

“拍照赚钱”的任务定价

摘要

日前,“拍照赚钱”——一种新兴的利用互联网赚钱的应用软件,受到了广大厂商和用户的欢迎。以此为背景,本文针对“拍照赚钱”的定价问题,采用定性与定量相结合和多元线性回归方法对任务定价规律进行了分析,找到了任务未完成的原因,进而改进了原定价方案存在的问题,提出了新的定价方案,并对此方案进行了实施效果的评价,对新项目的定价具有一定的参考意义。

针对问题一,为了获得项目的定价规律,采取定性分析和定量分析相结合的方法,以数据样本为对象,对各种影响因素进行了分析。通过分析各种影响因素与定价的相关系数,得出了任务定价规律的缺陷。在得出了项目的定价规律的基础上,再对任务未完成的原因进行分析。任务未完成的原因不单由价格这一因素决定,而是由多种因素(如任务位置,会员分布情况以及会员密度等)共同影响的。用 MATLAB 分析出各因素对任务完成情况的影响大小,得出任务没有完成的原因是因为定价方案仅考虑了会员密度这一个因素。

针对问题二,在设计新的定价方案时确立了更多影响因素,利用附件中已完成任务的数据建立多元回归数学模型,确定定价与影响因素之间的函数,做出新的定价方案。再用部分数据进行对方案的验证,与原方案进行比较。

针对问题三,对于之前的定价模型,采用 K-Means 聚类分析对现有的任务分布进行分类,以每一类任务的中心点作为圆心确定性能指标。再根据折扣和会员相对中心点的距离成反比的原则确定每一个任务的打折力度,建立新的多元回归数学模型,分析最终的任务完成情况。

针对问题四,利用既得的多元任务线性回归模型对多种因素进行了综合考虑后得出了最终定价。

关键词: 多元线性回归 K-Means 聚类 相关度分析

一、问题重述

1.1 问题背景

“拍照赚钱”是一种以互联网为载体的会员制 APP。用户通过下载 APP，注册成为会员，并领取拍照任务来利用软件获取相应的酬劳。对于企业而言，这种新型调查平台不仅可以快速直接地提供准确有效的信息，而且可以为企业节约成本。对于用户而言，这种自助式劳务众包软件也提供了新型的赚钱方式。因此，要想使 APP 更受欢迎，APP 中的任务定价是其核心要素。如果定价不合理，有的任务就会无人问津，进而导致企业无法获取商品信息，平台失去大量用户。

1.2 问题描述

题目中所给的附件一是已结束项目的任务数据，包含了每个任务的位置、定价和完成情况；附件二是会员信息数据，包含了会员的位置、信誉值、参考其信誉给出的任务开始预订时间和预订限额；附件三是一个新的检查项目任务数据，只有任务的位置信息。

请建立数学模型，完成下面的问题：

1. 研究附件一中项目的任务定价规律，分析任务未完成的原因。
2. 为附件一中的项目设计新的任务定价方案，并和原方案进行比较。
3. 实际情况下，多个任务可能因为位置比较集中，导致用户会争相选择，一种考虑是将这些任务联合在一起打包发布。在这种考虑下，如何修改前面的定价模型，对最终的任务完成情况又有什么影响？
4. 对附件三中的新项目给出你的任务定价方案，并评价该方案的实施效果。

二、模型假设

- (1)假设任务价格越低任务完成的越快；
- (2)假设获得参与任务机会的高低仅与会员的信誉值有关，会员的信誉值越高，获得任务参与机会的可能性越大；
- (3)假设所研究的地区没有发生严重的经济危机和社会动荡；
- (4)假设任务开始时间与定价无关；
- (5)假设平台的成本与费用与任务定价无关；
- (6)假设平台提供和数据真实可靠。

三、符号说明

符号	含义，单位
Y_i	各个完成任务的定价($i = 1, 2, 3, \dots, n$)，元
ρ	会员密度
T_i	会员与任务之间平均距离($i = 1, 2, 3, \dots, n$)，千米
E	会员的平均信誉值
k_1	会员密度比例系数
	平均距离比例系数
	平均信誉值比例系数
	折扣比例系数

k_2	折扣 距离, 千米
k_3	打折后的新价格, 元 打折前的原价格, 元
k_4	每个任务包的价格, 元
S	
d_{ij}	
R_N	
R_O	
R_D	

四、问题一的分析与解决

4.1 问题一的分析

(1)问题一要求研究项目的任务定价规律, 分析任务没有完成的原因。

先对任务定价进行定性分析: 当研究任务定价规律时, 由于产品的定价受多种因素 (如: 任务地理位置, 会员所处位置, 天气, 交通等) 的影响, 其中有主要影响因素也有次要影响因素。因此在分析的过程中, 本文对附件中体现的核心因素依次进行了定性分析, 忽略了附件中未体现的次要因素 (如: 天气, 交通等) 对价格的影响。运用 MATLAB 和智图编辑器绘制了价格与各主要影响因素之间的关系图, 最后通过对图像的分析 and 图像间的对比, 得出了各因素对价格的影响。

再对任务定价进行定量分析: 通过计算相关系数分析价格与会员密度, 价格与会员和任务间的距离和价格与会员信誉度之间的关系, 确定项目的任务定价规律

(2)在分析任务没有完成的原因时, 尽管任务没有完成与定价规律有一定的联系, 但是任务未完成原因也是受多种因素 (如任务地理位置, 会员密度, 交通, 天气等) 影响的。所以此时, 我们则不能忽略附件中没有体现的交通等因素的影响。用 MATLAB 和智图编辑器画出图像, 结合定价规律对任务的完成情况进行分析, 进而得出任务没有完成的原因。

4.2 问题一的结果分析

4.2.1 任务定价规律的结果分析

(1)先对任务定价进行定性分析

用 MATLAB 分别做出任务价格分布图和会员位置分布图。通过对比两图来进行任务价格定性的分析。

由图 4-1 可以看出任务定价较低的（用蓝色和绿色表示的）点主要分布在纬度 113.6° — 114.2° ，经度 22.8° — 23.4° ；纬度 113° — 112.4° ，经度 22.8° — 23.4° 两个范围内。而价格较高的（用绿色和褐色表示的）点主要分布在纬度 113° — 113.6° ，经度 22.6° — 23.2° 范围内。

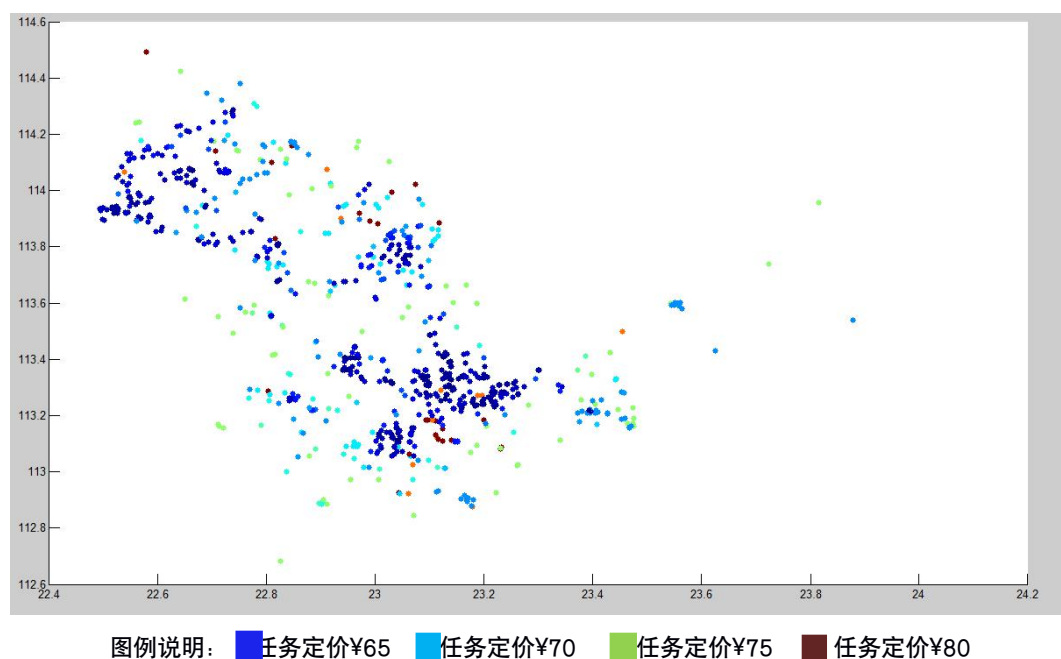


图 4-1 任务价格分布图

由图 4-2 可以看出会员主要分布在在纬度 113.6° — 114.2° ，经度 22.6° — 23.2° ；纬度 112.8° — 113.4° ，经度 22.8° — 23.4° 两个范围内。

对比图 4-1 和图 4-2 可以得出：会员密度较高的地区，任务定价较低；而会员密度较低的地方，任务定价相对较高。由此可知任务定价与会员密度呈负相关。

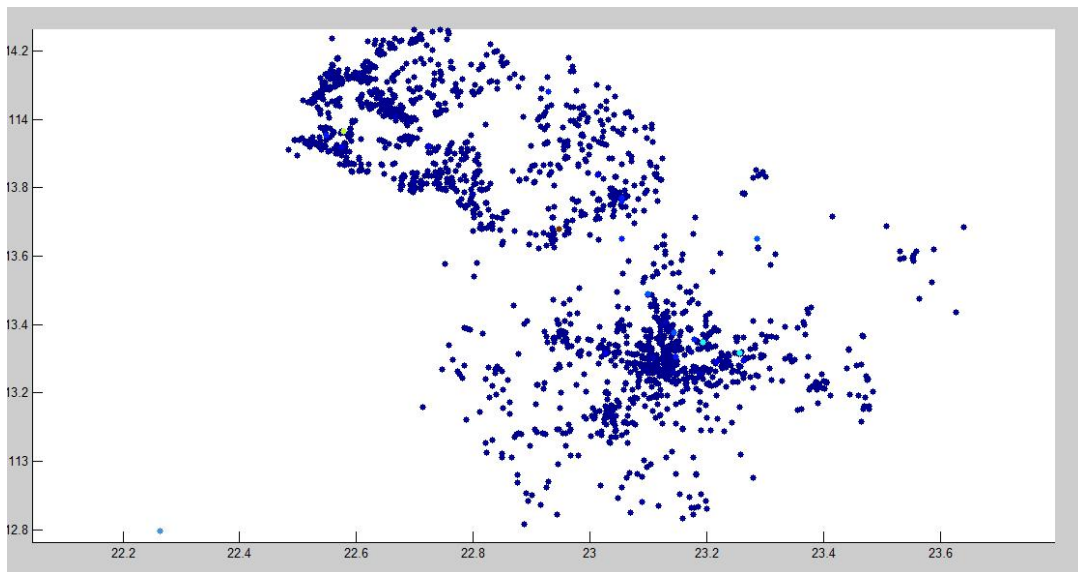


图 4-2 会员位置分布图

(2)再对任务定价进行定量分析

用 MATLAB 编程，通过计算相关系数分析价格与会员密度，价格与会员和任务间的距离和价格与会员信誉度之间的关系，确定项目的任务定价规律。

结果如图 4-3 所示

```
s1 =

    1.0000    -0.4024
   -0.4024     1.0000

s2 =

    1.0000     0.1326
     0.1326     1.0000

s3 =

    1.0000   -0.2056
   -0.2056     1.0000
```

图 4-3 定量分析结果图

对结果进行分析：s1 表示价格与会员人数分布的相关系数，s2 表示价格与会员和任务间距离的相关系数，s3 表示价格与会员信誉度的相关系数。相关系数越大表示相关性越强，因此价格与会员人数的相关性最强。会员人数越多的地方，任务的价格越低。

综合定性分析和定量分析的结果，任务的定价仅与会员人数的分布有关，且呈负相关关系。

4.2.2 任务未完成原因的结果分析

用 MATLAB 画出任务完成情况分布图，用智图编辑器画出任务坐标图，再结合图一任务价格分布图和图二会员位置分布图分析任务未完成的原因。

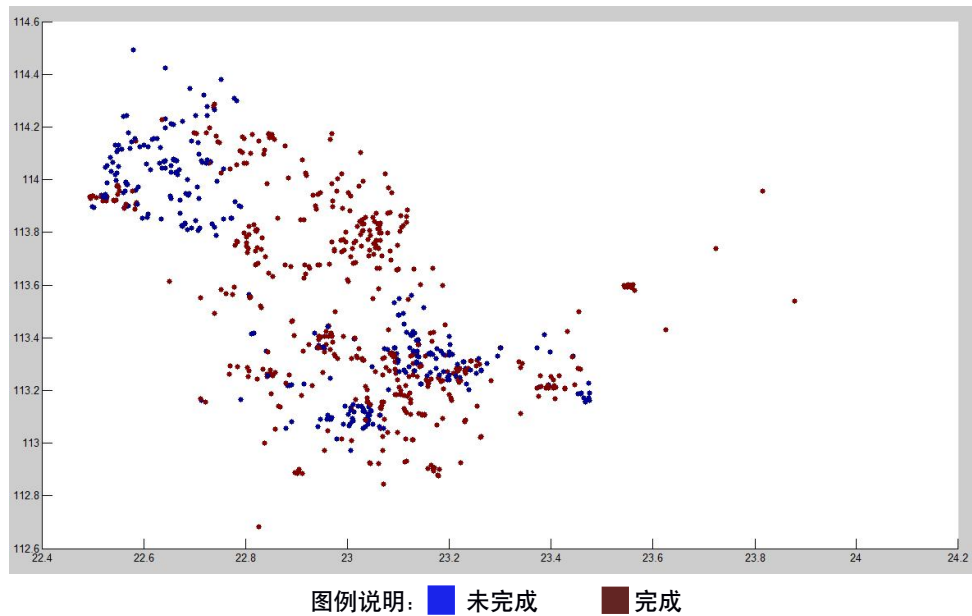
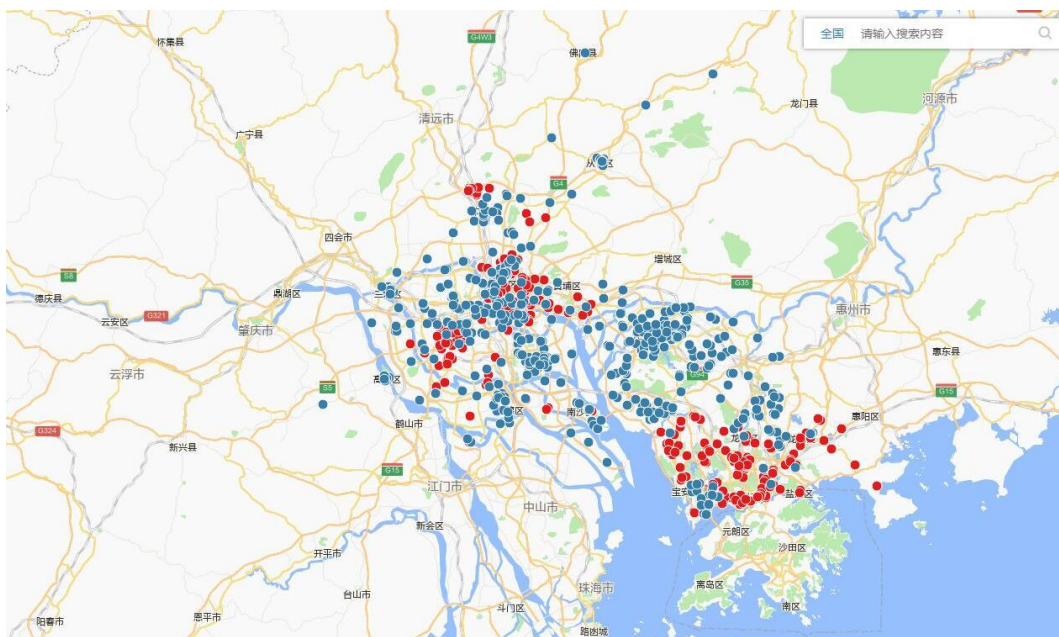


图 4-4 任务完成情况分布图

由图 4-4 可知,褐色点表示的已完成的任务集中分布在纬度 $113^{\circ} - 114^{\circ}$, 经度 $22.8^{\circ} - 23.2^{\circ}$ 区间内; 而蓝色点表示的未完成任务主要分布在纬度 $113.8^{\circ} - 114.4^{\circ}$, 经度 $22.6^{\circ} - 22.8^{\circ}$ 区间和纬度 $113^{\circ} - 113.4^{\circ}$, 经度 $23^{\circ} - 23.4^{\circ}$ 区间内。



图例说明: ■ 完成 ■ 未完成

图 4-5 任务坐标图

由图 4-4 可以更加直观的从地图上看出已完成任务和未完成任务的分布情况。

分析得出任务未完成的原因:

(1)由对任务价格规律的分析可知,任务价格仅与会员人数分布有关,而没有考虑其他因素的影响,因此该定价不合理,使很多任务无人问津,进而导致很多任务没有完成。

(2)对比图 4-1 任务价格分布和图 4-4 任务完成情况分布可以看出,任务价格较低的任务分布区域和未完成任务的分布区域有大部分的重合,而完成的任务则主要分布在中间价格点的区域。因此可以得出任务未完成的原因之一是有些任务的定价过低。结合实际情况进行分析,正是由于价格设置过低,很多会员不愿意选择这些项目,导致任务没有完成。

(3)对比图 4-2 和 4-5 得出会员分布与任务是否完成的关系。由图可看出,在会员分布较多的区域,任务未成功的情况较多。由此可以得出,会员人数多少与任务是否成功呈负相关关系。

(4)分析图 4-4 和图 4-5 可以得出已完成的任务主要分布在城市中心,交通便利的地方,而未完成的任务则大致分布在沿海和交通路线较少的地区。因此可以得出结论,交通的便利与否对任务的完成有影响。
综上,任务为完成的原因大致有四:项目的定价规律不合理,价格设置过低,会员密度过大和交通不便利。

五、问题二的模型建立、分析与解决

5.1 问题二的分析

本题要求为附件一中的项目设计新的任务定价方案,并将此方案与原方案进行比较。问题一中的方案只考虑了会员分布这一个因素,所以导致任务定价不合理,有的任务无人问津。本文认为设计定价方案应该考虑多种因素的影响,如会员密度、平均距离和平均信誉度等,建立多元线性回归数学模型,确定指标与定价函数,与原方案进行比较,确定最优方案。

第一步:由问题一的方案分析统计完成任务的定价,第二步:确定会员密度。不同的任务周边有不同数量的会员,且不同会员的信誉值不同,所以选取不同的任务与会员的信誉值样本,平均信用值与平均距离等参数,得到具有多样性和普遍性的样本。最后,建立多元线性回归数学模型,确定指标与定价函数,得出比较结果。

5.2 问题二模型的建立与解决

5.2.1 模型的准备与指标选择

(1)本题中运用了多选线性回归方程，下面对多元线性回归模型进行简单的分析介绍：

一元线性回归是因变量主要只受一个自变量影响，而在现实问题的研究中，因变量的变化往往受多个重要因素的影响，此时因变量的变化就要由两个或两个以上的影响因素来解释，这就是多元回归亦称多重回归。当多个自变量与因变量之间是线性关系时，所进行的回归分析就是多元线性回归。设 y 为因变量， x_1, x_2, \dots, x_k 为自变量，并且自变量与因变量之间呈线性关系时，则多元线性回归模型为

$$y = b_0 + b_1 x_1 + b_2 x_2 + e \quad (5-1)$$

建立多元线性回归模型时，为了保证回归模型具有优良的预测效果和解释能力，应注意自变量的选择。

多元性回归模型的参数用最小二乘法来估计。以二线性回归模型为例，求解回归参数的标准方程组为

$$\begin{cases} \sum y = nb_0 + b_1 \sum x_1 + b_2 \sum x_2 \\ \sum x_1 y = b_0 \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 \\ \sum x_2 y = b_0 \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2 \end{cases} \quad (5-2)$$

解此方程可求得 b_0, b_1, b_2 的数值，亦可用下列矩阵法求得

$$b = (x'x)^{-1} \cdot (x'y) \quad (5-3)$$

即

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum y \\ \sum x_1 y \\ \sum x_2 y \end{bmatrix} \quad (5-4)$$

(2)为了拟合分析多个变量，本文决定使用 MATLAB 中的 regress 函数，其调用格式为 $[b, bint, r, rint, stats] = \text{regress}(y, X)$ ，下面分别解释参数

B:回归系数，是个向量；
BINT:回归系数的区间估计；
R:残差；
RINT:置信区间；

STATS:用于检测回归模型的统计量，有四个数值：判定系数 R^2 (相关性函数)，F 统计量观测值，检测的 P 的值，误差方差的假设。

(3)在比较新的任务定价方案和原方案时，运用了鲁棒性。鲁棒性用以表示控制系统对特性或参数扰动的不敏感性，又称为抗变性^[2]。

5.2.2 数学建模的建立

数学模型框图如图 5-1 所示：

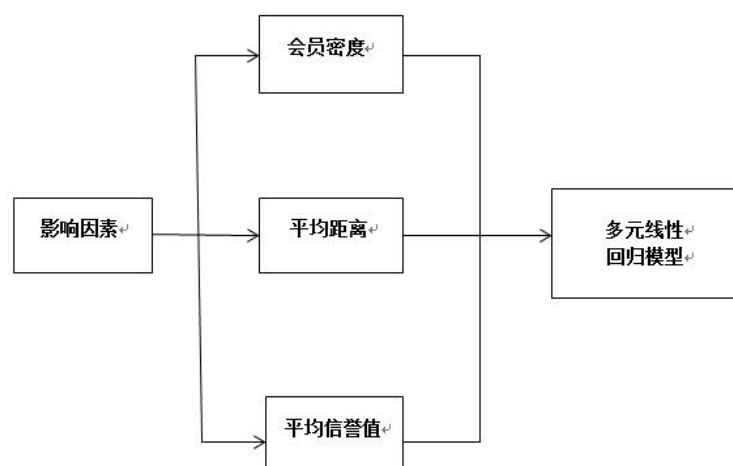


图 5-1 模型建立框图

步骤一：统计各个完成任务的定价 $Y_i (i = 1, 2, 3, \dots, n)$ ；

步骤二：计算会员密度，即以任务点为中心，固定半径（10km）的圆形区域范围内的会员人数，

$$\rho = \sum_{0}^{10} \text{人数} \quad (5-5)$$

计算会员与任务之间的平均距离，即以任务点为中心，半径为 10km 范围内的所有会员与该任务的平均距离，

$$\bar{l}_i = \frac{\sum_{0}^{10} (\text{第 } i \text{ 个任务与各个会员的距离})}{\text{会员人数}} \quad (5-6)$$

计算会员的平均信誉值，

$$E = \frac{\sum_i A_i}{\text{会员人数}} \quad (5-7)$$

其中， A_i 为第 i 个会员的信誉值。

步骤三：建立多元回归数学模型，

$$Y_i = k_1 \rho + k_2 \bar{l}_i + k_3 E + C \quad (5-8)$$

其中，C 为常数；K 分别为各个参量的系数，stats 中 R^2 为相关系数，体现了相关性的大小， R^2 越大相关性越大， R^2 越小相关性越小。

5.2.3 模型的求解

(1)在 MATLAB 中利用 regress 函数编写程序，即可求出比例系数 k_1, k_2, k_3 以及常数 C，得多元回归方程

$$Y_i = -0.0243\rho + 91.4458\bar{l}_i - 0.0023E + 67.0946 \quad (5-9)$$

相关系数为

$$R^2 = 0.4682 \quad (5-10)$$

(2)本文中 $R^2 = 0.4682$ 说明拟合模型效果较好。另外，解决本问时加入了对会员密度、平均距离、平均信誉值三个因素的考虑。由于考虑的因素较多，所以系统的鲁棒性较强，即系统的抗变换性较强，在一定的参数摄动下更能维持性能特性以适应环境的变化^[3]。综上，可以验证新的任务定价方案比原方案更好。

5.3 模型的评价与比较

多元回归模型的优点：

- 1、在分析多因素模型时，将会更加方便简单；
- 2、此模型可准确地计量回归拟合程度和各个因素之间的相关程度的高低，提高预测方程式的准确率。

多元回归模型的缺点：

在回归分析中，选用何种因子和该因子采用何种表达式有时只是一种推测，这影响了某些因子的多样性和不可测性，使得回归分析在某些情况下受到限制。

六、问题三的建立模型、分析与解决

6.1 问题三的分析

实际情况下，位置比较集中的若干任务会导致用户间的竞争力增大，对此考虑将这些任务联合打包。和前两问中的把每个任务用经纬度表示不同，本问利用 MATLAB 自带的 distance 函数通过参数调整，在保证每个任务周围有一定数量的

会员密度情况下，调整 distance 为 0.12，并对定价模型进行重新分析。以距离相近的任务区域内一点为初始簇中心，最短距离为依据打包每个任务，由此可将位置较为集中的任务划分为多个簇，抽取成一个模型。图 6-1 展示了本文对问题三 的模型建立方法。

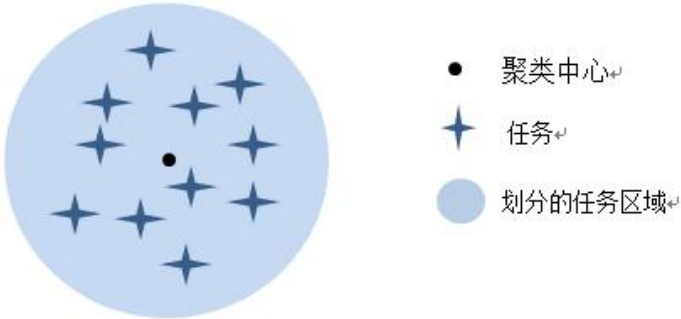


图 6-1 模型说明图

在对任务进行了打包后，每组任务的价格应该作相应的下调。而价格的折扣力度则仅仅与会员和聚类中心的距离有关，每组任务包的价格则是每组内所有任务的折扣后价格的总和。所以本文利用 K-Means 算法求出每组任务包数据的任务数和包的中心点，再求出不同任务的折扣，进行求和后得出了新的定价模型。

6.2 模型的建立与解决

6.2.1 模型的准备与介绍

由于本题采用了 K-Means 算法，在此我们对 K-Means 模型进行简单的介绍：K-Means 算法是输入聚类个数（用 k 表示），以及包含 N 个数据对象的数据库，同时输出满足方差最小标准 k 个聚类的一种算法。K-Means 算法接受输入量 k ；然后将 N 个数据对象分为 k 个聚类便于使所获得的各聚类满足：同一聚类中的对象相似度较高，而不同聚类中的对象相似度较小。

k-means 算法基本步骤

- (1) 从 N 个数据对象随机选择 k 个对象作为初始聚类中心；
- (2) 根据每个聚类对象的均值计算每个对象与这些中心对象的距离，并根据最小距离重新对这些对象进行划分；
- (3) 重新计算每个聚类的均值；
- (4) 计算标准测度函数，当满足一定条件（如函数收敛时），那么算法终止；如果条件不满足，则回到步骤（2）。

在本题中，把所有的任务分为 80 个簇，计算出每个任务包的任务个数，和每个任务包的中心点，从而确定对每个任务如何打折，最后算出每个任务包的新价格。

6.2.1 模型的建立

首先，我们使用上文提到的 K-Means 聚类算法对任务进行打包。以一点为初始簇中心，以最短距离为依据打包任务，这样就可以满足：同一聚类中的对象相似度较高；而不同聚类中的对象相似度较小的要求。本题中，输入聚类个数为 80 个。通过 MATLAB 编程可以计算出每个任务包的任务数 N 和每个任务包的中心点 A 。

接着，进行每个任务距离中心点的距离的计算。

我们可以知道对任务进行打包后，任务的定价会比打包前单个任务的总价格低，即会在原来定价的基础上打一定的折扣。因为折扣（用 S 表示）与任务包内的每个包到中心点的距离（用 d 表示）相关，所以折扣的计算公式为

$$S = 1 - 2d \quad (6-1)$$

计算出折扣后，就可以用原价格（用 R_o 表示）乘折扣求出每一个任务打包后的新价格（用 R_N 表示），公式为

$$R_N = R_o \times S \quad (6-2)$$

最后，对每个任务的新价格进行求和可以计算出每个任务包的价格

$$R_D = \sum_{N=1}^n R_o \quad (6-3)$$

6.2.2 模型的求解

(1)用 MATLAB 进行编程，运用 K-Means 函数得出类号和中心点。再通过公式 6-1 至 6-3 算出聚类后的最新定价(聚类后的最新定价的部分结果见表 6-1，其余价格见支撑材料)。

通过对结果进行分析可以看出，折扣与会员相对中心点的距离成反比，即会员距离中心点越远，所打的折扣就越小；反之，如果会员距离中心点较近，则所打的折扣相对较大。

表 6-1 聚类后的最新定价

中心位置（纬度）	中心位置（经度）	聚类后定价	所有任务新定价
23.08002847	113.9806002	214.4683523	64.68766503
22.88997713	112.8558417	392.9820048	64.12408032

22.68658673	113.8292514	708.8207692	61.44739961
23.13196304	113.4185674	1019.991442	64.99894288
23.12564372	113.6492187	527.0340657	61.63012573
22.71071204	113.9892717	196.3445581	64.27814154
22.95824422	113.256553	142.998081	63.92385914

(2)对最终任务完成情况的影响的分析

a.任务打包后使得会员竞争的次数降低;

b.使会员在完成相同的任务数量的情况下, 减少了行走距离, 同时获得了更多的利益;

c.便于公司发布任务;

d.减少费用: 减少了发布次数和后台维护管理费用, 同时由于打折也减少了发布者的花销。

6.3 模型的评价:

1.模型的优点:

K-Means 聚类算法的优点主要集中在:

(1)算法较为快速、简单;

(2)在大数据处理方面具有较高的效率并且是可伸缩的;

(3)时间复杂度近于线性, 而且适合挖掘大规模数据集。

2.模型的缺点:

聚类分析通常遇到的困难是:

(1)聚类个数 K 并不能预先确定。需要找到一个有效性指标,来确定最佳聚类个数;

(2)类中心点不能预先知道,本题中需要用 K-Means 函数求出。

八、问题四的模型建立、解决与分析

7.1 问题四的分析

利用既得的多元线性回归模型, 将附件三中计算生成的会员个数、平均距离和平均信誉值等多种因素进行了综合考虑, 并得出了最终定价。

7.2 问题四模型的建立与解决

多元线性回归模型为

$$Y_i = k_1 \rho + k_2 \bar{l}_i + k_3 E + C \quad (7-1)$$

其中, C 为常数, K 分别为各个参量的系数, ρ 为会员密度, \bar{l}_i 为平均距离, E 为平均信誉值。

$$Y_i = -0.024\rho + 91.45\bar{l}_i - 0.002E + 67.1 \quad (7-2)$$

将附件三中计算生成的会员个数、平均距离和平均信誉值等多种因素代入上述模型得出了最终定价，部分结果见表 7-1,其余结果见支撑材料。

表 7-1 新项目定价

任务号码	任务 GPS 纬度	任务 GPS 经度	新价格
C0001	22.73004117	114.2408795	70.54826
C0002	22.72704287	114.2996199	70.26776
C0003	22.70131065	114.2336007	70.08108
C0004	22.73235925	114.2866672	70.18162
C0005	22.71839144	114.2575495	70.18798
C0006	22.75392493	114.3819253	72.26662
C0007	22.72404221	114.2721836	69.98224

八、参考文献

- [1]姜启源，谢金星等.数学模型[M].4 版.北京：高等教育出版社，2001.1.
- [2]王忠,管章岑.图像数字水印抗几何变换方法的研究,《科学技术与工程》[J], 2007 年 09 期.
- [3]李蓓. 一种抗几何变换的鲁棒数字水印方案,《华中师范大学》[J], 2006 年.

附录

%%%%%%%% MATLAB 编程 %%%%%%%%%

%%%%%%%%
% 第一问 %
%%%%%%%%

%第一问需用到第二问生成表格，故将此文放到第二问后。

%%%%%%%%
% 第二问 %
%%%%%%%%

x=xlsread('附件一：已结束项目任务数据 111111',3,'B2:B523');%%任务的经纬度
y=xlsread('附件一：已结束项目任务数据 111111',3,'C2:C523');
p=xlsread('附件一：已结束项目任务数据 111111',3,'D2:D523');%%价格

m=xlsread('附件一：已结束项目任务数据 111111',2,'B2:B1878');%%用户的纬度
n=xlsread('附件一：已结束项目任务数据 111111',2,'C2:C1878');%%用户的经度
q=xlsread('附件一：已结束项目任务数据 111111',2,'E2:E1878');%%信誉值

lat=xlsread('附件一：已结束项目任务数据 111111',1,'B2:B836');% 所有任务的经度
lng=xlsread('附件一：已结束项目任务数据 111111',1,'C2:C836');% 所有任务的经度
%u=0;
%data=[];

for h=1:1:522 %%任务循环
 u(h)=0;
 xinyu(h)=0;
 d(h)=0; %%各个距离
 for j=h:1876 %%会员循环
 H=distance(x(h),y(h),m(t),n(t));%%任务与会员之间的距离
 if (H<0.1) %%画圈

 u(h)=u(h)+1; %%圈内个数之和

 d(i)=d(h)+H; %%各个任务与会员之间的距离之和

```

        xinyu(h)=xinyu(h)+q(t);%%圈内会员信誉值之和
    end
    if (u(h)==0)
        daver(h)=0;        %%%平均距离
        xinyuaver(h)=0;
    else
        daver(h)=d(h)/u(h);
        xinyuaver(h)=xinyu(h)/u(h);    %%%平均信誉值
    end
end
end
xlswrite('renwu',u,1,'B2');        %%将圈中会员个数输出到 renwu.xls 表格
xlswrite('renwu',daver,1,'C2');    %%将平均距离输出到 renwu.xls 表格
xlswrite('renwu',xinyuaver,1,'D2'); %%将平局信誉值输出到 renwu.xls 表格

f=[ones(1,522); u; daver; xinyuaver];%%将会员个数、平均距离和平局信誉值作为变量
F=f';
[b,bint,r,rint,stats]=regress(p,F); %%用 regress 函数进行多元线性拟合并返回值
num=[lat lng];
[idx,C]=kmeans(num,80);

%%第一问

geshu0=xlsread('renwu',1,'B2:B523');
juli0=xlsread('renwu',1,'C2:C523');
xinyu0=xlsread('renwu',1,'D2:D523');
jiage0=xlsread('renwu',1,'E2:E523');

s1=corrcoef(geshu,jiage)
s2=corrcoef(juli,jiage)
s3=corrcoef(xinyu,jiage)

%%%%%%%%%%
%        第三问        %
%%%%%%%%%%

jiage=xlsread('附件一：已结束项目任务数据 111111',1,'D2:D836'); % 所有任务的经度

for i=1:1:80    %分成 80 类,就有 80 个任务，对每个任务定价
    baojiage(i)=0;

```

```

for j=1:1:835
if (Idx(j)==i)
discount(i)=2*distance(C(i,1),C(i,2),lat(j),lng(j));    %%打包后，任务距离类中心点越
远，折扣越大
jiage(j)=jiage(j)*discount(i)*10;
baojiage(i)=baojiage(i)+jiage(j);
end
end
end
%%下一步,重新定义打包后的任务位置和任务价格。
xlswrite('dabaorenwu',C(:,1),1,'A1');
xlswrite('dabaorenwu',C(:,2),1,'B1');
xlswrite('dabaorenwu',jiage,1,'C1');
xlswrite('dabaorenwu',baojiage,1,'D1');

%%%%%%%%%%%%%%
%          第四问          %
%%%%%%%%%%%%%%

x=xlsread('附件三：新项目任务数据',1,'B2:B2067');    %% 所有任务的纬度
y=xlsread('附件三：新项目任务数据',1,'C2:C2067');    %% 所有任务的经度

for aa=1:1:2066    %%任务循环
uu(aa)=0;
xinyu(aa)=0;
dd(aa)=0;    %%各个距离
for bb=1:1877    %%会员循环
H=distance(x(aa),y(bb),m(aa),n(bb));    %%任务与会员之间的距离
if (H<0.1)    %%画圈
uu(aa)=uu(aa)+1;    %%%圈内个数
dd(aa)=dd(aa)+H;
xinyu1(aa)=xinyu1(aa)+q(bb);
end
if (uu(aa)==0)
ddaver(aa)=0;    %%%平均距离
xinyuaver1(bb)=0;
else
ddaver(aa)=dd(aa)/uu(aa);
xinyuaver1(aa)=xinyu1(aa)/uu(aa);    %%%平均信誉值
end
end
end
end

```

```
xinjiage=-0.024*uu+91.405*ddaver-0.002*xinyuaver+67.1;    %%新任务价格输出
```

```
xlswrite('xinrenwujiage',jiage,1,'A2');%%                输出到 xinrenwujiage.xls 表格
```