# 2017年夏季学期数学建模实践课程论文

**题目：** “拍照赚钱”任务定价模型

**作者信息：**

| 编号 | 学号 | 姓名 | 专业年级 | 分工 |
|---|---|---|---|---|
| 1 | 15020041044 | 牛小宇 | 海洋技术 2015 级 | 模型论文 |
| 2 | 15020041022 | 李晨光 | 海洋技术 2015 级 | 编程计算 |
| 3 | 15160021047 | 张寅丁 | 信息与计算科学 2015 级 | 模型算法 |

**评分指标及分值分布说明：**

| 指标 | 摘要 | 写作及整体 | 模型 | 解法及结果 | 分析检验 | 总分 |
|---|---|---|---|---|---|---|
| 分布 | 10 分 | 20 分 | 30 分 | 30 分 | 10 分 | 100 分 |
| 得分 | | | | | | |

# “拍照赚钱”的任务定价模型


## 摘要

“拍照赚钱”是一定互联网下的一种自助式服务模式。APP 平台的本质是众包情形下的劳动力中介。在整个的运营模式中，任务定价无疑是其核心要素。所以，本文旨在研究分析原始定价方案的规律，并在此基础上设计新的定价方案，并且给到方案效果的评价，最后带入新项目来实际检验方案的实施效果。

针对问题一，从定性和定量两个角度来研究分析任务的定价规律。首先，对附件一的数据进行三次样条法插值，得到任务标价随位置变化的空间三维曲面。从定性分析的角度，总结了任务标价随地理位置的变化规律然后，通过 K-Means 聚类分析来缩小研究范围进行定量研究。K-Means 聚类分析共划分了九个区域，聚类结果恰好反映了任务点所在的行政区分布。然后，针对每个区域选定任务数、会员信誉度、预定任务限额、会员数、众包劳动力供需比等五个指标进行多元线性回归分析，得到拟合函数关系，并检验其拟合效果，不同区域。最后，为了提高拟合精度更加接近于实际，采用主成分回归分析，在对指标进行评价分类后进行拟合，此时的结果更加准确，发现不同区域的任务定价规律不同。通过分析发现东莞市任务完成度很高，广州市及其周围价格完成度一般，深圳市任务完成度很低，我们认为任务未完成的主要原因是任务标价与当地的劳动力价格不匹配。

针对问题二，通过问题一的分析，我们认为主要的改进方向是使任务标价与当地的劳动力价格相匹配，查询相关资料后，我们使用了一套通过人口及经济发展指标来确定劳动力价格指数的模型，再通过劳动力价格匹配较好地区的任务标价数据找到劳动力价格指数与适合任务标价的关系。我们选取了任务价格、任务密度、会员密度、任务会员距离、劳动力供给需求比等指标，利用附件一的数据训练了支持向量机来预测在新定价下任务是否会被完成，支持向量机对测试集的判断成功率可达 78%以上，我们利用支持向量机判断了聚类分析划分出的一块区域在新定价下的任务完成度，预测任务完成度从原方案的 35.57%提高到了 56.7%，这是论文的亮点之一。

针对问题三，对任务进行打包，分打包方法模型和打包价格模型。对于打包方法模型，我们在 Q 型聚类算法的基础上进行改进，在聚类的过程中加入判断，使其为任务进行打包，这是论文的亮点之二。对于打包价格模型，通过分析，任务位置集中，使得用户完成任务方便省力，导致用户争相选择。这反映了 1.工作劳动量下降带来任务价值下降.2.劳动力供求关系供大于求。

对于 app 平台来说，打包以后将使得任务均价下降

对于用户来说之所以能够接受任务均价下降是由于节省了路途成本，所以任务价格下降=用户节省的路途成本。

针对问题四，我们使用问题三的任务打包模型对附件 4 的项目进行打包，利用已有的含打包情形的定价方法对其定价并用支持向量机模拟。



关键词：

众包定价　商品价值模型　K-Means 聚类分析　主成分分析　改进的 Q 聚类分析

# 一、 问题重述

## 1.1 问题背景

随着互联网的不断发展，网络对于人们的生活产生了越来越大的影响。目前网络甚至给大众提供了能够直接赚钱的方法，拍照赚钱 APP 就是一种新兴起来的，提供给大众的用于直接赚钱的网络工具。这类 APP 的代表有高德公司的高德淘金、百度公司的地图淘金以及拍吧科技公司的拍拍赚等，这类 APP 的出现不但为各种商业检查以及消费者调研提供了兼职劳务，而且助力于 o2o 大潮中的数据采集和产品推广，顺应了大数据时代的到来。与传统的检查、调研方式相比，此类基于智能手机应用所开发出的众包模式有着数据真实度高、数据采集机动性强、投入成本较低、覆盖范围广等优势。因此，越来越多的互联网公司，尤其是涉及到地图定位的公司纷纷参与到此类模式中来。但是，如何提高数据采集的完成率以及降低总的成本，依然需要进一步的理论研究和实践验证。

## 1.2 问题提出

APP 成为该平台运行的核心，而 APP 中的人物定价又是其核心要素。如果定价不合理，有的任务就会无人问津，从而导致商品检查或者信息获取的失败。

所以，本文将以“任务标价”为主线，进行一下问题的研究：

1. 研究分析附件一的定价规律，并分析任务未完成的原因。
2. 为附件一的项目设计新的定价方案，并和原方案进行比较。
3. 在某些位置任务可能会分布比较集中，这会导致用户的争相选择。在这种情况下，应如何将打包发布考虑进定价模型？这样新的定价模型又会对任务完成情况产生什么样的影响呢？
4. 对附件三的新项目给出合适的定价方案，并评价该方案的实施效果。

# 二、 问题分析

## 2.1 对于问题一的分析

问题一的主要目标是寻找任务定价规律，所以我们从价值规律入手。首先，确定了影响任务价值和任务供求关系的因素即为任务定价的影响因素。然后，整体上采用拟合的方式试图发现整体的定价规律，但整体的定价规律不好找到。接着，通过 K-Means 聚类分析把附件一的任务点分为 9 类，分别求解每一类的多元线性回归函数。最后，为了使结果更加接近于现实、更加精确，又采用主成分多元线性回归来拟合回归函数关系。在给出结果的评价分析后，主成分多元线性回归求得的结果最为准确。

## 2.2 对于问题二的分析

问题二的主要目标是设计新的标价方案（方法）,然后评价所设计新标价方案的效

果。

为了设计出新的标价方案，首先，应用 K-Means 聚类分析将任务点在空间中分成了 9 类；然后，在每一类中应用主成分多元线性回归分析来拟合每一类的原始标价方案。通过分析价格规律，联系实际情况，分析原始方案不准确的原因是，没有结合当实际的劳务价值，只是通过供求关系来制定标价方案。这就使得对于某些区域，原始标价方案是非常不适合的。

为了解决这样不匹配问题，我们从实际出发，搜集现实社会的资料信息，找到了劳务价值的影响函数关系。通过劳动力价值指数来求得劳务修正项，将该修正项带入原始方案的拟合函数关系就得到新标价方案的模型。

最后，用附件一的数据信息来训练支持向量机。训练好的支持向量机拿来分析新标价方案的任务完成度。通过研究新标价方案的任务完成度即可得到新方案定量评价。

### 2.3 对于问题三的分析

相较于问题二，问题三的特点在于任务点打包模型的建立。

运用 Q 型聚类分析的运行机理，结合题目的显示情况，我们创造性的提出了“针对打包问题改进型 Q 型聚类分析法”。通过程序流程图的方式，给出了“针对打包问题改进型 Q 型聚类分析法”的运行机理；并且分析了该方法的弱势和不足，有针对改不足，完成了 Q 型聚类分析修正方法。从而使得该方法更加贴近于实际，处理数据产生的结果更加精确。

设计了包含打包标价方案后，我们通过支持向量机来检验干方法的工作效果。从检验结果中可以看出，改进型聚类分析法使得任务完成度更高。从而说明，我们的任务点打包模型和任务点打包修正模型是相对合理的。

### 2.4 对于问题四的分析

首先，我们把附件三的任务点和附件一的任务点在二维平面地图上进行对比研究。明显能够发现的是，问题四的任务点分布更为集中，这说明问题四的任务价格分布受地区差异的影响较大。同时，我们发现问题四的任务点可以明显的分为三部分，而这三部分又刚好处在不同聚类区域里面。所以，我们已知取余的函数关系，求得初步标价方案；然后，再分析实际的社会数据给到修正项，从而建立出针对问题四任务点的标价方案；最后，利用前边已经训练好的支持向量机来给出标价方案下的任务完成度，从而给到了对设定标价方案的定量评价。

## 三、 基本假设

1. 所有任务在接单后，一定会被完成（会以某一概率完成）。
2. 所有对标价调整是建立在总标价不改变的前提下进行的。
3. 不考虑短期政策因素对劳动力价值成本的影响。
4. 在打包的任务点集合中，认为都能被完成，完成度为 1。

# 四、符号说明

本文中，用的符号见表 4.1

表 4.1 符号说明表

| 符号 | 说明 |
|---|---|
| $\boldsymbol{p_i^0}$ | i 号区域任务标价分界面 |
| $\overline{\boldsymbol{p^0}}$ | 总体区域任务标价分界面 |
| $\boldsymbol{p_{ji}}$ | J 号区域 i 号任务标价 |
| $\boldsymbol{L_i}$ | 第 i 号区域劳动力价值指数 |
| $\boldsymbol{x_i}$ | 劳动力价值指标影响因素 |
| $\boldsymbol{N_{0i}}$ | 修正前第 i 号区域劳动力价值修正项 |
| $\boldsymbol{N'_{0i}}$ | 修正后第 i 号区域劳动力价值修正项 |
| $\boldsymbol{N_1}$ | 任务数 |
| $\boldsymbol{N_2}$ | 会员信誉值 |
| $\boldsymbol{N_3}$ | 预定任务 |
| $\boldsymbol{N_4}$ | 会员数 |
| $\boldsymbol{N_5}$ | 供需比 |
| $\boldsymbol{r_0}$ | 会员基本活动能力半径 |
| $\boldsymbol{M_i}$ | 第 i 号的会员 |

# 五、问题一的模型建立与求解

## 5.1 任务标价的定性规律

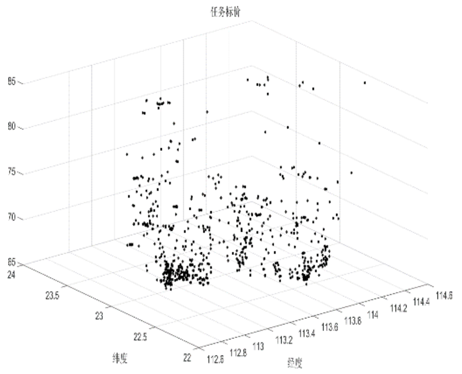

图 5.1.1 任务点的空间分布图

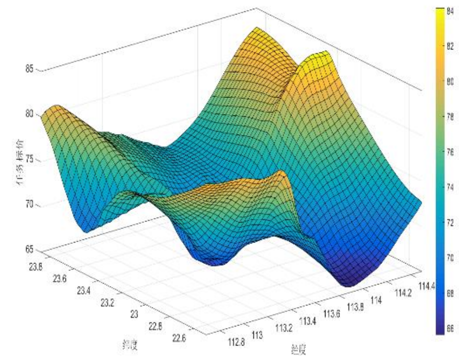

图 5.1.2 任务点拟合空间分布

首先，通过整体的拟合可以得到，所有任务标价随空间位置的分布。通过图 5.1.1 可以直观看到任务点标价随空间分布的情况。然后，又通过三次样条插值方法[4]，得到匹配程度较为理想的空间曲面来反映任务点标价随空间分布的直观表达。

在区间[a, b]内取 n+1 个离散节点，$a = x_0 < x_1 < \cdots < x_n = b$，且$y_i = f(x_i)$，若函数$S(x)$满足条件：

1. 子区间$[x_j,\ x_{j+1}]$上为三次多项式（$j = 0,1,\ldots,n-1$）.
2. 在整个区间[a, b]上二阶连续可导。
3. $S(x_j) = y_j$ ，则称$S(x)$为区间[a, b]上$f(x_j)$的三次样条函数。

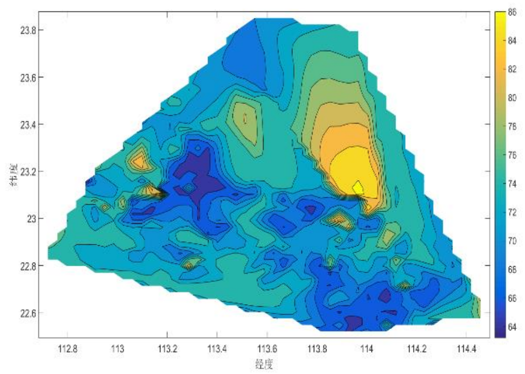

图 5.1.3 任务点标价平面分布图

在上述分析中，仅仅是根据附件一的资料很难找到任务标价随位置变化的具体规律。但是在局部观察中，可以看出还是具有比较明显的规律性。例如，在图 5.1.3 的右上角，黄色极大值区域，从中间到最上部，标价随位置依次均匀下降。但这只是定性分析。所以，要想得到局部较为明确的规律，需首先进行聚类法分类，再进行相关性分析。

## 5.2 任务标价的定量规律

### 5.2.1K-Means 聚类分析[5]划分区域

（1）K-Means 聚类算法

将物理或抽象对象的集合分成由类似的对象组成的多个类的过程被称为聚类。由聚类所生成的簇是一组数据对象的集合，这些对象与同一簇中的对象彼此相似，与其他簇中的对象相异。聚类分析又称群分析，它是研究（样品或指标）分类问题的一种统计分类方法。聚类分析计算方法主要有一下几种：划分方法、层次方法、基于密度方法、基于网格方法、基于模型方法。

K-Means 聚类算法是具名的划分聚类分割方法。

（2）K-Means 算法步骤

①从 N 个数据对象任意选择 K 个对象作为初始聚类中心。

②循环③到④直到每个聚类不在发生变化为为止。

③根据每个聚类对象的均值（中心对象），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相应对象进行划分。

④重新计算每个聚类的均值（中心对象），直到聚类中心不再变化。这种划分使得下式最小：

$$\mathrm{E}=\sum_{j=1}^{k}\sum_{x_i\in w_j}\left\|x_i-m_j\right\|^2 \qquad (5\text{-}2\text{-}1)$$

（3）K-Means 算法特点

①在 K-Means 算法中，K 是事先给定的，这个 K 值的选定是非常难以估计的。

②在 K-Means 算法中，首先需要根据初始聚类中心来确定一个初始划分，然后对初始划分进行优化。

③K-Means 算法需要不断地进行样本分类调整，不断地计算调整后的新的聚类中心，因此当数据量非常巨大时，算法的时间开销是非常巨大的。

④K-Means 算法对一些离散点和初始 K 值敏感，不同的距离初始值对同样的数据样本可能得到不同的结果。

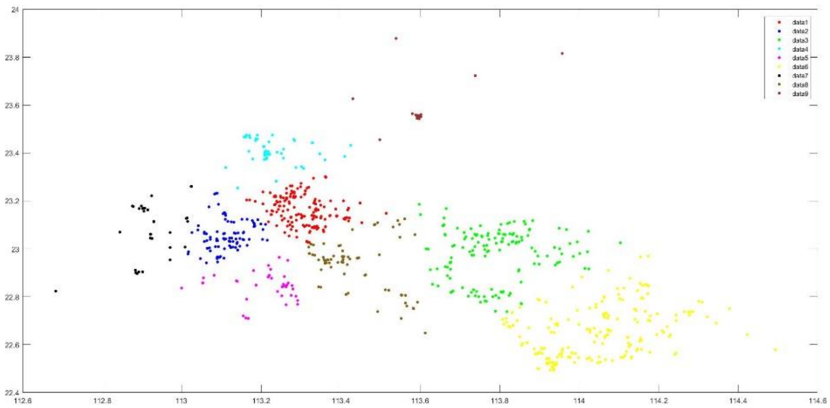


图 5.2.1.1K-Means 聚类分析结果

### 5.2.2 多元线性回归

（1）拟定回归函数

$$P = N_0 + \sum_{i=1}^{5} a_i * N_i \tag{5-2-2}$$

其中，

$$N_5 = N_3 * \frac{N_4}{N_1} \tag{5-2-3}$$

（2）多元线性回归

在回归分析中，如果有两个或两个以上的自变量，就称为多元回归。通过最小二乘法可以简便地求得未知的参数，并使得得到的数据与实际数据之间误差的平方和为最小，进而实现线性回归。

（3）拟合结果

表 5.2.2.1 多元线性回归拟合结果

| 区域 | 拟合结果 |
|---|---|
| 1 | $Y=69.1470 + 0.0111N_1 - 0.0014 N_2 - 0.5173 N_3 - 0.0266 N_4 +0.1470 N_5$ |
| 2 | $Y=84.8293 + 0.2929N_1 - 0.0991 N_2 - 0.9885 N_3 - 0.1809 N_4 +0.0322 N_5$ |
| 3 | $Y=73.9177 + 0.0434N_1 - 0.0014 N_2 - 0.1462 N_3 - 0.3063 N_4 +0.2606 N_5$ |
| 4 | $Y=74.8326 + 0.3360N_1 - 0.0008 N_2 - 0.3340 N_3 - 0.3850 N_4 +0.0674 N_5$ |
| 5 | $Y=72.7786 - 0.3722N_1 + 0.2582N_2 + 0.2309 N_3 - 0.5120 N_4 -0.1463 N_5$ |
| 6 | $Y=75.0578+ 0.1119N_1 + 0.0056N_2 – 1.0381N_3 - 0.1184 N_4 +0.0770 N_5$ |
| 7 | $Y=74.3161- 0.3538N_1 + 0.0006N_2 + 0.0774N_3 + 0.1086 N_4 -0.0200 N_5$ |
| 8 | $Y=73.7347- 0.1626N_1 - 0.0051N_2 - 0.0432N_3 - 0.1868 N_4 + 0.1236 N_5$ |
| 9 | $Y=75.2192- 0.2192N_1 - 0.0159N_2 + 0.0000N_3 +0.0000N_4 - 0.1152 N_5$ |

（4）拟合效果评价

①R2 表示方差解释率，R2 越接近 1 说明数据拟合程度越好。

②P 为 F 统计量对应的概率，越接近 0 越好，当 P<α（默认显著性水平α=0.05）时拒绝 H0，回归模型成立。

表 5.2.2.2 多元线性回归拟合效果评价

| 区域 | 方差解释率 R2 | 概率 P |
|---|---|---|
| 1 | 0.8357 | 0.0493 |
| 2 | 0.7925 | 0.2632 |
| 3 | 0.6711 | 0.0071 |
| 4 | 0.9353 | 0.0172 |
| 5 | 0.9284 | 0.4380 |
| 6 | 0.5146 | 0.0086 |
| 7 | 0.1114 | 0.9723 |
| 8 | 0.8147 | 0.0168 |
| 9 | 0.4167 | 0.6056 |

### 5.2.3 主成分[6]回归分析

主成分回归分析是为了克服最小二乘（LS）估计在数据矩阵 A 存在多重共线性时表现出的不稳定性而提出的。

主成分回归分析采用的方法是将原来的回归自变量变换到另一组变量，即主成分，选择其中一部分重要的主成分作为新的自变量，丢弃了一部分影响不大的自变量，实际上达到了降维的目的，然后用最小二乘法对选取主成分后的模型参数进行估

计，最后再变换回原来的模型求出参数估计。

表 5.2.3.1 主成分回归分析拟合结果

| 分区 | 拟合结果 | 原标准差 | 现标准差 |
|---|---|---|---|
| 1 | $Y=68.3137-0.0871N_1-0.0025N_2-0.2454N_3+0.0025N_4+0.1351N_5$ | 3.3946 | 3.1785 |
| 2 | $Y=81.1314+0.1484N_1-0.1521N_2-0.3808N_3+0.0036N_4+0.0059N_5$ | 3.5743 | 3.5736 |
| 3 | $Y=73.9177+0.0434N_1-0.0014N_2-0.1462N_3-0.3063N_4+0.2606N_5$ | 2.9535 | 2.9535 |
| 4 | $Y=74.8326+0.3360N_1-0.0008N_2-0.3340N_3-0.3850N_4+0.0674N_5$ | 0.0644 | 0.0644 |
| 5 | $Y=72.7786-0.3722N_1+0.2582N_2+0.2309N_3-0.5120N_4-0.1463N_5$ | 1.2147 | 1.2147 |
| 6 | $Y=75.6003-0.1157N_1+0.0122N_2-0.9368N_3-0.0854N_4+0.0396N_5$ | 3.8191 | 3.7901 |
| 7 | $Y=74.7183-0.1741N_1+0.0005N_2+0.0203N_3-0.0693N_4-0.0012N_5$ | 2.5048 | 2.3378 |
| 8 | $Y=73.7347-0.1626N_1-0.0051N_2-0.0432N_3-0.1868N_4+0.1236N_5$ | 1.8624 | 1.8624 |
| 9 | $Y=75.0245-0.0245N_1-0.0117N_2-0.1215N_3-0.1950N_4-0.0494N_5$ | 7.0711 | 4.0825 |

## 5.3 任务未完成原因分析

价格规律:

在一定或特定的时间内,某一商品或某类商品,所表现出来的价格波动,是由商品价值决定,并受供求关系影响围绕价值上下。

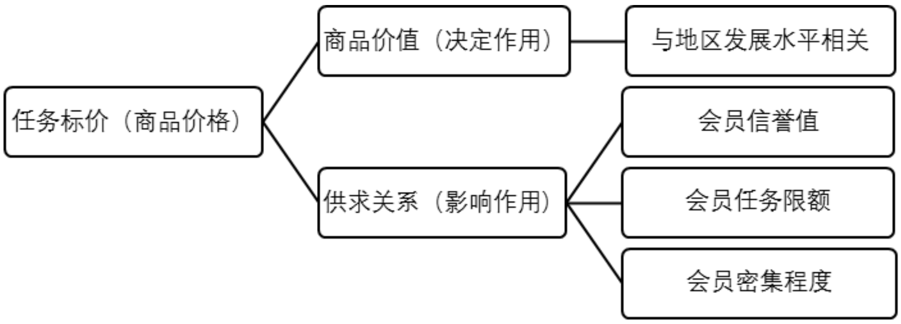


图 5.3.1 价值规律角度解释任务未完成原因分析

从价格规律入手，分析任务标价影响任务的完成。由供求关系的会员信誉值、会员任务限额、会员密集程度三方面因素拟合得到了总体区域任务标价分界面$\overline{p^0}$.

然而，实际情况存在$i \in \{1,2,3,4,5,6,7\}$

$$s,t\ \left|p_i^0-\overline{p^0}\right| \neq 0 \tag{5-3-1}$$

那么，不妨设$p_j^0, j \in \{1,2,3,4,5,6,7\}$

$$p_j^0-\overline{p^0} > 0 \tag{5-3-2}$$

则，j 号区域 i 号任务标价$p_{ji}$，满足

$$\overline{p^0} < p_{ji} < p_j^0 \tag{5-3-3}$$

会被误认为能够完成，实际上完不成。

$$\Delta = p_i^0-\overline{p^0} \tag{5-3-4}$$

表征了不同地区相较于总体分界面的差异

$$p_i^0 = \overline{p^0} + \left(p_i^0 - \overline{p^0}\right) = \overline{p^0} + \Delta \tag{5-3-5}$$

而$\Delta = \Delta\left(\text{地区发展水平}\right)$，所以，没有考虑地区发展水平建立的标价分界面会在 j 号区域的地方又较大的误差，这就导致了 j 号区域有许多任务在$\overline{p^0}$建立的标价方案下无法被完成。

# 六、问题二的模型建立与求解

## 6.1 问题二解题思路概括

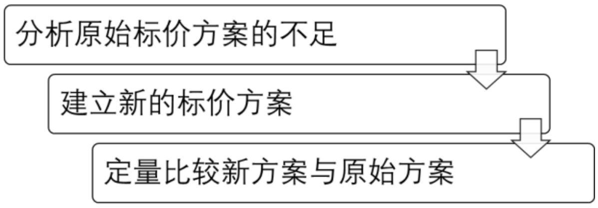


图 6.1.1 解题思路概括

## 6.2 分析原始方案的不足

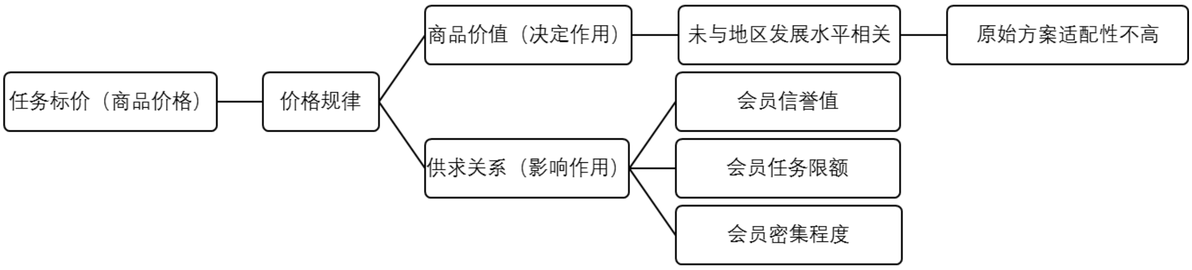


图 6.2.1 原始方案的不足分析

从本质上讲，“拍照赚钱”APP 是会员劳务商品的购买者（需求者），会员（用户）是劳务商品的提供者。所以，任务能否成功被执行也是符合经济市场的经典规律——价格规律：在一定或特定的时间内,某一商品或某类商品,所表现出来的价格波动,是由商品价值决定,并受供求关系影响围绕价值上下。那么，从价格规律入手讨论任务标价方案的不足就显得比较合理。

在图 6.2.1 中，原始标价方案考虑商品的价值和供求关系，但是仍旧具有不合理性。原因在于，商品的价值（劳务的价值）具有明显的地区性(对比观察图 6.2.2 和图 6.2.3)，而原始标价方案从整体均值入手考察了商品价值，并没有考察不同区域的劳务价值情况。这就使得，原始定价方案只能适配部分区域，不能适配所有的区域。故，原始方案的适配性不高。

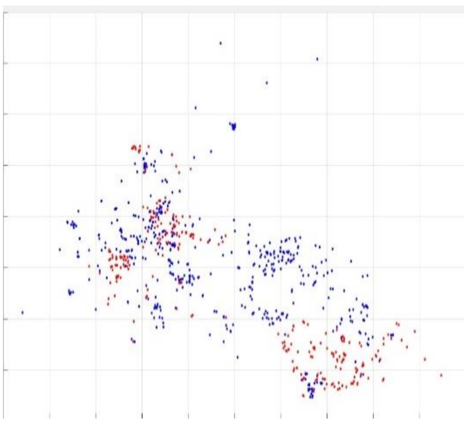

图 6.2.2 任务点在位置平面分布

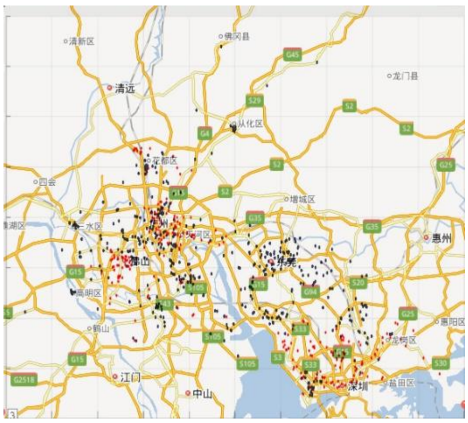

图 6.2.3 任务点在地图平面分布

## 6.3 建立新的标价方案

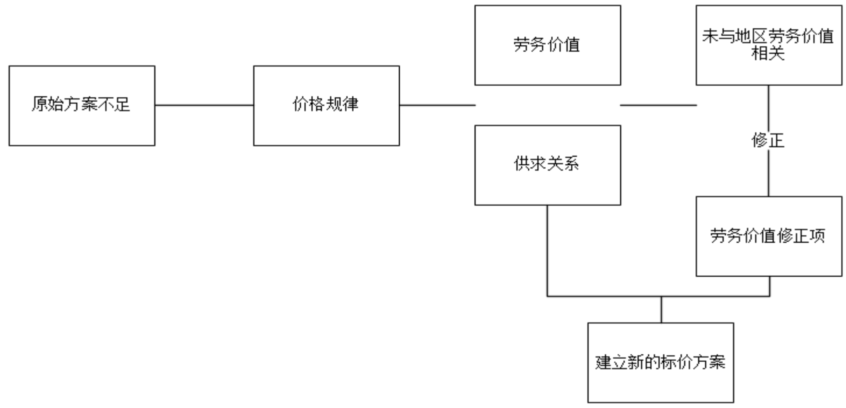

图 6.3.1 新方案建立总体思路

### 6.3.1 劳动力价值指数

表 6.3.2 劳动力价值指数影响因素[3]

| 编号 | 名称 | 符号 | 权重系数[3] |
|---|---|---|---|
| 1 | 人均 GDP | $x_1$ | 0.51766 |
| 2 | 第三产业占比 | $x_2$ | 0.00937 |
| 3 | 经济增长率 | $x_3$ | 0.00841 |
| 4 | 老龄化比例 | $x_4$ | 0.02519 |
| 5 | 人口增长率 | $x_5$ | 0.04207 |
| 6 | 公共财政教育支出 | $x_6$ | 0.06968 |
| 7 | 公共财政总支出 | $x_7$ | 0.02042 |

各数据以 2015 年为例

根据参考文献[2]，第 i 号区域劳动力价值指数的计算公式为

$$L_i = \sum_{i=1}^{7} a_{ij} x_i \text{ ,j} = 1,2,3,4,5,6,7,8 \tag{6-3-1}$$

$a_{ij}$为第 j 号影响因素在第 i 号区域权重系数。

而修正后的劳动力价值修正项$N'_{0i}$表达式为，

$$N'_{0i} = \alpha L_i \tag{6-3-2}$$

那么，任务标价（P）就应该修正为，

$$\mathrm{P} = N'_{0i} + \sum_{i=1}^{5} a_i * N_i \tag{6-3-3}$$

### 6.3.2 劳务价值修正项

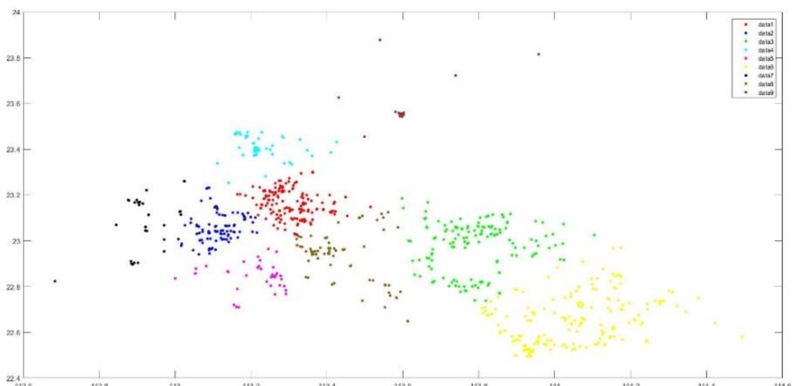
图 6.3.3K-Means 聚类分析结果编号

通过 K-Means 聚类分析得到了，得到了 9 个不同位置的区域依次标号,如图 6.3.3 所示。记$L_i$第 i 号区域所求的的劳动力价值指数，$N_{0i}$为第 i 号区域修正前劳动力价值修正项，$N_{0i}'$为第 i 号区域修正后劳动力价值修正项。在已有整体拟合的多元线性函数关系基础之上加入劳动力价值修正项，从而使得原有拟合函数得以修正，从而使任务定价更加适应所在区域的现实情况。进而，获得更大的任务完成度。

## 6.4 定量评价标价方案

### 6.4.1 支持向量机检验模型[7]

根据给定的训练集

$$\mathrm{T}=\{(x_1,y_1),(x_2,y_2),\ldots,(x_l,y_l)\}\in(X\times Y)^l$$

其中$x_i\epsilon X=R^n$, X 称为输入空间，输入空间中的每一个点$x_i$由 n 个属性特征组成，$y_i\in Y=\{-1,1\}, i=1,\ldots,l$。寻找$R_n$上的一个实值函数$\mathrm{g}(x)$, 以便用分类函数

$$\mathrm{f}(x)=sgn\big(g(x)\big),$$

推断任意一个模式 x 相对应的 y 值的问题为分类问题。

### 6.4.2 新的任务点定价

下面计算 6 号区域新的任务点定价：

通过参考文献[2]、表 6.3.2 和式（6-3-1）可以求得，

$$\begin{cases} N_{03}=73.9177 \\ L_3=13894 \\ N_{06}=75.6003 \\ L_6=14660 \end{cases} \tag{6-4-1}$$

又由公式

$$N_{0i}'=\alpha L_i \tag{6-3-2}$$

$$\mathrm{P}=N_{0i}'+\textstyle\sum_{i=1}^{5}a_i*N_i \tag{6-3-3}$$

可得到，

$$P_6=77.7123-0.1157N_1+0.0122N_2-0.9368N_3-0.0854N_4+0.0396N_5$$

至此，已求得 6 号区域新的标价方案。

### 6.4.3 评价新方案

$N_1$—$N_5$均为每个任务点的属性，可由附件一和附件二数据信息方便得到，这里便不再赘述。把 6 号区域每一任务点的属性信息带入$P_6$表达式，即可得到 6 号区域的每一点的任务价格。

计算结果如下，

表 6.4.3.1 新的标价方案评价结果

| | 原始方案 | 新方案 |
|---|---|---|
| 任务完成点的数量 | 69 | 110 |
| 任务未完成点的数量 | 125 | 84 |
| 总的任务数量 | 194 | 194 |
| 任务完成度 | 35.57% | 56.7% |

通过上述评价结果可以看出，相较于原始方案，新方案能使得任务完成度有大幅度提高。对于 APP 平台而言，设置任务标价，是为了外包平台上的任务尽可能被完

成。而新方案恰好能实现这样的功能，这就实现了我们最初制定新方案的目标。所以，我们设计的标价方案相对合理。

# 七、问题三的模型建立与求解

## 7.1 会员基本活动能力半径

在本文中，我们将会员（用户）基本活动能力半径$r_0$定义为 2 千米。

下面给出必要的说明和解释，

设定编号为 i 的会员的平面位置坐标$M_i(x_i,y_i)$, 编号为 j 的任务的平面位置坐标$A_j(u_j,v_j)$，遍历所有的$M_i(x_i,y_i)$，$A_j(u_j,v_j)$

$$d_i=\sqrt{\left(x_i-u_j\right)^2+\left(y_i-v_j\right)^2} \tag{7-1-1}$$

$$\min d_i \leq \mathrm{xkm} \tag{7-1-2}$$

满足上述条件，则称会员基本活动能力半径为 x 千米。而在该模型中 x=2。所以，会员的基本活动能力半径为 2 千米。

## 7.2Q 型聚类分析[8]

### 7.2.1 样本的相似性度量

要用数量化的方法对事物进行分类，就必须用数量化的方法描述事物之间的相似程度。一个食物常常需要用多个变量进行刻画。如果对于一群有待分类的样本点需用 p 个变度。一个事物常常需要用多个变量来刻画。如果对于一群有待分类的样本点需用 p 个变量描述，则每个样本点可以看成是$R^p$空间中的一个点。因此，很自然地想到可以用距离来度量样本点间的相似程度。

记$\Omega$是样本点集，距离$\mathrm{d}(*,*)$是$\Omega\times\Omega\to R^+$的一个函数，满足条件：

1. $\mathrm{d}(\mathrm{x},\mathrm{y})\geq 0,\mathrm{x},\mathrm{y}\in\Omega$
2. $\mathrm{d}(\mathrm{x},\mathrm{y})=0$当且仅当 x=y
3. $\mathrm{d}(\mathrm{x},\mathrm{y})=\mathrm{d}(\mathrm{y},\mathrm{x}),x,y\in\Omega$
4. $\mathrm{d}(\mathrm{x},\mathrm{y})\leq\mathrm{d}(\mathrm{x},\mathrm{z})+\mathrm{d}(\mathrm{z},\mathrm{y}),x,y,z\in\Omega$

这一距离的定义是我们所熟知的，它满足正定性、对称性和三角不等式。

### 7.2.2 绝对值距离

$$d_1(x,y)=\sum_{k=1}^{p}|x_k-y_k|$$

## 7.3 针对打包问题改进型聚类模型

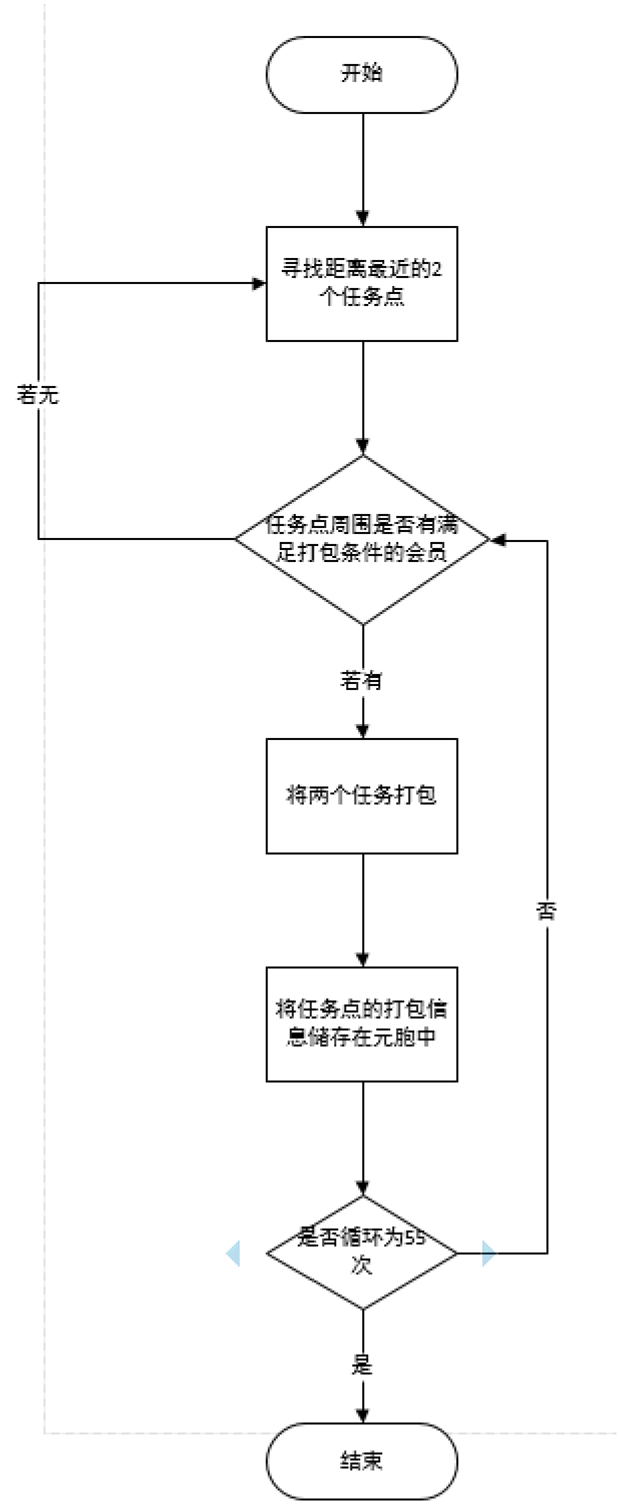

图 7.3.1 改进型 Q 聚类模型

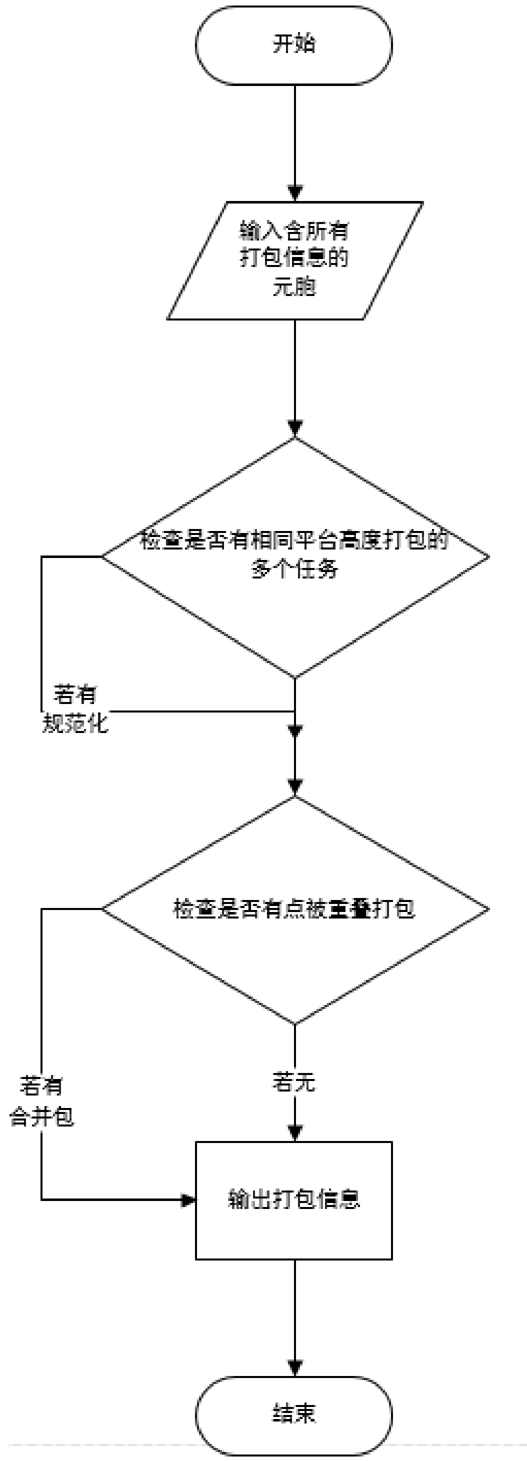

图 7.3.2 辅助修正模型

## 7.4 打包模型的结果

通过 matlab 编程计算，我们得到了打包模型的结果。

表 7.4.1 打包模型的结果

| 包裹编号 | 任务点号码 | | |
|---|---|---|---|
| 1 | 472 | 184 | 391 |
| 2 | 834 | 754 | |
| 3 | 378 | 180 | |
| 4 | 323 | 312 | |
| 5 | 391 | 184 | |
| 6 | 602 | 490 | |
| 7 | 693 | 598 | |
| 8 | 402 | 376 | |
| 9 | 197 | 135 | 183 |
| 10 | 701 | 618 | |
| 11 | 490 | 484 | |

根据题意，打包的原因是任务点距离比较近一些，会使得不同会员（用户）会对同一任务点造成竞争，从而降低了 APP 外包平台发布任务的完成度。所以，打包之后再确定所有任务的完成度，实际上是有未打包任务基本决定的。打包好的任务认为都完成，而剩余未打包的任务，仍旧通过之前训练好的支持向量机来预测任务的完成情况。Matlab 支持向量机的预测结果由表 7.5.1 展示，具体 matlab 编程和数据在附录中。

## 7.5 支持向量机模型检验

表 7.5.1 支持向量机模型检验结果

| | 原始方案 | 新方案 |
|---|---|---|
| 任务完成点的数量 | 89 | 119 |
| 任务未完成点的数量 | 125 | 84 |
| 总的任务数量 | 204 | 193 |
| 任务完成度 | 45.57% | 59.7% |

通过这个表 7.5.1，我们可以清晰看出，经过打包之后所有任务的完成度有了大约 10%的提高。这说明我们的打包模型是合理有效的。

# 八、问题四的模型建立与求解

在解决附件一的标价方式，多元线性回归方程找到原始方案标价规律。然后，通过分析每一区域的现实情况，在原始方案标价规律的基础上加入修正项，从而给出了新的标价方案。最后，通过训练好的支持向量机，划分新方案标价的任务点的可完成的和不可完成的，从而求得新方案的任务完成度，给到新方案优劣的定量评价。

在解决问题四也是类似的思路：

1. 用 K-Means 聚类分析法将附件三的任务点分成 3 个区域。
2. 将地区现实差异性考虑在内，给到 3 个区域含修正项的标价方案。
3. 用训练好的支持向量机，将每个区域的任务点分类，计算完成度，从而万郴对标价方案的评价。

## 8.1 确定新项目任务点的标价方案

通过分析任务点的位置，可知在该位置处前面已经解决多元线性回归曲线：

表 8.1.1 任务点多元线性回归曲线

| | | | |
|---|---|---|---|
| 1 | $Y=73.9177+0.0434N_1-0.0014N_2-0.1462N_3-0.3063N_4+0.2606N_5$ | 2.9535 | 2.9535 |
| 2 | $Y=74.8326+0.3360N_1-0.0008N_2-0.3340N_3-0.3850N_4+0.0674N_5$ | 0.0644 | 0.0644 |
| 3 | $Y=72.7786-0.3722N_1+0.2582N_2+0.2309N_3-0.5120N_4-0.1463N_5$ | 1.2147 | 1.2147 |

然后，联系该地区的现实社会发展情况计算每个区域的修正项，

表 8.1.2 劳动力价值指标影响因素

| 编号 | 名称 | 符号 | 权重系数[3] |
|---|---|---|---|
| 1 | 人均 GDP | $x_1$ | 0.51766 |
| 2 | 第三产业占比 | $x_2$ | 0.00937 |
| 3 | 经济增长率 | $x_3$ | 0.00841 |
| 4 | 老龄化比例 | $x_4$ | 0.02519 |
| 5 | 人口增长率 | $x_5$ | 0.04207 |
| 6 | 公共财政教育支出 | $x_6$ | 0.06968 |
| 7 | 公共财政总支出 | $x_7$ | 0.02042 |

根据参考文献[2]，第 i 号区域劳动力价值指数的计算公式为

$$L_i=\sum_{i=1}^{7}a_{ij}x_i\,,\mathrm{j}=1,2,3,4,5,6,7,8 \tag{6-3-1}$$

$a_{ij}$为第 j 号影响因素在第 i 号区域权重系数。

而修正后的劳动力价值修正项$N_{0i}'$表达式为，

$$N_{0i}'=\alpha L_i \tag{6-3-2}$$

那么，任务标价（P）就应该修正为，

$$\mathrm{P}=N_{0i}'+\sum_{i=1}^{5}a_i*N_i \tag{6-3-3}$$

从而确定任务点的含修正项的标价方案，

表 8.1.3 任务点多元线性回归曲线

| | | | |
|---|---|---|---|
| 1 | $Y=74.9677+0.0434N_1-0.0014N_2-0.1462N_3-0.3063N_4+0.2606N_5$ | 2.9535 | 2.9535 |
| 2 | $Y=76.866+0.3360N_1-0.0008N_2-0.3340N_3-0.3850N_4+0.0674N_5$ | 0.0644 | 0.0644 |
| 3 | $Y=76.789-0.3722N_1+0.2582N_2+0.2309N_3-0.5120N_4-0.1463N_5$ | 1.2147 | 1.2147 |

## 8.2 打包模型

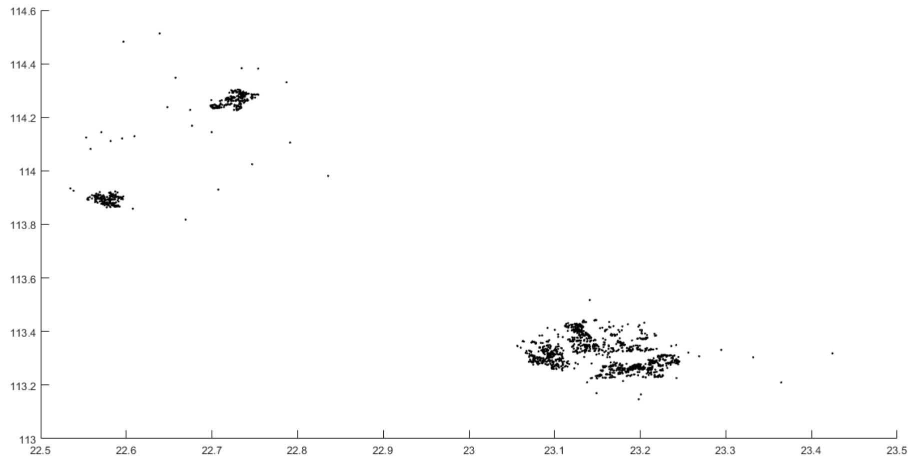


图 8.2.1 新项目任务点位置分布图

要用数量化的方法对事物进行分类，就必须用数量化的方法描述事物之间的相似程度。一个食物常常需要用多个变量进行刻画。如果对于一群有待分类的样本点需用 p 个变度。一个事物常常需要用多个变量来刻画。如果对于一群有待分类的样本点需用 p 个变量描述，则每个样本点可以看成是$R^p$空间中的一个点。因此，很自然地想到可以用距离来度量样本点间的相似程度。

记$\Omega$是样本点集，距离$\mathrm{d}(*,*)$是$\Omega\times\Omega\rightarrow R^{+}$的一个函数，满足条件：

5. $\mathrm{d}(\mathrm{x},\mathrm{y})\geq 0,\mathrm{x},\mathrm{y}\in\Omega$
6. $\mathrm{d}(\mathrm{x},\mathrm{y})=0$当且仅当 x=y
7. $\mathrm{d}(\mathrm{x},\mathrm{y})=\mathrm{d}(\mathrm{y},\mathrm{x}),x,y\in\Omega$
8. $\mathrm{d}(\mathrm{x},\mathrm{y})\leq\mathrm{d}(\mathrm{x},\mathrm{z})+\mathrm{d}(\mathrm{z},\mathrm{y}),x,y,z\in\Omega$

这一距离的定义是我们所熟知的，它满足正定性、对称性和三角不等式。

绝对值距离

$$d_1(x,y)=\sum_{k=1}^{p}|x_k-y_k|$$

## 8.3 打包模型运行结果

表 8.3.1 新项目打包结果

| 组号 | 任务点序号 | | | 组号 | 任务点序号 | | | 组号 | 任务点序号 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 67 | 61 | 1113 | 20 | 1663 | 556 | | 39 | 2028 | 952 | |
| 2 | 786 | 777 | | 21 | 2063 | 2059 | | 40 | 669 | 658 | |
| 3 | 1464 | 327 | | 22 | 1016 | 3 | | 41 | 161 | 160 | |
| 4 | 1989 | 920 | | 23 | 1996 | 929 | | 42 | 1269 | 1380 | |
| 5 | 942 | 936 | | 24 | 1632 | 524 | | 43 | 1427 | 482 | |
| 6 | 1352 | 240 | | 25 | 1288 | 180 | | 44 | 1755 | 750 | |
| 7 | 76 | 75 | | 26 | 397 | 325 | 1548 | 45 | 1828 | 958 | |
| 8 | 1275 | 1272 | | 27 | 1362 | 252 | | 46 | 2029 | 452 | |
| 9 | 1095 | 1093 | | 28 | 2055 | 987 | | 47 | 1590 | 771 | |
| 10 | 1898 | 838 | | 29 | 2038 | 972 | | 48 | 774 | 135 | 1218 |
| 11 | 1916 | 1898 | | 30 | 1861 | 722 | | 49 | 1222 | 37 | |
| 12 | 1234 | 1231 | | 31 | 1492 | 231 | | 50 | 38 | 120 | |
| 13 | 1044 | 35 | | 32 | 342 | 229 | | 51 | 1197 | 343 | |
| 14 | 2000 | 1821 | | 33 | 390 | 342 | | 52 | 459 | 132 | |
| 15 | 176 | 175 | 1292 | 34 | 393 | 343 | | 53 | 1211 | 417 | |
| 16 | 474 | 473 | | 35 | 1656 | 515 | | 54 | 1545 | 995 | |
| 17 | 1202 | 119 | | 36 | 1081 | 3 | | 55 | 2063 | 989 | |
| 18 | 1685 | 612 | | 37 | 1912 | 840 | | 56 | 992 | 90 | |
| 19 | 1010 | 1009 | | 38 | 1435 | 295 | | 57 | 1167 | 295 | |

根据题意，打包的原因是任务点距离比较近一些，会使得不同会员（用户）会对同一任务点造成竞争，从而降低了 APP 外包平台发布任务的完成度。所以，打包之后再确定所有任务的完成度，实际上是有未打包任务基本决定的。打包好的任务认为都完成，而剩余未打包的任务，仍旧通过之前训练好的支持向量机来预测任务的完成情况。Matlab 支持向量机的预测结果由表 7.5.1 展示，具体 matlab 编程和数据在附录中。

## 8.4 支持向量机模型检验

表 8.4.1 支持向量机模型检验结果

| | 新方案 |
|---|---|
| 任务完成点的数量 | 961 |
| 任务未完成点的数量 | 262 |
| 总的任务数量 | 1223 |
| 任务完成度 | 78.5% |

通过表 8.4.1 可以看出在标价方案的作用下，任务的完成率能达到 78.5%。这样数据能够充分说明了该标价方案的合理性。

# 九、模型总结

模型的优点：

1. 解答层次分明逻辑清晰。在研究分析标价规律部分，首先，进行整体上的拟合进行直观定性上的分析和思考；稍微有所思路后，聚类分割研究目标，从小处着眼研究对象的发展变化规律，进行定量分析；最后，设计定量分析后还给出了效果的评价，用主成分的方法来优化算法，使其更为精确。这种由直观到抽象、由主要到次要、由表面到本质、由粗糙到精确的思想，是本文的一大亮点。
2. 紧密联系实际。在自己设计标价方案部分，从实际出发优化原始方案。因为，建模问题从实际中来，最后也应该回到实际中去。
3. 图文并茂。本文注重用图片、图表、框图来表达。

模型缺点：

1. 单目标规划。本文自始至终都是围绕任务执行的完成度这条主线来的。但在实际情况中，APP 外包平台外包劳务工作，不光要考虑任务的可执行的完成度，还要考虑成本问题。而本文没有重点谈及成本问题。
2. 实际信息充分度不够。本文涉及了从实际出发去修正做法，但是在短时间没能获得足够多的实际信息，使得由实际信息分析出的结论并不那么充分。

# 九、参考文献

[1]王小川，史峰，郁磊，李洋，MATLAB 神经网络 43 个案例分析，北京市，北京航空航天大学出版社，2013 年，第 12 页
[2]李华，黄碧玲，广州统计信息手册 2017，广州市统计局，2017
[3] 陈昌华，中国劳动力价格影响因素分析 [J]，重庆理工大学学报: 社会科学，2014( 4) : 38-44
[4]dreamcs,三样条插值 http://blog.csdn.net/dreamcs/article/details/7884045，2017/9/18
[5]yqtaowhu,深入理解 K-Means 聚类算法，http://blog.csdn.net/taoyanqi8932/article/details/53727841,2017/9/17
[6]司守奎，孙兆亮，数学建模算法与应用（第二版），北京，国防工业出版社，2016，P231-P239
[7] 司守奎，孙兆亮，数学建模算法与应用（第二版），北京，国防工业出版社，2016，P201-P214
[8]郭仁忠，张克权，Q 型聚类分析中变量相关性的处理方法分析[J]，武汉测绘科技大学学报，1987(3),64-78

附录

1.读取数据

```
clear all;clc;

%%读取附件一已结束项目任务数据%%
a1=xlsread('已结束项目任务数据.xls');%读取数据
Task_La=a1(:,1);%任务纬度
Task_Lo=a1(:,2);%任务经度
Task_price=a1(:,3);%任务标价
Task_finish=a1(:,4);%任务执行情况
save data1 Task_La Task_Lo Task_price Task_finish;%保存附件一中的数据为.mat 格式
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%读取附件二会员信息数据%%
a1=xlsread('会员信息数据.xlsx');%读取数据
Member_La=a1(:,1);%会员纬度
Member_Lo=a1(:,2);%会员经度
Member_value=a1(:,5);%信誉值
Member_task=a1(:,3);%预定任务限额
save data2 Member_La Member_Lo Member_value Member_task;%保存附件二中的数据为.mat 格式
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

2.画任务价格的空间和平面拟合分布图

```
clear all;clc;
%%所有任务点的空间分布散点图%%
load('data1.mat');
x=Task_Lo;
y=Task_La;
z=Task_price;
scatter3(x,y,z,200,'k','.');%画散点图
xlabel('经度');ylabel('纬度');zlabel('任务标价');
set(gca, 'FontSize', 15);
set(get(gca,'YLabel'),'Fontsize',18);
set(get(gca,'XLabel'),'Fontsize',18);

%%所有任务点的空间拟合分布图%%
```

```
createFit(x, y, z);%createFit是以lowess拟合的空间分布图
xlabel('经度');ylabel('纬度');zlabel('任务标价');
set(gca, 'FontSize', 15);
set(get(gca,'YLabel'),'Fontsize',18);
set(get(gca,'XLabel'),'Fontsize',18);

%%所有任务点的平面拟合分布图%%
createFit1(x, y, z);%createFit是以cubicinterp拟合的空间分布图
xlabel('经度');ylabel('纬度');
set(gca, 'FontSize', 15);
set(get(gca,'YLabel'),'Fontsize',18);
set(get(gca,'XLabel'),'Fontsize',18);


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%拟合图的函数createFit和createFit2%%
function [fitresult, gof] = createFit(x, y, z)
%CREATEFIT(X,Y,Z)
%  Create a fit.
%
%  Data for 'untitled fit 1' fit:
%      X Input : x
%      Y Input : y
%      Z Output: z
%  Output:
%      fitresult : a fit object representing the fit.
%      gof : structure with goodness-of fit info.
%
%  另请参阅 FIT, CFIT, SFIT.

%  由 MATLAB 于 16-Sep-2017 01:09:31 自动生成


%% Fit: 'untitled fit 1'.
[xData, yData, zData] = prepareSurfaceData( x, y, z );

% Set up fittype and options.
ft = fittype( 'lowess' );

% Fit model to data.
[fitresult, gof] = fit( [xData, yData], zData, ft );
```

```
% Plot fit with data.
figure( 'Name', 'untitled fit 1' );
h = plot( fitresult)%, [xData, yData], zData );
legend( h, 'untitled fit 1', 'z vs. x, y', 'Location', 'NorthEast' );
% Label axes
xlabel x
ylabel y
zlabel z
grid on

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [fitresult, gof] = createFit1(x, y, z)
%CREATEFIT1(X,Y,Z)
%  Create a fit.
%
%  Data for 'untitled fit 2' fit:
%      X Input : x
%      Y Input : y
%      Z Output: z
%  Output:
%      fitresult : a fit object representing the fit.
%      gof : structure with goodness-of fit info.
%
%  另请参阅 FIT, CFIT, SFIT.

%  由 MATLAB 于 16-Sep-2017 11:02:13 自动生成


%% Fit: 'untitled fit 2'.
[xData, yData, zData] = prepareSurfaceData( x, y, z );

% Set up fittype and options.
ft = 'cubicinterp';

% Fit model to data.
[fitresult, gof] = fit( [xData, yData], zData, ft );

% Make contour plot.
figure( 'Name', 'untitled fit 2' );
%plot( fitresult, [xData, yData], zData, 'Style', 'Contour' );
plot( fitresult,'Style', 'Contour' );
% Label axes
```

```
xlabel x
ylabel y
grid on
```

3. k-means 聚类

```
%%k-means 聚类方法%%
clc;clear;
%加载数据
load data1
%Task_input=[Task_Lo,Task_La,Task_price];
Task_input=[Task_Lo,Task_La];
x=Task_Lo;
y=Task_La;
z=Task_price;
X=[x,y];

%Kmeans 分类
a1=[113.304,23.19];
%a2=[113.933,23.539];
a2=[113.105,23.02];
%a4=[113.929,23.962];
a3=[113.785,23.06];
a4=[113.205,23.44];
a5=[113.267,22.883];
a6=[114.04,22.674];
a7=[113.0155,23.14];
a8=[113.393,22.961];
%a10=[113.572,23.766];
a9=[113.598,23.56];

start=[a1;a2;a3;a4;a5;a6;a7;a8;a9];
[IDleibei,center,sumLength,Length]=kmeans(Task_input,9,'start',start);
```

```
%画出聚类为1的点。X(IDleibei==1,1),为第一类的样本的第一个坐标;
%X(IDleibei==1,2)为第二类的样本的第二个坐标
size=14;
plot(Task_input(IDleibei==1,1),Task_input(IDleibei==1,2),'r.','MarkerSize',
size)
hold on
plot(Task_input(IDleibei==2,1),Task_input(IDleibei==2,2),'b.','MarkerSize',
size)
hold on
plot(Task_input(IDleibei==3,1),Task_input(IDleibei==3,2),'g.','MarkerSize',
size)
hold on
plot(Task_input(IDleibei==4,1),Task_input(IDleibei==4,2),'c.','MarkerSize',
size)
hold on
plot(Task_input(IDleibei==5,1),Task_input(IDleibei==5,2),'m.','MarkerSize',
size)
hold on
plot(Task_input(IDleibei==6,1),Task_input(IDleibei==6,2),'y.','MarkerSize',
size)
hold on
plot(Task_input(IDleibei==7,1),Task_input(IDleibei==7,2),'k.','MarkerSize',
size)
%plot(x,y,'Color',[1 0 0])
plot(Task_input(IDleibei==8,1),Task_input(IDleibei==8,2),'.','Color',[0.5
0.4 0.1],'MarkerSize',size)
hold on

plot(Task_input(IDleibei==9,1),Task_input(IDleibei==9,2),'.','Color',[0.6
0.2 0.2],'MarkerSize',size)
hold on
%plot(Task_input(IDleibei==10,1),Task_input(IDleibei==10,2),'.','Color',[0.
2 0.6 0.2],'MarkerSize',size)
hold on
%plot(Task_input(IDleibei==11,1),Task_input(IDleibei==11,2),'.','Color',[0.
2 0.2 0.6],'MarkerSize',size)

%%保存不同聚类的样本
task1_Lo=X(IDleibei==1,1);task1_La=X(IDleibei==1,2);task1_price=Task_price(
IDleibei==1);
task2_Lo=X(IDleibei==2,1);task2_La=X(IDleibei==2,2);task2_price=Task_price(
```

```
IDleibei==2);
task3_Lo=X(IDleibei==3,1);task3_La=X(IDleibei==3,2);task3_price=Task_price(
IDleibei==3);
task4_Lo=X(IDleibei==4,1);task4_La=X(IDleibei==4,2);task4_price=Task_price(
IDleibei==4);
task5_Lo=X(IDleibei==5,1);task5_La=X(IDleibei==5,2);task5_price=Task_price(
IDleibei==5);
task6_Lo=X(IDleibei==6,1);task6_La=X(IDleibei==6,2);task6_price=Task_price(
IDleibei==6);
task7_Lo=X(IDleibei==7,1);task7_La=X(IDleibei==7,2);task7_price=Task_price(
IDleibei==7);
task8_Lo=X(IDleibei==8,1);task8_La=X(IDleibei==8,2);task8_price=Task_price(
IDleibei==8);
task9_Lo=X(IDleibei==9,1);task9_La=X(IDleibei==9,2);task9_price=Task_price(
IDleibei==9);
save data11 task1_Lo task1_La task2_Lo task2_La task3_Lo task3_La task4_Lo
task4_La task5_Lo task5_La task6_Lo task6_La task7_Lo task7_La task8_Lo
task8_La task9_Lo task9_La task9_price task8_price task1_price task2_price
task3_price task4_price task5_price task6_price task7_price;
```

4.多元回归（对划分的 9 个区域分别进行线性回归）

```
clear;clc;
load('data11.mat')
x1=task1_Lo;y1=task1_La;z1=task1_price;
%%处理任务数，任务标价等数据%%

m=length(x1);
T_Lo=zeros(20);
T_La=zeros(20);
T=zeros(20);
T_v=zeros(20);
for i=1:m
  if((y1(i)>=22.4)&(y1(i)<=23.9)&(x1(i)>=112.5)&(x1(i)<=114.5))
```

```
  xx(i)=ceil((x1(i)-112.5)/0.1);
  yy(i)=ceil((y1(i)-22.4)/0.1);
  T(xx(i),yy(i))=T(xx(i),yy(i))+1;
  T_v(xx(i),yy(i))=T_v(xx(i),yy(i))+z1(i);
  T_Lo(xx(i),yy(i))=xx(i)*0.1+112.5;
  T_La(xx(i),yy(i))=yy(i)*0.1+22.4;
  end
end
save data9 T_v T T_Lo T_La

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%处理会员数，预定任务限额，信誉值等数据%%
clc;clear;
load('data2.mat');
x1=Member_Lo;y1=Member_La;z1=Member_value;t=Member_task;

m=length(x1);
M_n=zeros(20);
M_t=zeros(20);
M_v=zeros(20);
for i=1:m
  if((y1(i)>=22.4)&(y1(i)<=23.9)&(x1(i)>=112.5)&(x1(i)<=114.5))
  xx(i)=ceil((x1(i)-112.5)/0.1);
  yy(i)=ceil((y1(i)-22.4)/0.1);
  M_n(xx(i),yy(i))=M_n(xx(i),yy(i))+1;
  M_v(xx(i),yy(i))=M_v(xx(i),yy(i))+z1(i);
  M_t(xx(i),yy(i))=M_t(xx(i),yy(i))+t(i);
  end

end
save data8 M_n M_t M_v

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%利用 regress 函数进行多元线性回归%%
clear;clc;
load('data8.mat');
load('data9.mat');
jj=0;
for i=1:20
    for j=1:20
        if(T(i,j)~=0)
            jj=jj+1;
```

```
            TT(jj)=T(i,j);
            aT_v(jj)=T_v(i,j)/T(i,j);
            La(jj)=T_La(i,j)-22.4;
            Lo(jj)=T_Lo(i,j)-112.5;
            d1(jj)=(La(jj)-23.1).^2+(Lo(jj)-113.2).^2;
            d2(jj)=(La(jj)-22.8).^2+(Lo(jj)-114).^2;
            if(M_n(i,j)~=0)
            aM_v(jj)=M_v(i,j)/M_n(i,j);
            aM_t(jj)=M_t(i,j)/M_n(i,j);
            aM_n(jj)=M_n(i,j);
            else
            aM_v(jj)=0;
            aM_t(jj)=0;
            aM_n(jj)=M_n(i,j);
            end

        end
    end
end
x=[ones(11,1),TT',aM_v',aM_t',aM_n',((aM_t.*aM_n)./TT)'];
[b,bint,r,rint,stats]=regress(aT_v',x);
S_d=((aM_t.*aM_n)./TT)';
xx=[TT' aM_v' aM_t' aM_n' S_d  aT_v'];
save data10 xx;
```

5.主成分回归分析

```
clear;clc;
load('data10.mat');
[m,n]=size(xx);
x0=xx(:,1:n-1);y0=xx(:,n);
r=corrcoef(x0) %计算相关系数矩阵
xb=zscore(x0); %对设计矩阵进行标准化处理
yb=zscore(y0); %对 y0 进行标准化处理
%以下命令利用设计矩阵进行主成分分析，返回值 c 的列向量对应着主成分的系数
%返回值 s 对应着式（26）中的 Z 矩阵，t 返回的是特征值
[c,s,t]=princomp(xb)
```

```
contr=cumsum(t)/sum(t) %计算累积贡献率，第 i 个分量表示前 i 个主成分的贡献率
num=input('请选项主成分的个数:') %通过累积贡献率交互式选择主成分的个数
hg1=[ones(m,1),x0]\y0; %计算普通最小二乘法回归系数
hg1=hg1'
%下面显示普通最小二乘法回归结果
fprintf('y=%f',hg1(1));
for i=1:n-1
 fprintf('+%f*x%d',hg1(i+1),i);
end
fprintf('\n')
hg=s(:,1:num)\yb; %主成分变量的回归系数
hg=c(:,1:num)*hg; %标准化变量的回归方程系数
%下面计算原始变量回归方程的系数
hg2=[mean(y0)-std(y0)*mean(x0)./std(x0)*hg, std(y0)*hg'./std(x0)]
%下面显示主成分回归结果
fprintf('y=%f',hg2(1));
for i=1:n-1
 fprintf('+%f*x%d',hg2(i+1),i);
end
fprintf('\n')
%下面计算两种回归分析的剩余标准差
rmse1=sqrt(sum((x0*hg1(2:end)'+hg1(1)-y0).^2)/(m-n))
rmse2=sqrt(sum((x0*hg2(2:end)'+hg2(1)-y0).^2)/(m-num-1))
```

6. 城区分析

```
clc
```

```
clear
A= xlsread ('【29】各城区分析 3');
OD=A(:,1);N=A(:,2);GDP1=A(:,3);GR=A(:,4);S=A(:,5);EF1=A(:,6);PGS=A(:,7);
GDP(:,1)=log(GDP1(:,1)); EF(:,1)=log(EF1(:,1));
M=zeros(14,1);
for i=1:14

m(i)=exp(0.02519*OD(i)+0.04207*N(i)+0.51766*GDP(i)+0.00841*GR(i)+0.00937*S(
i)+0.06968*EF(i)+0.02042*PGS(i)); %引用模型十的系数
end
m=m'
```

7.支持向量机

```
%SVM 支持向量机预测在某个价格下任务是否被完成
clear,clc

%%
%输入数据
%data1 是来自附件 1 的已完成任务的数据，一共 835 个数据
errorRate1=0.6;errorRate2=0.6;
while((errorRate1>0.3)||(errorRate2>0.3))
load('data.mat')
load('task_density.mat')
load('Member_density.mat')
load('data1.mat')
d_min=d_min';
B1=B1';
B2=B2';
%输入典型点数据：
Task_input=[d_min,Task_Lo,Task_La,Task_price,B1,B2];
%Task_input=[Task_Lo,Task_La,Task_price];
Task_finish=[Task_finish];
%从 835 组个数据中随机选择 785 组训练数据和 50 组预测数据
n=randperm(835);
n=n';
input_train=Task_input(n(1:785),:);
output_train=Task_finish(n(1:785),1);
input_test=Task_input(n(785:835),:);
output_test=Task_finish(n(785:835),:);
%训练数据标准化
[inputn,Sin]=mapstd(input_train);
```

```
%%
%训练支持向量机
%SMO序列最小优化 OP二次规划法 LS最小二乘法
%核函数，quadratic二次方程 polynomial多项式 rbf高斯径向基函数 mlp多层
%kktviolationlevel允许违反KKT条件的程度 ,'kktviolationlevel',0.1
s=svmtrain(input_train',output_train','Method','LS','kernel_function','polynomial');   %训练,为了使用支持向量机，对矩阵进行了转置
sv_index=s.SupportVectorIndices;  %返回支持向量机的标号

%%
%检验
check1=svmclassify(s,input_train) ;%验证已知样本点（训练集（还未验证测试集
errorRate1=1-sum(output_train==check1)/length(output_train);%计算错判率
%检验2
check2=svmclassify(s,input_test) ;%验证已知样本点（测试集集
errorRate2=1-sum(output_test==check2)/length(output_test);%计算错判率
end

%%
%预测
load('d_min（10）.mat')
load('task6_density(10).mat')
load('Member6_density(10).mat')
load('price_6.mat')
load('weizhi6.mat')
B1_6=(B1_6)';B2_6=(B2_6)';d6_min=(d6_min)';
task6_p=[task6_p;task6_p];task6_Lo=[task6_Lo;task6_Lo];task6_La=[task6_La;task6_La];d6_min=[d6_min;d6_min];B2_6=[B2_6;B2_6];B1_6=[B1_6;B1_6];
Task_input2=[d6_min,task6_Lo,task6_La,task6_p,B1_6,B2_6];

w=svmclassify(s,Task_input2);
num=sum(w)
```

8. 根据训练出的网络画出拟合效果图

```
A=[22.4,112.6];
for x=22.3:0.01:24.4
    for y=112.3:0.01:114.4
        A=[x,y;A];
    end
end
A=A';
A1=mapminmax('apply',A,Sin); %数据归一化
```

```
z=sim(net,A1);
z=mapminmax('reverse',z,Sout); %反归一化
X=A(1,:);
Y=A(2,:);
[XX,YY]=meshgrid(22.4:0.01:24.4,112.3:0.01:114.6);
ZZ=griddata(X,Y,z,XX,YY);
surf(XX,YY,ZZ)
shading interp
hold on
```