

# 基于运筹学的“拍照赚钱”任务定价设计模型

## 摘要

本文针对“拍照赚钱”任务的定价方案制定问题，通过分析任务与会员的分布规律以及联系，主要利用了多目标优化的方法，建立了任务不打包与打包两种情况下的定价模型，并与原定价方案进行对比，证明设计的定价方案的有效性。

针对任务定价规律的研究问题，首先限定了主要的分布区域，剔除异常数据。然后根据任务位置分布进行 K-Means 聚类，将已结束任务分到十五个区域内。通过观察已结束项目定价分布，先对该平台的定价规律进行定性的分析，然后综合考虑定义了七个影响任务定价的因素：区域会员数、任务数、会员数与任务数的比值、人均任务限额、人均任务限额、人均信誉值、距离区域中心的距离和人均 GDP。接着建立灰色关联分析的模型求解，得出对任务定价影响最大的是人均任务限额，任务限额越大，任务定价越低。而对于任务未完成的原因的探究，通过观察任务完成的分布情况，初步分析出影响任务完成率的六个指标：任务均价、任务限额与任务数之比、人均任务限额、距离区域中心的距离、人均 GDP 和任务均价与人均 GDP 的匹配程度。接着通过皮尔逊相关系数检验它们的相关性，选择相关性较大的匹配程度、人均 GDP 和任务均价建立任务完成率函数并求解。最后通过对比完成率函数个影响因素的系数大小，说明任务完成与否主要由区域任务均价与人均 GDP 的匹配程度决定。

针对已结束项目新定价方案的设计问题，首先对最优定价方案的特征进行了分析，将设计定价方案的主要目标放在总任务完成量最大和平均任务价格最低上。以不同区域的定价为决策变量，最低定价为约束条件，总任务完成量最大和任务价格最低为目标函数建立多目标优化模型。为了对模型进行求解，则对两个目标进行无量纲化后线性加权的处理，将多目标化为单目标，对不同的权值组合进行求解，得到权值为 0.2、0.5、0.8 三种情况下的定价方案（具体各区域定价间正文），与原定价方案进行比较，当  $k = 0.2$  时，平台总支出由原来的 53730 元减少到 53176.51 元，并且任务完成率由 68.02% 提高到了 76.07%。当  $k = 0.5, 0.8$  时，虽然任务均价都有所提高，但均价提高幅度较小，且任务完成率的提高的幅度比较大，分别提高了 12.27%，17.53%。

针对任务打包的定价方案的设计问题，首先需要考虑任务打包的方式，制定了综合任务包内任务间距离以及任务数量的限制的任务打包模型。接着结合前两问提出的定价模型进行改进，引入等效任务数以及等效定价的概念。将等效任务均价、人均 GDP 和区域等效任务均价与人均 GDP 匹配程度作为新的影响因素，并重新根据任务打包的情况建立多目标优化模型求解出最优定价模型，具体各区域定价间正文。最后与打包前进行对比，得出结论：在绝大多数个区域内打包后的平均定价均低于未打包时的平均定价，且打包后任务的完成率普遍高于未打包时任务的完成率。说明对任务进行打包既降低了被打包任务的定价，从而减少平台支出。打包后平均定价下降 14.53 元，并且完成率上升了 10.34%，相比打包前有很好的改善效果。

针对新项目定价方案的设计问题，只需要首先对任务进行 K-Means 聚类，然后利用第三问的方法对任务进行打包，接着同样在考虑三个任务完成率影响因素（等效任务均价、人均 GDP 和区域等效任务均价与人均 GDP 匹配程度）的情况下，建立多目标优化函数求解最优定价方案，具体各区域定价间正文。通过最终结果可知：新任务经打包后，所有区域的任务完成率均很高，任务的平均定价也在 58 元以下，甚至一部分平均定价在 52 元以下。相比于打包前来说，任务完成率提高了 10% 到 20%，而任务的平均定价减少了 5 到 15 元，可见对于新任务打包定价方案实施的效果显著。

**关键字：**K-Means 聚类 灰色关联度 多目标优化 蒙特卡洛法

## 一、问题重述

“拍照赚钱”是移动互联网下的一种自助式服务模式。用户通过下载 APP 领取任务完成指定的商品检查拍照任务，赚取对应的酬金。此平台上任务的定价是其核心要素。如果定价不合理，有的任务就会无人问津，而导致商品检查的失败。

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、问题分析

随着互联网的发展，出现了“拍照赚钱”这种自助式服务模式。它通过在官方平台发布拍照任务，让用户接单拍照完成任务来运行。这种基于移动互联网的劳务众包平台，为企业提供各种商业检查和信息收集。这个平台运行的核心就是需要有一套良好的定价体系，如果定价不合理，便会造成有些任务无人问津，导致商品检查失败。只有通过良好的定价体系，才能使平台长久运营下去。

为了方便进一步研究定价方案，研究环境因素对某任务定价的影响，可以先利用 K-Means 聚类的方法对已结束项目数据进行分类。

### 2.1 任务定价规律以及任务未完成原因的分析

问题一首先要求我们根据一项已完成项目的任务数据中的每个任务的位置、定价和完成情况来分析任务定价规律。在经济学中，一个竞争性市场上的商品价格规律受到供求关系影响而上下波动，因此，要分析定价规律，便需要找到影响定价的因素，对于每一个任务而言，它的定价与完成情况会受到其它任务与会员的影响，我们将从这两大方面考虑任务定价的影响因子，定义影响任务定价的多个影响因子。接着可以进一步结合附件一给出的数据将影响因子量化。最后，利用灰色关联度矩阵计算各影响因子与定价之间的相关度，定量分析任务的定价规律。

对于任务未完成原因的探讨，同样先从任务定价、限额、距离、GDP 等方面定义多个任务完成的影响因素。接着考虑使用皮尔逊相关系数对各影响因素的相关性进行检验，选出相关性较大的影响因素进行任务完成率模型的建立，最后便能够通过比较拟合的任务完成率表达式中各影响因素的系数大小来找出对任务完成率影响最大的因素。

### 2.2 关于对原项目定价方案制定的分析

针对原项目进行重新设计定价方案的问题，可以从第一问分析得到的具体的任务完成率的具体表达式着手做。考虑从两个方面寻找优化目标：平台的酬金支出最少，以及发布的任务完成量最大。对于平台的酬金，为了商业利益，肯定是希望越低越好，酬出。且考虑每个区域的定价金则可以通过完成的任务数求出。而任务总量可以通过第一问求解出的任务完成率求不能低于最小值。最后通过将多目标函数归一成为单目标函数即可求解最佳定价方案，并与原设计方案进行横向比较得出结论。

## 2.3 打包发布任务情况定价模型的分析

在对任务进行打包处理后，能够提高任务完成率，而打包处理任务还能够有效降低定价。为了修改定价模型，可以考虑打包操作对会员的影响。然后考虑打包任务的条件和约束，使打包后的任务距离和数量都能适中。接着利用同样的方法建立新的最佳定价目标约束模型求解出最好的定价模型，且对任务打包前后进行比较分析。

## 2.4 新定价方案设计的分析

对于新项目定价方案的设计问题，可以结合前面三问的求解分析。先将新项目的利用 K-Means 聚类。接着需对任务进行打包处理，之后便可以利用和第三问同样的方法建立定价方案的目标优化模型，接着便能求解出最佳定价方案，最后对方案的实施效果进行检验并评价。

# 三、模型假设

1. 论文使用的数据真实可靠；
2. 假设会员在预定任务时，优先预定对其吸引力最大的任务；
3. 假设当多个会员同时想预定同一任务时，信誉值最高的会员优先获得该任务；
4. 假设任务完成难度相同。

# 四、变量说明

变量名称	含义
$y$	任务完成率
$a$	任务完成率影响因素系数或常数
$x$	任务完成率的影响因素
$A$	区域内总任务数
$B$	地区总任务完成数
$W$	但各地区平均每个任务所给酬金
$A'$	区域的等效任务数
$Z_c$	区域内任务的打包率

# 五、数据预处理

“拍照赚钱”平台运营主要依靠此 APP 的会员的数量及分布。为了有效的研究此平台任务的定价规律，以及为了有效的对主要区域进行新定价方案的制定，需要对已知数据进行一定的处理和筛选。

论文附件是大量数据的集合，为了提高数据挖掘的质量，因此采用拉以达准则算法进行异常数据的剔除。

对附件一已完成项目来说，以定置信概率为 99.7% 为标准，以三倍测量列标准偏差极限为依据，凡超过此界限的误差，就认为它不属于随机误差的范畴，而是粗大误差，即需要被剔除的异常数据。

对附件二会员的数据来说，同样采用拉以达准则进行异常数据的剔除，通过 matlab 算法对算数平均值和标准偏差的计算从而对数据进行筛选。筛选结果过表示：

表 1 筛选结果表示

会员编号	会员位置 (GPS) 纬度	会员位置 (GPS) 经度	预订任务限额	预订任务开始时间	信誉值
B0005	33.65205	116.9705	66	6:30:00	20919.07
B1175	113.1315	23.03182	1	6:36:00	19.923

结果显示附件一数据无异常，附件二中 5 号和 1175 号数据异常，5 号会员位置相对较为偏远，因此在问题中的影响较小可以直接忽略，而 1175 号会员位置经纬度数据明显有误，因此将剔除附件二中异常的两个数据。

## 六、模型的建立与求解

### 6.1 问题一

#### 6.1.1 定价规律定性研究

为了研究初始项目定价的规律，需要从题目所给的数据进行分析。每个任务都有对应经纬度坐标，以及其定价。所以首先利用 MATLAB，以经度为横坐标，纬度为纵坐标，绘制任务定价分布的热力图，将不同价格的任务用不同颜色的点表示，如图 1，源代码见附录 B。

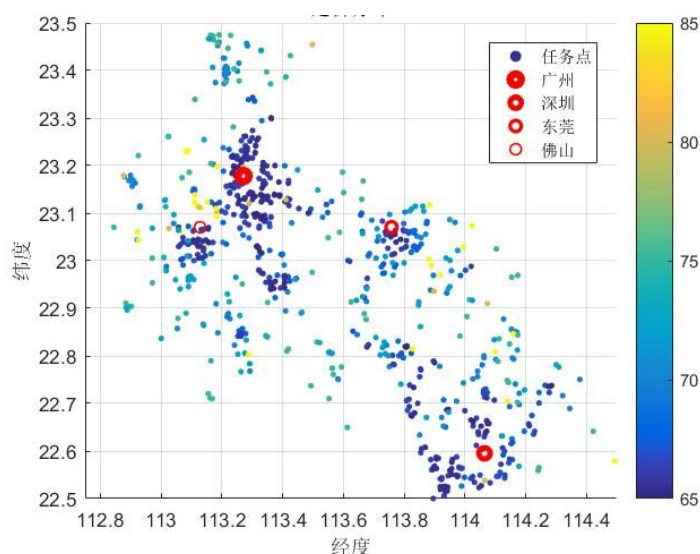


图 1 已结束项目定价分布图

其中，四个红色空心点分别代表此项目分布区域周围的四个主要城市：广州、深圳、东莞和佛山。其余实心点表示任务分布，从蓝色到黄色定价依次对应 65 到 85 的任务定价。从图中可以看成出，任务分布主要还是围绕在城市中心周围，此区域任务分布非常密集，反之离城市越远的地方任务分布越稀疏。而相对于密集区域，距离市中心较远区域的任务定价都较市中心定价高，而任务分布密集的市中心价格都是接近最低价。

所以，可以定性分析判断该平台的定价规律：

1. 任务大多集中分布在广东省的广州市、深圳市、东莞市以及佛山市等城市的繁华地带，在这些区域人流量较大，交通便利，会员分布也较为密集。

- 2. 任务的分布越密集、周围的任务数量越多，该区域的颜色较深，代表任务定价越低。
- 3. 较少的任务分布在远离市区的郊区，在这些区域，任务分布稀疏，人口密度与人流量较小，交通不便，会员数量也较少。
- 4. 将任务的发布看做需要被完成的任务需求方，会员的人数看做完成任务的供给方。在交通便利的市区，会员人数大于任务的数量，供大于求，任务定价较低。同样，当供小于求时，任务定价较高。通过图像观察到的定价规律符合经济学的供求定理。

6.1.2 定价规律定量分析模型的建立

6.1.3.1 根据任务位置分布进行 K-Means 聚类分析

根据在百度中标出的任务的位置分布可以看出任务大致分布在广东省的南部，呈现出部分集中，相对分散的特点。位置的地理位置分布在很大程度上会影响其定价，在城市中心地带，交通便利，人口密集，商业繁华，因而分布着大量的任务，并且任务定价相对较低。据此可以推断，任务的位置距离其所处区域的密集分布中心点的距离会影响其定价。

通过观察地图可知任务的分布大概有十五个分布群以及十五个中心点。根据任务的经纬度坐标，利用 K-Means 聚类分析法将任务分为十五类，分类图如图所示：

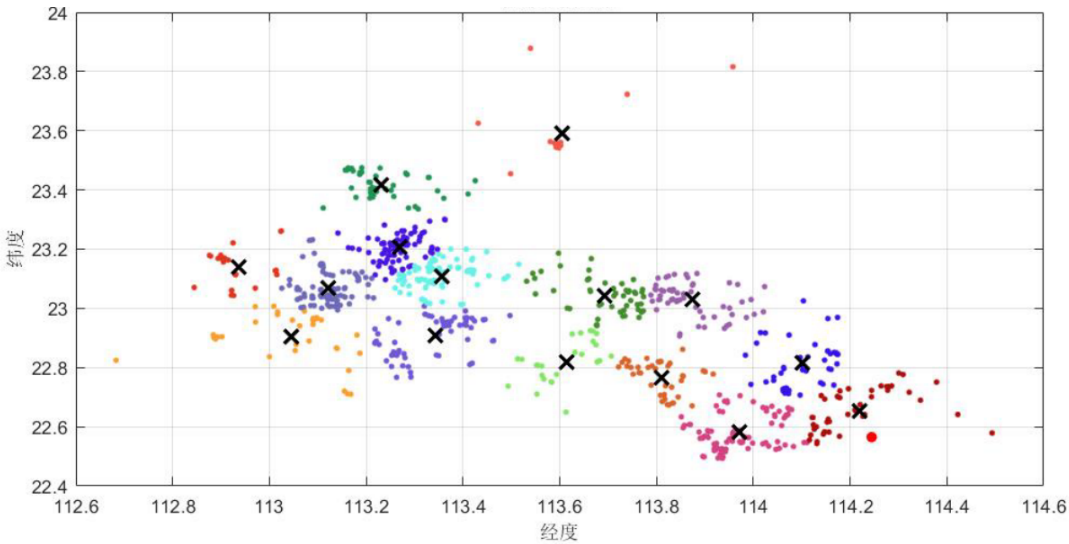


图 2 任务位置分布以及十五个分类中心点

根据 K-Means 分析结果，得到了十五个中心点的经纬度坐标，每个任务都属于这十五个分区之一，每个区域的中心以及实际位置如下表：

表 2 聚类区域分布具体信息

编号	聚类中心		实际位置
	经度/°E	纬度/°N	
1	113.9713	22.5820	广东省深圳市南山区
2	113.3429	22.9085	广东省广州市番禺区大巷涌路 13 号
3	113.0451	22.9044	广东省佛山市南海区 S269(沙龙路)
4	113.2309	23.4171	广东省广州市花都区三东大道西
5	113.6927	23.0424	广东省东莞市
6	114.2196	22.6532	广东省深圳市龙岗区保康路 32 号
7	113.2692	23.2066	广东省广州市白云区机场路 1737 号 102
8	113.8104	22.7657	广东省深圳市宝安区新和大道
9	113.1217	23.0688	广东省佛山市禅城区
10	113.3562	23.1081	广东省广州市海珠区
11	113.6049	23.5913	广东省广州市从化区
12	112.9366	23.1394	广东省佛山市南海区
13	113.6142	22.8184	广东省东莞市
14	114.1014	22.8157	广东省东莞市大岭古街 5 号
15	113.8741	23.0302	广东省东莞市

### 6.1.3.2 确立影响因素

一项任务所处位置的周围环境包括其他任务的定价、数量，会员的分布与数量以及这些会员对应的信誉值。在已知会员的相关信息的前提下，假设所有任务的定价被同时发布，因此，每个任务定价不受其周围的任务价格的影响。但是，一项任务周围的任务数量、会员数量与会员的信誉情况将会影响到该任务的定价。我们将从其周围任务与会员的分布状况与特定属性来考虑影响任务定价的因素。

综合考虑每项任务的位置与周围环境，定义影响任务定价的七个因子：

#### (1) 区域会员数

通过对任务数据的 K-Means 聚类处理，将总区域分为十五个小的区域，每个  $\square$  任务都处在一个网格内，将这个网格区域内分布的会员总数记为  $Q_i$  人， $i=1,2,...15$ 。

#### (2) 区域任务数

通过对任务数据的 K-Means 聚类处理，将总区域分为十五个小的区域，每  $\square$  个任务都处在一个区域内，将这个网格区域内存在的任务总数记为  $q_i$  个， $i=1,2,...15$ 。

#### (3) 区域会员数与任务数比值

区域人数与任务数的比值能够一定程度上表现出当前区域任务的“抢手”程度。任务数不变时，比值越高，会员越多，对任务的需求量就越大，趋向于供不应求的状态发展；而比值越小，代表会员越少，供需情况趋向于供大于求的方向变化。计算公式如下：

$$\text{区域会员数与任务数比值} = \frac{\text{区域会员总数}}{\text{区域任务总数}} \quad (1)$$

#### (4) 区域人均任务限额

人均任务限额指的是在当前区域内，平台会员预定任务限额的平均值，它等于当前区域所有会员任务预定限额之和除以当前区域内总会员数。它代表当前区域可以预定任务数的一个平均上限。人均任务限额越大，代表单个会员可同时接受任务越多，相对而言，单个任务的定价便可以稍低。其计算公式如下：

$$\text{人均任务限额} = \frac{\sum \text{预定任务限额}}{\text{会员总数}} \quad (2)$$

#### (5) 区域人均信誉值

区域人均信誉值表示当前区域内平均每个会员的信誉均值，它代表了这个区域的一个整体的信誉水平。计算公式如下：

$$\text{人均信誉值} = \frac{\sum \text{单个会员信誉值}}{\text{会员总数}} \quad (3)$$

#### (6) 距离区域中心距离

任务距离区域中心的位置同样对任务的定价有影响，区域中心的人口数量较多，会员数量也同样较多；而距离较远的地方会员数量相对较少。

通过图 2 得到影响任务定价的影响因子  $R_i$ ，表示任务  $i$  与其所属区域中心点的距离，根据图像可以得出距离越远，定价越高的定性规律。

#### (7) 区域人均 GDP

一个区域的发展水平对任务的定价也有规律，GDP 能够有效的表现出当前区域的人民生活质量以及消费水平。GDP 越高的区域，任务定价也相应的应该越高，否则对本区域会员没有足够的吸引力。

### 6.1.3.3 灰色关联分析求解模型相关性

至此，通过对附件一中已完成项目进行空间离散化处理与 K-Means 聚类分析，得到影响任务定价规律的七个因素，建立起定价规律影响模。

在定价规律影响模型中，被解释变量即任务的定价的数值由附件一给定，通过 K-Means 分析可以将影响因子量化，计算出其数值。

### 6.1.3 任务完成情况初步分析

通过已结束项目任务完成的数据，可以对任务完成与否进行分析。首先将已结束项目任务完成与否的数据通过 MATLAB 绘出，可以得到图 3，源代码见附录 C。



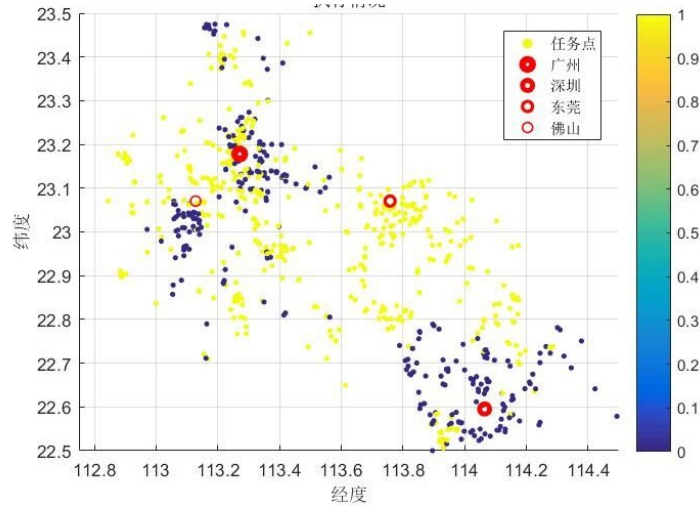


图 3 已结束项目完成情况

图中蓝色的点表示未完成任务，黄色的点表示已完成的任务。任务的完成率可以表示为：

$$\text{任务完成率} = \frac{\text{任务完成数量}}{\text{任务总数}} \quad (4)$$

于是可以初步分析出，广州佛山区域未完成数量处于正常水平，而东莞区域的完成率非常高，接近 100%，深圳区域的完成率则很低，大部分任务都没有完成。主要的因素可能是定价不合理，或者会员以及任务分布情况不能满足需求或者过大。

#### 6.1.4 任务完成率预测模型的建立

##### 6.1.5.1 确立任务完成的影响因素

通过对任务和会员的关系的分析，最终提炼出了三个与每个区域任务完成率有关的因素：

##### (1) 区域任务均价

任务均价指当前区域内所有任务的平均价格，它等于当前区域内所有任务价格之和除以当前区域内任务总数。任务价格主要影响了会员接受任务并完成任务的意愿，价格越高的任务越容易被会员接受并完成；反之价格越低，会员接受任务的意愿越低，任务完成率自然低。具体计算公式如下：

$$\text{任务均价} = \frac{\sum \text{任务价格}}{\text{任务总数}} \quad (5)$$

##### (2) 区域任务限额与任务数之比

任务限额与任务数之比指的便是当前区域内，会员预定任务限额之和与任务数的比值。预定任务限额表示每个会员预定任务数量的最大值，任务限额与任务数之比表示平均每个任务可以被预约的次数。当这个比值较高，则表示任务数较少，任务总限额较大，此时平台处于优势，任务被接受的概率大大提升，因而任务完成率也相比较较高。因为会员接受任务，所以这个比例可以类比为供需比率，平台提供任务，会员有接受任务赚取现金的需求。当这个比值大于 1，代表供不应求的状况，会员数大于此区域内任务数，此时平台占优势，可以通过降低价格来调整供需关系，使之接近 1；当比值小于 1



时，任务数大于会员数，供过于求，平台需要提高任务定价来吸引更多的会员加入。如果不这样做则会导致任务没有会员接受，导致任务完成率下降。具体的计算公式如下：

$$\text{平均预定任务限额} = \frac{\sum \text{预定任务限额}}{\text{任务总数}} \quad (6)$$

### (3) 区域人均任务限额

区域人均任务限额同之前的分析。人均任务限额越高，说明当前区域每个会员可接任务数越多，而可接任务多的原因便是之前任务的完成率较高，平台相对于其他地方增加了此地的预定任务上限，所以人均任务限额能体现出任务完成率，人均任务限额越高，说明任务完成率越高。具体计算公式见 (2)。

### (4) 距离区域中心距离

任务距离区域中心的位置同样对任务的完成率有影响，区域中心的会员较多，相对于偏远地区，高质量的会员也会较多，他们的任务完成率远比偏远地区的会员的任务完成率高，所以距离区域中心的距离也会完成率有影响。

### (5) 区域人均 GDP

一个区域的发展水平对任务的定价也有规律，GDP 能够有效的表现出当前区域的人民生活质量以及消费水平。GDP 越高的区域，会员的质量也相对较高，自然可以推出任务完成率较高的结果。

### (6) 区域任务均价与人均 GDP 匹配程度

区域内任务均价与人均 GDP 的匹配程度能够有效的反应当前区域的任务完成情况，它等于人均 GDP 除以任务均价，得到结果再开 8 次方。在任务均价不变的情况下，人均 GDP 越高，人们生活水平越高，此类任务对人群的吸引力就下降，任务完成率也就随之受影响而下降。具体的计算公式如下：

$$\text{匹配程度} = \left( \frac{\text{人均 GDP}}{\text{任务均价}} \right)^{\frac{1}{8}} \quad (7)$$

得到六个任务完成率的影响因素之后，首先利用皮尔逊相关系数的求解方法对每一个指标进行相关性检验，接着选出相关系数大的影响因素，利用这些影响因素对曲线进行拟合并且计算拟合优度和置信度。

## 6.1.5 模型算法

### (1) 算法思想

对定价模型来说，当指标体系建立之后，还需科学地对任务定价模型影响因子进行确定。考虑到各个影响因子数值的合理性求解，采用搜索算法对模型建立的各种因子进行求解。确定影响因子数值之后，为了研究附件一中部分任务未完成的原因，因此考虑通过灰色相关度矩阵，对每个任务定价与各影响因子的相关度进行分析。统一算法思想如下：对每个任务进行遍历，确定一个任务下相对应的网格区域，进而对所有任务或者会员进行遍历搜索，得到该网格内相应任务或会员数量以及会员能力。

对任务完成率模型来说，在对区域进行 K-Means 聚类之后，首先对各个影响因素进行皮尔逊相关系数的计算，选出相关性强的三个影响因素，接着计算聚类后对应的每个

区域的三个影响因素的数值大小。由于影响因素量纲不同，需要先对影响因素去量纲化的操作，通过调用 MATLAB 的线性回归拟合函数“regress”对多个系数和一个常数进行拟合，通过函数输出的拟合优度和置信区间判断拟合程度。

## (2) 算法步骤

### 定价模型：

**Step1** 确定反映系统行为特征的参考数列和影响系统行为的比较数列。反映系统行为特征的数据序列，称为参考数列，即任务定价。影响系统行为的因素组成的数据序列，称比较数列，即影响因素。

**Step2** 对人物定价和影响因素进行无量纲化处理。由于多个影响因素的意义不同，导致数据的量纲也不一定相同，不便于比较，或在比较时难以得到正确的结论。因此在进行灰色关联度分析时，一般都要进行无量纲化的数据处理。设第  $k$  个影响因素的变化区间为  $[j_{k1}, j_{k2}]$ ,  $j_{k1}$  为第  $k$  个影响因素在所有任务定价中的最小值， $j_{k2}$  为第  $k$  个指标在所有任务定价中的最大值，则可以用下式将上式中的原始数值变成无量纲值  $C_k \in (0, 1)$ 。

$$C_k = \frac{j_k - j_{k1}}{j_{k2} - j_{k1}}, k = 1, 2, \dots, 7 \quad (8)$$

**Step3** 求任务定价与影响因素的灰色关联系数  $\xi(k)$ 。将  $C^* = [C_1^*, C_2^*, \dots, C_7^*]$  作为参考数列，将  $C = [C_1, C_2, \dots, C_7]$  作为被比较数列，则用关联分析法分别求得人物定价的第  $k$  个影响因素与第  $k$  个影响因素最优指标的关联系数，即

$$\xi(k) = \frac{\min(k)|C_k^* - C_k| + \rho \max(k)|C_k^* - C_k^l|}{|C_k^* - C_k| + \rho \max(k)|C_k^* - C_k^l|} \quad (9)$$

式中  $\rho \in (0, 1)$ ，一般取  $\rho = 0.5$ 。

**Step4** 求关联度。因为关联系数是任务定价与影响因素在各个时刻（即曲线中的各点）的关联程度值，所以它的数不止一个，而信息过于分散不便于进行整体性比较。因此有必要将各个时刻（即曲线中的各点）的关联系数集中为一个值，即求其平均值，作为任务定价与影响因素间关联程度的数量表示，关联度  $r_i$  公式如下：

$$r = \frac{1}{7} \sum_{k=1}^7 \xi(k), k = 1, 2, \dots, 7 \quad (10)$$

**Step5** 关联度排序。影响因素间的关联程度，主要是用关联度的大小次序描述，而不仅是关联度的大小。将七个影响因素对任务定价关联度按大小顺序排列起来，便组成了关联序，它反映了对于任务定价来说各影响因素的影响大小。

### 任务完成率模型：

**Step1** 计算每一个影响因素的皮尔逊相关系数，比较后选出相关性强的三个影响因素  $x_1$ 、 $x_2$ 、 $x_3$ 。

**Step2** 对聚类后每个区域的三个强相关性影响因素  $x_1$ 、 $x_2$ 、 $x_3$  进行标准化处理：

$$\bar{x} = \frac{x - x_{max}}{x_{max} - x_{min}}$$

**Step3** 将任务完成率关于多个影响因素的函数表达式  $y = 1 - e^{-(a_0 + a_1x_1 + a_2x_2 + a_3x_3)}$  化为  $\ln y = -(a_0 + a_1x_1 + a_2x_2 + a_3x_3)$ 。

**Step4** 遍历该区域的  $y$  的值，若  $y$  等于 1，则将其值修改为 0.9999，并计算  $\ln(1 - y)$ 。  
**Step5** 调用 MATLAB 中的 “regress” 函数做  $\ln(1 - y)$  与  $x_1$ 、 $x_2$ 、 $x_3$  线性回归系数  $a_1$ 、 $a_2$ 、 $a_3$  的值，并根据返回的参数检验拟合效果。

### 6.1.6 模型结果及分析

#### (1) 定价模型求解

基于 K-Means 聚类分析，分别确定每个任务所在区域内的会员数、任务数、会员数与任务数的比值、人均任务限额、人均信誉值、距离区域中心的距离和人均 GDP。通过 matlab 对任务和会员进行遍历从而确定相应数值。得到七个影响因素数值如下：

表 3 七个影响因素部分数值

区域编号	影响因素						
	会员数	任务数	会员数/任务数	人均限额数	人均信誉值	距离区域中心距离	人均 GDP
1	304	88	0.2895	5.3191	269.9468	9.7164	33.9269
2	81	73	0.9012	7.0123	37.1777	28.4175	11.33
3	56	37	0.6607	12.1607	416.4020	20.3325	12.02
4	59	46	0.7797	8.0847	415.3679	26.8842	11.87
5	104	53	0.5096	15.7212	1133.7294	7.4142	9.1
6	185	45	0.2432	5.6649	86.7034	17.1337	17.999
7	213	93	0.4366	7.5164	464.1830	3.1678	6.93
8	170	45	0.2647	3.0588	87.4318	32.2527	11.4294
9	103	84	0.8155	8.8447	75.6868	0.6962	9.9
10	293	94	0.3208	5.6109	287.7007	11.6881	10.46
11	17	19	1.1176	12.3529	44.8622	57.1929	6.24
12	31	24	0.7742	6.8710	93.5218	21.0333	12.02
13	30	32	1.0667	4.8333	51.0027	31.6788	9.1
14	97	46	0.4742	4.9175	119.8973	24.8010	9.1
15	133	56	0.4211	7.6466	87.5364	12.6561	9.1

接着根据灰色相关度的定义，通过算法对定义的关联系数和关联度进行计算，得到灰色关联度矩阵：

表 4 灰色关联度矩阵

	距离类城市中心距离	人均 GDP	人数	任务数	人数/任务数	人均限额数	人均信誉值
关联度	0.877	0.9238	0.7667	0.9087	0.9475	0.9621	0.8343

根据灰色关联分析结果，可以得到如下结论：七个影响因素的关联度几乎都大于 0.8，半数以上的影响因素的关联度大于 0.9。有力的说明了所选的影响因素的有效性和合理性，说明这些影响因素能够很好的对定价规律进行分析。其中，对任务定价影响最大的是人均任务限额，任务限额越大，任务定价越低。

(2) 任务完成率模型求解

通过皮尔逊相关系数检验，得到六个影响任务完成率的相关系数如下：

表 5 皮尔逊相关系数

	限额数与任务数之比	任务均价	人均限额数	距离区域中心距离	人均 GDP	匹配程度
皮尔逊相关系数	-0.2862	0.5297	0.3426	0.3631	-0.6085	-0.6152

对于皮尔逊相关系数来说，绝对值越接近 1，相关性越大，且正负号代表正相关和负相关。由上表可以看出相关性最大的因素是区域任务均价与人均 GDP 的匹配程度、区域人均 GDP 和区域任务均价，其中区域任务均价和任务完成率正相关；区域任务均价与人均 GDP 的匹配程度、区域人均 GDP 与任务完成率成负相关。所以选择这三个影响因素进行任务完成率函数的拟合。

调用 MATLAB 中“regress”函数拟合出的系数和常数值，以及拟合优度和置信区间的值如下表，源代码见附录 D：

表 6 任务完成率函数拟合结果

$R^2$	P	$a_0$	$a_1$	$a_2$	$a_3$
0.4948	0.0499	171.304	0.6959	0.7834	-254.2186

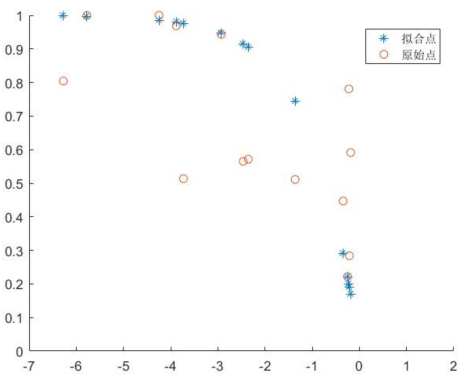


图 4 各区域原始任务完成率与拟合结果比较

其中拟合优度  $R^2$  较大，置信度  $P < 0.05$ ，所以拟合效果较好。则最终得到的任务完成率函数为：

$$y = 1 - e^{-171.304 - 0.6959x_1 - 0.7834x_2 + 254.2186x_3} \tag{11}$$

三个系数的大小代表其决定完成率的占比大小。 $|a_3| > |a_2| > |a_1|$ ，即占比较大的是区域任务均价与人均 GDP 的匹配程度，说明任务完成与否主要由区域任务均价与人均 GDP 的匹配程度决定。选取了两个任务完成率最高和最低的地区进行比较分析。

表 7 部分区域比较

区域	实际位置	任务均价	人均 GDP	匹配程度	完成率
11	广东省广州市从化区	72.11	6.2400	0.0865	100.00%
12	广东省佛山市南海区	73.56	12.0200	0.1634	100.00%
1	广东省深圳市南山区	66.42	33.9269	0.5108	28.41%
6	广东省深圳市龙岗区保康路 32 号	69.57	17.9990	0.2587	22.22%

通过四个区域对比，发现完成率低的地区匹配程度都较完成率高的地方高；人均 GDP 也较高，说明人均 GDP 较高的地区人们生活水平普遍较高，对拍照赚钱的兴趣都不大，导致完成率偏低；对于任务均价，价格较高的两个区域的完成率都很高，而两个价格低的区域完成率低，说明任务定价高低对完成率的影响也是较大的。

## 6.2 问题二

针对原定价方案的改进问题，分别从平台支出最少和任务完成数最多两个约束条件出发，建立多目标优化模型，寻找每个区域内对应任务的最佳定价。以设计出新的定价方案，然后通过与原定价方案比较平台支出量以及平均定价的大小来评价新定价方案的好坏。

### 6.2.1 模型准备

通过第一问的求解，得到了任务完成率计算公式。为了对已结束项目任务设计新的定价方案，需要从平台支出和任务完成率两个方面进行考虑。首先设单个区域的任务数为  $A_i$ 。总的任务数设为  $A$ ，可得：

$$A = \sum A_i, i = 1, 2, \dots, 15$$

对于单个具体的区域，其任务完成率则可以通过将此区域对应的  $x_1$ 、 $x_2$ 、 $x_3$  求出后，带入任务完成率函数为：

$$y_i = 1 - e^{-(a_0 + a_1 x_{1i} + a_2 x_{2i} + a_3 x_{3i})} \quad (12)$$

则每个区域的任务完成总数等于当前区域任务完成率乘以当前区域的任务总数。通过将单个地区所有区域的完成的任务数相加，等于地区总的任务完成数  $B$ ：

$$B = \sum A_i y_i = \sum A_i - A_i e^{-(a_0 + a_1 x_{1i} + a_2 x_{2i} + a_3 x_{3i})} \quad (13)$$

### 6.2.2 最优定价多目标优化模型的建立

为了设计相比于原方案更加合理的定价方案，需要合理考虑每个区域的定价，基于对平台支出最少和任务完成量最大，建立了多目标优化模型。

#### 目标一：地区总任务完成量最大

按照题目描述，为了提高平台的收入，平台希望发布的任务都能够被接受并且完成。由于单个地区总任务量不变，所以最大的任务完成量便是最优定价方案的一个重要组成方面。单个地区的总任务量可由式 (13) 求出，其中的区域人均 GDP  $x_{2i}$  与单个任务定价无关，只需要通过查阅相关文献即可得到具体的 GDP 数据。而对于匹配程度  $x_{3i}$  来

说，通过其计算公式，只需要知道区域任务平均定价  $x_{1i}$ ，即可算出。所以，任务完成数只与区域任务平均定价  $x_{1i}$  有关。

故优化目标一可以表示为：

$$\max B(x_{1i}) = \sum A_i - A_i e^{-(a_0 + a_1 x_{1i} + a_2 x_{2i} + a_3 x_{3i})} \quad (14)$$

其中，匹配程度  $x_{3i}$  的计算公式为：

$$x_{3i} = \left( \frac{x_{2i}}{x_{1i}} \right)^{\frac{1}{8}}$$

## 目标二：平台在单个区域平均每个任务的酬金最低

一个最优的定价方案不仅需要使任务完成量最大，还得使平台支出的酬金量尽量少。设平台对单个地区平均每个任务所给的酬金为  $W$ ， $W$  只与完成的任务数量有关，因为对于没有完成的任务，平台不对会员进行酬金奖励。通过每个区域任务完成数乘以当前区域任务平均定价再除以总任务数，即得到单个地区平均每个任务所给的酬金  $W$ ，即：

$$W = \frac{\sum A_i y_i x_{1i}}{B} \quad (15)$$

而平均酬金  $W$  与总任务量  $B$  的关系一样，都只与单个地区任务平均定价  $x_{1i}$  有关。故优化目标二可以表示为：

$$\min W(x_{1i}) = \frac{\sum A_i y_i x_{1i}}{\sum A_i - A_i e^{-(a_0 + a_1 x_{1i} + a_2 x_{2i} + a_3 x_{3i})}} \quad (16)$$

## 决策变量与约束条件：

通过对优化目标一和优化目标二的分析可知，决策变量即为单个区域任务均价  $x_{1i}$ 。

每个区域都存在一个最低定价，低于此定价则此区域的任务会无人问津，决策变量的约束条件便是大于等于最低定价。当区域任务完成率恰好为 0 时，其对应的任务定价即是此区域的最低任务定价。于是可以得到：

$$0 = 1 - e^{-(a_0 + a_1 x_{1imin} + a_2 x_{2i} + a_3 x_{3i})}$$

通过上式即可求解出每个不同区域的最低定价  $x_{1imin}$ 。因此得到限定  $x_{1i}$  如下：

$$x_{1imin} \leq x_{1i} \quad (17)$$

### 6.2.3 模型汇总

通过对定价情况的影响因素和约束条件的分析，最终得到多目标约束模型如下：

$$\begin{aligned} \min \quad & W(x_{1i}) = \frac{\sum A_i y_i x_{1i}}{\sum A_i - A_i e^{-(a_0 + a_1 x_{1i} + a_2 x_{2i} + a_3 (\frac{x_{2i}}{x_{1i}})^{\frac{1}{8}})}} \\ \max \quad & B(x_{1i}) = \sum A_i - A_i e^{-(a_0 + a_1 x_{1i} + a_2 x_{2i} + a_3 (\frac{x_{2i}}{x_{1i}})^{\frac{1}{8}})} \\ s.t. \quad & x_{1imin} \leq x_{1i} \end{aligned}$$

其中  $x_{1imin}$  表示不同区域的最低定价。

由于目标函数不唯一，需要将多目标函数化为单目标函数 [1][2]。先对两个目标函数进行无量纲化处理，再采用线性加权法对两个目标函数进行归一化处理，将多目标函数化为单目标函数：

$$\max H(x_{1i}) = k\left(\frac{B}{B_{max}}\right) + (1 - k)\frac{W_{min}}{W} \quad (18)$$

其中  $k$  和  $1 - k$  代表了每个目标函数的重要程度，权重之和为 1。 $B_{max}$  和  $W_{min}$  分别代表再所选取的自变量范围内，地区总任务量可能达到的最大值、单个地区任务平均酬金可能达到的最小值。 $B_{max}$  可以由函数关系式得知为任务总量， $W_{min}$  则为上述不同区域的最低定价  $x_{1imin}$  中最小的取值。对不同的权重组合进行求解并对结果进行比较和分析。

#### 6.2.4 任务吸引度模拟模型的建立

在利用目标优化模型求解出各区域的定价之后，我们便可以利用每个任务对不同会员的吸引力来模拟会员选择任务的过程。为了使得问题的分析适当简化以及保证模型的合理性，模型的建立基于以下规则：

1. 每项任务对于每一位会员都有特定的吸引力值，任务的定价越高、与会员的距离越近，其对会员的吸引力就越大。
2. 当会员在某时间点开始预定任务时，假设一次只能预定一个任务，不能同时预定多个任务，并且每次预定时都选择对自己吸引力最大的任务。
3. 当多个会员在同一时间点选取同一任务时，信誉最高者优先获得该任务。
4. 如果会员的预定次数多于一次，在一次任务预订成功后，可以继续预定任务。
5. 根据附件一中已完成项目的任务数据，可以求出任务吸引度的阈值，只有一项任务对会员的吸引度不小于阈值时，会员才愿意预定任务并完成。

##### (1) 任务吸引度定义

在问题一未完成原因的分析中，可以得到距离、价格这两个影响因子的影响作用显著。因此，会员在选择是否预定并完成某一单时，会重点考虑这一任务与自己的距离，同时，任务的定价直接影响到会员的收益，也会作为重要的考虑因素。由第一问分析结果可知，吸引度与任务的价格  $p$  成正相关，与任务的距离  $L$  成负相关，并且我们假设吸引度的范围为  $[0, 1]$ 。所以可以类似第一问的模型建立任务吸引度函数如下：

$$w_{ij} = 1 - e^{-\left(\frac{a}{L_{ij}} + bP_j\right)} \quad (19)$$

其中  $L_{ij}$  表示  $j$  会员与  $i$  任务的距离， $P_j$  表示  $j$  任务的定价， $a, b$  为任务吸引度函数的两个参数， $w_{ij}$  表示  $i$  任务对  $j$  会员的吸引度。当吸引度的值接近于零时，说明这个任务对会员毫无吸引力；当吸引度的值接近于 1 时，说明这个任务非常具有吸引力。

##### (2) 参数 $a, b$ 求解

为了确定任务吸引度中的两个参数  $a, b$ ，我们需要两组已知数据。在最高价且已完成的任务中，选取会员与任务距离最短的两组数据。对于距离最短的那一组来说，任务对相应的会员的吸引度也相应的最大，所以将这个任务对这个会员的吸引度设为 0.95；而对于距离第二短的数据，将其吸引度设为 0.9。通过这两组数据，建立方程即可求出两个参数的值。



### (3) 阈值的确定

在得到具体的任务吸引度函数表达式之后，需要将每个人物对每个会员的吸引度  $w_{ij}$  与这个任务的吸引度阈值  $w_i$  作比较，用以判断这个任务是否具有足够的吸引度被完成。当吸引度大于等于阈值时，该人物具备足够的吸引度被会员接受并完成，即：

$$C_i = \begin{cases} 1, & w_{ij} > w_i \\ 0, & w_{ij} \leq w_i \end{cases}$$

对于未完成任务，该任务的吸引度阈值为对周围会员的吸引度最高的任务对应的阈值；对于已经完成任务，该任务的吸引度阈值为对周围会员吸引度最低的任务对应的阈值。

利用吸引度  $w_{ij}$  的计算公式与附件一中的任务数据计算出各个人物对各个会员吸引度，当第  $i$  个任务被完成时，其阈值至少低于其对一个会员的吸引度；当第  $i$  个任务未被完成时，其阈值不低于任何其对一个会员的吸引度。

### 6.2.5 模型求解算法

#### (1) 算法分析

对于最优定价方案的问题，由于目标函数为非线性函数，且总共有 15 个决策变量，所以利用传统的 MATLAB 函数求解较为复杂。于是，决定使用蒙特卡洛投点的方式进行随机搜索，以寻找目标函数最优解。

求解出每个情况的最优定价之后，利用会员的信誉值将会员排序，接着通过各任务对会员的吸引度结合阈值，将会员按照所排顺序进行任务的选择。

#### (2) 算法步骤

最优定价求解算法：

分别对两个地区三种不同权重值  $k = 0.2, 0.5, 0.8$  进行以下计算。

**Step1** 设置重复投点次数 = 1000000，当前投点次数  $n=0$ ，权重值  $k$ ，判断标志设为初始值无限大  $\min=\inf$ 。

**Step2** 在决策变量约束下随机生成 15 个定价  $x_{1i}$ ，利用第一问中任务完成率的表达式进行计算。

**Step3** 计算目标函数  $H(x_{1i})$ ，对  $H(x_{1i})$  进行判断。如果  $H(x_{1i}) < \min$ ，则令  $\min = H(x_{1i})$ ，并保存此组  $x_{1i}$  的值，投点次数  $n$  加 12。如果  $\min \leq H(x_{1i})$ ，则投点次数  $n$  加 1。

**Step4** 若投点次数  $n \leq N$ ，回到 Step2，否则结束循环，输出  $x_{1i}$  和  $H(x_{1i})$  的值。

模拟任务选择算法：

**Step1** 时间从 6:30 开始，8:00 结束，3 分钟一轮，一轮里会员选任务的优先顺序按信誉值进行排序；

**Step2** 按照优先顺序遍历会员，计算会员接任务范围内的每个任务对会员的吸引度；

**Step3** 会员从吸引度最高的任务开始挑选，判断任务对会员的吸引度是否大于任务的阈值，若大于阈值，则任务被会员接走；

**Step4** 若小于阈值，则会员挑选下一个吸引度大的任务，判断任务是否大于任务的阈值，若大于阈值，则会员接走任务，返回 step2，若小于阈值，则重复该步骤，直至选取了一个任务或者放弃接任何任务，返回 step2。

### 6.2.6 模型的求解结果及分析

通过 MATLAB 编程求解得到各区域的最低价格，权值  $k = 0.2, 0.5, 0.8$  时的平均定价，源代码见附录 F，结果如下：

表 8 各区域不同权重值拟合结果

区域	原始任务均价	任务数	最低价	修改后任务均价		
				k=0.2	k=0.5	k=0.8
1	66.42	88	65.89	68.67	68.86	69.47
2	68.81	73	67.40	71.07	70.85	70.88
3	72.64	37	66.73	70.40	70.20	70.41
4	71.51	46	66.96	69.52	70.75	70.57
5	68.71	53	65.50	69.49	68.39	68.04
6	69.57	45	69.24	71.36	73.21	71.46
7	66.69	93	63.10	66.44	66.74	66.73
8	67.94	45	67.73	70.83	71.02	71.69
9	70.35	84	65.82	69.61	69.22	69.47
10	66.36	94	67.49	70.26	71.09	71.36
11	72.11	19	60.27	63.88	63.00	62.71
12	73.56	24	66.47	68.88	68.84	69.56
13	70.83	32	64.89	68.42	67.03	67.94
14	71.83	46	64.61	67.92	67.73	66.71
15	71.03	56	64.84	68.75	67.31	68.05
总计		835	53730	53176.51	54012.16	57850.24

接着将不同权值对应的不同区域的定价带入任务选择的模拟系统，接着计算各个区域的任务完成率，得到如下结果：

表 9 模拟系统任务完成率统计

区域编号	原始	k=0.2	k=0.5	k=0.8
1	28.41%	43.18%	52.27%	55.68%
2	78.08%	84.93%	84.93%	86.30%
3	51.35%	56.76%	67.57%	75.68%
4	56.52%	50.00%	50.00%	76.09%
5	94.34%	94.34%	96.23%	96.23%
6	22.22%	93.33%	95.56%	95.56%
7	59.14%	55.91%	73.12%	75.27%
8	51.11%	86.67%	86.67%	93.33%
9	57.14%	48.81%	57.14%	72.62%
10	44.68%	56.38%	61.70%	67.02%
11	100.00%	100.00%	100.00%	100.00%
12	100.00%	91.67%	95.83%	100.00%
13	96.88%	100.00%	100.00%	100.00%
14	80.43%	82.61%	86.96%	91.30%
15	100.00%	96.43%	96.43%	98.21%
总计	68.02%	76.07%	80.29%	85.55%

将两张表的总体情况作柱状图，得到如下：

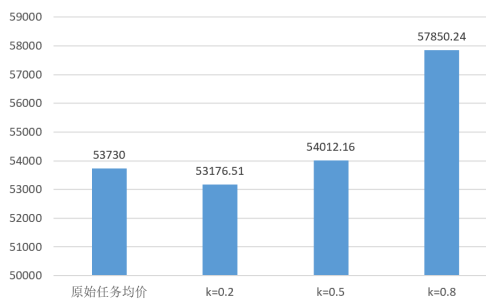


图 5 任务均价统计

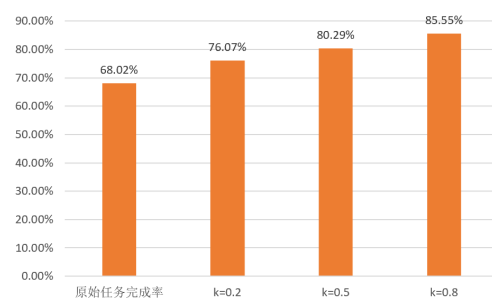


图 6 任务完成率统计

通过两张图，可以观察到当  $k = 0.2$  时，平台总支出由原来的 53730 元减少到 53176.51 元，并且任务完成率由 68.02% 提高到了 76.07%。当  $k = 0.5, 0.8$  时，虽然任务均价都有所提高，但均价提高幅度较小，且任务完成率的提高的幅度比较大，分别提高了 12.27%，17.53%。所以新设计的任务定价方案有效，且三个方案中最优定价方案为  $k = 0.2$  时的定价方案。

### 6.3 问题三

在平台的实际运营中，由于任务分布可能会不均，多个任务可能因为位置比较集中，导致用户会争相选择，一种有效的方法便是将这些任务联合在一起打包发布。打包发布的任务能够有效减少用户抢任务的情况，因为任务的打包提高了做任务的成本，便将一部分会员排除在外了。

### 6.3.1 任务打包模型的建立

一个优秀的任务打包方案不仅能够将参与的会员数控制在合理范围内，还能保证每个任务包内的任务数量合理且距离合适。为了合理的制定打包方案，决定使用遍历搜索的方法对每个点进行遍历组合成任务包，首先需要考虑打包搜索的范围。

#### (1) 任务打包方式搜索范围

为了通过遍历的方式寻找合适的打包方案，首先需要确定搜索的范围，即搜索的条件。对于整个区域所有任务来说，需要对每一个单独的任务进行所属任务包的确认。在确认完当前任务所属任务包后，判断遍历任务个数与总任务数的关系。如果当前遍历任务数小于总任务数，则继续进行下一个任务的任务所属任务包的确认；如果遍历任务数等于总任务数，则结束遍历。

#### (2) 任务打包判定

对于遍历中的单个任务来说，需要从此任务周围开始进行搜索。一个任务包的设定必须满足两个条件：任务包内任务间的距离不能相隔太远；任务包内任务总数不能太多。如果任务包内任务相隔太远会导致会员完成一个任务包的时间成本过高，数量太多也会导致会员完成一个任务包的成本高，并且不同地区会员的信誉值也不同，会员信誉值越高，则能更加放心推出大型任务包，会员信誉值越低，则考虑到如果任务包的难度过大，会造成会员放弃执行任务。所以需要对任务数量和任务距离进行限定。

首先对任务包距离进行考虑。任务包内各任务间的距离是首先需要考虑的，所以引入包内平均离差平方和的概念来作为表示包内各任务距离的一个指标 [3]。平均离差平方和表示当前任务包内所有任务到所有任务包的中心位置的距离之和的平方，它反应了各任务与任务包中心的一个偏离程度，包内平均离差平方和越大，说明任务包内各任务间的总距离越大。包内平均离差平方和  $d^2$  的计算公式如下：

$$d^2 = \frac{1}{N_i} \sum_{j=1}^{N_i} (x_j^{(i)} - m^{(i)})(x_j^{(i)} - m^{(i)})^T \quad (20)$$

其中， $N_i$  表示当前任务包内任务的数量， $x_j^{(i)}$  表示第  $k$  个任务的坐标  $(x, y)$ ， $m^{(i)}$  表示当前任务包的所有任务的中点坐标。

由于会员对不同大小的任务包的接受程度不同，并且不同信誉值会员分布情况也不同，所以在考虑任务包的距离大小时，需要结合当前任务包区域内的会员信誉值的平均值进行考虑。首先定义信誉等级，不同信誉分数对应的信誉等级不同，具体等级划分如下：

表 10 信誉等级划分

信誉值	0 至 99	100 至 199	200 至 299	300 至 399	大于等于 400
信誉等级	1	2	3	4	5

为了综合考虑信誉对任务包范围的影响，设计了信誉系数的概念，信誉系数  $k$  为当前任务包范围内会员的平均信誉值对应的信誉等级的倒数，即：

$$k = \frac{1}{\text{当前任务包范围内会员的信誉等级}}$$

综上，任务包范围搜索大小由包内平均离差平方和与信誉系数共同决定，于是便可以将这两个值相乘定义为当前任务所属任务包的打包范围  $D$ ：

$$D = d^2 \cdot k$$

对于不同的任务都有不同的打包范围最大值  $D_{max}$ ，此任务的打包范围肯定小于最大值。

对于一个任务包除了距离还有另外一个约束条件，即任务包内任务的总数量  $C$ 。

综上得到两个对单个任务打包的停止条件：

$$\begin{cases} d^2 \cdot k \leq D_{max} \\ C \leq C_{max} \end{cases} \quad (21)$$

最后只需要对地区内所有的任务进行遍历判断即可得到最终的任务包分部。

### 6.3.2 打包定价模型的建立

由于任务打包后问题二中建立的最有定价模型不再适用，需要对其进行修改。

首先引入任务等效均价的概念。在任务包中单个任务的均价比不打包任务单价低的情况，两者对任务完成率的贡献可能相等。则称打包前的任务均价  $x_1$  为打包后单个任务的任务等效均价，设为  $x'_1$ 。对于一个未打包前定价为  $q$  的任务，设打包后的任务定价为  $\alpha q$  且打包前的定价和打包后的定价对完成率的贡献相等，则未打包的定价  $q$  为打包后的定价  $\alpha q$  的等效定价，其中  $\alpha \in (0, 1]$ 。打包后的任务的任务等效定价之间可以由  $\alpha$  进行转换。而等效任务总数表示将打包后的任务包看作单个任务计算得到的总任务数。

修改问题一中三个影响定价的主要因素，建立打包后的新多目标规划模型。

#### 1. 等效任务均价

由于有的任务被打包，所以在计算任务均价时需要将这部分价格按照打包后的平均价格计算。任务均价则等于未打包的任务总价乘以未打包比例加上打包任务总价乘上打包比率：

$$x_1 = x'_1(1 - Z_c) + \alpha x'_1 \cdot Z_c$$

其中  $x'_1$  表示任务等效均价，即任务未打包的价格， $\alpha x'_1$  则为任务打包的价格，取  $\alpha = 0.8$ 。

#### 2. 人均 GDP

人均 GDP 没有变化，通过查阅相关文献获取。

#### 3. 区域等效任务均价与人均 GDP 匹配程度

区域等效任务均价与人均 GDP 匹配程度  $x_3$  的定义为：

$$x'_3 = \left(\frac{x_2}{x'_1}\right)^{\frac{1}{8}} \quad (22)$$

#### (3) 定价模型修改

在重新确定三个影响因素之后，便可以建立新的多目标优化模型。将式 (12) 中的任务数  $x_1$  换成等效任务数  $x'_1$ ，即可得到新的任务完成率函数：

$$y_i = 1 - e^{-(a_0 + a_1 x'_{1i} + a_2 x_{2i} + a_3 (\frac{x_{2i}}{x'_{1i}})^{\frac{1}{8}})} \quad (23)$$

### 目标一：地区总任务完成量最大

对于任务打包后的情况，地区总任务完成量也需要尽可能达到最大值，利用同样的求法，得到优化目标一为：

$$\max B'(x'_{1i}) = \sum A'_i - A'_i e^{-(a_0 + a_1 x'_{1i} + a_2 x_{2i} + a_3 (\frac{x_{2i}}{x'_{1i}})^{\frac{1}{8}})} \quad (24)$$

其中， $A'_i$  表示等效任务总数。

### 目标二：平台在单个区域平均每个任务的酬金最低

由于任务打包后导致等效任务均价情况的出现，平台的实际支出便变成了未打包任务支出的酬金和打包任务支出的酬金之和。则优化目标可以表示为：

$$\min W'(x'_{1i}) = \frac{\sum A_i(1 - Z_c)y_i x'_{1i} + A_i Z_c y_i x'_{1i} \alpha}{\sum A_i(1 - Z_c)y_i x'_{1i} + a_i Z_c y_i} \quad (25)$$

其中  $A_i$  表示任务总数。

### 决策变量与约束条件：

通过前面优化目标的分析，决策变量为等效任务均价  $x'_{1i}$ 。

对于约束条件，也可以通过问题二中类似的方法求解等效任务均价的最小值  $x'_{1imin}$ ，即令任务完成率为 0：

$$0 = 1 - e^{-(a_0 + a_1 x'_{1imin} + a_2 x_{2i} + a_3 (\frac{x_{2i}}{x'_{1i}})^{\frac{1}{8}})}$$

## (4) 模型汇总

通过对定价情况的影响因素和约束条件的分析，最终得到多目标约束模型如下：

$$\begin{aligned} \min \quad & W'(x'_{1i}) = \frac{\sum a_i(1 - Z_c)y_i x'_{1i} + a_i Z_c y_i x'_{1i} \alpha}{\sum a_i(1 - Z_c)y_i x'_{1i} + a_i Z_c y_i} \\ \max \quad & B'(x'_{1i}) = \sum A_i - A_i e^{-(a_0 + a_1 x'_{1i} + a_2 x_{2i} + a_3 (\frac{x_{2i}}{x'_{1i}})^{\frac{1}{8}})} \\ s.t. \quad & x'_{1imin} \leq x'_{1i} \end{aligned}$$

接着对目标函数进行无量纲化处理，再采用线性加权法对其进行归一化处理，化为单目标函数：

$$\max \quad H'(x'_{1i}) = k' \left( \frac{B'}{B'_{max}} \right) + (1 - k') \frac{W'_{min}}{W'} \quad (26)$$

其中， $k$  和  $1 - k$  代表了每个目标函数的权重。

## 6.3.3 模型求解算法

### (1) 算法思想

对于打包模型的求解，则是通过对地区中每一个任务进行遍历。而对于每一个任务，分别在满足两个约束条件的情况下一直寻找相邻任务点，直到不满足其中一个约束条件为止。若在建立任务包的时候碰到已被归到其他包的任务，则对这个任务到两组任务包的距离进行求解比较，归并到距离较小的一组，若导致其他任务包发生变化，则对变化的任务包进行更新，直到所有的连带变化的组更新完毕，进行下一个任务包的寻找。

对于最优定价方案的求解同第二问的算法思想和步骤，不同的地方时将原来的任务总数替换成等效任务总数。

## (2) 算法步骤

### 任务打包算法

- step1** 将任务分为已打包任务和未打包任务；
- step2** 从未打包任务中随机选取一个任务，将此任务作为一个类；
- step3** 计算类中任务的横、纵坐标的平均值，并将平均值作为类的中心；
- step4** 以类的中心为圆心搜索，找出距离中心最近的任务；
- step5** 计算假设将该任务归入类后的中心，计算类周围区域的人均信誉值，根据人均信誉值确定平均类内离差平方的上限值，并判断此时平均类内离差平方与类内任务数是否满足约束；
- step6** 若平均类内离差平方与类内任务数满足约束，则判断该任务是否已经属于其他类，若不属于，则归入该类，更新类的中心，返回 **step4**；
- step7** 若该任务已经属于其他类，则分别计算该任务在原类中与原类中心的距离和假设该任务归入新类后与新类中心的距离，若在原类中与原类中心的距离小于在新类中与新类中心的距离，则不归入该任务，并且跳过该任务，返回 **step4** 继续搜索。反之归入该任务，更新中心，返回 **step4** 进行搜索；
- step8** 若平均类内离差平方与类内任务数不满足约束，则结束搜索，也不归入该任务，将此时类中的任务打为一个包；
- step9** 若类更新过程中归入属于其它类的任务，则返回到这些类重新继续搜索更新，结束后返回 **step2**。

### 修改后任务完成率模型算法

- step1** 时间从 6:30 开始，8:00 结束，3 分钟一轮，一轮里会员选任务的优先顺序按信誉值进行排序；
- step2** 按照优先顺序遍历会员，计算会员接任务范围内的每个任务对会员的吸引度，任务距离为会员从起点出发做完所有包里任务并回到原点的最短路径除以 2 倍的任务数量，定价为包任务的均价；
- step3** 会员从吸引度最高的任务开始挑选，判断任务对会员的吸引度是否大于任务的阈值，若大于阈值，则任务被会员接走；
- step4** 若小于阈值，则会员挑选下一个吸引度大的任务，判断任务是否大于任务的阈值，若大于阈值，则会员接走任务，返回 **step2**，若小于阈值，则重复该步骤，直至选取了一个任务或者放弃接任何任务，返回 **step2**。

#### 6.3.4 模型求解结果及分析

利用 MATLAB 编写任务包遍历寻找的算法（源代码见附录 H 和附录 I），成功在保持数量和距离合适的情况下将已完成项目任务进行了打包操作。具体的打包结果如下：



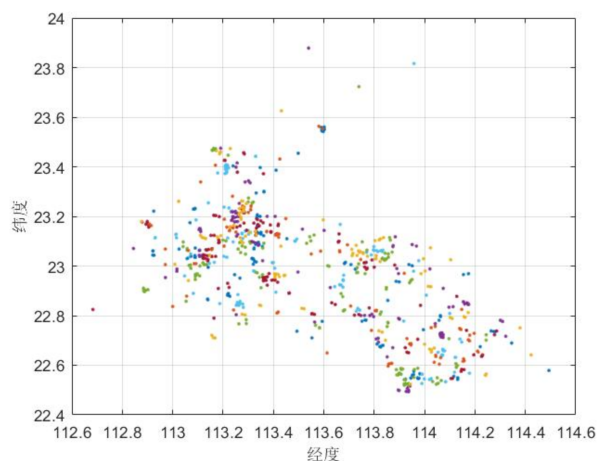


图 7 任务打包结果

图中位置相近且颜色相同的任务属于同一个任务包，从绘图结果能够明显的观察到任务打包的效果非常好。每一个任务包中的任务距离都很近，对于偏远的任务没有进行打包的操作。在任务分布密集的区域可以观察到，任务包的个数明显增多，这样很有效的保持了每个任务包内任务的个数。总体来说，对任务打包的结果很有效。

最终求解得到打包前后最优定价方案的具体数据如下，其中权值  $k = 0.2$ ：

表 11 打包前后定价与完成率对比

区域编号	打包前最优情况		打包后最优情况		
	平均定价	完成率	平均定价	平均等效定价	完成率
1	68.67	52.27%	70.32	56.25	79.55%
2	71.07	84.93%	65.12	52.09	93.15%
3	70.40	67.57%	68.70	54.96	94.59%
4	69.52	50.00%	71.75	57.40	95.65%
5	69.49	96.23%	65.40	52.32	90.57%
6	71.36	95.56%	70.75	56.60	86.67%
7	66.44	73.12%	66.00	52.80	93.55%
8	70.83	86.67%	68.69	54.95	71.11%
9	69.61	57.14%	68.91	55.13	82.14%
10	70.26	61.70%	71.84	57.47	82.98%
11	63.88	100.00%	68.12	54.49	100.00%
12	68.88	95.83%	66.39	53.11	100.00%
13	68.42	100.00%	65.02	52.02	100.00%
14	67.92	86.96%	72.51	58.01	91.30%
15	68.75	96.43%	65.39	52.31	98.21%
总计	69.19	80.29%	68.47	54.66	90.63%

对表格中的数据作直方图：

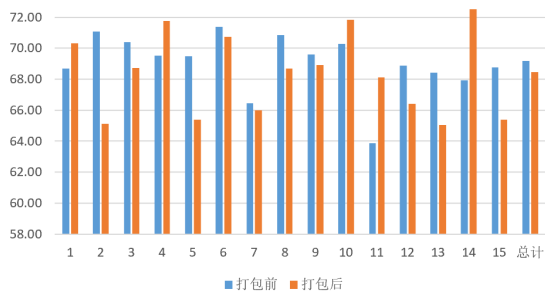


图 8 各地区不同情况定价比较

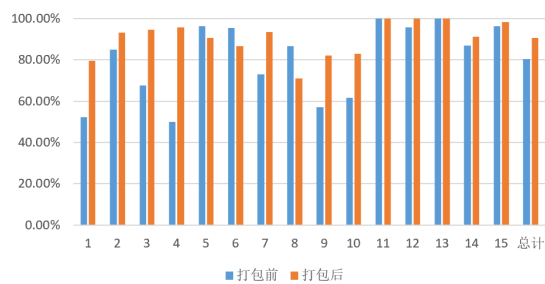


图 9 各地区不同情况完成率比较

由上图分析可得：任务经打包后，在绝大多数个区域内打包后的平均定价均低于未打包时的平均定价，而且打包后任务的完成率普遍高于未打包时任务的完成率。说明对任务进行打包既降低了被打包任务的定价，从而减少平台支出，又激发了会员的积极性，提高了任务的完成率，这些均有利于平台。通过最后一栏总体情况可以看出，打包后平均定价下降 14.53 元，并且完成率上升了 10.34%，相比打包前有很好的改善效果。

## 6.4 问题四

对于新项目定价方案的求解问题，利用相同的方案，结合前三问的求解。首先对任务分布区域进行 K-Means 聚类，将任务分为多个区域。然后将各任务按照第三问的方式进行打包操作，接着利用第三问中三个具体的任务完成度影响因素来建立多目标优化模型求解，即可得出对新项目的定价方案，最后方案进行评价。

首先，将新项目的任务分布点用 MATLAB 绘图，得到大致的分布情况，以及与城市的相对位置等信息，如图。

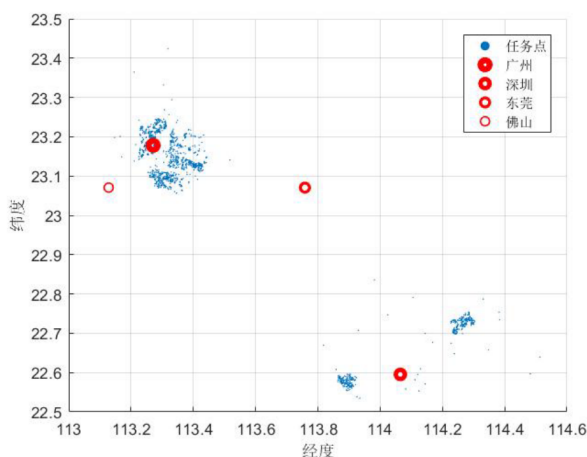


图 10 新项目任务分布图

### 6.4.1 K-Means 聚类

同样的，对新项目任务进行聚类分析，可以得到：

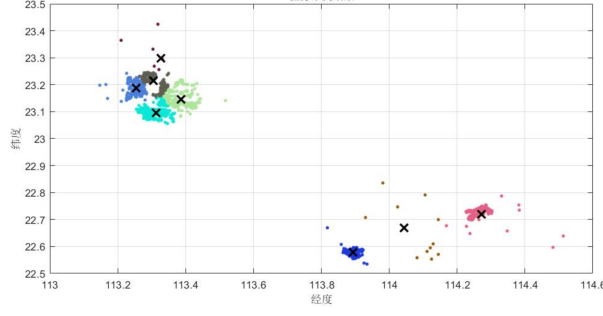


图 11 聚类分布图

#### 6.4.2 任务打包

完成任务区域的划分过后，接着利用第三问的任务打包模型对所有任务进行打包处理，得到结果如图：

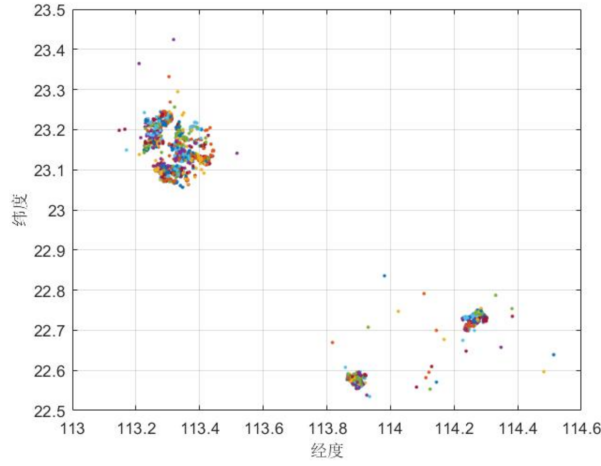


图 12 新项目任务打包分布图

通过上图可以看出打包效果非常好，每一个任务包内任务的距离没有出现很大的，并且在任务密集分布的区域也没有出现一个任务包任务数量过多的情况。

最后需要计算每个地区的打包率、每个区域的等效任务数和总任务数。

#### 6.4.3 最优定价模型的建立

利用第三问中对任务完成率影响的三个因素：等效任务均价、人均 GDP 和区域等效任务均价与人均 GDP 匹配程度，同样可以建立新的多目标优化模型，得到新的任务完成率函数：

$$y_i = 1 - e^{-(a_0 + a_1 x'_{1i} + a_2 x_{2i} + a_3 (\frac{x_{2i}}{x'_{1i}})^{\frac{1}{8}})} \quad (27)$$

接着建立多目标优化定价模型：

##### 目标一：地区总任务完成量最大

对于任务打包后的情况，地区总任务完成量也需要尽可能达到最大值，利用同样的求法，得到优化目标一为：

$$\max B'(x'_{1i}) = \sum A'_i - A'_i e^{-(a_0 + a_1 x'_{1i} + a_2 x_{2i} + a_3 (\frac{x_{2i}}{x'_{1i}})^{\frac{1}{8}})} \quad (28)$$

其中  $A'_i$  表示等效任务总数。

## 目标二：平台在单个区域平均每个任务的酬金最低

平台的实际支出为未打包任务支出的酬金和打包任务支出的酬金之和。优化目标可以表示为：

$$\min W'(x'_{1i}) = \frac{\sum A_i(1 - Z_c)y_i x'_{1i} + A_i Z_c y_i x'_{1i} \alpha}{\sum A_i(1 - Z_c)y_i x'_{1i} + A_i Z_c y_i} \quad (29)$$

其中， $A_i$  表示任务总数。

## 决策变量与约束条件：

通过前面优化目标的分析，决策变量为等效任务均价  $x'_{1i}$ 。

对于约束条件和之前相同，也可以通过问题二中类似的方法求解等效任务均价的最小值  $x'_{1imin}$ ，即令任务完成率为 0：

$$0 = 1 - e^{-(a_0 + a_1 x'_{1imin} + a_2 x_{2i} + a_3 (\frac{x_{2i}}{x'_{1i}})^{\frac{1}{8}})}$$

### 6.4.4 模型汇总

通过对定价情况的影响因素和约束条件的分析，最终得到多目标约束模型如下：

$$\begin{aligned} \min \quad & W'(x'_{1i}) = \frac{\sum a_i(1 - Z_c)y_i x'_{1i} + a_i Z_c y_i x'_{1i} \alpha}{\sum a_i(1 - Z_c)y_i x'_{1i} + a_i Z_c y_i} \\ \max \quad & B'(x'_{1i}) = \sum A_i - A_i e^{-(a_0 + a_1 x'_{1i} + a_2 x_{2i} + a_3 (\frac{x_{2i}}{x'_{1i}})^{\frac{1}{8}})} \\ s.t. \quad & x'_{1imin} \leq x'_{1i}, i = 1, 2, \dots, 15 \end{aligned}$$

接着对目标函数进行无量纲化处理，再采用线性加权法对其进行归一化处理，化为单目标函数：

$$\max \quad H'(x'_{1i}) = k' \left( \frac{B'}{B'_{max}} \right) + (1 - k') \frac{W'_{min}}{W'} \quad (30)$$

其中， $k$  和  $1 - k$  代表了每个目标函数的权重。

最后只需要将新项目中划分的各区域的信息带入模型求解，即可得到最佳定价方案。

### 6.4.5 模型算法

网格划分的算法同第一问；任务打包和目标规划的算法同第三问。

### 6.4.6 模型求解结果及分析

对目标函数求解，得到各个区域最优的定价方案，以及任务完成率如下：

表 12 附件三任务打包前后任务定价完成率对比

区域	打包前		打包后		
	平均定价	完成率	平均定价	等效平均定价	完成率
1	65.49	72.29%	65.49	52.39	91.57%
2	70.18	63.50%	69.52	55.62	94.07%
3	69.19	82.88%	68.54	54.84	97.26%
4	65.22	75.88%	66.45	53.16	94.55%
5	66.98	92.47%	64.46	51.57	92.47%
6	66.83	86.11%	65.16	52.13	94.70%
7	65.41	85.00%	66.46	53.17	95.71%
8	67.50	62.50%	71.69	57.35	100.00%

由上表分析可得，新任务经打包后，所有区域的任务完成率均很高，任务的平均定价也在 58 元以下，甚至一部分平均定价在 52 元以下。相比于打包前来说，任务完成率提高了 10% 到 20%，而任务的平均定价减少了 5 到 15 元，可见对于新任务打包定价方案实施的效果显著。

## 七、模型的评价及改进

### 7.1 模型优点

1. 使用 K-Means 聚类的方法进行研究，能够较为精细地对定价进行研究。
2. 对不同的城市分别进行了研究，能针对不同城市给出不同的定价方案
3. 提取了三个指标，并对任务完成度进行了定量的分析。
4. 从任务可行性、区域信誉值、包内任务距离这三个方面进行综合考虑，给出一种可行有效的打包方案。
5. 考虑了打包后的任务价格对完成率的影响。

### 7.2 模型的缺点

1. 尽管使用了网格化的方法，但还不能具体到对一个特定的任务进行定价。
2. 未考虑低信誉值的人群的流动。

### 7.3 模型的改进

对单个任务进行分析，可以更加精细和全面地给出定价方案。

## 参考文献

- [1] 司守奎. 孙玺菁. 数学建模算法与程序. 海军航空工程学院, 9:95–98, 2007.
- [2] 杨桂元 and 郑亚豪. 多目标决策问题及其求解方法研究. PhD thesis, 2012.
- [3] 王小平. 利用样本方差计算离差平方和. 数理医药学杂志, 16(5):394–395, 2003.

## 附录清单

编号	名称
A	问题一已结束项目位置分布图的 MATLAB 代码
B	问题一已结束项目定价规律分布的 MATLAB 代码
C	问题一已结束项目完成情况的 MATLAB 代码
D	问题一广州、佛山完成度函数拟合的 MATLAB 代码
E	问题一深圳完成度函数拟合的 MATLAB 代码
F	问题二不同权重下广州、佛山任务最优定价的 MATLAB 代码
G	问题二不同权重下深圳任务最优定价的 MATLAB 代码
H	问题三广州、佛山任务打包区域的 MATLAB 代码
I	问题三深圳打包区域划分的 MATLAB 代码
J	问题三广州、佛山转移人数的 MATLAB 代码
K	问题三深圳转移人数的 MATLAB 代码
L	问题三两地区各项指标

## 附录 A 问题一已结束项目位置分布图的 MATLAB 代码

```
%任务分布
clc,clear
data=xlsread('附件一.xls');
y=data(:,1);%经度
x=data(:,2);%纬度
scatter(x,y, '.')
hold on
plot(113.270963693286,23.1781324014828,'ro','LineWidth',4);%广州
plot(114.06485880878,22.5952184287333,'ro','LineWidth',3);%深圳
plot(113.758425499061,23.0705111619974,'ro','LineWidth',2);%东莞
plot(113.128194445181,23.0705276914163,'ro','LineWidth',1);%佛山
legend('任务点','广州','深圳','东莞','佛山');
grid on
xlabel('经度');
ylabel('纬度');
title('任务分布')
```

## 附录 B 问题一已结束项目定价规律分布的 MATLAB 代码

```
%任务定价规律分布
clc,clear
data=xlsread('附件一.xls');
y=data(:,1);%经度
x=data(:,2);%纬度
price=data(:,3);%定价
scatter(x,y,10,price,'filled')
hold on
plot(113.270963693286,23.1781324014828,'ro','LineWidth',4);%广州
plot(114.06485880878,22.5952184287333,'ro','LineWidth',3);%深圳
plot(113.758425499061,23.0705111619974,'ro','LineWidth',2);%东莞
plot(113.128194445181,23.0705276914163,'ro','LineWidth',1);%佛山
axis([112.75 114.5 22.5 23.5])
legend('任务点','广州','深圳','东莞','佛山');
grid on
xlabel('经度');
ylabel('纬度');
title('定价分布')
```

## 附录 C 问题一已结束项目完成情况的 MATLAB 代码

```
%执行情况
clc,clear
data=xlsread('附件一.xls');
y=data(:,1);%经度
x=data(:,2);%纬度
z=data(:,4);%执行情况
scatter(x,y,10,z,'filled')
hold on
plot(113.270963693286,23.1781324014828,'ro','LineWidth',4);%广州
plot(114.06485880878,22.5952184287333,'ro','LineWidth',3);%深圳
plot(113.758425499061,23.0705111619974,'ro','LineWidth',2);%东莞
plot(113.128194445181,23.0705276914163,'ro','LineWidth',1);%佛山
axis([112.75 114.5 22.5 23.5])
legend('任务点','广州','深圳','东莞','佛山');
```



```

grid on
xlabel('经度');
ylabel('纬度');
title('执行情况')

```

## 附录 D 问题一广州、佛山完成度函数拟合的 MATLAB 代码

```

%分区域
clc,clear
data1=xlsread('订单分布_广州、佛山.xlsx');
data2=xlsread('会员分布_广州、佛山.xlsx');
length1=size(data1);
length2=size(data2);
k_x=7;%经度间隔
k_y=4;%纬度间隔
x=linspace(112.75,113.6,k_x);%经度分成6份
y=linspace(22.8,23.5,k_y);%纬度分成4份
result=zeros((k_x-1)*(k_y-1),5);%每一列分别表示:任务数 价格 任务完成总个数 人数 限额总和
for i=1:length1(1)
    for m=1:k_x-1
        for n=1:k_y-1
            if data1(i,1)<y(n+1) && data1(i,1)>y(n) &&...
                data1(i,2)<x(m+1) && data1(i,2)>x(m)
                result(n+(k_y-1)*(m-1),1)=result(n+(k_y-1)*(m-1),1)+1;
                result(n+(k_y-1)*(m-1),2)=result(n+(k_y-1)*(m-1),2)+data1(i,3);
                result(n+(k_y-1)*(m-1),3)=result(n+(k_y-1)*(m-1),3)+data1(i,4);
            end
        end
    end
end

for i=1:length2(1)
    for m=1:k_x-1
        for n=1:k_y-1
            if data2(i,1)<y(n+1) && data2(i,1)>y(n) &&...
                data2(i,2)<x(m+1) && data2(i,2)>x(m)
                result(n+(k_y-1)*(m-1),4)=result(n+(k_y-1)*(m-1),4)+1;
                result(n+(k_y-1)*(m-1),5)=result(n+(k_y-1)*(m-1),5)+data2(i,3);
            end
        end
    end
end
RESULT=zeros((k_x-1)*(k_y-1),4);%每一列分别表示:任务均价 任务限额/最大限额/任务数 完成率
for i=1:(k_x-1)*(k_y-1)
    if result(i,1)==0;
        RESULT(i,1)=0;
        RESULT(i,2)=0;
        RESULT(i,4)=0;
    else
        RESULT(i,1)=result(i,2)/result(i,1);
        RESULT(i,2)=result(i,5)/(232*result(i,1));
        RESULT(i,4)=result(i,3)/result(i,1);
    end
    if result(i,4)==0
        RESULT(i,3)=0;
    else
        RESULT(i,3)=result(i,5)/result(i,4);
    end
end
for i=(k_x-1)*(k_y-1):-1:1

```

```

if result(i,1)==0
RESULT([i],:)=[];
end
end
D=RESULT;
%拟合
x1=D(:,1);
x2=D(:,2);
x3=D(:,3);
%标准化
x1=(x1-min(x1))/(max(x1)-min(x1));
x2=(x2-min(x2))/(max(x2)-min(x2));
x3=(x3-min(x3))/(max(x3)-min(x3));
y=D(:,4);
yy=D(:,4);
length=size(y);
for i=1:length(1)
if y(i)==1
y(i)=0.9999;
end
y(i)=log(1-y(i));
end
X=[ones(length(1),1),x1,x2,x3];
[b,bint,r,rint,stats]=regress(y,X);
z=(b(1)+b(2)*x1+b(3)*x2+b(4)*x3);
Z=1-exp(z)
scatter(-z,Z,'*')
hold on
scatter(-z,yy,'o')
axis([0 12 0 1])
legend('拟合点','原始点');
xlabel('数据点');
ylabel('完成率');
title('执行率-广州、佛山')

```

## 附录 E 问题一深圳完成度函数拟合的 MATLAB 代码

```

A=[3 3 9 22 39 52 10 27 109 32 51 61 13 11 14 1]; %各区域任务数
a=[-79.8481,1.1254,41.5263,0.013]; %参数
x2=[0.073275862 0.103448276 0.055555556 0.02507837 0.057471264 0.045507294...
0.030172414 0.055395913 0.056785827 0.054822198 0.033384043 0.092707744...
0.038461538 0.038009404 0.047721675 0.004310345]';
%各区域平均任务限额
x3=[12.75 7.2 8.285714286 6.736842105 10.19607843 9.465517241 8.75 8.897435897...
5.653543307 7.980392157 6.810344828 6.56 5.043478261 16.16666667 4.558823529 ...
1 ]'; %人均限额
y=zeros(16,0);
Bmax=457;Wmin=67.05054;
xmin=[68.09976415 67.05053835 68.80519729 69.94767117 68.71244183 69.16234115...
69.7364504 68.80402171 68.79020673 68.83578449 69.64034986 67.45424774...
69.47340449 69.36159892 69.1373118 70.78026241];
k=0.5; %权重
N=1000000;
min=inf;
for j=1:N
x=zeros(1,16);
%任意投点
for i=1:16
x(i)=xmin(i)+3*rand;
end
for i=1:16

```

```

y(i)=1-exp(-(a(1)+a(2)*x(i)+a(3)*x2(i)+a(4)*x3(i)));
end
f=-(k*sum(A.*y)/Bmax+(1-k)*Wmin/(sum(A.*y.*x)/sum(A.*y)));
if f<min
min=f;
xx=x;
yy=y;
end
end

```

## 附录 F 问题二不同权重下广州、佛山任务最优定价的 MATLAB 代码

```

%分区域
clc,clear
data1=xlsread('订单分布_深圳.xlsx');
data2=xlsread('会员分布_深圳.xlsx');
length1=size(data1);
length2=size(data2);
k_x=5;%经度间隔
k_y=3;%纬度间隔
x=linspace(113.7,114.5,k_x);%经度分成4份 约2km
y=linspace(22.5,22.8,k_y);%纬度分成2份 约2km
result=zeros((k_x-1)*(k_y-1),5);%每一列分别表示:任务数 价格 任务完成总个数 人数 限额总和
for i=1:length1(1)
for m=1:k_x-1
for n=1:k_y-1
if data1(i,1)<y(n+1) && data1(i,1)>y(n) &&...
data1(i,2)<x(m+1) && data1(i,2)>x(m)
result(n+(k_y-1)*(m-1),1)=result(n+(k_y-1)*(m-1),1)+1;
result(n+(k_y-1)*(m-1),2)=result(n+(k_y-1)*(m-1),2)+data1(i,3);
result(n+(k_y-1)*(m-1),3)=result(n+(k_y-1)*(m-1),3)+data1(i,4);
end
end
end
end

for i=1:length2(1)
for m=1:k_x-1
for n=1:k_y-1
if data2(i,1)<y(n+1) && data2(i,1)>y(n) &&...
data2(i,2)<x(m+1) && data2(i,2)>x(m)
result(n+(k_y-1)*(m-1),4)=result(n+(k_y-1)*(m-1),4)+1;
result(n+(k_y-1)*(m-1),5)=result(n+(k_y-1)*(m-1),5)+data2(i,3);
end
end
end
end

RESULT=zeros((k_x-1)*(k_y-1),4);%每一列分别表示: 任务均价 任务限额/最大限额/任务数 完成率
for i=1:(k_x-1)*(k_y-1)
if result(i,1)==0;
RESULT(i,1)=0;
RESULT(i,2)=0;
RESULT(i,4)=0;
else
RESULT(i,1)=result(i,2)/result(i,1);
RESULT(i,2)=result(i,5)/(232*result(i,1));
RESULT(i,4)=result(i,3)/result(i,1);
end
if result(i,4)==0
RESULT(i,3)=0;
else

```

```

RESULT(i,3)=result(i,5)/result(i,4);
end

end
for i=(k_x-1)*(k_y-1):-1:1
if result(i,1)==0
RESULT([i],:)=[];
end
end
D=RESULT;
%拟合
x1=D(:,1);
x2=D(:,2);
x3=D(:,3);
%标准化
x1=(x1-min(x1))/(max(x1)-min(x1));
x2=(x2-min(x2))/(max(x2)-min(x2));
x3=(x3-min(x3))/(max(x3)-min(x3));
y=D(:,4);
yy=D(:,4);
length=size(y);
for i=1:length(1)
if y(i)==1
y(i)=0.9999;
end
y(i)=log(1-y(i));
end
X=[ones(length(1),1),x1,x2,x3];
[b,bint,r,rint,stats]=regress(y,X);
z=(b(1)+b(2)*x1+b(3)*x2+b(4)*x3);
Z=1-exp(z)
scatter(-z,Z,'*')
hold on
scatter(-z,yy,'o')
% axis([0 1 0 1])
legend('拟合点','原始点');
xlabel('数据点');
ylabel('完成率');
title('执行率-深圳')

```

## 附录 G 问题二不同权重下深圳任务最优定价的 MATLAB 代码

```

%优化目标
%参数
clc,clear
A=[11 25 51 36 23 23 3 5]; %各区域任务数
a=[-1.4307 0.0186 3.5914 0.0363]; %参数
x2=[0.032523511 0.05362069 0.104208925 0.059626437...
0.086956522 0.099137931 0.001436782 0.075]';
%各区域平均任务限额
x3=[3.192307692 3.141414141 6.774725275 3.387755102...
4.10619469 7.450704225 1 4.833333333]'; %人均限额
y=zeros(16,0);

xmin=[64.40937061 60.43512476 43.57642684 58.79473677 ...
52.11556342 43.23631566 74.69031942 53.00510753];
Bmax=sum(A);Wmin=min(xmin);
k=0.1; %权重
N=1000000;
min=inf;
for j=1:N

```

```

x=zeros(1,8);
%任意投点
for i=1:8
x(i)=xmin(i)+20*rand;
end
for i=1:8
y(i)=1-exp(-(a(1)+a(2)*x(i)+a(3)*x2(i)+a(4)*x3(i)));
end
f=-(k*sum(A.*y)/Bmax+(1-k)*Wmin/(sum(A.*y.*x)/sum(A.*y)));
if f<min
min=f;
xx=x;
yy=y;
end
end

```

## 附录 H 问题三广州、佛山打包区域划分的 MATLAB 代码

```

%广州、佛山打包区域划分
clc,clear
%参数
Dmax=1;%最大5km
Nmax=10;%最多联合五个任务
%订单周围信誉值
x=xlsread('订单分布_广州、佛山.xlsx');
y=xlsread('会员分布_广州、佛山.xlsx');
length1=size(x);
length2=size(y);
for i=1:length1(1)
sum=0;
num=0;
for j=1:length2(1)
dist(i,j)=distance(x(i,1),x(i,2),y(j,1),y(j,2))*pi/180*6370;
if dist(i,j)<10
num=num+1;
sum=sum+y(j,5);
end
end
x(i,5)=sum/num;
end

%分配信誉系数 第六列
length=size(x);
for i=1:length(1)
if x(i,5)>=400
x(i,6)=5;
elseif x(i,5)>=300
x(i,6)=4;
elseif x(i,5)>=200
x(i,6)=3;
elseif x(i,5)>=100
x(i,6)=2;
else
x(i,6)=1;
end
end

%打包区域划分 第七列
x(:,7)=zeros(length(1),1);
seq=1;%打包序号
bao=[];

```

```

for i=1:length(1)
if x(i,7)~=0
continue;
else
x(i,7)=seq;
end
for j=i+1:length(1)
if x(j,7)~=0
continue
else
x(j,7)=seq;
bao=find(x(:,7)==seq);%寻找包序为seq的包内任务
length_bao=size(bao);%计算包内的任务总数
avg_w=sum(x(bao,1))/length_bao(1);%计算包内平均纬度
avg_j=sum(x(bao,2))/length_bao(1);%计算包内平均经度
xin=sum(x(bao,6))/length_bao(1);%计算包内平均信誉度
d=0;
for k=1:length_bao
d=d+(distance(x(bao(k),1),x(bao(k),2),avg_w,avg_j)*pi/180*6370)^2;
end
d=d/length_bao(1);%计算类内平均离差平方和
if length_bao(1)>Nmax
x(j,7)=0;
continue;
else
if (d/xin)>Dmax
x(j,7)=0;
end
end
end
end
seq=seq+1;
end

%打包可视化
seq=max(x(:,7));
for i=1:seq
[m,n]=find(x(:,7)==i);
data=x(m,:);
plot(data(:,2),data(:,1),'.')
hold on
end
grid on
plot(113.270963693286,23.1781324014828,'ro','LineWidth',4);%广州
plot(113.128194445181,23.0705276914163,'ro','LineWidth',1);%佛山
legend('任务点','广州','佛山');
axis([112.75 113.6 22.8 23.5])
xlabel('经度');
ylabel('纬度');

```

## 附录 I 问题三深圳打包区域划分的 MATLAB 代码

```

%深圳打包区域划分
clc,clear
%参数
Dmax=1;%最大5km
Nmax=10;%最多联合五个任务
%订单周围信誉值
x=xlsread('订单分布_深圳.xlsx');
y=xlsread('会员分布_深圳.xlsx');
length1=size(x);

```

```

length2=size(y);
for i=1:length1(1)
sum=0;
num=0;
for j=1:length2(1)
dist(i,j)=distance(x(i,1),x(i,2),y(j,1),y(j,2))*pi/180*6370;
if dist(i,j)<10
num=num+1;
sum=sum+y(j,5);
end
end
x(i,5)=sum/num;
end

%分配信誉系数 第六列
length=size(x);
for i=1:length(1)
if x(i,5)>=400
x(i,6)=5;
elseif x(i,5)>=300
x(i,6)=4;
elseif x(i,5)>=200
x(i,6)=3;
elseif x(i,5)>=100
x(i,6)=2;
else
x(i,6)=1;
end
end

%打包区域划分 第七列
x(:,7)=zeros(length(1),1);
seq=1;%打包序号
bao=[];
for i=1:length(1)
if x(i,7)~=0
continue;
else
x(i,7)=seq;
end
for j=i+1:length(1)
if x(j,7)~=0
continue
else
x(j,7)=seq;
bao=find(x(:,7)==seq);%寻找包序为seq的包内任务
length_bao=size(bao);%计算包内的任务总数
avg_w=sum(x(bao,1))/length_bao(1);%计算包内平均纬度
avg_j=sum(x(bao,2))/length_bao(1);%计算包内平均经度
xin=sum(x(bao,6))/length_bao(1);%计算包内平均信誉度
d=0;
for k=1:length_bao
d=d+(distance(x(bao(k),1),x(bao(k),2),avg_w,avg_j))*pi/180*6370)^2;
end
d=d/length_bao(1);%计算类内平均离差平方和
if length_bao(1)>Nmax
x(j,7)=0;
continue;
else
if (d/xin)>Dmax
x(j,7)=0;
end
end
end
end

```



```

end
seq=seq+1;
end

%打包可视化
seq=max(x(:,7));
for i=1:seq
[m,n]=find(x(:,7)==i);
data=x(m,:);
plot(data(:,2),data(:,1),'.')
hold on
end
grid on
plot(114.06485880878,22.5952184287333,'ro','LineWidth',3);%深圳
legend('任务点','深圳');
axis([113.75 114.25 22.5 22.8])
xlabel('经度');
ylabel('纬度');

```

## 附录 J 问题三广州、佛山转移人数的 MATLAB 代码

```

%转移人数
clc,clear
k=0.5;
%每一列： 区域 打包率 高信誉人数 等效任务数
x=[1 1 0 1
2 0 1 3
3 0.5666 2 6
4 0.8636 1 9
5 0.8205 8 15
6 0.8654 19 20
7 0.7 0 4
8 0.8519 4 12
9 0.945 31 28
10 0.75 5 15
11 0.902 9 15
12 0.9194 32 15
13 0.6154 3 7
14 0.9091 1 4
15 0.7143 4 6
16 0 0 1

];
length1=size(x);
y=zeros(16);
for i=1:length1(1);
for j=1:length1(1)
y(i,j)=x(i,2)-x(j,2);
end
end
%每列： 原地 目的地
zhuanYiRen=[2 1
2 4
3 1
3 4
3 5
4 6
5 6
5 8
6 9

```

```

7 6
7 10
8 9
8 11
10 9
11 12
11 14
12 9
13 10
13 12
15 12
15 14
16 13
16 15
];
length2=size(zhuanYiRen);
for i=1:length2(1)
    zhuanYiRen(i,3)=y(zhuanYiRen(i,2),zhuanYiRen(i,1));%打包率差值
    zhuanYiRen(i,4)=x(zhuanYiRen(i,1),3);%原地高信誉人数
    zhuanYiRen(i,5)=x(zhuanYiRen(i,2),4);%目的地等效任务数
    zhuanYiRen(i,6)= zhuanYiRen(i,3)*zhuanYiRen(i,5);%权值
end
for i=1:length2(1)
    sum=0;
    for j=1:length2(1)
        if zhuanYiRen(j,1)==zhuanYiRen(i,1)
            sum=sum+zhuanYiRen(j,6);
        end
    end
    zhuanYiRen(i,7)=k*zhuanYiRen(i,6)*zhuanYiRen(i,4)/sum;%转移人数
end
%转移后人数
r=x(:,3);
for i=1:length1(1)
    for j=1:length2(1)
        if zhuanYiRen(j,1)==i;
            r(i)=r(i)-zhuanYiRen(j,7);
        end
        if zhuanYiRen(j,2)==i;
            r(i)=r(i)+zhuanYiRen(j,7);
        end
    end
end
end

```

## 附录 K 问题三深圳转移人数的 MATLAB 代码

```

%转移人数
clc,clear
k=0.5;
%每一列: 区域 打包率 高信誉人数 等效任务数
x=[1    0.8 1  2
2    0.8333 7  9
3    1    13 6
4    0.8333 4  5
5    0.8261 27 11
6    0.8889 11 9
7    0.875  13 8
8    0.75   6  10

];
length1=size(x);

```

```

y=zeros(8);
for i=1:length1(1);
for j=1:length1(1)
y(i,j)=x(i,2)-x(j,2);
end
end
%每列： 原地 目的地
zhuanYiRen=[1 2
1 3
4 3
4 6
5 3
5 6
5 7
8 7

];
length2=size(zhuanYiRen);
for i=1:length2(1)
zhuanYiRen(i,3)=y(zhuanYiRen(i,2),zhuanYiRen(i,1));%打包率差值
zhuanYiRen(i,4)=x(zhuanYiRen(i,1),3);%原地高信誉人数
zhuanYiRen(i,5)=x(zhuanYiRen(i,2),4);%目的地等效任务数
zhuanYiRen(i,6)= zhuanYiRen(i,3)*zhuanYiRen(i,5);%权值
end
for i=1:length2(1)
sum=0;
for j=1:length2(1)
if zhuanYiRen(j,1)==zhuanYiRen(i,1)
sum=sum+zhuanYiRen(j,6);
end
end
zhuanYiRen(i,7)=k*zhuanYiRen(i,6)*zhuanYiRen(i,4)/sum;%转移人数
end
%转移后人数
r=x(:,3);
for i=1:length1(1)
for j=1:length2(1)
if zhuanYiRen(j,1)==i;
r(i)=r(i)-zhuanYiRen(j,7);
end
if zhuanYiRen(j,2)==i;
r(i)=r(i)+zhuanYiRen(j,7);
end
end
end
end

```

## 附录 L 两地区各项指标

表 13 广州、佛山地区任务打包后的各项数据指标

区域	普通总限额	平均等效任务数	任务限额总量	高信誉人数	普通人数	各区域人数
1	51	1	52.8190	0.1197	4	4.1197
2	21	3	28.6000	0.5000	9	9.5000
3	62	6	77.2000	1.0000	12	13.0000
4	114	9	134.2093	1.3296	18	19.3296
5	354	15	423.1717	4.5508	43	47.5508
6	284	20	478.8290	12.8177	47	59.8177
7	70	4	70.0000	0.0000	8	8.0000
8	251	12	299.3710	3.1823	35	38.1823
9	749	28	1669.3973	60.5525	224	284.5525
10	298	15	342.9970	2.9603	46	48.9603
11	149	15	224.2027	4.9475	48	52.9475
12	462	15	806.9471	22.6939	177	199.6939
13	76	7	98.8000	1.5000	10	11.5000
14	39	4	67.0559	1.8458	5	6.8458
15	54	6	84.4000	2.0000	30	32.0000
16	1	1	1.0000	0.0000	1	1.0000

表 14 深圳地区任务打包后的各项数据指标

区域	普通总限额	平均等效任务数	任务限额总量	高信誉人数	普通人数	各区域人数
1	57	2	65.7900	0.5000	13	13.5000
2	186	9	310.8166	7.0999	84	91.0999
3	284	6	666.8362	21.7768	76	97.7768
4	162	5	197.1600	2.0000	62	64.0000
5	878	11	1115.3300	13.5000	112	125.5000
6	683	9	955.1809	15.4824	88	103.4824
7	472	8	799.7064	18.6409	60	78.6409
8	258	10	310.7400	3.0000	34	37.0000