

3. 广告费 T_3 : 为第三方平台向商户收取的广告费用。通过调查广告网站设定广告单价 $adFree = 40RMB/CPM$, 单位为每一千人浏览广告的价格。广告费 T_3 的表达式为:

$$T_3 = Pay_{sum} * adFree/1000 \quad (3)$$

4. 沉淀资金 T_4 : 沉淀资金即为备付金, 为办理客户委托的支付业务而实际收到的预收货币代付资金。其中风险准备金比例不得低于其银行账户利息所得 10% , 这也就意味着第三方支付机构最多可以获得 90% 的利息收入。在以活期存款形式的客户备付金满足日常支付业务的需要后, 其他客户备付金可以“以活期存款、单位定期存款、单位通

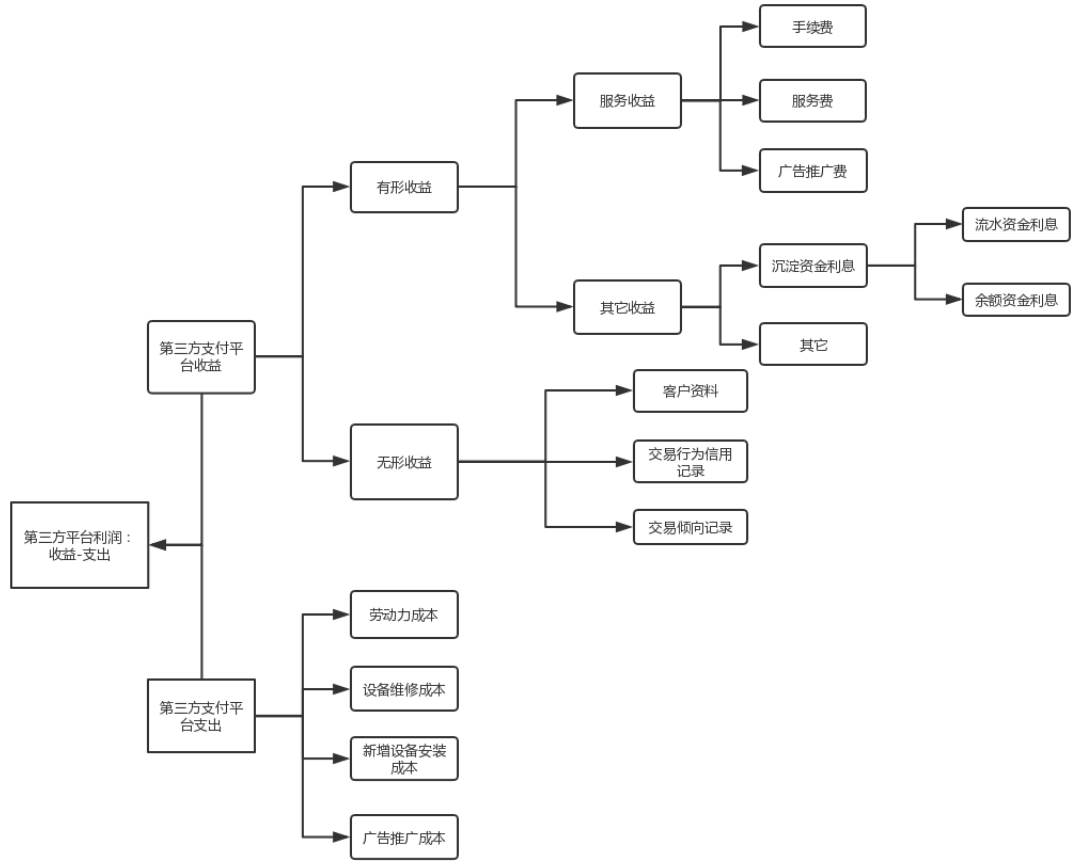


图 1 盈利模型

知存款、协定存款或经中国人民银行批准的其他形式”存放，但”期限不得超过 3 个月”。这意味着，部分客户备付金可转成为期 3 个月的单位定期存款。协议存款率为 4% ~ 5%，手续费估算为 0.78%，沉淀金包含长期沉淀金与短期沉淀金两个部分。长期沉淀金主要为用户在第三方平台的余额，而短期沉淀金为第三方平台的流水金额。

沉淀资金增长率的计算：设 r_1 为风险准备金比例， a_1 为活期资金占总资金比例， r_2 为三个月的活期年利率， r_3 为协定年利率。沉淀资金的增长率

$$r = (1 - r_1)(a_1 * r_2 + (1 - a_1) * r_3)/12 \quad (4)$$

设 a_3 为发卡额的保有比例， $UserBalance(i)$ 表示经过 i 月以后，沉淀资金中用户余额总数的值，其中 $UserBalance(0)$ 表示未放入银行获益前的用户余额总数。

$$UserBalance(0) = (\overline{remian} * NumberOfPeople) * a_3 \quad (5)$$

第 i 月的流动资金用 $SedMoney(i)$ 表示， $SedMoney^j(i)$ 表示第 i 月的流动资金经

过 j 月以后的值, 则 $SedMoney^j(i)$ 的表达式为

$$SedMoney^j(i) = \begin{cases} (1+r) * (1-t_1) SedMoney^{j-1}(i) & j = 3n-2 \quad j! = 1 \\ (1+r) * SedMoney^{j-1}(i) & j = 3n-1 \text{ or } 3n \\ SedMoney(i) & j = 0 \end{cases} \quad (6)$$

$$UserBalance(i) = \begin{cases} (1+r) * (1-t_1) UserBalance(i-1) & j = 3n-2 \quad j! = 1 \\ (1+r) * UserBalance(i-1) & j = 3n-1 \text{ or } 3n \end{cases} \quad (7)$$

计算沉淀资金的年利润:

$$T_4 = \sum_{i=0}^{12} SedMoney^{12-i}(i) - \sum_{i=0}^{12} SedMoney(i) + UserBalance(12) - UserBalance(0) \quad (8)$$

其中, $\sum_{i=0}^{12} SedMoney^{12-i}(i)$ 表示年底时, 每个月流水金额的值。 $UserBalance(12)$ 表示年底时用户总余额的值。

5.2.2 第三方支付平台的支出

将平台的支出分为固定成本和推广成本

1. 固定成本 O_1 : 主要体现为新增设备安装成本、维护设备成本以及劳动力成本三个部分。

设平均一台支付机器的价格为 $V_{machine}$, 机器数量为 $N_{machine}$, 机器维护均价为 $V_{maintain}$ 员工工资支出为 S , C_1 的表达式为:

$$C_1 = N_{machine} * (V_{machine} + V_{maintain}) + S \quad (9)$$

2. 推广费用 O_2 : 主要体现为平台在各个媒体渠道投放广告用来提高平台的知名度以及市场份额

设投放广告数量为 N_{ad} , 广告均价为 V_{ad} , C_2 的表达式为:

$$C_2 = N_{ad} * V_{ad} \quad (10)$$

5.2.3 第三方支付平台的盈利模型

第三方平台的总盈利 W : 收入-支出

$$W = T_1 + T_2 + T_3 + T_4 - C_1 - C_2 \quad (11)$$

结合表达式 1 ~ 10 可得：

$$\begin{aligned}
 W = & S_{sum} * FreeRate + S_{sum} * ServiceRate + Pay_{sum} * adFree/1000 \\
 & + \sum_{i=0}^{12} SedMoney^{12-i}(i) - \sum_{i=0}^{12} SedMoney(i) + UserBalance(12) - UserBalance(0) \\
 & - (N_{machine} * (V_{machine} + V_{maintain}) + S) - N_{ad} * V_{ad}
 \end{aligned} \tag{12}$$

5.2.4 问题二模型的求解

5.2.5 问题二模型的分析

5.3 问题三的建模与解答

当该市的公交移动支付覆盖率从四分之一变为全部覆盖时，移动支付的人数会提高，通过附件一二的的数据可以求出实施移动支付的公交地铁站的移动支付占比，但由于附件 1 与附件 2 是该市四分之一的公交与地铁安装移动支付设备后试营运期间的数据，想要得到移动支付占比则需要知道 1/4 试运营站点中的移动支付的客流量占比 $O_{1/4}$ ，现通过分析南京市地铁的客流量来计算 $O_{1/4}$ ，其流程图如下：

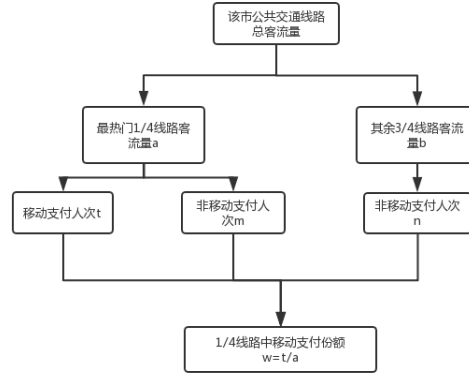


图 2 移动支付占比计算

以南京市为例，我们通过查阅南京地铁集团跟踪评级报告获取到 2017 年度南京各地地铁线路的客流量和里程数，我们假设移动支付试运营的 1/4 的线路取地铁线路客流量最大的 1/4 线路，由下表 2 给出了 2017 年南京各个线路的里程数与客流量的数据：

根据表 2 中的数据绘制出各线路客流量比例和里程数比例。

由图 3 我们可以发现一号线和二号线为客流量最多的两条铁路线路，且一二号地铁线的公里数总和占地铁总公里数的 26%，近似于 1/4 的铁路线路，符合模型中的场景，

表 2 南京地铁线路

线路	里程 (km)	客流量 (万人)	线路	里程 (km)	客流量 (万人)
1 号线	37.5	31241.59	10 号线	21.2	4602.58
2 号线	37.1	26288.11	S1 号线	33.0	2267.29
3 号线	43.9	24752.45	S3 号线	35.5	117.33
4 号线	32.6	4890.78	S8 号线	44.4	3574.44

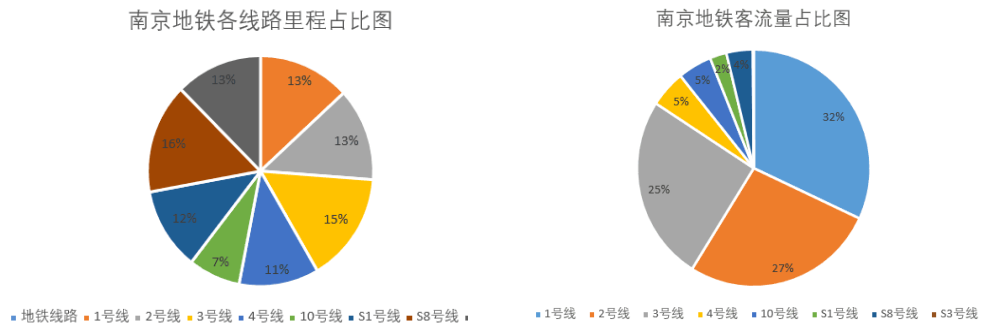


图 3 南京地铁客流量占比

所以取一号线与二号线为试运营的 1/4 线路。设 P_i 为八条线路中第 i 条线路的客流量， $O_{1/4}$ 就等于计算一号线二线线的客流量占比即：

$$O_{1/4} = \frac{P_1 + P_2}{\sum_{i=1}^8 P_i} \quad (13)$$

求出 $O_{1/4}$ 后，将南京市客流量占比应用到该市，可得到该市 1/4 试运营地点的客流量。设第 i 月第 j 天 1/4 线路中移动支付占比为 $MobileProportion(i, j)$ ，则根据附件一，二的统计数据可得到 $MobileProportion(i, j)$ 的表达式：

$$MobileProportion(i, j) = DailyCharge_{(i,j)}^{MobilePay} / (DailyCharge_{(i,j)} * O_{1/4}) \quad (14)$$

5.3.1 问题三模型的建立

5.3.2 问题三模型的求解与分析

六、 模型的评价

6.1 模型的优点

6.2 模型的缺点

6.3 模型的改进

七、 模型的推广与应用

参考文献

- [1] 王琛. 第三方支付的主要盈利模式及存在的风险分析——以支付宝为例 [J]. 商业文化,2015(18):166-168.
- [2] 钱凯凯, 蒋秀. 从支付宝微信支付看第三方支付的盈利模式 [J]. 商场现代化,2016(30):77-78.
- [3] 孙吉贵, 刘杰, 赵连宇. 聚类算法研究 [J]. 软件学报,2008(01):48-61.
- [4] 陈佳琪. 第三方支付平台的风险评估 [D]. 湖南大学,2018.
- [5] 姚宁. 基于移动平台的第三方支付模式研究 [D]. 华东理工大学,2016.
- [6] 张念, 王毅磊. 中国移动支付发展的现状及问题——以支付宝为例 [J]. 现代营销 (下旬刊),2019(03):50.
- [7] 易涌征. 移动支付消费者接受的影响因素研究 [D]. 浙江大学,2012.

附录 A 遗传算法–python 源程序

```
print helloworld!
```

附录 B 程序–X 代码

```
cout<<"helloworld";
```