



東南大學
Southeast University

第十三届大学生数学建模竞赛

(2019.05.15-2019.05.20)

主办：东南大学教务处
承办：东南大学数学学院
东南大学数学建模竞赛组委会

论文题目： 公交移动支付问题的评估方案

参赛队号： 20191095

参赛队员信息：

	组长	合作队员1	合作队员2
姓名	喻泽弘	杜昕昱	卢立强
学号	71117415	71117121	71117417
手机	15651833082	15651833110	17612582337
Email	1161018109@qq.com	2656308604@qq.com	Lu-Liqiang@outlook.com

公共移动支付问题的评估

摘要

本文通过 K-means 聚类分析以及数据统计的方法，得出了该市乘车人出行的支付特征，构建了第三方支付平台的盈利模型，预测了该市全面运营移动支付以后，第三方支付平台的月盈利额随时间的变化情况。

对于问题一，根据附件一的数据来分析用户支付行为特征。附件一中的数据量较大，而且在大量数据中存在数据丢失与数据错误的情况，需要我们在进行特征分析前对数据进行清理。对于 UPTIME 字段的缺失，本文根据该用户个人乘车的历史记录来对这部分数据进行补充；对于 PAYTYPE 字段的数据缺失，由于 NULL 表示未刷卡，本文通过分析该用户的移动支付方式与其他支付方式的比例来对 NULL 数据进行填充；对于乘车次数极少的用户，对他们的数据进行清理，防止对特征分析产生干扰。在进行数据清洗得到有效数据之后，首先建立统计模型，通过为数据建立 ID, 支付时间, 支付类型等索引进行整理，得到用户的支付特征与时间，乘车次数等变量之间的关系，分析其中原因。再建立 K-means 模型对用户支付特征进行聚类分析，得到了五类特征较为独立的用户群体，并对这五类用户群体的支付特征进行描述，分析其中原因。

对于问题二，根据附件三以及相关资料的查询，得出第三方支付平台 4 种主要的收入方式：手续费、广告费、沉淀资金利息和服务费，以及 2 种主要的支出的方式：固定成本和推广费用。对于广告费，我们采取 CPM 的方式进行计算，即每千次浏览的计费方法，较为精确的算出了第三方平台的广告收益。根据上述收入以及支出方式建立第三方支付平台的盈利模型。查阅支付宝、微信等各类第三方支付平台资料，分析以及预测出模型中各系数的参数，从而代入相关数据定量分析得出第三方支付平台的年利润：W=1648.6 万元。最后，本文对模型二中的可变参数进行敏感性分析，判断出其对盈利模型的影响程度。

对于问题三，首先查阅南京地铁运营数据，估算出南京市热门 1/4 线路的客流量占比。将南京地铁的客流占比推广到该市的公共交通中，得出该市 1/4 线路内移动支付份额与时间的柱形图。根据该图，预测出该市公共交通全面运营移动支付以后，该市公共交通移动支付份额随时间的变化关系，从而估算出移动支付在 2018 年 2 月、5 月、8 月以及 11 月的交易额。将其带入模型二，得出第三方平台月利润随时间的变化图像，进而估算出第三方支付平台在 2017 年 11 月、2018 年 2 月、2018 年 5 月、2018 年 8 月以及 2018 年 11 月的利润分别为：55.75 万元、114.5 万元、124 万元、127.6 万元以及 131.25 万元。

关键字： 聚类分析 K-means 方法 南京地铁 CPM 敏感性分析

目录

一、问题提出与重述	3
1.1 问题的提出	3
1.2 问题的重述	3
二、问题分析	3
2.1 问题的分析	3
三、模型的假设	4
3.1 模型的假设	4
四、符号说明	5
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 支付行为模型	6
5.1.3 支付行为特征分析	8
5.1.4 聚类统计分析	9
5.2 问题二的建模与解答	11
5.2.1 第三方支付平台的收入	11
5.2.2 第三方支付平台的支出	13
5.2.3 第三方支付平台的盈利模型	14
5.2.4 盈利模型定量分析	14
5.2.5 盈利模型的敏感性分析	14
5.3 问题三的建模	15
5.3.1 问题三模型的求解与分析	17
六、模型的评价	18
6.1 模型的优点	18
6.2 模型的不足	18
6.3 模型的改进	19
七、模型的推广与应用	19
参考文献	19

附录 A k-means 算法	20
附录 B conda 环境列表	21

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。其中公交移动支付不仅可以避免现金支付，刷卡支付等带来的不便，还可减少人工成本。

在公共出行领域相比于打车与共享单车等出行方式，公交地铁的运量更大，交易量也更可观，移动支付在该行业的使用和推广会产生可观的利润，因此通过分析市民乘车的支付行为特征建立模型来分析与评估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：

为了对用户的支付行为特征进行分析，本文用统计方法提取用户的支付行为特征并进行聚类分析。分析附件一二给出的公交支付信息，发现其中存在的缺失数据与错误数据，并进行数据清洗与数据填充，得到有效数据。从而，本文可以建立统计模型，通过为数据建立 ID, 支付时间、支付类型等索引进行整理，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分析其原因。建立聚类分析模型，使用 k-means 方法对用户群体进行聚类，分析用户的出行时间、次数等特征。

问题二分析：

附件三给出了第三方支付平台的盈利模式，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的支出。平台盈利即为平台收入减平台支出。构建好平台盈利模型，通过更改影响模型的因素进行敏感性分析，绘制盈利变化的曲线，分析收支与盈利情况。

问题三分析：

随着移动支付覆盖率的上升，移动支付的市场份额也会有相应的变化，进而影响第三方支付平台的盈利。针对客流量变化的情况，可以通过真实城市的运营数据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支付平台盈利数据。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) 假设同一个人只能拥有一个 ID。
- (2) 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。
- (3) 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。
- (4) 假设该市乘坐公交以及地铁的均价为 2 元。
- (5) 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	PCS
$DailyCharge_{(i,j)}$	第 i 月星期 j 的公共交通支付数目	PCS
$DailyCharge_{(i,j)}^{MobilePay}$	第 i 月星期 j 当天移动支付公共交通的笔数	PCS
$CustomerValidSum(i)$	乘客 i 有效公共交通支付总量	PCS
S_{sum}	销售总额	RMB
Pay_{sum}	用户支付笔数	PCS
$FreeRate$	第三方平台对用户的手续费率	PCS
$ServiceRate$	第三方平台对商户的服务费率	PCS
$adFree$	第三方平台向商户收取的广告费	RMB/CPM
\overline{remain}	用户人均余额	RMB
$NumberOfPeople$	使用该第三方支付平台的人数	People
$SedMoney(i)$	第 i 月的存入银行的沉淀资金	RMB
$SedMoney^j(i)$	第 i 月的沉淀资金经过 j 月以后的总额	RMB
r	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，分别为 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字段。每天大概有一百多万条支付数据，大量数据读取起来消耗时间，本文模型采用了

jupyter、parso 等各种强大的 python 库进行处理，效果较好，遍历数据耗时较短。大量数据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，现对以下三个类型的数据进行清洗：

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，其他数字代表其他方式，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，本文把支付类型为 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 $CustomerSum(i) \leq 4$ 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 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。

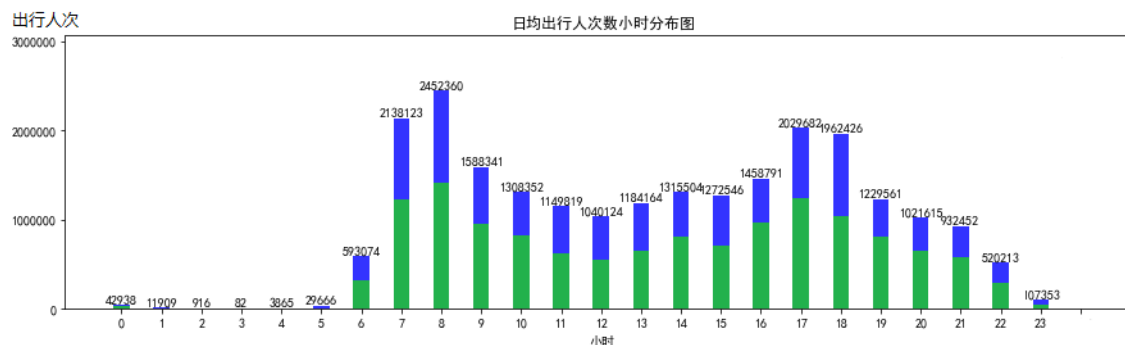


图 1 公交移动支付与总支付笔数按小时分布图

通过小时分布图我们可以发现在 7~9 点和 16 ~ 19 点的时候出现了两个峰值，表明在这两个时

间段内用户公交支付次数有较为明显的增长，而 7~9 点和 16 ~ 19 恰好是早高峰与晚高峰时间段，上班人群与学生人群主要集中于这段时间出行，该曲线符合现实情况。

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柱状图。

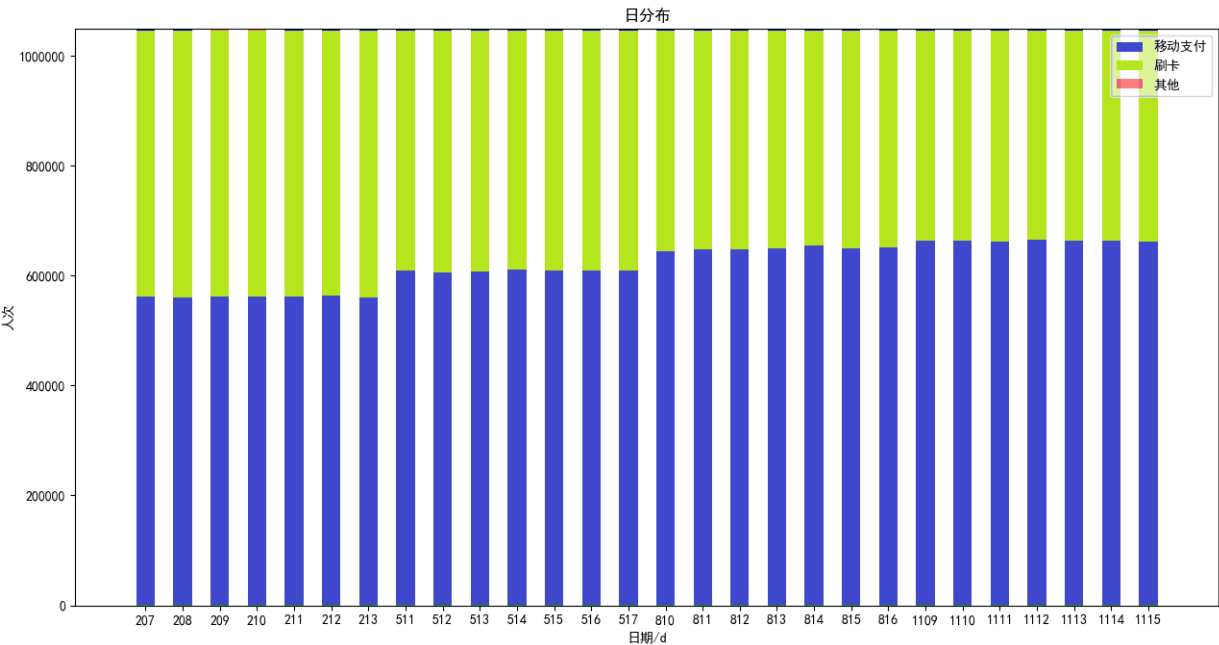


图 2 公交移动支付与总支付笔数按日期分布图

通过日期分布图，我们发现在同一个月的一周七天中，移动支付的比例几乎维持稳定，工作日与双休日的移动支付占比几乎没有差异，横向对比四个月，发现在试运营期间随时间的增加，公交移动支付的占比有所提升并最终趋于稳定。分析原因是由于移动支付试运营期间，用户对于一个新的支付方式的尝试与接受需要时间，所以在移动支付试运营的几个月中，移动支付的占比量会随着时间的积累而提高。

3. 人数与公交移动支付占比随移动支付次数分布图

计算乘车次数为 i 的乘客的人数与他们的移动支付占比，绘制人数随乘车次数增长与的散点图，存在次数过多的乘客（两万次）与次数过少（小于 4 次）。我们对这些数据进行裁剪得到有效数据作图。

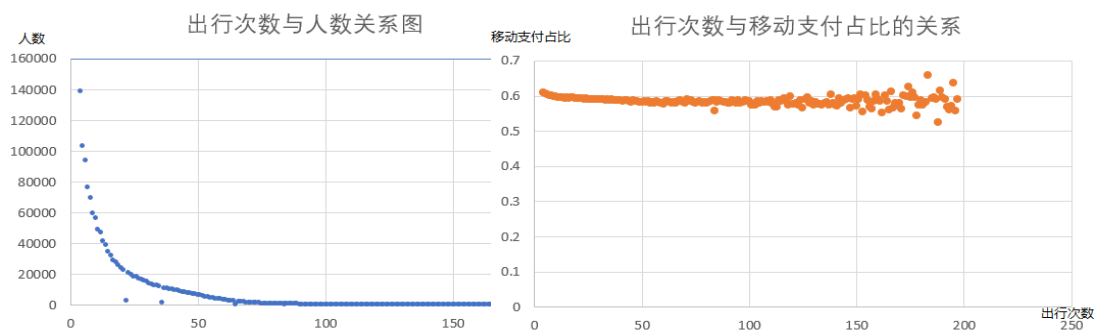


图 3 南京地铁客流量占比

通过人数随出行次数分布图反映出人数随出行次数而递减，大致符合双曲函数的分布特征。移动支付占比随出行次数分布图反映出全部用户群体的移动支付占比都稳定在 60% 左右，当次数很大的时候，波动较大的原因是由于频繁乘坐地铁的人数较少，导致数据较少，移动支付占比差异较大。

4. 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）

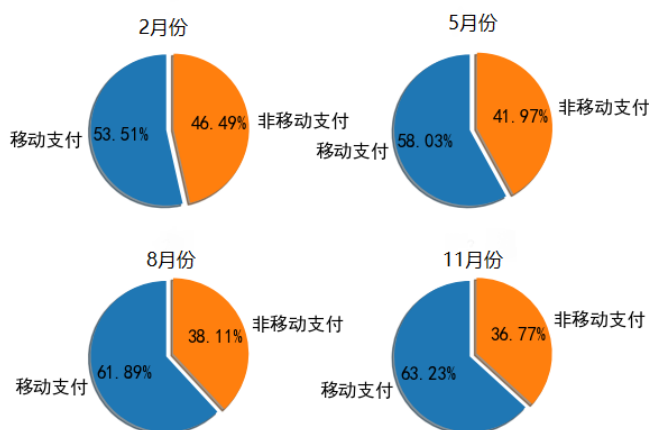


图 4 四个月支付占比

根据四个月的四张移动支付占比饼图我们可以看出四个月的移动支付占比均超过 50%，并且随着时间的增加，移动支付占比增多，增长趋势逐渐平缓，最后一个月为 63%。

5.1.3 支付行为特征分析

整理统计模型的分析，我们得到如下行为特征分析：

1. 7~9 点和 16 ~ 19 点两个早晚高峰时间段的用户支付笔数出现峰值。分析原因为上班人群与学生人群的出行集中于早晚两高峰时间。

2. 公交移动支付的占比量会随时间累计提高，最终趋于稳定。分析原因为用户尝试和接受新的支付方式需要一定的时间，最终倾向使用移动支付的人群逐渐稳定。

3. 人数随公交支付笔数的增加而下降，呈双曲函数形状。
4. 一个星期内工作日与双休日并未发现用户的移动支付占比产生太大变化，用户使用移动支付的倾向受到假期与工作日不同的影响较小。
5. 由于附件一给出的数据中每天的公交支付总笔数是固定的，所以无法得到用户支付总笔数的特征。

5.1.4 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。本文采用 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：

本题构建 K-means 模型，每一个样本点为一个用户，表示为一个 $item[24]$ 的 24 维数组，数组中 0 ~ 23 表示该用户 24 小时中每小时的公交支付总次数。通过聚类模型我们可以根据用户一天中不同时间的支付规律对用户群体进行划分。我们选择海明距离作为相似度函数计算，计算两个用户 $item_1$ 与 $item_2$ 的每小时的乘车次数差的绝对值的总和作为两个样本之间的距离 $distance$ ， $distance$ 的值越大则表示两个用户支付行为特征差异越大：

$$distance = \sum_{i=0}^{23} |item_1[i] - item_2[i]| \quad (1)$$

K-means 模型将用户样本集划分为 K 个簇，每个簇之间的 $distance$ 尽可能大，簇内的点尽可能紧密地连在一起，从而得出用户的 K 种支付行为特征，我们通过修改 K 的值，经过多次训练，找到了 K=5 时，用户行为具有较为明显的划分，用户的 5 种支付行为描绘如下图，每个子图的横坐标为小时，纵坐标为公交支付次数：

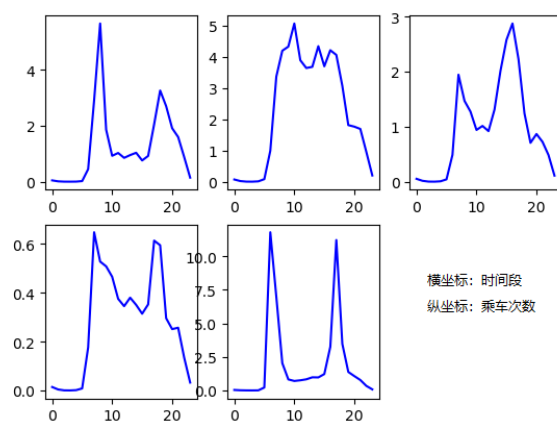


图 5 聚类分析模型

基于 K-means 模型的支付行为特征描述：

由 K-means 模型生成的五种用户支付特征图，发现五类特征较为明显的群体，我们对这五种特征进行描述：

1. 用户群体在 8 点与 18 点左右出现公交支付峰值，其中 8 点的峰值最高，约为 6 次，总出行次数不多。
2. 用户群体在 9 点至 16 点区间内出现公交支付峰值，约为 4 次，总出行次数较多。
3. 用户群体在 8 点与 15 点左右出现峰值，约为 2 次与 3 次，总出行次数不多。
4. 用户群体公交支付次数极少，几乎不乘坐公共交通。
5. 用户群体在 7 点与 18 点左右出现明显峰值，约为 12 次，总出行次数较多。

基于 K-means 模型的支付行为特征分析：

1. 综合五个群体我们发现所有用户群体基本都在 7-9 点早高峰时间段与 17-19 点晚高峰时间段出现公交支付峰值，这与我们数据统计模型所得出的特征是一致的，这一部分峰值的贡献者主要是乘坐公共交通上下班的上班族与乘公共交通上放学的学生。

2. 用户群体 1, 3, 5 具有相似的形状，其中用户群体 5 的特征最为鲜明，在早晚高峰出现了很高的支付次数，在其他时间段很少有支付记录，可推测这类用户主要为上班族与学生群体，具有规律的出行时间，并且倾向于使用公共交通作为出行工具。用户群体 1, 3 的峰值较小，推测其为偶尔使用公共交通出行的上班与学生群体。

3. 用户群体 2 在 9 点到 17 点的区间有峰值，恰好和早晚高峰错开，可以推测这类用户群体为并不需要上班与上学的用户群体，以退休的老年人为主，避开早晚高峰进行活动，主要公交出行时间为中午时分。

4. 用户群体 4 的公交支付次数极少，可推测其为非本市居民，以游客为主。此类用户由于记录很少，无法通过变化图表分析出其行为特征。

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。

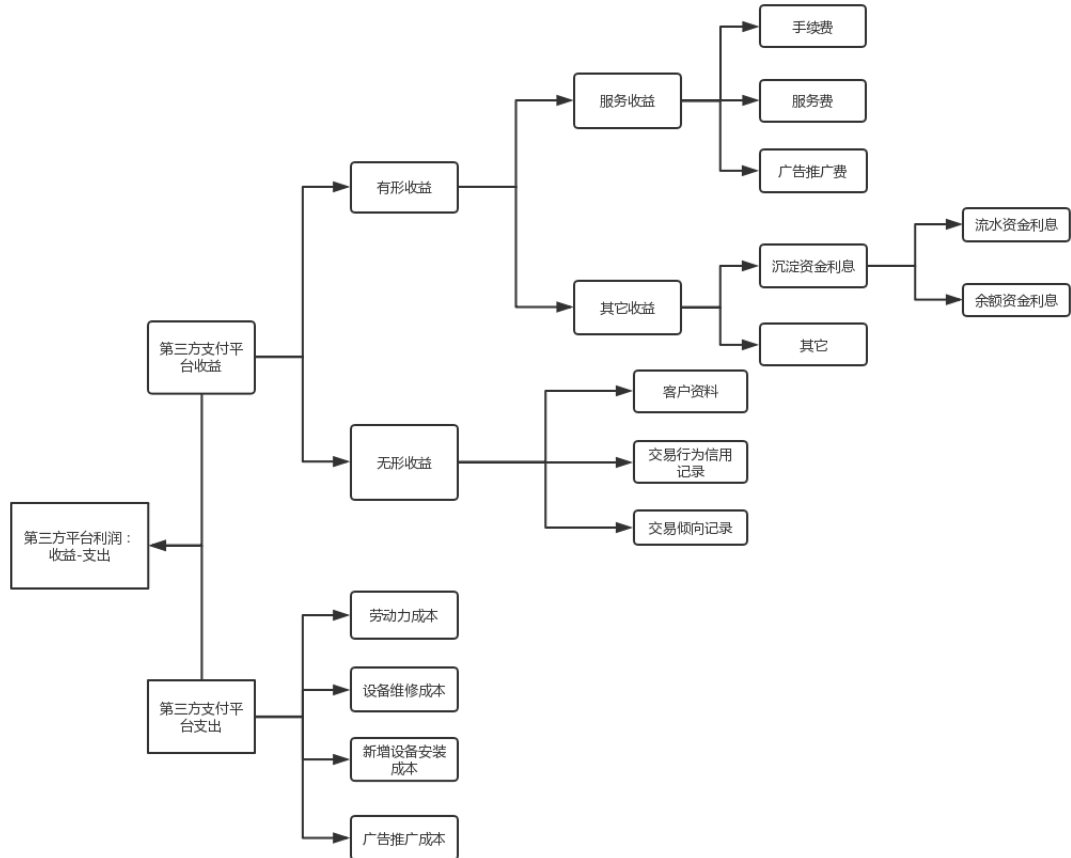


图6 盈利模型

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：

$$T_1 = S_{sum} * FreeRate \quad (2)$$

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:

$$T_2 = S_{sum} * ServiceRate \quad (3)$$

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 $adFree = 40RMB/CPM$, 广告费单位 (RMB/CPM) 单位表示的是平均每一千人分别听到或者看到某广告一次一共需要多少广告成本。广告费 T_3 的表达式为:

$$T_3 = Pay_{sum} * adFree/1000 \quad (4)$$

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。下图为我们构建的沉淀资金盈利模式:

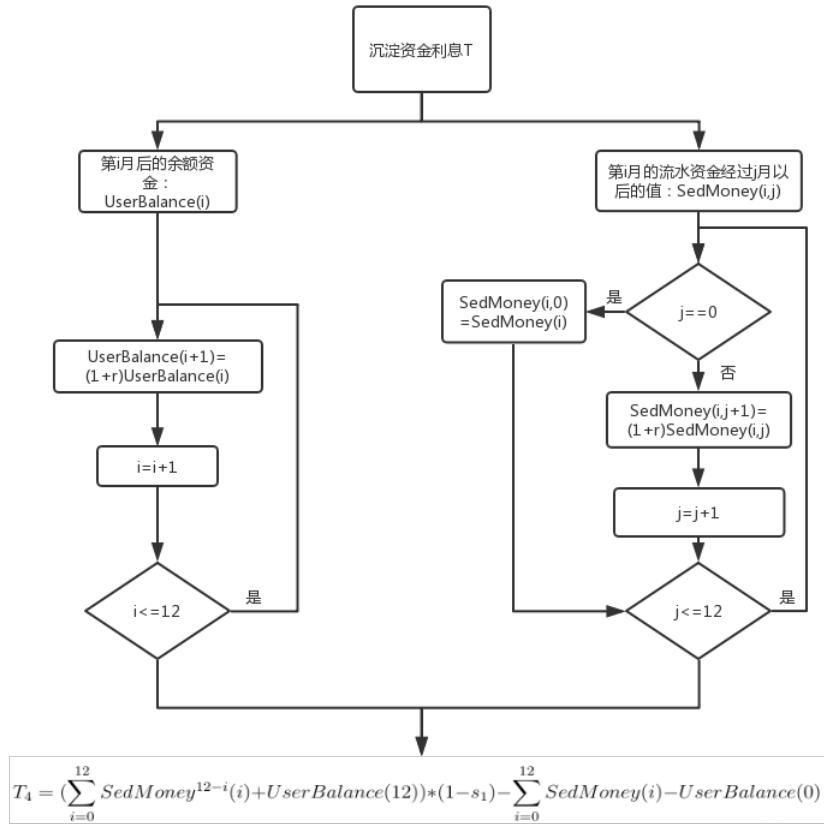


图 7 沉淀资金盈利模型

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

$$r = (1 - r_1)(a_1 * r_2 + (1 - a_1) * r_3)/12 \quad (5)$$

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。

$$UserBalance(0) = (\overline{remian} * NumberOfPeople) * a_3 \quad (6)$$

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

$$SedMoney^j(i) = \begin{cases} SedMoney(i) & j = 0 \\ (1 + r) * SedMoney^{j-1}(i) & j \neq 0 \end{cases} \quad (7)$$

$$UserBalance(i) = (1 + r) * UserBalance(i - 1) \quad (8)$$

计算沉淀资金的年利润，设银行收取的手续费率 s_1 ，由于取出资金的时候要收取手续费，所以沉淀资金在年底积累的总额除去手续费之后为原来资金的 $(1 - s_1)$ 倍：

$$T_4 = \left(\sum_{i=0}^{12} SedMoney^{12-i}(i) + UserBalance(12) \right) * (1 - s_1) - \sum_{i=0}^{12} SedMoney(i) - UserBalance(0) \quad (9)$$

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 C_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：

$$C_1 = N_{machine} * (V_{machine} + V_{maintain}) + S \quad (10)$$

2. 推广费用 C_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：

$$C_2 = N_{ad} * V_{ad} \quad (11)$$

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

$$W = T_1 + T_2 + T_3 + T_4 - C_1 - C_2 \tag{12}$$

结合表达式 1 ~ 10 可得：

$$\begin{aligned} W = & S_{sum} * FreeRate + S_{sum} * ServiceRate + Pay_{sum} * adFree/1000 \\ & + \sum_{i=0}^{12} SedMoney^{12-i}(i) - (\sum_{i=0}^{12} SedMoney^{12-i}(i) + UserBalance(12)) * (1 - s_1) \\ & - \sum_{i=0}^{12} SedMoney(i) - UserBalance(0) \\ & - (N_{machine} * (V_{machine} + V_{maintain}) + S) - N_{ad} * V_{ad} \end{aligned} \tag{13}$$

5.2.4 盈利模型定量分析

现对盈利模型中各参数赋值，进行定量分析：

表 2 盈利模型参数表

年度营业额/万元	手续费率	服务费率	交易笔数	广告费用/元每 CPM	总用户余额/万元	月均营业额/万元	机器数量/万台
43800	0.001	0.006	21900	60	1300	3650	0.1
机器单价/元	维护单价/元	劳动力成本/万元	投放广告均价/万元	投放广告数量	沉淀资金增长率	银行手续费率	
500	20	200	10	5	0.03	0.0078	

将以上数据带入盈利模型表达式 (13) 中，计算第三方支付平台的盈利，得到第三方支付平台盈利为 1648.6 万元。

5.2.5 盈利模型的敏感性分析

对模型进行敏感性分析，分析盈利模型中各个参数的取值对最终盈利造成的影响。

1. 分析利润受协定利率变化与用户余额变化的影响程度，由于附件三中说明协定变化率为 4% ~ 5%，所以变化范围限定在 4% ~ 5%。

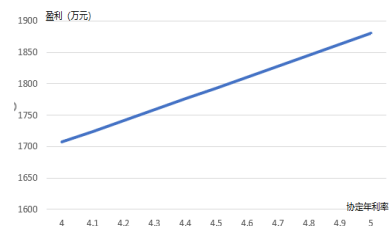


图 8 利润随协定利率增长曲线

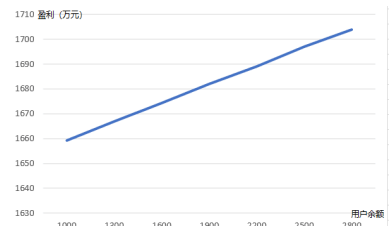


图 9 利润随用户余额增长曲线

通过两个图的分析我们可以看出，范围内的协定利率的增长并不会带来利润的大幅提高，对利润影响较小。用户余额的提升甚至翻倍也不会对利润产生过大的影响，主要是受到协定年利率的影响，所以协定年利率与用户余额二者的变化对盈利不会产生太大影响。

2. 分析利润受服务费影响程度，由于主流的支付平台的服务费率基本在 0.5% ~ 0.6%，所以绘制 0.5% ~ 0.6% 范围内服务费的变化对利润产生的影响。

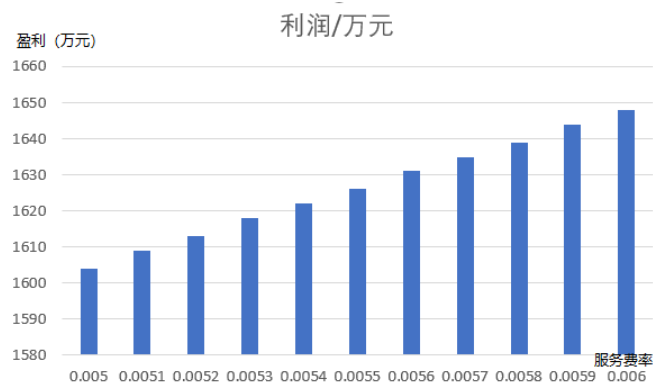


图 10 利润随服务费增长图

分析柱形图得出，范围内服务费的变化不会对利润造成太大影响。

3. 分析利润受广告费用影响程度

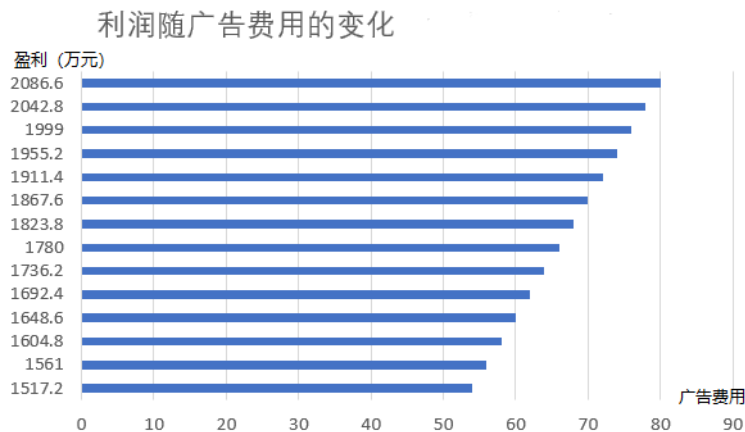


图 11 利润随广告费增长图

分析利润随广告费增长图，发现广告费的增加会对利润造成较大影响，分析其原因主要是由于用户量较大，广告浏览量，从而导致广告方面的盈利占比大。

5.3 问题三的建模

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，移动支付的人数会相应提高，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：

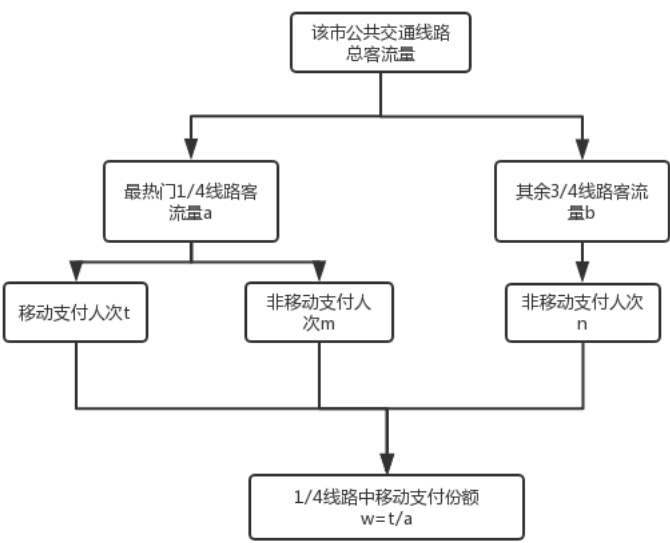


图 12 移动支付占比计算

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 3 给出了 2017 年南京各个线路的里程数与客流量的数据：

表 3 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	32441.59	S1 号线	33.0	2267.29
2 号线	37.1	27488.11	S3 号线	35.5	117.33
3 号线	43.9	24752.45	S7 号线	28.2	10.4
4 号线	32.6	3490.78	S8 号线	44.4	3574.44
10 号线	21.2	3602.58	S9 号线	52.4	6.8

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。

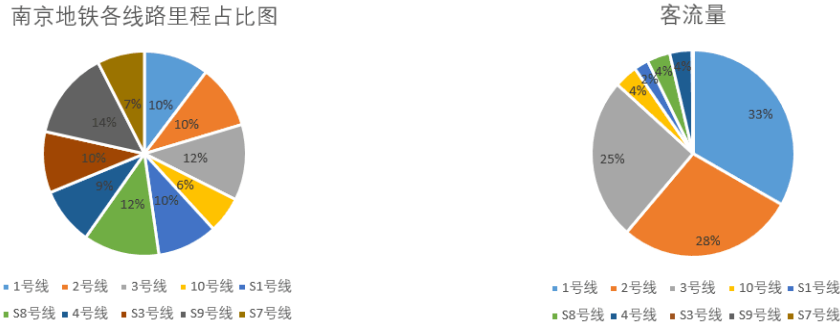


图 13 南京地铁客流量占比

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，且一二号地铁线加上三号线线路的 43% 的公里数总和占地铁总公里数的 1/4，符合模型中的场景，所以取一二号地铁线加上三号线线路的部分做为试运营的 1/4 线路。设 P_i 为十条线路中第 i 条线路的客流量， $O_{1/4}$ 就等于计算一号线二号线部分三号线的客流量占比即：

$$O_{1/4} = \frac{P_1 + P_2 + P_3 * 43\%}{\sum_{i=1}^{10} P_i} \quad (14)$$

求出 $O_{1/4}$ 后，将南京市的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：

$$MobileProportion(i, j) = \frac{DailyCharge_{(i,j)}^{MobilePay}}{(DailyCharge_{(i,j)} * O_{1/4})} \quad (15)$$

5.3.1 问题三模型的求解与分析

根据构建模型带入南京地铁的数据可得 $O_{1/4} = 72.3\%$ ，再将 $O_{1/4}$ 与该市数据带入表达式 (15) 中可得到每个月的 $MobileProportion(i, j)$ ，根据求出的每个月的 $MobileProportion(i, j)$ 绘制柱状图：

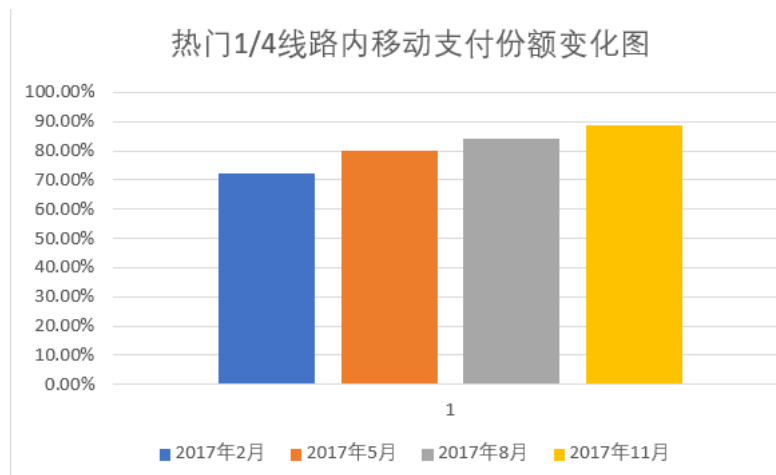


图 14 移动支付占比计算

得到每个月的移动支付占比后, $SumCharge * MobileProportion(i, j)$ 即为该市在全部覆盖公交移动支付后的移动支付用户数量, 将其带入问题二中构建的盈利模型中, 可以得到该市在实现公交移动支付全覆盖之后的第三方支付平台的盈利, 绘制图表:

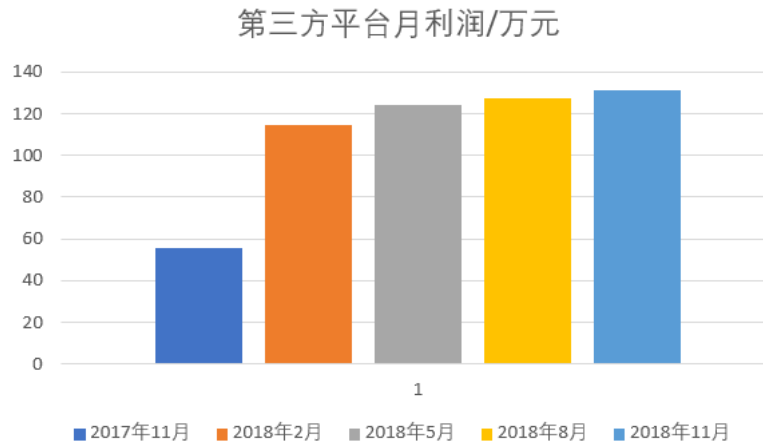


图 15 第三方支付平台每月盈利

由图表可看出, 该市在公交移动支付全部覆盖后, 短时间内第三方支付平台的盈利会出现大幅提升, 主要原因在于短期内使用移动支付的用户大量提升, 之后趋于稳定。

六、模型的评价

6.1 模型的优点

1. 本模型尽量的搜集了能获取的真实资料, 包括南京市地铁各线路的客流量与里程、各大支付平台 2017 年手续费、服务费等相关数据, 从而将实际情况大致的进行的还原。
2. 模型一使用 K-means 方法进行聚类分析得出不同的用户群体进行特征分析, 针对数据量较大的支付数据使用计算量较小的 K-means 算法, 效果较好。
3. 模型三给出了试运营期间 1/4 线路内移动支付份额随时间的变化图, 根据图片, 我们拟合了移动支付正式运营以后, 移动支付份额的变化。

6.2 模型的不足

1. 直接将 1/4 热门线路内移动支付份额推广到整体市场, 忽视了移动支付覆盖率的提升对用户消费行为的影响。
2. 模型二中对于无形收益难以给出一个定量的评估, 给盈利模型带来误差。对广告推广以及用户数量, 没有给出一个定量的关系。

6.3 模型的改进

1. 在模型求解的过程中存在许多不确定因素，包括推广费用对用户数量的影响、不同年龄段，职业对移动支付的态度等等。
2. 模型二中的只考虑了 CPM 形式广告费带来的收益，但现实情况中，可能存在 CPC、PPC、CPL 等多种形式的广告费，我们可以对其进行综合考虑，得出第三方平台的日均广告收益。

七、模型的推广与应用

1. 本文中模型三可以应用出不同城市中，估算出此城市运营第三方公共交通支付能带来的收益。
2. 本文中模型二充分考虑各种盈利方式以及支出方式，其它关于求解利润的问题也可以参考模型二，具有较强的可推广性。
3. 本文中的模型二可以不仅仅用于求解有关公交的第三方支付平台，也可以适用于求解校园生活等主题有关的第三方支付平台。

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 k-means 算法

```
def distance(item0, item1):
    res = 0
    for i in range(len(item0)):
        res += abs(item0[i] - item1[i])
    return res

def generateKernel(data, kernelNumber):
    kernelList = []
    for _ in range(kernelNumber):
        kernelList.append(data[random.randint(0, len(data))])
    return kernelList

def choose(data, kernelList):
    res = 1000000000000000000000000
    for i in kernelList:
        temp = distance(data, i)
        if temp < res:
            key = i
            res = temp
    return kernelList.index(key)

def processData():
    kernelNumber = 6
    print('[info] processing data')
    dataList = []
    originList = []
    for i in range(kernelNumber):
        dataList.append(list())
    with open(dataRoot, 'r') as raw:
        input = csv.reader(raw)
        for i in input:
            originList.append(i)
    for i in range(len(originList)):
        for j in range(len(originList[i])):
            originList[i][j] = int(originList[i][j])
    kernelList = generateKernel(originList, kernelNumber)
    for i in originList:
        dataList[choose(i, kernelList)].append(i)
    return dataList

def average(src):
    print('[info] calculating average')
    res = []
    for i in src:
```

```

    temp = []
    for _ in range(24):
        temp.append(0)
    for j in i:
        for key in range(24):
            temp[key] += j[key]
    for k in range(24):
        temp[k] /= len(i)
    res.append(temp)
return res

def draw(src):
    print('[info] drawing')
    from matplotlib import pyplot as plt
    import math
    for i in range(len(src)):
        plt.subplot(2, math.ceil(len(src) / 2), i+1)
        plt.plot([x for x in range(24)], src[i], 'b-')
    plt.show()

```

附录 B conda 环境列表

```

name: mcm
channels:
  - defaults
dependencies:
  - backcall=0.1.0=py36_0
  - certifi=2019.3.9=py36_0
  - colorama=0.4.1=py36_0
  - decorator=4.4.0=py36_1
  - ipykernel=5.1.0=py36h39e3cac_0
  - ipython=7.5.0=py36h39e3cac_0
  - ipython_genutils=0.2.0=py36h3c5d0ee_0
  - jedi=0.13.3=py36_0
  - jupyter_client=5.2.4=py36_0
  - jupyter_core=4.4.0=py36_0
  - libsodium=1.0.16=h9d3ae62_0
  - parso=0.4.0=py_0
  - pickleshare=0.7.5=py36_0
  - pip=19.1.1=py36_0
  - prompt_toolkit=2.0.9=py36_0
  - pygments=2.4.0=py_0
  - python=3.6.8=h9f7ef89_7
  - python-dateutil=2.8.0=py36_0
  - pyzmq=18.0.0=py36ha925a31_0

```

```
- setuptools=41.0.1=py36_0
- six=1.12.0=py36_0
- sqlite=3.28.0=he774522_0
- tornado=6.0.2=py36he774522_0
- traitlets=4.3.2=py36h096827d_0
- vc=14.1=h0510ff6_4
- vs2015_runtime=14.15.26706=h3a45250_4
- wcwidth=0.1.7=py36h3d5aa90_0
- wheel=0.33.2=py36_0
- wincertstore=0.2=py36h7fe50ca_0
- xlwt=1.3.0=py36h1a4751e_0
- zeromq=4.3.1=h33f27b4_3
- pip:
  - autopep8==1.4.4
  - cycpler==0.10.0
  - entrypoints==0.3
  - flake8==3.7.7
  - h5py==2.9.0
  - joblib==0.13.2
  - kiwisolver==1.1.0
  - matplotlib==3.0.3
  - numpy==1.16.3
  - pandas==0.24.2
  - pycodestyle==2.5.0
  - pyflakes==2.1.1
  - pyparsing==2.4.0
  - pytz==2019.1
  - scikit-learn==0.21.1
  - scipy==1.3.0
  - sklearn==0.0
  - xlrd==1.2.0
```

prefix: D:\programFiles\Anaconda\envs\mcm