

“拍照赚钱”的任务定价

摘要

“拍照赚钱”是移动互联网下的一种自助式服务模式，本文通过建立数学模型，对任务定价问题进行研究。

针对问题一，我们使用**K-均值聚类法**对任务位置、会员位置进行聚类，基于此，定义任务和会员的距离计算公式，分广州、佛山、东莞、深圳四个地区进行回归拟合，得出每个地区距离（任务和会员距离）和定价之间的函数关系。

针对问题二，我们将**市场供求比**和**会员信誉**两个因素纳入定价模型，对问题一中的模型作出修正，建立**分段函数综合定价模型**；基于项目总额最低和完成度最高两个目标建立**多目标优化模型**，使用**基于实数编码加速遗传算法**对该非线性优化模型进行求解，求出分段函数综合定价模型**待估参数的最优值**，从而完成定价；定义**项目性价比**这一评价指标对修正的定价模型进行对比评价。实施方案后，广州、深圳两个地区项目性价比提高，佛山、东莞两个地区项目性价比降低，总体项目性价比由 1.082×10^{-5} 提高为 1.15×10^{-5} 。

针对问题三，我们建立**基于任务匹配程度打包模型**，将匹配度介于 0 到 0.3 的任务进行打包合并；根据**公平性费率分摊**的思想，利用 100%，75%，60%三种比率分别对任务价格进行独立的费用折减，建立**基于公平性费率分摊定价模型**，实现对稠密任务的打包定价，并对模型进行检验。

针对问题四，我们首先将附件三中的数据在实际地图中模拟，确立任务分布的地区为广州城区和深圳地区，分别建立分段函数综合定价模型和打包定价模型，通过模型检验可以得出：广州城区应采用打包定价模型，深圳地区应采用分段函数综合定价模型。

关键词： K-均值聚类 分段函数综合定价模型 多目标规划
任务匹配程度打包模型 基于公平性费率分摊定价模型

1. 问题重述

1.1 问题背景

伴随着互联网发展，信息的传递变得更加高速、容易，人们不必用重金的方式促使信息传得更快更远，公司和机构开始更多的尝试将过去由员工执行的任务以自由自愿的形式外包给非特定的大众网络的做法。这种伴随互联网发展而兴起的新兴网络经济被称之为众包，“拍照赚钱”是在众包基础上发展形成的一种移动互联自助式服务模式。它的运营流程为，用户下载有关 APP，注册成为 APP 的会员，然后从 APP 上领取需要拍照的任务（如上超市去检查某种商品的上架情况），完成后即可赚取 APP 对任务所标定的酬金。任务的发布和定价将会受到任务位置、任务预定时间，会员位置、会员信誉值以及会员配额等因素的影响。相比传统的市场调查方式，这种基于移动互联网的自助式劳务众包平台，能为企业提供各种商业检查和信息搜集，大大节省调查成本，而且有效地保证了调查数据真实性，缩短了调查的周期，受到了企业的欢迎。以 2016 年为例，一年中以“美团拍店”，“高德淘金”为代表的一批“拍照赚钱”APP，下载量突破百万次，成交金额突破千万元。但在此劳务众包平台中，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，而导致商品检查的失败。制定合理的出价策略对于任务发布者提高产出并相对地降低成本有着重要的现实意义，对网站的运营管理具有不可忽视的现实指导意义。

1.2 目标任务

问题一：通过研究附件一，探索“拍照赚钱”的任务定价规律，分析任务未完成的原因。

问题二：为附件一设计新的任务定价方案，并和原方案进行比较。

问题三：将任务联合发布的实际情况纳入考虑，修改定价模型，分析打包发布对最终任务完成情况的影响。

问题四：为附件三设计任务定价方案，并评价该方案的实施效果。

2. 问题分析

问题一，主要研究附件一中的任务定价规律。通过对任务位置与会员位置进行 K-均值聚类，分广州、佛山、东莞、深圳四个地区进行观察分析，得到基于距离因素的定价规律。

问题二，主要研究如何修正第一问的定价模型，并对比效果。将第一问中未考虑到的会

员信誉和市场供求比因素加入到模型中去,建立分段函数综合定价模型。并定义性价比概念,对比效果。

问题三,主要研究如何将位置较近的任务进行联合分布打包,并修改定价模型问题。由于难以对所有任务较近的位置进行计算,所以建立了基于任务匹配程度的打包模型,将匹配度较高的任务进行打包。并基于打包任务,分别按对应比例进行独立折减,以得到定价改进模型,计算任务完成情况并与第二问进行对比。

问题四,主要研究在前三问中得到的定价模型的基础上,如何给出附件三中新项目的定价模型问题。分别参考第二问的分段函数综合定价模型和第三问中的任务分配打包模型,基于附件三中的数据,给出新项目的两个模型,并参照附件一中已完成任务的完成度,对比完成情况。

3. 模型假设和符号说明

3.1 模型假设

- (1) 假设任务的定价与任务的难易程度无关。
- (2) 假设划分地区后某一地区内的任务只能由该地区的会员完成。
- (3) 假设任务完成度只与任务定价高低有关,不考虑会员的突发因素。
- (4) 每个会员都接受任务联合打包策略。
- (5) 假设样本数据真实有效,能够反映具体情况
- (6) 假设会员不会违反规定进行人为刷单

3.2 符号说明

符号	含义
D_{Gi}	第 <i>i</i> 个任务与与对应地区会员类中心的距离
P_i	第 <i>i</i> 个地区的线性拟合优度
S_i	第 <i>i</i> 个地区的会员总数
D_I	第 <i>i</i> 个地区的总任务量
δ	会员总数与相应总任务数的供求比
δ_{Gi}	第 <i>i</i> 个地区的供求比
k	前述模型比例系数
C_0	任务定价的基本价格
Q_{Gi}	某地区内会员任务预定限额
α	项目完成度

β	项目总额
φ	项目完成度与项目总额之比
t_i	任务定价和任务完成度之间的拟合优度

4. 问题一的解决（定价规律的探索）

“拍照赚钱”作为众包基础上的一种开放式创新，其成功与否与任务发布者的定价密切相关。合理的出价策略有利于任务发布者获取更好的解决方案、降低成本（时间成本和经济成本），也能使接受任务者获得满意的酬劳，从而提升任务的完成质量。研究表明任务的出价策略受到任务自身属性以及市场竞争性状况的影响，不同条件下的任务完成度相应的也有所不同。能否更好通过合理的指标，改善不同条件下的任务完成度关系着任务双方的切身利益。

4.1 定性描述

利用附件一中的数据，我们在地图中标出任务的位置，分别画出价格分布图和任务完成情况分布图，如图 4.1 和图 4.2 所示：

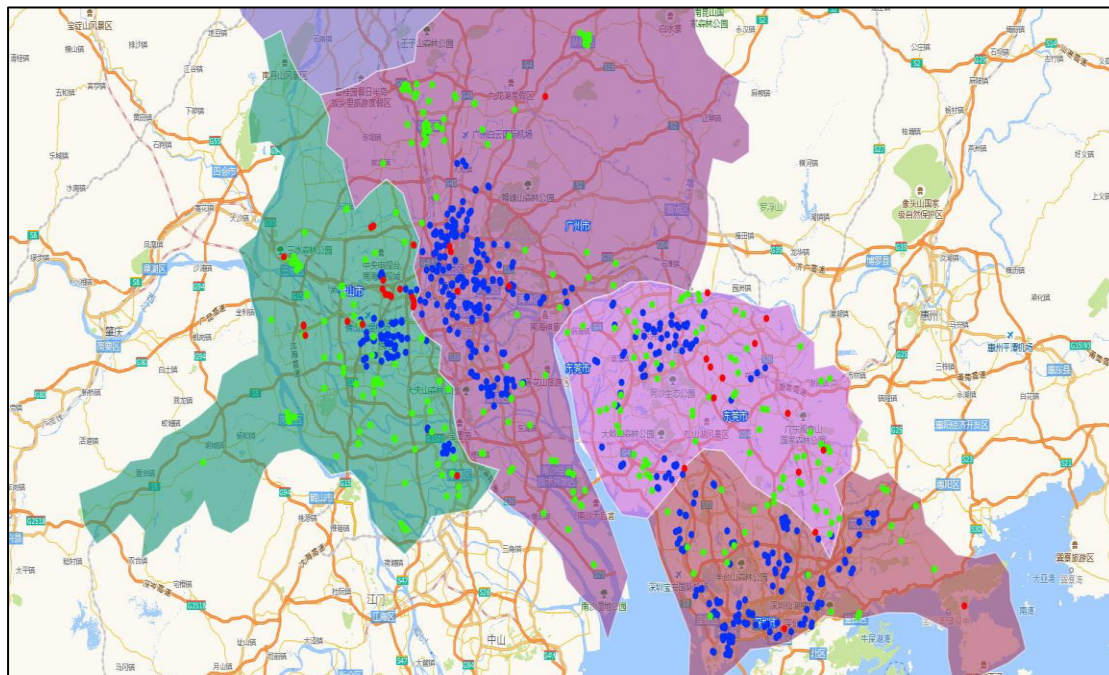


图 4.1 价格分布图（注：蓝色（65-69.5）绿色（70-75）红色（80 以上））

图中四个阴影地区依次是佛山市、广州市、东莞市、深圳市，仅从图中看出，四个地区的总体价格大致为佛山>东莞>深圳>广州，可以初步判断任务的定价与地区有关，因此在后续的定价规律研究中我们选择分地区进行研究。

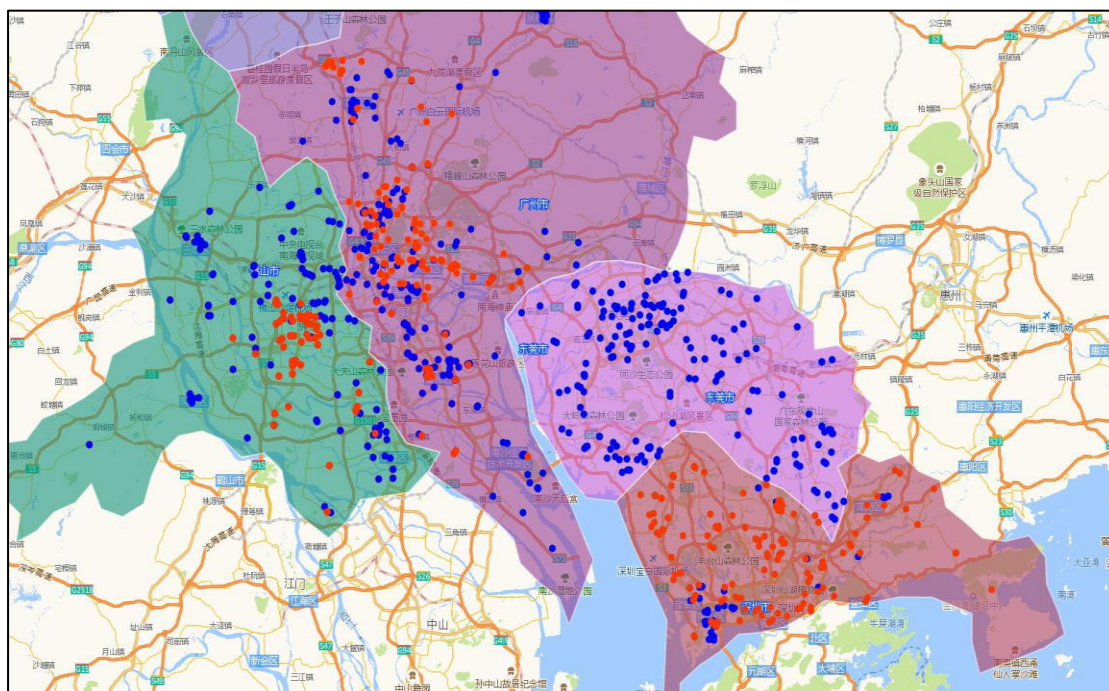


图 4.2 任务完成情况分布图（注：红色未完成，蓝色代表已完成）

上图反映的是任务的完成情况，也是地区定价规律的合理反映，图中大致可以观察出任务完成率的大小关系为东莞>佛山>广州>深圳。图中显示的地域差异较为明显，因此我们更加肯定的认为任务的定价规律应该分地区进行讨论。可知任务完成项目 522 项，任务未完成项目 313 项，任务完成率为 62.51%。绘制相关散点图，观察可发现，地理位置与任务完成度有着紧密的联系。未完成任务项目主要集中在两个区域，分别是纬度北纬 22.6-22.8°，经度东经 113.8-114.4°，和纬度北纬 22.8-23.2°，经度东经 113-113.4° 两个地区。在这样的情况下，如何合理分配任务资源，高效完成出行任务是每一个企业和任务接受者的需求。

4.2 基于距离因素的分地区定价规律讨论

鉴于经验，我们认为距离是进行任务定价的重要影响因素，因此推测附件一中的定价规律是距离（会员位置和任务位置之间的距离）和价格之间的函数关系式。

4.2.1 距离的定义

任务和会员均有数量大，分布广的特点，要想计算每个任务到会员之间的距离计算量较大，因此我们使用 K-均值聚类的方法将所有任务和会员分别聚为 4 类，聚类结果分别如表 4-1 和图 4-3、4-4 所示。

表 4-1 任务位置、会员位置聚类结果

	类别	聚类中心	数量
基于任务位置聚类结果	1	(22.6631, 114.0464)	196
	2	(22.9561, 113.7491)	173
	3	(23.2589, 113.3166)	229
	4	(22.9991, 113.1757)	237
基于会员位置聚类结果	1	(22.6418, 114.0707)	577
	2	(22.9943, 113.1032)	212
	3	(22.9119, 113.8321)	422
	4	(23.1644, 113.3335)	655

从任务位置的聚类结果和会员位置的聚类结果来看，两次聚类的类中心大致吻合，而且也呈现出明显的地区轮廓，从左到右依次可以看作佛山市、广州市、东莞市、深圳市。

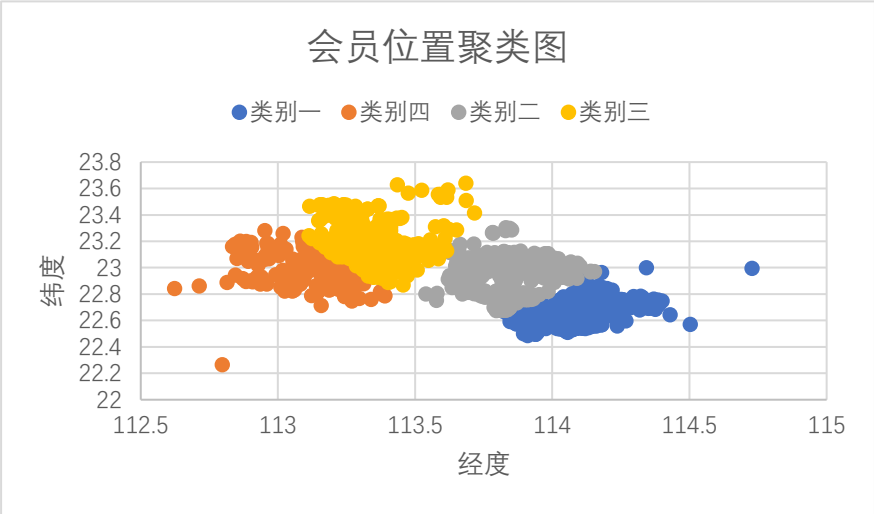


图 4-3 会员位置聚类图

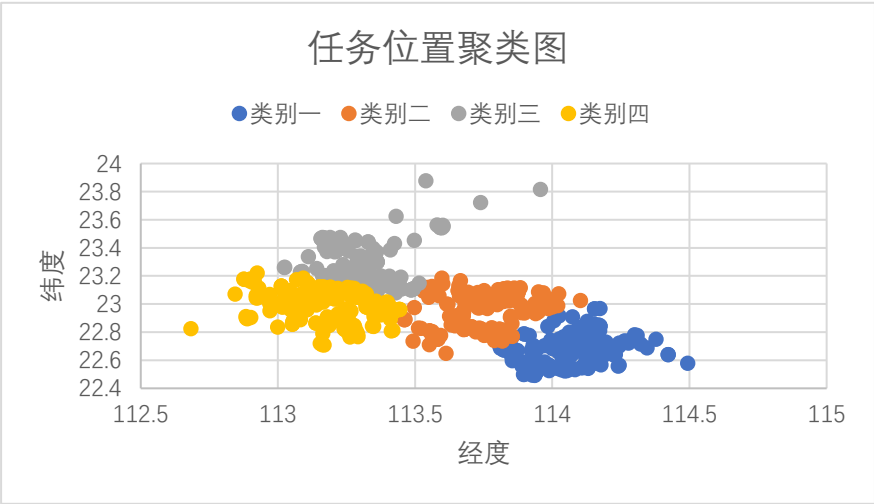


图 4-4 任务位置聚类图

考虑到地球是一个近乎标准的椭圆体，它的赤道半径为 6378.140 千米，极半径为 6356.755 千米，平均半径为 6371.004 千米。如果我们假设地球是一个完美的球体，那么它的半径就是地球的平均半径，记为 R 。如果以 0 度经线为基准，那么根据地球表面任意两点的经纬度就可以计算出这两点的地表距离（这里忽略地球表面地形对于计算带来的误差，仅仅是理论上的估算值）。

设某地区任意任务位置点 A_{Gi} 的经纬度为 $(LonA, LatA)$ ，该地区会员类中心的位置 B 的经纬度为 $(LonB, LatB)$ ，按照 0 度经线的基准，东经取经度的正值（Longitude），西经取经度的负值（-Longitude），北纬取 90+经纬值（90 - Latitude），南纬取 90+纬度值（90 + Latitude），则经过上述处理过后的两点被记为 $(MLonA, MLatA)$ 和 $(MLonB, MLatB)$ 。那么根据三角推导，可以得到某地区任意任务位置到会员位置之家的距离公式为：

$$C_{Gi} = \sin(MLatA_{Gi}) \times \sin(MLatB) \times \cos(MLonA_{Gi} - MlonB) + \cos(MLatA_{Gi}) \times \cos(MLatB) \quad (4.1)$$

$$D_{Gi} = \frac{R \times \text{Arcocos}(C_{Gi}) \times \pi}{180} \quad (4.2)$$

4.2.2 线性回归模型

我们假设该公司制定定价策略时仅考虑距离与价格之间的线性函数关系，利用分区数据，对方程进行拟合，结果如下。

广州地区：

$$P = 64.3 + 0.1815(d - 3.6) \quad (4.3)$$

$$R^2 = 0.501$$

佛山地区：

$$P = 64.1 + 0.1177(d - 2.47) \quad (4.4)$$

$$R^2 = 0.631$$

东莞地区：

$$P = 65.5 + 0.1125(d - 6.95) \quad (4.5)$$

$$R^2 = 0.702$$

深圳地区：

$$P = 64.7 + 0.1796(d - 2.85) \quad (4.6)$$

$$R^2 = 0.602$$

从拟合结果来看，各地区的定价方程拟合优度不高，均在 0.5-0.7 之间，这说明任务与会员之间的距离是影响价格制定的重要因素，但不是唯一因素，会员

信誉值、限定配额等同样可能影响价格的制定。单纯考虑距离的定价策略将容易造成资源的不合理配置，阻碍信息的传递。此外，结果显示不同城市的拟合优度差异有所不同，东莞地区的拟合优度为 0.702，而广州地区仅有 0.501，两地拟合优度差距较大，佛山地区和深圳地区拟合优度相近分别为 0.631 和 0.602. 对此可能的原因是，距离对价格的影响被供给削弱了。对企业而言，平台中会员数量越多对任务发布者越有利，任务发布者将可以用更低的价格来完成同样多的任务，并且可以适当减少对距离因素的考虑。通过观察可以发现，广东地区会员人数最多，任务相对少，供给最为充足，因而距离对价格的影响有所降低。东莞地区则正好相反，会员人数少，任务多，公司需要更多的考虑距离因素对任务完成度的影响。

5. 问题二的解决

问题一中所讨论的定价模型只是关于任务点与会员之间距离模型，并未考虑到市场的供求比以及会员的预定限额等因素，基于此，我们将供求比以及会员信誉因素纳入到定价模型中，对前述模型做出修正，从而建立分段函数综合定价模型。

在假设任务完成度只与任务定价高低并不考虑会员的突发因素的前提下，任务完成度是会员对价格意愿的直接反映，因此使项目完成度和项目总额达到均衡是定价策略的核心，在考虑实现任务高完成度的同时，也要兼顾项目实行总额，基于此种思想，我们建立以任务完成度最高和项目实行总额最小的多目标优化。

5.1 模型准备

1、距离

修正模型中的距离仍然使用附件一所研究的距离，在此不做赘述，详见公式 (4.1)、(4.2)。

2、供求比

供求比被看作是衡量供需平衡状态的指标，基于移动互联网的自助式劳务众包平台任务分配问题是一种典型的供需问题，因此，在定价时我们将此因素考虑在内。

通过查阅文献，供求比的公式可以定义为：

$$\text{供求比 } \delta = \frac{\text{一定时间内某地区会员总数 (S)}}{\text{相应的总任务数 (D)}} \times 100\% \quad (5.1)$$

$$\begin{cases} \text{当 } \delta > 1 \text{ 时, 供过于求, 此时市场呈现供过于求状态} \\ \text{当 } \delta < 1 \text{ 时, 供不应求, 此时市场呈现供不应求状态} \\ \text{当 } \delta = 1 \text{ 时, 供求平衡, 此时市场呈现供求平衡状态} \end{cases}$$

为了能够准确而形象地描述供求比和定价的关系, 我们利用指数函数来定义对二者之间的数量关系。对于每一地区 $G_i (i = 1, 2, 3, 4)$ 而言, 该地区内的所有会员为供应侧, 其总数记为供应量 $S_i (i = 1, 2, 3, 4)$ 所有发布的任务为需求侧, 总数为需求量 $D_i (i = 1, 2, 3, 4)$

每个地区的供求比表示为:

$$\delta_{Gi} = \frac{S_{Gi}}{D_{Gi}} \quad (i = 1, 2, 3, 4) \quad (5.2)$$

其中:

当 $\delta_{Gi} > 1$ 时, 该地区任务与会员配置不均匀, 呈现供过于求状态, 为保证任务完成度和成本最小两目标优化的目的, 此时可以采取适当降低任务价格的定价策略。

当 $\delta_{Gi} < 1$ 时, 该地区任务与会员配置不均匀, 呈现供不应求状态, 为保证任务完成度和成本最小两目标优化的目的, 此时可以采取适当提高任务价格的定价策略。

当 $\delta_{Gi} = 1$ 时, 该地区任务与会员配置大致均匀, 呈现供求平衡状态, 可以假设该种情况下的定价策略不须考虑供求状况。

综上所述, 某一时间内地区内的价格趋势应与供求比呈反比例函数关系。

3、会员质量

会员的任务限额参考会员的信誉给出, 并且原则上会员信誉越高, 越优先挑选任务, 在任务分配时根据预定限额所占比例进行配发。在实际建模过程中, 我们计算得出会员信誉值和会员任务限额之间的相关系数为 0.6570, 二者呈显著相关, 因此在考虑影响低价策略的因素时, 只对会员的预定任务限额比例做出讨论。其公式可定义为:

$$\theta_{Gi} = \frac{Q_{Gi}}{\sum_{i=1}^n Q_{Gi}} \quad (i = 1, 2, 3, 4) \quad (5.3)$$

其中, Q_{Gi} 为某个地区内的会员任务预定限额。

θ_{Gi} 能够反映出地区内会员总体信度情况, 会员总体信度对任务完成可以起

到积极作用，因此，对于总体信度高的地区可以采取提高价格的策略，对会员适当激励；对于总体信度不高的地区可以采取适当降低任务定价的策略，有效降低任务成本，进而提高项目的可行度。

综上所述，项目的定价与地区会员总体信度呈正向相关的关系，在此，我们用自然对数函数来量化任务定价和会员总体信度之间的数量关系。

基于此，可以给出距离定价的修正模型，即分段函数综合定价模型：

$$p = \begin{cases} C_0, & d \leq d_0 \\ C_0 + (k + e^{-\delta_{Gi}} - \ln \theta_{Gi}) \times (d - d_0), & d \geq d_0 \end{cases} \quad (5.4)$$

其中， p 为任务定价， k 为前述模型比例系数；

C_0 为基本价格， d 为某项任务到会员类中心的相对距离；

d_0 为相对基本距离， $(k + e^{-\delta_{Gi}} - \ln \theta_{Gi})$ 为模型的修正系数。

式中的 C_0 d_0 k 为待估参数，下面建立多目标优化模型求得最优参数。

5.2 多目标优化模型的建立

5.2.1 任务定价与完成度数量关系的确立：

在本问题中，得到任务定价与项目完成情况之间的数量关系是建立多目标优化模型的关键，任务完成情况是二分类变量，因此我们建立逻辑回归模型得到距离（任务与类中心）、任务定价和任务完成度之间的关系。拟合方程如下：

$$t_i = \frac{1}{1 + e^{-0.32d_0 - 0.014p_i}} \quad (5.5)$$

其中， d_0 表示距离， p_i 表示任务定价。

$$\left\{ \begin{array}{l} \max f_1 = \frac{1}{N} \sum_{i=1}^N t_i \\ \min f_2 = \sum_{i=1}^N p_i \\ t_i = \frac{1}{1 + e^{-0.32d_0 - 0.014p_i}} \\ p_i = \begin{cases} C_0, & d \leq d_0 \\ C_0 + (k + e^{-\delta_{Gi}} - \ln \theta_{Gi}) \times (d - d_0), & d \geq d_0 \end{cases} \end{array} \right. \quad (5.6)$$

$$\text{s.t.} \begin{cases} C_0 \geq 0 \\ d_0 \geq 0 \\ k \geq 0 \end{cases} \quad (5.7)$$

5.3 模型求解：

这是一个以 (C_0, d_0, k) 为优化变量的复杂优化问题，用传统的优化算法处理较难, 我们使用基于实数编码加速遗传算法对待估参数的最优值进行求解。

算法流程:

1) 编码。采用实数编码, 利用公式:

$$a(j) = c(j) + y(j)(d(j) - c(j)) (j = 1, 2, 3)$$

将初始变化区间为 $[c(j), d(j)]$ 区间的第 j 个优化变量 $a(j)$ 对应到 $[0, 1]$ 区间上的实数 $y(j)$ 。

2) 父代群体初始化。设群体规模为 n , 生成 n 组 $[0, 1]$ 区间上的均匀随机数

$$u(j, i) (j = 1 \sim 3, i = 1 \sim n)。$$

3) 适应度评价。

任务定价模型中的参数包括 C_0 , d_0 , k , 因此定义遗传个体为 (C_0, d_0, k) , 为区别多目标问题个体的优劣性, 个体的适应度函数采取权重加权法, 任意个体的多目标适应度函数如下:

$$F(x_i) = \lambda_1 \frac{f_{1,g} - f_1(x_i)}{|f_{1,g}|} + \lambda_2 \frac{f_2(x_i) - f_{1,g}}{|f_{1,g}|}$$

式中: $f_{i,g}$ 为单目标 f_i 的当前最优值; $f_i(x_i)$ 为个体 x_i 的单目标值, λ_i 为各单目标的权重, $\sum_{i=1}^3 \lambda_i = 1$

将 $u(j, i)$ 作为初始群体的父代个体值 $y(j, i)$ 带入实数编码公式(7), 再经公式得到父代群体的适应度函数值, 并将 $F(i)$ 按从大到小排序, 对应个体 $y(j, i)$ 也跟着排序, 称排序后前面的个体为优秀个体。

4) 选择操作。取比例选择方式, 父代个体 $y(j, i)$ 的选择概率 $ps(i)$ 为

$$ps(i) = \frac{F(i)}{\sum_{i=1}^n F(i)}$$

令 $p(i) = \sum_{k=1}^i ps(k)$, 序列 $\{p(i) | i = 1 \sim n\}$ 把 $[0, 1]$ 区间分成 n 个子区间:

$[0, p(i)], (p(1), p(2)], \dots, (p(n-1), p(n)]$, 这些子区间与 n 个父代个体 $\{y(j, i)\}$ 建立一一对应关系。生成 $n-m$ 个随机数 $\{u(k) | k = 1, 2, \dots, n-m\}$, 若 $u(k)$ 在 $(p(i-1), p(i)]$ 中, 则第 i 个个体 $y(j, i)$ 被选中, 即 $y_1(j, k) = y(j, i)$ 。这样从父代群体中选择 $n-m$ 个个体, 为增强RAGA全局优化搜索能力, 将 m 个父代优秀个体直接移入子代群体中, 得到第一个子代群体。

5) 交叉操作。根据(7)随机选择一对父代个体 $y(j, i1)$ 和 $y(j, i2)$ 作为双亲, 并进行如下随机线性组合, 产生一个子代个体 $y_2(j, i)$:

$$\begin{cases} y_2(j, i) = u_1 y(j, i1) + (1 - u_1) y(j, i2), u_3 < 0.5 \\ y_2(j, i) = u_2 y(j, i1) + (1 - u_2) y(j, i2), u_3 \geq 0.5 \end{cases}$$

式(9)中, u_1 、 u_2 、 u_3 都是随机数。

6) 变异操作。采用 p 个随机数以 $pm(i) = 1 - ps(i)$ 的概率来代替个体 $y(j, i)$, 从而得到子代个体 $y_3(j, i)$, $j = 1 \sim p$ 。即:

$$\begin{cases} y_3(j, i) = u(j), um < pm(i) \\ y_3(j, i) = y(j, i), um \geq pm(i) \end{cases}$$

式(10)中, $u(j)$ ($j = 1 \sim p$)、 um 均为随机数。

7) 迭代、加速遗传。由 4) -6) 步得到 $3n$ 个子代个体, 按投影函数值从大到小排序, 取前 n 个个体作为新的父代群体进行下一次迭代, 直至达到最大迭代数 I 。

在实际计算中, 设置参数为 $n = 300$, 优秀个体数 $m = 30$, 循环次数为 $I = 200$, 运行程序, 可得到最优参数估计值, 从而得到分地区分段综合定价模型:

广州地区:

$$p_1 = \begin{cases} 64.3021, d \leq 3.8 \\ 64.3021 + (0.1815 + e^{-0.5616} - \ln 0.2893) \times (d - 3.8), d > 3.8 \end{cases} \quad (5.8)$$

佛山地区:

$$p_2 = \begin{cases} 64.1214, d \leq 2.48 \\ 64.1214 + (0.1177 + e^{-0.2641} - \ln 0.4255) \times (d - 2.48), d > 2.48 \end{cases} \quad (5.9)$$

东莞地区:

$$p_3 = \begin{cases} 65, d \leq 6.95 \\ 65 + (0.1125 + e^{-1.0802} - \ln 0.0659) \times (d - 6.95), d > 6.95 \end{cases} \quad (5.10)$$

深圳地区:

$$p_4 = \begin{cases} 65.5, d \leq 2.85 \\ 65.5 + (0.1796 + e^{-0.3396} - \ln 0.2193) \times (d - 2.85), d > 2.85 \end{cases} \quad (5.11)$$

5.4 模型检验

检验指标:

项目性价比, 项目实行总额和项目完成度的比值, 从成本和完成满意度两个层次综合反映项目实施的意义, 更能反映出修正前后两个模型的优良性。

$$\varphi = \frac{\alpha}{\beta} \quad (5.12)$$

式中 α 代表项目完成度, β 代表项目总额。

基于求得最优结果, 得到未修正模型和修正后的模型检验指标如 5-1 所示:

表 5-1 模型修正前后检验指标

	地区	项目总额	项目完成率	项目性价比
未修正模型	广州城区	20632.5	65%	3.15×10^{-5}
	佛山地区	11564	67.5%	5.837×10^{-5}
	东莞地区	12105.5	95.38%	7.88×10^{-5}
	深圳地区	13405.5	36.2%	2.70×10^{-5}
	总计	57707.5	62.44%	1.082×10^{-5}
修正后模型	广州城区	23631.5	75.5%	3.19×10^{-5}
	佛山地区	12423.3	70.3%	5.65×10^{-5}
	东莞地区	11902.4	92.3%	7.75×10^{-5}

深圳地区	16702.6	60.1%	3.59×10^{-5}
总体	64659.8	74.64%	1.15×10^{-5}

5.5 结果分析:

检验结果表明, 实施方案后, 广州、深圳两个地区项目性价比提高, 佛山、东莞两个地区项目性价比降低; 总体项目性价比由 1.082×10^{-5} 提高到 1.15×10^{-5} , 实现了项目总额和完成度两个目标优化。我们将市场供求比和会员信誉度两个因素纳入到定价模型中, 较为充分的考虑了定价的影响因素, 分段函数综合定价模型取得较好的效果。

6. 问题三的求解（打包定价）

6.1 基于任务匹配度打包模型

6.1.1 模型准备

为了研究任务合并问题, 本文借助匹配的思想, 将任务打包问题转化为任务匹配。即通过计算目标任务与其他任务之间距离匹配度的高低, 来决定是否将任务进行合并。若两任务之间匹配度较高, 则将任务合并形成新的目标任务, 并依次与剩余任务进行匹配, 直至无法完成匹配, 将目标任务进行打包。

通过分析和借鉴已有研究, 引入图论中的概念, 将任务与任务之间的距离匹配用图进行表达, 提出基于抽象方法的任务匹配算法。

（1）设定相关术语和符号

用带权图 $G = (V, E, w)$ 抽象的表示任务间的距离, 式中 V 为结点的集合, E 代表边的集合, w 表示边的权重集合, $w(i, j)$ 表示从任务 i 到任务 j 之间的距离。选择待匹配的目标任务 $o \in V$ 作为匹配路径的源点, 任务 $d \in V$ 表示匹配路径的目的点。 O 和 D 分别表示与 o 和 d 距离接近的点的集合, 将 O 称为源团, D 称为目的团。团实质上就是带权图 G 的一个顶点子集, 在这个子集中, 距中心点 o 或 d 距离最大的结点与中心点间的距离称为团的半径。

（2）对任务进行筛选

由于任务数量较大, 为了避免将所有任务进行遍历, 计算目标任务与其余任务匹配度的复杂过程, 提高运算的效率, 缩短匹配时间。在进行任务匹配之前, 先对任务进行筛选。此处定义一个目标半径, 以目标任务为圆心, 利用该目标半径画圆。将位于目标圆外的距离较远的任务剔除, 不参与匹配度的计算。

6.1.2 模型建立

(1) 匹配度的计算

本问中匹配度用来衡量两任务是否应该合并问题。将一定数值范围内匹配度高的任务作为目标任务的匹配对象，同目标任务进行合并，打包发布。

充分考虑任务之间的相关性，将打包任务距离与单个任务距离作为匹配因子，将打包任务距离与多任务单独完成距离总和的比值作为匹配度，计算公式为

$$f_i = \begin{cases} \max \frac{l_i + l_j}{l_{ij}} \times Q_I & m = 2 \quad i, j \in N \text{ 且 } i \neq j \\ \max \frac{l_i + l_k + l_j}{l_{ikj}} \times Q_I & m = 3 \quad i, k, j \in N \text{ 且 } i \neq j \neq k \end{cases} \quad (6.1)$$

其中： f_i 表示第 i 个任务的匹配度；

N 为待选乘客集合；

i, j, k 表示任务编号；

m 表示任务包中任务数量；

d_{ij}, d_{ikj} 表示 m 个任务合并后状态下打包任务距离；

d_i 表示第 i 个任务单独完成的距离。

(2) 粗糙处理

为了简化计算流程，对打包任务距离 d_{ij}, d_{ikj} 进行粗糙处理，如下：

$$\begin{cases} l_{ij} = o_i o_j + d_i d_j + o_i d_j & m = 2 \\ l_{ij} = o_i o_k + o_i o_j + d_i d_k + o_i d_i + d_i d_j & m = 3 \end{cases} \quad (6.2)$$

粗糙处理虽然存在一定范围内的误差，但仍保持匹配结果的参考价值。

(3) 匹配度的取值范围

当匹配度等于 1 时，说明任务合并后总距离与任务未合并距离的和相同。此时，任务失去合并意义，不对任务进行合并。

当匹配度等于 0 时，说明目标任务附近未找到与它较临近的点，没有打包任务。

当匹配度位于 0 到 1 之间，说明任务合并后距离较单独完成时减小。因此，匹配度越接近 0，表明匹配效果越好。通过数据分析，我们选取匹配度在 0-0.3 之间的任务进行合并。

(4) 目标半径的确定

对于目标半径的确定，是以目标任务的位置为圆心，以不同任务间的距离为半径，

通过圆域实现有效邻点的检索。取值过大, 则圆域范围过大, 造成计算的繁琐与复杂。取值过小, 则会缩小匹配范围, 使得匹配度下降, 难以完车任务合并。因此, 在取值过程中要经过不断地计算和比较, 确定半径范围。

6.1.3 模型求解

该模型算法流程如下:

<p>STEP1: 初始化参数源点 o, 目的点 d 及定义 $dis(x,y)$ 距离函数。</p> <p>STEP2 : 确定数组 N 中的记录点数 n 值, 若 $n \geq 2$, 令 $i=1, j=2$, 计算匹配度:</p> $F_0 = \frac{L_i O_j}{L_i} + L_0 + L_j = \frac{dis(oL, d1) + dis(o, d) + dis(o2, d2)}{dis(o, o1) + dis(o, o2) + dis(d, d1) + dis(d, d2) + dis(o, d)} \times Q$ <p>STEP3: 对 $i, j = 1 \dots n$ 且 $i \neq j$, 遍历组合, 求得:</p> $F = \frac{dis(oi, di) + dis(o, d) + dis(oj, dj)}{dis(o, o1) + dis(o, o2) + dis(d, d1) + dis(d, d2) + dis(o, d)} \times Q$ <p>STEP4: 若 $F_{max} < F$, 则令 $F_{max} = F$, 将 F_{max} 输出</p> <p>STEP5: 若 $n < 2$, 且 $n = 1$, 则计算匹配度</p> $F_{max} = \frac{L10}{L0} + L1 = \frac{dis(o, d) + dis(o1, d1)}{dis(o, o1) + dis(d, d1) + dis(o, d)} \times Q$ <p>并转至第三步, 否则令 $F_{max} = 0$ 并输出。</p>

6.2 基于公平性费率分摊模型

6.2.1 模型准备

打包任务费率设计原则:

(1) 打包发布必须达到双赢的目的。即从用户的角度, 打包后任务价格必须较单独完成任务时有所增加, 且满足一定期望。从任务发布者角度, 打包后任务价格应小于单独任务价格之和, 使得竞争率下降, 成本降低。

(2) 科学合理, 充分考虑任务与会员的位置分布, 会员信誉值以及预定限额等因素。

(3) 费率等级明确, 规则简单易懂。

6.2.2 模型建立:

将任务数作为分类对象, 按照问题二中设置的定价模型作为标准, 在任务打包情况下对每个任务进行独立的费用折减。

(1) 单独任务定价公式:

$$p = \begin{cases} C_0 \times A_1\% , d \leq d_0 \\ C_0 \times A_1\% + (k + e^{-\delta G_i} - \ln \theta_{Gi}) \times (d - d_0) \times A_1\% , d > d_0 \end{cases} \quad (6.3)$$

其中: p 为任务定价;

k 为前述模型比例系数;

C_0 为基本价格；

d 为某项任务到会员类中心的相对距离；

d_0 为相对基本距离；

$(k + e^{-\delta_{Gi}} - \ln\theta_{Gi})$ 为模型的修正系数。

(2) 任务打包定价公式：

$$p_n' = \begin{cases} \sum_{i=1}^n \frac{C_0}{d_0} \times d_{0i} \times A_i\% , \sum_{i=1}^n d_{0i} \leq d_0 \\ \sum_{i=1}^n \left[\frac{C_0}{d_0} \times d_{0i} + (k + e^{-\delta_{Gi}} - \ln\theta_{Gi}) \times d_i \right] \times A_i\% , d > d_0 \end{cases} \quad (6.4)$$

其中： p_n' 表示每个任务包的定价；

n 表示任务包中任务数目；

A_i 表示比例因子；

d_{0i} 表示在 i 状态下，在相对基本距离内的距离；

d_i 表示在 i 状态下，超出相对基本距离范围的距离；

(3) 比例系数的确定：

①任务包中有一个任务。则比例系数 $A_1=100\%$ ；

②任务包中有两个任务。若两个任务的费率太高，如 90%，则对于用户来说，吸引率较大，争相选择没有起到改善作用；若两任务的费率太低，如 50%，则对于用户来说，没有达到他们的期望价格，可能导致任务的完成度不高。当比例系数 $A_2=75\%$ 时，吸引度适中，容易达成目标。

③任务包中有三个任务。完成此任务包，对于任务发布者来说，可以降低 40% 的成本输出；对于用户来讲，也可以获得相对于单个任务来说 1.8 倍的收益，双方达到共赢。此时，可取 $A_3=60\%$ 。

6.2.3 模型求解：

在这里，我们仍然分地区，即：广州、佛山、东莞、深圳四个地区进行定价讨论，使用基于项目总额和任务完成度两个目标的多目标优化模型进行待估参数的最优值求解。

(1) 广州地区

单任务定价模型：

$$p_1 = \begin{cases} 64.3021 & d \leq 3.8 \\ 64.3021 + (0.1815 + e^{-0.5416} - \ln 0.2893) \times (d - 3.8), & d > 3.8 \end{cases} \quad (6.5)$$

据此得到广州地区的多任务定价模型为：

$$p_n' = \begin{cases} \sum_{i=1}^n 16.9216 \times d_{0i} \times A_i\% & \sum_{i=1}^n d_{0i} \leq 3.8 \\ \sum_{i=1}^n [16.9216 \times d_{0i} + (0.1815 + e^{-0.2301} - \ln 0.2893) \times d_i] \times A_i\%, & \sum_{i=1}^n d_{0i} > 3.8 \end{cases} \quad (6.6)$$

式中： $A_1 = 100\%$, $A_2 = 75\%$, $A_3 = 60\%$ 。

(2) 佛山地区 u

单任务定价模型：

$$p = \begin{cases} 64.1214, d \leq 2.48 \\ 64.1214 + (0.1177 + e^{-0.2641} - \ln 0.4255) \times (d - 2.48), d \geq 2.48 \end{cases} \quad (6.7)$$

据此得到佛山地区的多任务定价模型为：

$$p_n' = \begin{cases} \sum_{i=1}^n 25.8554 \times d_{0i} \times A_i\% & \sum_{i=1}^n d_{0i} \leq 2.48 \\ \sum_{i=1}^n [25.8554 \times d_{0i} + (0.1177 + e^{-0.2531} - \ln 0.4255) \times d_i] \times A_i\%, \sum_{i=1}^n d_{0i} > 2.48 \end{cases} \quad (6.8)$$

式中： $A_1 = 100\%$, $A_2 = 75\%$, $A_3 = 60\%$ 。

(3) 东莞地区

单任务定价模型：

$$p = \begin{cases} 65, d \leq 6.95 \\ 65 + (0.1125 + e^{-1.0802} - \ln 0.0659) \times (d - 6.95), d \geq 6.95 \end{cases} \quad (6.9)$$

据此得到东莞地区的多任务定价模型为：

$$p_n' = \begin{cases} \sum_{i=1}^n 9.3525 \times d_{0i} \times A_i\% & \sum_{i=1}^n d_{0i} \leq 6.95 \\ \sum_{i=1}^n [9.3525 \times d_{0i} + (0.1125 + e^{-0.9874} - \ln 0.2893) \times d_i] \times A_i\%, \sum_{i=1}^n d_{0i} > 6.95 \end{cases} \quad (6.10)$$

式中： $A_1 = 100\%$, $A_2 = 75\%$, $A_3 = 60\%$ 。

(4) 深圳地区

单任务定价模型：

$$p = \begin{cases} 65.5, d \leq 2.85 \\ 65.5 + (0.1796 + e^{-0.3396} - \ln 0.2193) \times (d - 2.85), d \geq 2.85 \end{cases} \quad (6.11)$$

据此得到深圳地区的多任务定价模型为：

$$p_n' = \begin{cases} \sum_{i=1}^n 22.9824 \times d_{0i} \times A_i\% & \sum_{i=1}^n d_{0i} \leq 2.85 \\ \sum_{i=1}^n [22.9824 \times d_{0i} + (0.1815 + e^{-0.3089} - \ln 0.2193) \times d_i] \times A_i\%, \sum_{i=1}^n d_{0i} > 2.85 \end{cases} \quad (6.12)$$

式中： $A_1 = 100\%$, $A_2 = 75\%$, $A_3 = 60\%$ 。

6.3 打包方案模型检验

基于上述两种模型(基于任务匹配度打包模型和基于公平性费率分摊模型)的建立,我们实现了对任务集中地区的联合打包任务,并且给出定价方案,与前述两种方案作出对比如表 6-1 所示:

表 6-1 模型检验指标对照表

	地区	项目总额	项目完成率	项目性价比
未修正模型	广州城区	20632.5	65%	3.15×10^{-5}
	佛山地区	11564	67.5%	5.837×10^{-5}
	东莞地区	12105.5	95.38%	7.88×10^{-5}
	深圳地区	13405.5	36.2%	2.70×10^{-5}
	总计	57707.5	62.44%	1.082×10^{-5}
分段综合定价模型	广州城区	23631.5	75.5%	3.19×10^{-5}
	佛山地区	12423.3	70.3%	5.65×10^{-5}
	东莞地区	11902.4	92.3%	7.75×10^{-5}
	深圳地区	16702.6	60.1%	3.59×10^{-5}
	总体	64659.8	74.64%	1.15×10^{-5}
打包定价模型	广州城区	21051.6	76.1%	3.61×10^{-5}
	佛山地区	12006.2	70.9%	5.91×10^{-5}
	东莞地区	11813.2	87.43%	7.40×10^{-5}
	深圳地区	14521.3	59.5%	4.10×10^{-5}
	总体	59392.3	73.31%	1.23×10^{-5}

检验结果表明：打包定价模型实施下项目总体性价比为 1.23×10^{-5} ，该结果要更加优于分段综合定价模型；具体到地区而言，该方案实施后项目后广州、佛山、深圳三地区项目性价比明显比分段综合定价模型的结果，而东莞地区项目性价比明显降低。原因在于广州、佛山、深圳三个地区的任务密度大，而东莞地区任务密度较小，该结果验证了打包定价模型更加适应任务分布稠密的地区。

7 问题四的解决（新任务定价策略）

7.1 定价方案准备

对于附件三给出的数据，我们首先将其描绘在地图中，得到任务分布情况，如图 7-1 所示：

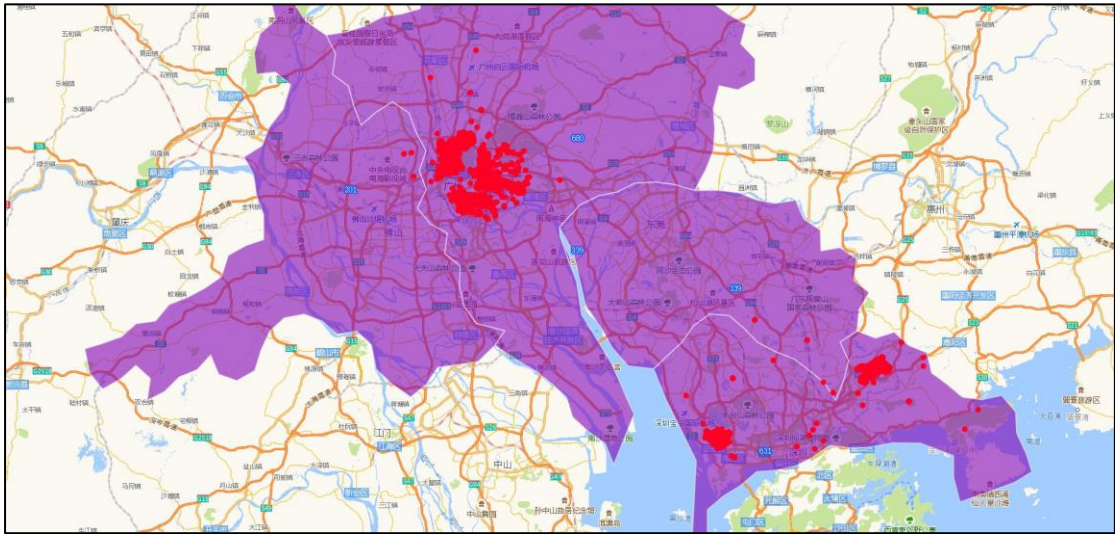


图 7.1 附件三任务分布图

从图中可以看出，任务分布主要集中在广州城区和深圳地区，而且分布较

为密集，因此我们根据问题二和问题三中分别建立的分段综合定价模型和打包定价模型研究其定价问题，根据模型检验结果最终确定新任务的定价方案。

7.2 分段综合定价方案

根据 4.4 中确立的广州城区分段综合定价模型，为：

$$p = \begin{cases} 64.3021, d \leq 3.8 \\ 64.3021 + (0.1815 + e^{-0.5616} - \ln 0.2893) \times (d - 3.8), d \geq 3.8 \end{cases}$$

深圳地区分段综合定价模型：

$$p = \begin{cases} 65.5, d \leq 2.85 \\ 65.5 + (0.1796 + e^{-0.3396} - \ln 0.2193) \times (d - 2.85), d \geq 2.85 \end{cases}$$

计算附件三中给出的任务的定价，部分结果如表 7-1 所示：

表 7-1 分段综合定价方案定价方案

任务号码	任务 GPS 经度	任务 GPS 纬度	价格
C0001	114.2408795	22.73004117	65
C0002	114.2996199	22.72704287	64.3
C0003	114.2336007	22.70131065	72.6
C0004	114.2866672	22.73235925	68.9
C0005	114.2575495	22.71839144	67.5
C0006	114.3819253	22.75392493	84.3
C0007	114.2721836	22.72404221	67.5
C0008	114.2732478	22.71937803	68.5
C0009	114.2304955	22.73028254	72.4
C0010	114.267027	22.7187968	73.6

7.3 打包定价方案

根据 6.1 基于任务匹配度打包模型和 6.2 基于公平性费率分摊模型，即广州城区多任务打包定价模型为：

$$p_n' = \begin{cases} \sum_{i=1}^n 16.9216 \times d_{0i} \times A_i\% & \sum_{i=1}^n d_{0i} \leq 3.8 \\ \sum_{i=1}^n [16.9216 \times d_{0i} + (0.1815 + e^{-0.2301} - \ln 0.2893) \times d_i] \times A_i\%, & \sum_{i=1}^n d_{0i} > 3.8 \end{cases}$$

式中： $A_1 = 100\%$ ， $A_2 = 75\%$ ， $A_2 = 60\%$ 。

深圳市多任务定价模型：

$$p_n' = \begin{cases} \sum_{i=1}^n 22.9824 \times d_{0i} \times A_i\% & \sum_{i=1}^n d_{0i} \leq 2.85 \\ \sum_{i=1}^n [22.9824 \times d_{0i} + (0.1815 + e^{-0.3089} - \ln 0.2193) \times d_i] \times A_i\%, & \sum_{i=1}^n d_{0i} > 2.85 \end{cases}$$

式中： $A_1 = 100\%$ ， $A_2 = 75\%$ ， $A_2 = 60\%$ 。

基于此，附件三中给出的任务的定价，部分结果如表 7-2 所示：

表 7-2 打包定价模型定价方案

任务号码	任务 GPS 经度	任务 GPS 纬度	价格
C0001	114.2408795	22.73004117	63.5

C0002	114.2996199	22.72704287	61.2
C0003	114.2336007	22.70131065	65.8
C0004	114.2866672	22.73235925	68.7
C0005	114.2575495	22.71839144	66.8
C0006	114.3819253	22.75392493	82.1
C0007	114.2721836	22.72404221	67.0
C0008	114.2732478	22.71937803	68.0
C0009	114.2304955	22.73028254	69.5
C0010	114.267027	22.7187968	72.0

7.4 模型检验:

在两种定价模型下,我们分别给出附件三中的每项任务的定价,在此基础上进行模型验证,如表 7-3 所示:

表 7-3 模型检验指标:

		项目总额	项目完成度	项目性价比
分段综合定价 模型	广州城区	103761.8	72.7%	7.006×10^{-6}
	深圳地区	20198.6	68.3%	3.381×10^{-5}
	总体	123960.4	65.8%	5.308×10^{-6}
打包定价模型	广州城区	95674.4	71.4%	7.462×10^{-6}
	深圳地区	19754.3	61.2%	3.09×10^{-5}
	总体	115428.7	66.8%	5.78×10^{-6}

从检验结果来看,因为广州城区的任务密度较大,所以基于打包定价模型下的项目定价方案要明显优于分段综合定价模型下的定价方案;深圳地区任务密度较小,分段综合定价模型下的定价方案要明显优于打包定价模型。因此对于附件三中的任务定价,我们首先将其分为广州和深圳两地区,广州城区使用打包定价模型定价,深圳地区使用分段综合定价模型定价。

8、模型评价

8.1 模型优点

(1) 建立的分段函数综合定价模型,实用价值高,具有一定创新性。使用 K-均值聚类法对任务位置、会员位置进行聚类,分广州、佛山、东莞、深圳四个地区进行回归拟合,更贴近客观实际,更为具体。

(2) 问题二增添了市场供求率和会员信誉值两个指标,对原模型进行了修正,提高原模型的实用价值与真实性。

(3) 将任务合并问题转化为基于任务匹配的打包模型,具有创意,且误差较小。建立的打包任务费率计算模型,可以进行推广,以用于其他组合类产品的定价。

8.2 模型的缺点

(1) 由于时间限制,问题二中的定价模型中用会员信誉度代表了预定任务限额和预定任务开始时间对于任务定价的影响。未对后两个因素进行详细分析。

(2) 在问题三中的任务打包问题中,未用会员的可预定限额条件对任务包中任务数目进行限制,可能产生误差。

9、模型的改进与推广

9.1 模型的改进

模型没有对预定任务限额与预定任务开始时间对于定价的影响进行详细分析，可以将这两个因素考虑进去，对模型进行改进。

9.2 模型的推广

本文是对拍照任务的定价问题进行分析，可以推广到其他任务或者出租车等服务行业的定价问题中，并加以改进，具有较强的实际意义和使用价值。

参考文献

- [1]王千,王成,冯振元,叶金凤. K-means 聚类算法研究综述[J]. 电子设计工程,2012
- [2]雷小锋,谢昆青,林帆,夏征义. 一种基于 K-Means 局部最优性的高效聚类算法[J]. 软件学报,2008
- [3]何清,李宁,罗文娟,史忠植. 大数据下的机器学习算法综述[J]. 模式识别与人工智能,2014
- [4]黄丽. BP 神经网络算法改进及应用研究[D]. 重庆师范大学,2008
- [5]刘帅. 量化投资:若干金融衍生品的定价模型及投资策略研究[D]. 上海大学,2016.
- [6]张利花. 路径依赖型期权定价模型和方法研究[D]. 华南理工大学,2013.
- [7]Min-Jeong Kim. A field study using an adaptive in-house pricing model for commercial and industrial customers in Korea[J]. Energy Policy,2017
- [8]Fatemeh Karandish, Šimůnek. A comparison of numerical and machine-learning modeling of soil water content with limited input data[J]. Journal of Hydrology,2016,:
- [9]Paulo Fernandez,Sandra Mourato,Madalena Moreira,Luís Pereira. A new approach for computing a flood vulnerability index using cluster analysis[J]. Physics and Chemistry of the Earth,2016,:
- [10]Ye Young Lee,Jeong Tai Kim,Geun Young Yun. The neural network predictive model for heat island intensity in Seoul[J]. Energy & Buildings,2016,110:.
- [11]Yacine Hammami,Faouzi Jilani. Testing Factor Pricing Models in Tunisia: Macroeconomic Factors vs. Fundamental Factors[J]. Review of Middle East Economics and Finance,2012,7(2):

附录:

1、R 代码

```
## 聚类
url1 <- "C:\\Users\\zhouhao1\\Desktop\\document1.csv"
data1<- read.csv(url1,header = TRUE, sep = ",", stringsAsFactors = FALSE)
data11 <-data1[,2:3]
url2 <- "C:\\Users\\zhouhao1\\Desktop\\document2.csv"
data2<- read.csv(url2,header = TRUE, sep = ",", stringsAsFactors = FALSE)
data21 <- data2[,c(2:3)]
library(fpc)
set.seed(252964) # 设置随机值, 为了得到一致结果。
km1 <- kmeans(na.omit(data11), 4)# 显示 K-均值聚类结果
plotcluster(na.omit(data11), km1$cluster) # 生成聚类图
a <- km1$centers
m <- km1$cluster
clu1 <- data1[which(m==1),]
clu2 <- data1[which(m==2),]
clu3 <- data1[which(m==3),]
clu4 <- data1[which(m==4),]
library(fpc)
set.seed(252964) # 设置随机值, 为了得到一致结果。
km2 <- kmeans(na.omit(data21), 4)# 显示 K-均值聚类结果
plotcluster(na.omit(data21), km2$cluster) # 生成聚类图
a1 <- km2$centers
m1 <- km2$cluster
clu11 <- data2[which(m1==1),]
clu22 <- data2[which(m1==2),]
clu33 <- data2[which(m1==3),]
clu44 <- data2[which(m1==4),]

## 计算供求比
D <- c(nrow(clu1),nrow(clu2),nrow(clu3),nrow(clu4))
S <- c(nrow(clu11),nrow(clu44),nrow(clu22),nrow(clu33))
Y <-D/S

## 计算距离
rshen <- clu1
rfo <- clu2
rdong <- clu3
rguang <- clu4
hshen <- clu11
hfo <- clu44
hdong <- clu22
```

```

hguang <- clu33

b1 <- as.matrix(a1[3,])
q1 <- rep(b1, each = 1, times = 237)
d1 <- matrix(q1, 237, 2, byrow = TRUE)
c1 <- as.matrix(rguang[, c(2:3)])
e1 <- c1-d1
di1 <- sqrt(rowSums(e1*e1))
dis1 <- matrix(di1, 237, 1)
diss1 <- data.frame(dis1)
write.csv(diss1, "C:\\Users\\zhouhao1\\Desktop\\guangzhou.csv")

b2 <- as.matrix(a1[4,])
q2 <- rep(b1, each = 1, times = 173)
d2 <- matrix(q2, 173, 2, byrow = TRUE)
c2 <- as.matrix(rfo[, c(2:3)])
e2 <- c2-d2
di2 <- sqrt(rowSums(e2*e2))
dis2 <- matrix(di2, 173, 1)
diss2 <- data.frame(dis2)
write.csv(diss2, "C:\\Users\\zhouhao1\\Desktop\\foshan.csv")

b3 <- as.matrix(a1[2,])
q3 <- rep(b3, each = 1, times = 229)
d3 <- matrix(q3, 229, 2, byrow = TRUE)
c3 <- as.matrix(rdong[, c(2:3)])
e3 <- c3-d3
di3 <- sqrt(rowSums(e3*e3))
dis3 <- matrix(di3, 229, 1)
diss3 <- data.frame(dis3)
write.csv(diss3, "C:\\Users\\zhouhao1\\Desktop\\dongguan.csv")

b4 <- as.matrix(a1[1,])
q4 <- rep(b4, each = 1, times = 196)
d4 <- matrix(q4, 196, 2, byrow = TRUE)
c4 <- as.matrix(rshen[, c(2:3)])
e4 <- c4-d4
di4 <- sqrt(rowSums(e4*e4))
dis4 <- matrix(di4, 196, 1)
diss4 <- data.frame(dis4)
write.csv(diss4, "C:\\Users\\zhouhao1\\Desktop\\shenzhen.csv")

```

2、MATLAB 代码

```

popsize = 200; % 种群数量
variablecount = 3; % 变量个数
elitecount = 15; % 精英数量
pop0 = initial(popsize, variablecount); % 产生初始种群
F = fitness([pop0], popsize, DD); % 计算适应度

```

```
[Fnew pop] = paixu([F],[pop0]); % 排序

for i=1:100
newpop1 = selection(pop,Fnew,popsize,elitecount); % 选择 产生第一代
newpop2 = crossover(pop,Fnew); % 交叉 产生第二代
newpop3 = mutation(pop,Fnew,popsize,variablecount); % 变异产生第三代
newpop = [newpop1,newpop2,newpop3];
Fnew1 = fitness(newpop,3*popsize,DD);
[bestfit0 bestvidual0] = best( Fnew1,newpop );
bestfit(i) = bestfit0;
bestvidual(:,i) = bestvidual0;
Fnew = Fnew1(1:popsize);
pop = newpop(:,1:popsize);
t=i
end
```