

“拍照赚钱”的任务定价

摘要

本文针对“拍照赚钱”的任务定价问题建立基于二元非等距插值的多元二项式回归分析模型、基于意愿半径的任务未完成原因分析模型，研究了任务的定价规律并分析任务未完成的原因；建立基于弹性函数的需求导向定价模型、基于支持向量机的完成率评判模型，设计了新的任务定价方案并与原方案比较；建立基于 Q 型聚类分析和非线性规划的定价优化模型，在任务联合打包发布的情况下修改了定价模型并分析对任务完成情况的影响；建立基于修改圈近似算法的旅行商模型、基于模糊层次分析法和熵值法的综合评价模型，确定了新项目的定价方案并评价该方案实施效果。

针对问题一，建立基于二元非等距插值的多元二项式回归分析模型、基于意愿半径的任务未完成原因分析模型，研究了任务的定价规律并分析任务未完成的原因。首先，用二元非等距插值补充数据。其次，建立 4 种定价关于经纬度的多元二项式回归模型，选取剩余标准差最小的模型作为定价函数。再次，引入意愿半径，计算已完成和未完成的任务定价平均值、意愿半径内会员数平均值和会员平均信誉度，通过对比大小关系分析任务未完成原因。最后，用 Q-Q 图检验回归分析模型。灵敏度分析发现，定价极值附近经纬度坐标的轻微变化极大影响定价。定价规律为佛山地区定价最高且狮子洋、珠江流域旁的定价波动显著；任务未完成的原因为任务定价低，周边会员数少且信誉值低。

针对问题二，建立了基于弹性函数的需求导向定价模型、基于支持向量机的完成率评判模型，设计了新的定价方案并与原方案比较。首先，构造与销售范围内会员数、会员的平均信誉度有关的弹性函数，利用需求导向定价法重新定价。其次，用支持向量机预测新方案的任务完成情况。再次，通过比较任务完成率，判断新方案的优劣。最后，构造了新的弹性函数检验模型的定价效果。发现任务的新定价普遍高于原定价，新方案完成情况优于原方案。

针对问题三，建立了基于 Q 型聚类分析和非线性规划的定价优化模型，在任务联合打包情况下修改了定价模型并分析了对任务完成情况的影响。首先，采用 Q 型聚类分析对相邻任务进行打包，并通过 AIC 准则对聚类效果进行检验。其次，引入打包后包内的任务数作为定价因素之一。再次，以任务发布者的佣金成本最小为优化目标，采用非线性规划，修改问题二的定价模型。最后，针对修改后的定价预测完成情况。与问题二对比，发现佣金成本明显下降，完成率略低。

针对问题四，建立了基于修改圈近似算法的旅行商模型、基于模糊层次分析法和熵值法的综合评价模型，确定了新项目的定价方案并评价该方案的实施效果。首先，利用支持向量机预测新项目在各定价方案下的任务情况，选择最佳定价方案。其次，选取佣金成本、最短路程、完成率、时间完成度 4 个评价指标，结合模糊层次分析法和熵值法建立综合评价模型。再次，用相对比较赋权法构建新的综合评价模型进行检验。最后，给予各指标相同的增量进行灵敏度分析。最优方案为问题三的方案，其评价值为 0.7038。

关键词：多元二项式回归 需求导向定价法 支持向量机 Q 型聚类 模糊层析分析

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。用户下载 APP，注册成为 APP 的会员，然后从 APP 上领取需要拍照的任务（比如上超市去检查某种商品的上架情况），赚取 APP 对任务所标定的酬金。这种基于移动互联网的自助式劳务众包平台，为企业提供各种商业检查和信息搜集，相比传统的市场调查方式可以大大节省调查成本，而且有效地保证了调查数据真实性，缩短了调查的周期。因此 APP 成为该平台运行的核心，而 APP 中的任务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，而导致商品检查的失败。

附件一是一个已结束项目的任务数据，包含了每个任务的位置、定价和完成情况（“1”表示完成，“0”表示未完成）；附件二是会员信息数据，包含了会员的位置、信誉值、参考其信誉给出的任务开始预订时间和预订限额，原则上会员信誉越高，越优先开始挑选任务，其配额也就越大（任务分配时实际上是根据预订限额所占比例进行配发）；附件三是一个新的检查项目任务数据，只有任务的位置信息。请完成下面的问题：

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

二、问题分析

2.1 问题一的分析

问题一要求研究附件一中项目的任务定价规律并分析任务未完成的原因。此问题是整个问题的开端和基础。其解题思路将启发本文其余问题的解决方法。

问题一的第一小问要求研究附件一中项目的任务定价规律，由于附件一仅包含任务的位置、定价和完成情况，因此研究项目的任务定价规律即研究任务的定价与经度、纬度之间的数量关系，数量关系用函数表达式体现，故第一小问属于确定函数表达式类问题。此类这类问题通常采用曲线拟合法，多元回归分析法，微分方程等方法解决。本题将采用二元非等距插值扩充数据。故本文采用多元二项式回归确定函数表达式。用此方法解决问题，需要了解各任务的定价、经纬度坐标。故第一小问采用二元非等距插值法扩充数据，建立基于二元非等距插值的多元二项式回归分析模型。

问题一的第二小问要求分析附件一中任务未完成原因。根据廖什的市场区位理论^[1]，产品存在一定的销售范围，而一般将销售范围定义为圆心为产地，半径为最大销售直线距离的圆。根据用户参与众包平台发布任务的实际完成情况，发现用户完成任务的意愿与用户所在位置有很大的关联性。故本题在此基础上采用意愿半径确定任务“销售范围”，建立基于意愿半径的任务未完成原因分析模型。

求解过程：这个求解一般要经过二元非等距插值扩充数据，建立 4 个多元二项式回归模型，选取剩余标准差最小的模型，计算该模型的回归系数，确定任务的定价和经纬度的函数表达式，计算已完成和未完成任务的标价平均值、销售半径内会员数的平均值和会员的平均信誉值，并比较大小关系确定任务未完成的原因等步骤得到。最后，可绘制 Q-Q 或 P-P 图检验多元二项式回归分析的拟合优

度。灵敏度分析：定价极值附近经纬度坐标的轻微变化将极大影响定价。

通过这些分析，得出求解此问题的最佳模型为基于二元非等距插值的多元二项式回归分析模型、基于意愿半径的任务未完成原因分析模型。

2.2 问题二的分析

问题二要求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，此问题是问题一的拓展，又是问题三的基石，在整个赛题中是承上启下的衔接过渡作用。

问题二的第一小问要求制定新的定价方案属于定价类问题，众包平台属于服务型企业，目前市场上服务型产品常用的定价方式为：成本定价法、需求导向定价法和竞争定价法。^[2]由于众包平台“拍照赚钱”不属于完全竞争市场和寡头垄断市场。而竞争定价法又以随行就市法为主，随行就市法主要的适用的市场有两种，为完全竞争市场和寡头垄断市场。^[3]因此本题宜采用需求导向定价法设计新的任务定价方案。用此方法解决问题，需要确定需求价格弹性与销售范围内会员数、会员的平均信誉值的数量关系。通过引入弹性函数的概念，可确定数量关系。

问题二的第二小问要求和原方案进行比较，定义任务的完成率作为衡量定价优劣的标准，而附件一中任务的完成情况由 0-1 整数变量表示，联想到支持向量机，可统计新定价方案下已完成的任务数和未完成的任务数，从而模拟计算新定价方案下的任务完成率。因此本问题采用支持向量机法统计任务完成情况、任务完成率。在此基础上利用弹性函数方法，建立基于弹性函数的需求导向定价模型；将任务完成率作为衡量定价优劣的标准，建立基于支持向量机的任务完成率评判模型。求解过程：先利用需求导向定价法定价，采用支持向量机法统计任务完成情况，将原定价方案和新方案的任务完成率比较。构造新的弹性函数检验模型。

通过这些分析，得出求解此问题的最佳模型为基于弹性函数的需求导向定价模型、基于支持向量机的完成率评判模型。根据实际情况预期新的定价应普遍高于原定价。

2.3 问题三的分析

问题三要求在位置较集中的任务联合打包发布的情况下，修改定价模型并分析对任务完成情况影响，此问题是问题二的延伸问题，需用到问题二建立的模型。并且新制定的定价方案将作为问题四中新项目的定价方案的备选方案之一。

问题三的综合性强，首先需要解决所有任务如何打包的分类问题。聚类分析是多元统计学中进行定量分类的有效手段，主要有 Q 型和 R 型聚类分析两种，Q 型聚类分析依据样本间的相似度实现样本的分类，而 R 型聚类分析依据同一样本的不同变量的亲疏关系对变量进行分类。^[4]而本题的目标是对不同任务进行打包分类，故采用 Q 型聚类分析对任务打包分类。

其次，需要解决在任务联合打包后修改定价的问题，此问题在问题二已建立的基于弹性函数的需求导向定价模型的基础上，引入联合打包后包内的任务数，考虑包内任务数对任务定价的影响，从而修改定价。最后，为优化模型，可以从任务发布者角度考虑，以任务发布平台的酬金成本最小为目标，采用非线性规划进行优化。求解过程：采用 Q 型聚类分析对任务打包，确定修改后的标价计算表达式，确定非线性规划模型优化定价方案。利用 AIC 准则对聚类进行了检验。

通过以上分析，得出求解此问题的最佳模型为基于 Q 型聚类分析和非线性规划的定价优化模型。预期结果为新的标价将使得任务发布平台的酬金成本最小。

2.4 问题四的分析

问题四要求确定附件三中新项目的任务定价方案并评价该方案实施效果。为

新项目选定定价方案需根据上面三个问题的结果，故此问题是整个赛题的总结，并提出“对于新项目，该如何制定定价方案”的一般性问题，这与实际情况联系密切，为众包平台“拍照赚钱”任务的发展和推广，提出了制定价格的示范方案。

问题一的第一小问为新项目选定定价方案，利用问题二中基于支持向量机的完成率评判模型，计算新项目在各自的定价方案下任务的完成率，选定完成率高的定价方案即可。问题四的第二小问要求评价该方案实施效果，属于综合评价类问题。这类问题通常采用层次分析法，熵值法，模糊评价法等方法解决。传统的层次分析法存在判断矩阵一致性检验易引起思维错误、检验准则客观科学性差等缺点。^[5]于是本文采用了模糊层次分析法。而模糊层次分析法并未解决层次分析法在赋权过程中主观随意性较强的缺点，故将模糊层次分析法和熵值法相结合建立综合评级模型。

用上述方法解决问题，需要知道在该定价方案下，平台酬金成本、包内任务的最短路程、任务的完成率和时间完成度 4 个评价指标的指标值。其中任务数最大的包内完成所有任务的最短路程的指标值，需要建立基于修改圈近似算法的旅行商模型。求解过程：选取平台的酬金成本、任务数最大的包内任务的最短路程、任务的完成率和时间完成度 4 个评价指标，模糊层次分析法计算权值，熵值法计算差异性系数，计算组合权重。并采用相对比较赋权法构建新的综合评价模型对已建的模型检验。分别给予 4 个指标相同增量，比较对评价值影响大小进行灵敏度分析。

通过这些分析，得出求解此问题的最佳模型为基于修改圈近似算法的旅行商模型、基于模糊层次分析法和熵值法的综合评价模型。根据实际情况预期选用问题三中基于 Q 型聚类分析非线性规划定价法。

三、基本假设

- (1) 假设所有数据真实可信。
- (2) 由于附件一仅包含任务的位置、定价和完成情况，因此假设附件一中项目的任务定价仅与任务的经度、纬度有关。
- (3) 假设本题意愿半径 r 为常数， $r = 3$ 千米。
- (4) 假设任务联合打包后包内的任务数 m 与定价之间属于非线性相关。
- (5) 本题定价方案不考虑任务的难易程度、竞争因素。

四、符号说明

- x : 任务的经度坐标。
 y : 任务的维度坐标。
 c : 附件一中任务的标价。
 r : 意愿半径，取值为 3 千米。
 n : 销售范围内会员数。
 s : 销售范围内会员的平均信誉值。
 \bar{c}_1 : 已完成的任务组所有任务的标价平均值。
 \bar{c}_2 : 未完成的任务组所有任务的标价平均值。
 \bar{n}_1 : 已完成的任务组所有任务的任务销售半径内会员数的平均值。
 \bar{n}_2 : 未完成的任务组所有任务的任务销售半径内会员数的平均值。
 \bar{s}_1 : 已完成的任务组所有任务的任务销售半径内会员的平均信誉值。

\bar{s}_2 : 未完成的任务组所有任务的任务销售半径内会员的平均信誉值。

c' : 问题二中各任务新的标价。

E : 需求价格弹性。

g : 调价率。

μ : 任务的完成率。

a : 已完成的任务数。

b : 未完成的任务数。

μ_1 : 附件一中原定价方案的任务完成率。

μ_2 : 附件一中新定价方案的任务完成率。

c'' : 问题三中各任务新的标价。

m : 任务联合打包后包内的任务数。

p : 任务发布平台的佣金成本。

L : 任务数最大的包内完成所有任务的最短路程。

T : 时间完成度。

N : 总的任务数。

d_{max} : 任务数最大的包内两个任务标点间距离的最大值。

t_i : 任务 i 是否完成的 0-1 变量。

五、模型建立与求解

5.1 问题一的模型建立和求解

通过建立基于二元非等距插值的多元二项式回归分析模型、基于意愿半径的任务未完成原因分析模型, 解决研究附件一中项目的任务定价规律并分析任务未完成原因的问题。最后绘制 Q-Q 图检验多元二项式回归分析模型。灵敏度分析: 定价极值附近经纬度坐标的轻微变化将极大影响定价。

5.1.1 模型建立

(1) 建模准备

1) 了解众包的相关概念

众包的起源^[6]: 新的“众包模式”思想来源于加拿大 Goldcorp 公司的老板。

众包的定义: 众包^[7]是指企业或公共机构将原本应由本单位员工完成的工作, 外包给社会人群自发自愿地处理并解决。

2) 绘制任务的地理位置分布情况图

为从附件一中的每个任务的经度和纬度信息研究任务的地理位置分布情况, 现分别绘制全部任务、已完成的任务、未完成的任务在全国地图上的大致分布情况图, 通过网络工具导入附件一中的数据得到此图, 对下图进行观察, 不难得出: 附件一中的任务主要分布在佛山、广州、东莞和深圳四市。

任务分布情况图如下:

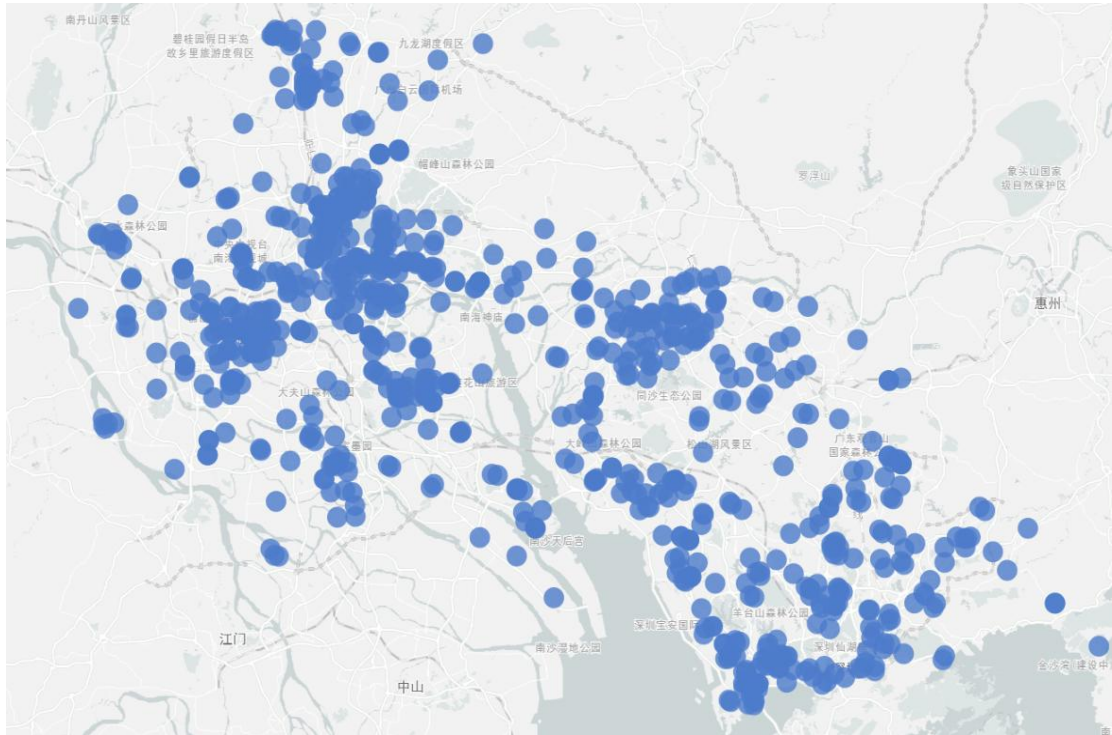


图 5.1.1.1 全部任务在全国的分布情况图

通过未完成任务分布图可以看出：东莞地区的任务基本都被完成，未完成任务几乎没有。该地区的任务完成率比广州、深圳两地高很多。

(2) 建立基于二元非等距插值的多元二项式回归分析模型

定理 5.1.1: 在仅已知任务的位置、标价和完成情况的情况下研究项目的任务定价规律，假定任务定价仅与任务的经度、纬度有关，先采用二元非等距插值扩充任务的标价、经度和纬度 3 方面的数据，并采用多元二项式回归确定任务定价与经度、纬度之间的具体函数关系式，表达式如下：

$$c = 255273.297 - 3996.15x - 2470.67y + 19.3xy + 15.65x^2 + 6.14y^2$$

(其中： x 、 y 、 c 分别代表任务的经度、纬度、定价)，即通过建立基于二元非等距插值的多元二项式回归分析模型可研究项目的任务定价规律。

证明：

由于附件一仅包含任务的位置、定价和完成情况，并且全国各地区的经济状况不同，劳动力的酬金也存在显著差异。因此为了研究项目的任务定价规律，假设项目的任务定价仅与任务的经度、纬度有关。

1) 二元非等距插值扩充数据

以附件一中项目的全部任务的经维度为自变量，任务标价为因变量，绘制散点图。由于数据有限，散点图无法体现任务的定价、经维度之间的数量关系。而插值^[8]能够利用已知离散点确定在其余未知点处的取值。因此采用二元非等距插值扩充任务的标价、经维度 3 方面的数据。

2) 多元二项式回归分析

为了研究项目的任务定价规律，将任务的经纬度作为自变量，定价为因变量，采用多元回归分析的方法确定函数关系式。由于因变量和自变量之间是否存在线性关系未知，故而舍弃传统的多元线性回归法，采用多元二项式回归法。

多元二项式回归的建模步骤如下：

①建立如下 4 个多元回归模型^[9]:

$$\text{线性: } y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m,$$

$$\text{纯二次: } y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m + \sum_{j=1}^m \beta_{jj} x_j^2,$$

$$\text{交叉: } y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m + \sum_{1 \leq j < k \leq m} \beta_{jk} x_j x_k,$$

$$\text{完全二次: } y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m + \sum_{1 \leq j \leq k \leq m} \beta_{jk} x_j x_k.$$

②带入数据拟合, 计算 4 个模型的剩余标准差。剩余标准差最小的模型最好。

③选定最佳回归模型后, 计算回归系数, 确定具体表达式。

在多元二项式回归确定函数表达式后, 即可研究任务定价和经纬度间的规律。

综上, 已证明上述定理。

(3) 建立基于意愿半径的任务未完成原因分析模型

定理 5.1.2: 在已知任务标价、位置、完成情况和会员的位置、信誉值的条件下, 首先以任务位置为圆心和意愿半径做出任务“销售范围”。其次, 分别计算已完成和未完成的任务的标价平均值 \bar{c}_1 、 \bar{c}_2 、任务意愿半径内会员数平均值 \bar{n}_1 、 \bar{n}_2 和会员的平均信誉值 \bar{s}_1 、 \bar{s}_2 。最后, 比较 \bar{c}_1 、 \bar{c}_2 , \bar{n}_1 、 \bar{n}_2 , \bar{s}_1 、 \bar{s}_2 两两间的大小关系, 即可分析出任务未完成的原因。

证明:

根据廖什的市场区位理论^[1], 最初, 每个服务型任务的服务范围是以发布地为圆心, 以一定的距离为半径所形成的圆。本题发现用户完成任务意愿与用户所在位置有很大关联性, 一般用户距离任务越近, 用户完成任务的意愿越高, 即若将用户完成任务视作众包平台的“任务产品成功销售”, 则任务的完成也是存在一定的销售范围。该题销售范围取 3 千米。

首先, 以各项任务为圆心, 作 3 千米意愿半径, 计算销售范围内会员数 n , 并且根据意愿半径内的会员的信誉值计算销售范围内会员的平均信誉值 s 。然后按任务完成情况对附件一中的任务进行分类。并计算已完成和未完成任务标价平均值 \bar{c}_1 、 \bar{c}_2 , 平均会员数 \bar{n}_1 、 \bar{n}_2 , 平均信誉值 \bar{s}_1 、 \bar{s}_2 。比较 \bar{c}_1 、 \bar{c}_2 , \bar{n}_1 、 \bar{n}_2 , \bar{s}_1 、 \bar{s}_2 大小, 研究任务任务未完成的原因。

综上, 已证明上述定理。

5.1.2 模型求解

(1) 求解基于二元非等距插值的多元二项式回归分析模型

1) 二元非等距插值扩充数据

设任务的经度 x 和维度 y 为自变量, 标价 c 为因变量, 采用二元非等距插值扩充任务的标价、经度和维度 3 方面的数据, 并绘制任务的标价关于经度和维度的三维曲面图。

2) 多元二项式回归分析

①建立 4 个多元回归模型。

②代入数据, 分别计算 4 个模型的剩余标准差, 选择剩余标准差最小的模型。

③选定最佳回归模型后，计算回归系数，确定回归模型的具体表达式。

3) 研究项目的任务定价规律

在利用多元二项式回归确定模型的具体表达式后，研究任务定价与经纬度之间的数量关系，即研究项目的任务定价规律。

(2) 求解基于意愿半径的任务未完成原因分析模型

1) 计算销售范围内会员数 n 和会员的平均信誉值 s

以每个任务为主体，根据附件二中会员的位置计算销售范围内会员数 n ，并且根据意愿半径内的会员的信誉值计算销售范围内会员的平均信誉值 s 。

2) 按任务完成情况对附件一中的任务进行分类

按任务完成情况分为已完成的任务组和未完成的任务组。

3) 对已完成的任务组和未完成的任务组分别计算

对已完成的任务组和未完成的任务组分别计算所有任务的标价平均值

\bar{c}_1 、 \bar{c}_2 ，意愿半径内会员数的平均值 \bar{n}_1 、 \bar{n}_2 ，会员的平均信誉值 \bar{s}_1 、 \bar{s}_2 。

4) 分析任务未完成原因

比较 \bar{c}_1 、 \bar{c}_2 ， \bar{n}_1 、 \bar{n}_2 ， \bar{s}_1 、 \bar{s}_2 的大小关系，研究任务任务未完成的原因。

5.1.3 模型检验与分析

Q-Q图是检验有关拟合类问题拟合优度的一种成熟、简单和直观的方法，其基本思想^[9]是：在直角坐标系上分别绘制经验分布函数和分布模型的分位数点

（其中分布模型的分位数点称为**理论分位数点**），将分位数点和对应的理论分位数点作为一堆数组，若多个观测数据对应的点在坐标系中近似分布于一条直线，则说明模型精度高。本文采用**Q-Q图**检验回归分析的拟合优度。

5.1.4 灵敏度分析

利用已确定的任务定价与经纬度之间的函数表达式，求解任务定价极值的经纬度位置。在定价极值附近，给经纬度坐标一个轻微增量，将极大影响定价。

5.1.5 问题答案

(1) 基于二元非等距插值的多元二项式回归分析模型的答案

1) 二元非等距插值扩充数据拟合作三维图

绘制任务的标价 c 关于经度 x 和维度 y 的三维曲面图

起先，将附件一中给出数据绘制散点图，发现效果并不理想，于是对数据插值再绘制任务定价关于经纬度的三维图。此图中， x,y 坐标轴为任务的经纬度， z 轴为任务的定价。由图形可看出：定价与任务所处位置有一定的关系，结合下面的等高线图，能够更清晰的看出各地方任务定价的高低。

定价与任务所处位置有一定关系，结合三维图和等高线图可发现：纬度23.3~23.7，经度112.7~112.8范围（即佛山地区）内，定价最高。而在广州和深圳之间的狮子洋、珠江流域两旁的任务定价大致相同，但变化波动明显，邻近的地区定价也有些微差异。其他地区定价平滑，变化较小。

图形如下所示：

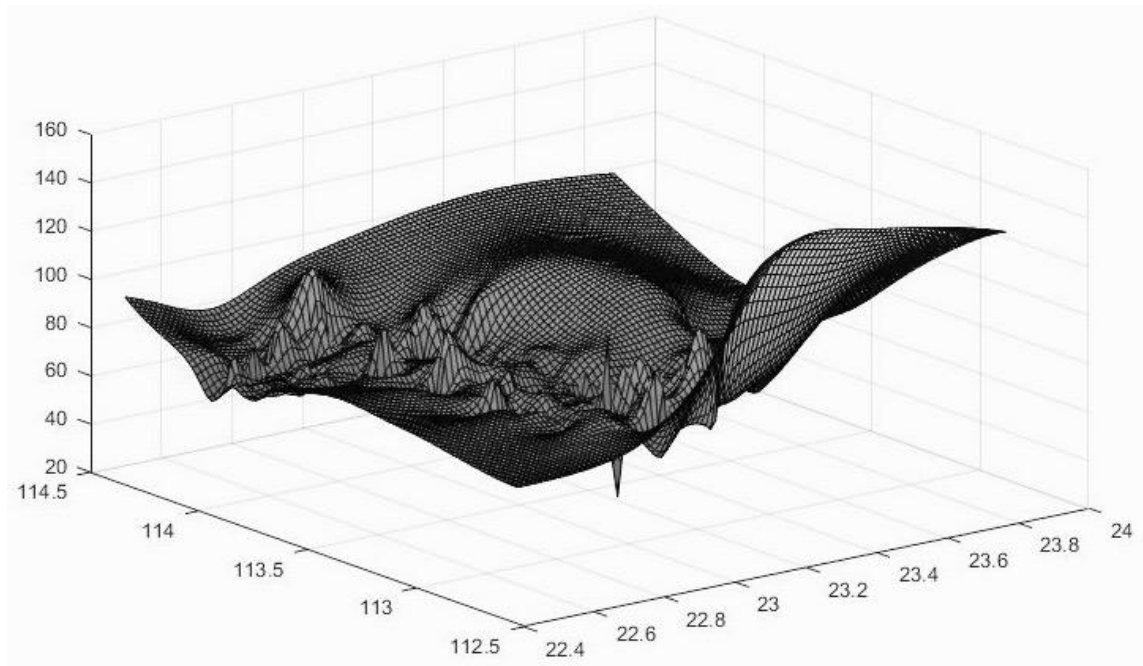


图 5.1.5.1 插值后的标价关于经度、纬度的三维曲面图

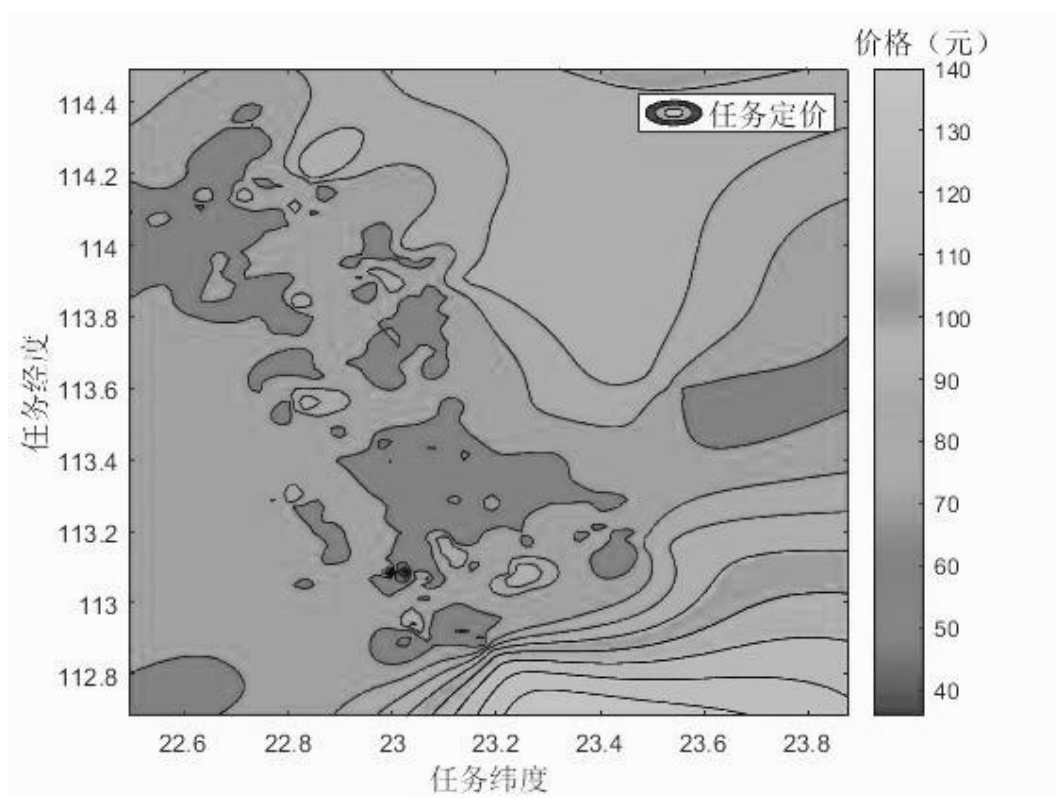


图 5.1.5.2 插值后的任务定价关于经度、纬度的等高线图

2) 多元二项式回归分析

① 带入数据拟合，分别计算 4 个模型的剩余标准差

带入数据拟合，分别计算 4 个模型的剩余标准差，结果如下：

表 5.1.5.1 4 个模型对应的剩余标准差

模型类型	线性多项式	纯二次多项式	交叉多项式	完全二次多项式
剩余标准差	4.4845	4.2854	4.4819	4.1652

由上表可见，多元完全二次多项式回归模型在上述 4 个模型中剩余标准差最小，故应选用多元完全二次二项式回归模型。

②计算回归系数，确定相应回归模型的具体表达式

计算多元完全二次二项式回归模型的回归系数，确定回归模型的表达式如下：

$$c = 255273.297 - 3996.15x - 2470.67y + 19.3xy + 15.65x^2 + 6.14y^2,$$

3) 研究项目的任务定价规律

使用上述表达式研究任务定价与经度、纬度之间的关系，发现任务定价与经度、纬度之间大致存在如下数量关系：

①纬度 23.3~23.7，经度 112.7~112.8 范围内，定价最高。

而在广州和深圳之间的狮子洋、珠江流域两旁的任务定价大致相同，但变化幅度明显，邻近的地区定价也有些微差异。其他地区定价平滑，变化较小。

②在经纬度坐标 (22.5, 114.4) 和 (23.2, 112.9) 两点间形成的带状区域内，定价波动明显，波动范围约为 50~80 元。

③在经纬度坐标 (23, 114.4)、(23.4, 113.6) 和 (23.9, 114.4) 三点间形成的片状区域内，定价均在 100 元左右。

根据上述数量关系，分析出以下的任务定价规律：

①在附件一现有数据中，佛山地区内，任务定价最高

②在广州和深圳之间的狮子洋、珠江流域两旁的任务定价波动明显，但大致分布在一定范围内，邻近的地区定价都存在些微差异。

③在东莞地区，任务定价偏高，均在 100 左右。

(2) 基于意愿半径的任务未完成原因分析模型的答案

1) 计算 \bar{c}_1 、 \bar{c}_2 ， \bar{n}_1 、 \bar{n}_2 和 \bar{s}_1 、 \bar{s}_2

表 5.1.5.1 完成和未完成任务的三个指标值

	平均价格 \bar{c}_1	意愿半径内平均会员数 \bar{n}_1	意愿半径内平均信誉值 \bar{s}_1
完成的任务	69.8199233	18.0440613	219.69856
未完成任务	67.9281150	17.5632183	152.90809

通过表格，可明显看出：完成任务的平均定价稍高于未完成任务；完成任务的意愿半径内平均会员数也多于未完成任务；而完成任务的平均信誉值远高于未完成任务。任务未完成的原因：任务定价低、任务周边会员人数少、且周边会员平均荣誉值低。

2) 分析任务未完成原因

比较已完成的任务和未完成的任务 \bar{c}_1 、 \bar{c}_2 ， \bar{n}_1 、 \bar{n}_2 ， \bar{s}_1 、 \bar{s}_2 的大小关系，发现：前者均大于后者。分析出任务未完成的原因如下：

①任务定价过低；

②任务周边范围内会员数偏少；

③任务周边范围内会员平均信誉值较低。

(3) 灵敏度分析结果

任务定价的极大值的地理位置主要分布在纬度 23.3~23.7，经度 112.7~112.8 范围内，在此地理范围内经纬度坐标的改变将会极大地影响任务的定价。

(4) Q-Q 图检验回归分析的拟合优度

为检验基于二元非等距插值的多元二项式回归分析模型，绘制 Q-Q 图检验

拟合优度。所得的 Q-Q 图如下所示：

由图可分析得：Q-Q 图中的所有数据散点分布在 45 度角的斜线周围，所以完全二次二项式回归效果较好，故模型整体上通过检验。

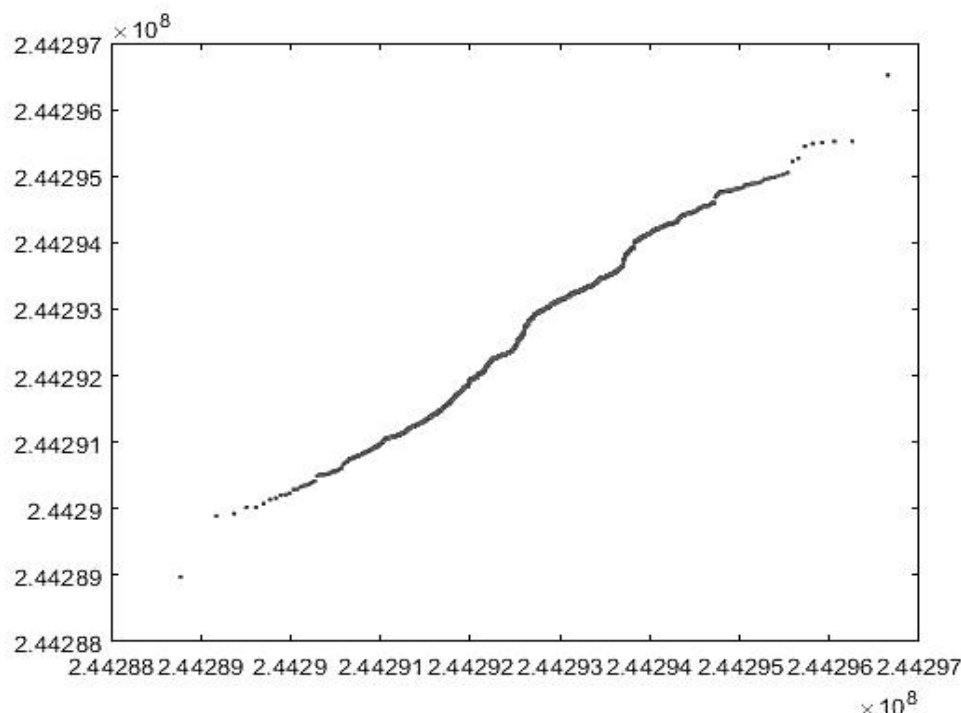


图 5.1.5.3 多元完全二次二项式回归模型检验的 Q-Q 图

5.1.6 问题总结

本节通过建立了基于二元非等距插值的多元二项式回归分析模型、基于意愿半径的任务未完成原因分析模型，解决了研究任务定价规律并分析任务未完成原因的问题。采用二元非等距插值扩充数据、多元二项式回归分析法求解，基于意愿半径确定任务的“销售范围”。最后绘制 Q-Q 图检验。灵敏度分析：定价极值附近经纬度坐标的轻微变化将极大影响定价。求得答案：（1）定价规律：1）佛山地区内，任务定价最高；2）狮子洋、珠江流域两旁的任务定价波动明显；3）东莞地区，任务定价偏高。（2）任务未完成的原因：1）任务定价低；2）任务周边范围内会员数少；3）任务会员平均信誉值低。

本节的创新之处在于：（1）在数据不足的情况下，采用二元非等距插值扩充数据，方便后续计算。（2）在定价和经纬度间是否存在线性关系未知时，舍弃传统的多元线性回归法，而采用多元二项式回归法，避免了出现拟合精度低、模型适用性弱等缺点。（3）在确定任务未完成原因时，认为任务的完成存在一定的销售范围。本文将用户完成任务的意愿半径作为销售半径，建模思维极富创造性。

5.2 问题二的模型建立和求解

通过建立基于弹性函数的需求导向定价模型、基于支持向量机的完成率评判模型，解决为附件一中的项目设计新定价方案并和原方案进行比较的问题。最后，构造新的弹性函数检验需求导向定价模型。这样就能求得合理的问题答案。

5.2.1 模型建立

（1）建立基于弹性函数的需求导向定价模型

定理 5.2.1： 在已知各任务销售范围内会员数 n 和会员的平均信誉值 s 的情况

下，附件一中各任务新的标价为 c' 可按以下公式确定：

$$c' = \frac{1}{2} \left(3 - \frac{1}{s + \frac{1}{n}} \right) c。$$

证明：

众包平台属于服务型企业，现有服务型和有形两种产品，在产品性质上，两者存在很大差异，故两者的都有各自常用的定价方式。^[10]目前市场上服务型产品常用的定价方式为：成本定价法、需求导向定价法和竞争定价法。^[2]

由于众包平台“拍照赚钱”的任务形式发展阶段，不属于完全竞争市场和寡头垄断市场。而竞争定价法又以随行就市法为主，随行就市法主要的适用的市场有两种，为完全竞争市场和寡头垄断市场。^[3]故也不采用竞争定价法。因此本将采用需求导向定价法为附件一中的项目设计新的任务定价方案。

1) 确定新的标价与需求价格弹性、调整前的价格的数量关系

需求导向定价法的价格计算公式推导过程如下：

设附件一中各任务新的标价为 c' ，需求价格弹性为 E ，附件一中各的标价 c 为调整前的价格，调价率为 g ，则调价率和需求价格弹性存在以下等量关系^[11]：

$$g = \frac{1}{2} \left(1 - \frac{1}{|E|} \right)。$$

由于附件一中任务新的标价、调整前的价格和调价率存在以下等量关系：

$$c' = (1 + g)c。$$

结合以上两个公式确定新的标价与需求价格弹性、调整前的价格的关系为：

$$c' = \frac{1}{2} \left(3 - \frac{1}{|E|} \right) c。$$

2) 确定需求价格弹性与销售范围内会员数、会员的平均信誉值的数量关系

为附件一中的项目设计更佳的任务定价方案，将考虑各任务销售范围内会员数 n 、会员的平均信誉值 s 两个因素对各任务新的标价 c' 的影响。为激励会员完成任务，各任务新的标价 c' 必然和销售范围内会员的平均信誉值 s 成正比。为避免“供不应求”现象，新的标价 c' 必然和销售范围内会员数 n 成反比，即：

$$c' \propto s, \frac{1}{n},$$

而根据公式：

$$c' = \frac{1}{2} \left(3 - \frac{1}{|E|} \right) c,$$

可得新定价 c' 正比于 $|E|$ ， $|E|$ 正比于 $s, \frac{1}{n}$ ，即：

$$c' \propto |E|, |E| \propto s, \frac{1}{n},$$

人们对于非基本生活必需品的需求量随着价格的变动会产生较大变化，即人们对此类产品的需求属于弹性需求。“拍照赚钱”任务作为非基本生活必需品，则人们对“拍照赚钱”任务的需求也属于弹性需求。根据上述研究结果，不放构造弹性需求函数，如下所示：

$$|E| = s + \frac{C}{n},$$

上式中 C 为常数，本文取值为1，即：

$$|E| = s + \frac{1}{n}。$$

3) 确定新的标价的具体计算公式

根据公式：

$$\begin{cases} c' = \frac{1}{2} \left(3 - \frac{1}{|E|} \right) c, \\ |E| = s + \frac{1}{n} \end{cases},$$

可以确定新的标价 c' 与销售范围内会员数 n 、会员的平均信誉值 s 的等量关系：

$$c' = \frac{1}{2} \left(3 - \frac{1}{s + \frac{1}{n}} \right) c。$$

综上，已证明在已知各任务销售范围内会员数 n 和会员的平均信誉值 s 的情况下，附件一中各任务新的标价为 c' 可按以下公式确定：

$$c' = \frac{1}{2} \left(3 - \frac{1}{s + \frac{1}{n}} \right) c。$$

(2) 建立基于支持向量机的完成率评判模型

定理 5.2.2: 在利用只求向量机输出在新的定价方案下任务的完成情况，即已知新定价方案下已完成的任务数和未完成的任务数，根据任务完成率的计算公式：

$$\mu = \frac{a}{b},$$

分别计算原定价方案和新定价方案的任务完成率 μ_1 、 μ_2 ，比较 μ_1 、 μ_2 的大小关系，即可确定新的定价方案与原方案相比的优劣情况。

证明：

1) 采用支持向量机确定新的定价方案下任务的完成情况

对于样本数据量有限的情况下模式识别中存在的本质性问题，支持向量机已有了系统性的理论研究成果，其结构风险最小化准则不仅实现了样本点误差值最小，而且在模型的泛化能力上有了极大的突破，与传统地机器学习相比，克服了“过学习”、“非线性”、“维数灾难”等缺点。^[12]

因此本问题将采用支持向量机法确定在新的任务定价方案下附件一中任务的完成情况。采用支持向量机确定任务完成情况的步骤如下：

①确定训练集

假设训练集 T ，含有 l 个样本 $\{\alpha_i, y_i\}$ ， $T \in (\Omega \times Y)^l$ ， $i = 1, \dots, l$ 。其中， Ω 为输入空间， $\alpha_i \in \Omega \in R^l$ ， $y_i \in Y = \{0, 1\}$ ， Y 为输出空间，则 α_i 与 y_i 分别为第 i 个输入与输出模式。^[13]

②训练支持向量机分类器

支持向量机的核心为训练支持向量机分类器,利用线性内核函数的支持向量机模型进行分类,求得支持向量和线性分类函数。

③用训练好的支持向量机预测新的任务定价方案下的任务完成情况。

2) 定义任务完成率比较两种定价方案

设任务的完成率 μ ,已完成的任务数为 a ,未完成的任务数为 b ,则任务的完成率的计算公式如下:

$$\mu = \frac{a}{b}。$$

分别计算附件一原定价方案和新定价方案的任务完成率,其中新定价方案下已完成的任务数和未完成的任务数来自于支持向量机的输出。

比较两种定价方案任务完成率的大小关系,确定新定价方案与原方案的优劣。

综上,已证明在利用支持向量机输出在新定价方案下任务的完成情况,即已知新定价方案下已完成和未完成的任务数,根据任务完成率的计算公式:

$$\mu = \frac{a}{b},$$

分别计算原定价方案和新定价方案的任务完成率 μ_1 、 μ_2 ,比较 μ_1 、 μ_2 的大小关系,即可确定新的定价方案与原方案相比的优劣情况。

5.2.2 模型求解

(1) 求解基于弹性函数的需求导向定价模型

利用问题一求得的任务的意愿半径内会员数 n 和会员的平均信誉值 s ,按公式:

$$c' = \frac{1}{2} \left(3 - \frac{1}{s + \frac{1}{n}} \right) c$$

确定各任务新定价 c' 。

(2) 求解基于支持向量机的完成率评判模型

1) 支持向量机求解新的定价方案下任务的完成情况

①确定训练集

训练集中含有 835 个样本数据,输入空间由 835 个任务样本的价格 c 、销售范围内会员数 n 、会员的平均信誉值 s 方面属性组成。

②训练支持向量机分类器

利用线性内核函数的支持向量机模型,求得支持向量和线性分类函数。

③预测新的任务定价方案下的任务完成情况

训练好的支持向量机输出值为 0 表示任务未完成,输出值为 1 表示任务已完成。统计新定价方案下已完成的任务数 a 和未完成的任务数 b 。

2) 计算任务完成率比较两种定价方案

利用任务的完成率的计算公式如下:

$$\mu = \frac{a}{b}。$$

分别计算附件一原定价方案和新定价方案的任务完成率 μ_1 、 μ_2 ,比较 μ_1 、 μ_2 的大小关系,确定新的定价方案与原方案相比的优劣情况。

5.2.3 模型检验与分析

由于本文在模型建立中构造弹性函数 $|E|$ 时使用的是 s 和 $\frac{1}{n}$ 的算数和的形式,

可使用 s 和 $\frac{1}{n}$ 的乘积形式构造新的弹性函数，构造新的弹性函数表达式如下：

$$|E| = \frac{s}{n}。$$

在此基础上制定新的定价方案，利用基于支持向量机的完成率评判模型，计算此方案下的任务完成率 μ'_2 ，比较 μ_1 、 μ_2 、 μ'_2 的大小关系，检验需求导向定价模型。

5.2.4 问题答案

(1) 新的任务定价方案

新的定价方案为：

$$c' = \frac{1}{2} \left(3 - \frac{1}{s + \frac{1}{n}} \right) c$$

上式中， c' 为各任务新的标价， n 为任务的销售范围内会员数， s 为平均信誉值。

(2) 和原方案比较的结果

1) 新的任务定价方案下的任务完成情况

统计新定价方案下已完成的任务数 a 和未完成的任务数 b 结果如下：

$$a = 781、b = 54。$$

2) 任务完成率的比较结果

计算并比较所得附件一原方案和新定价方案的任务完成率 μ_1 、 μ_2 结果如下：

$$\mu_1 = 62.51\%、\mu_2 = 93.53\%，$$

$$\mu_1 < \mu_2。$$

发现在新的定价方案下，任务的完成率显著提高，故新定价方案优于原方案。

(3) 模型检验的结果

在新的弹性函数表达式的基础上计算所得的任务完成率结果如下：

$$\mu'_2 = 87.07\%，$$

则 μ_1 、 μ_2 、 μ'_2 的大小关系如下：

$$\mu_1 < \mu'_2 < \mu_2。$$

结果表明，新的弹性函数制定的定价方案的任务完成率仍高于附件一中的原定定价方案，但低于原弹性函数制定的定价方案的任务完成率，说明及建立的基于弹性函数的需求导向定价模型定价效果极好，模型整体上通过了检验。

5.2.5 问题总结

本节通过建立了基于弹性函数的需求导向定价模型、基于支持向量机的完成率评判模型，解决了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的问题。最后，构造了新的弹性函数并利用基于支持向量机的完成率评判模型计算此定价方案下的任务完成率，从而检验了基于弹性函数的需求导向定价模型。

(和灵敏度分析)。求得合理的问题答案：(1)概括新的定价方案的计算公式如下：

$$c' = \frac{1}{2} \left(3 - \frac{1}{s + \frac{1}{n}} \right) c$$

上式中, c' 为各任务新的标价, n 为各任务的销售范围内会员数, s 为会员的平均信誉值。(2) 附件一原定价方案的任务完成率小于新定价方案的任务完成率, 说明新的定价方案优于原方案。(3) 经检验后发现模型的定价效果极好。本节的创新之处在于: 为研究各任务销售范围内会员数、会员的平均信誉值两个因素对各任务新的标价的影响, 引入弹性函数的概念, 具体确定了如何从会员数、会员的平均信誉值两个方面具体计算定价方案。建模思维活跃且极富创造力, 根据给定的数据确定每个任务新的标价的计算公式。

5.3 问题三的模型建立和求解

通过建立基于 Q 型聚类分析和非线性规划的定价优化模型, 解决在位置较集中的任务联合打包发布的情况下修改定价模型并分析对任务完成情况影响的问题。最后用 AIC 准则对已建模型进行检验。这样就能求得合理的问题答案。

5.3.1 模型建立

建立基于 Q 型聚类分析和非线性规划的定价优化模型

定理 5.3.1.1: 在任务打包的情况下, 已知各任务销售范围内会员数 n 、会员的平均信誉度 s 和任务打包后包内的任务数为 m , 任务的定价模型调整为:

$$c'' = \frac{1}{2} \left(3 - \frac{a_1 m^2 + a_2 m + a_3}{s + \frac{1}{n}} \right)$$

证明:

(1) 采用 Q 型聚类分析对位置较集中的任务联合打包

聚类分析是多元统计学中进行定量分类的有效手段, 主要有 Q 型和 R 型聚类分析两种, 两种方法的分类对象不同, Q 型聚类分析依据样本间的相似度实现样本的分类, 而 R 型聚类分析依据同一样本的不同变量的亲疏关系对变量进行分类。^[4]而本题的目标是对不同任务进行打包分类, 故采用 Q 型聚类分析对位置较集中的任务联合打包。

由于题意“对位置较集中的任务联合打包发布”, 自然地, 采用距离作为样本间相似度的衡量标准。对不同任务聚类要求计算两个任务点间的距离, 采用欧几里得距离计算, 计算公式^[9]如下:

$$d_2(x, y) = \left[\sum_{k=1}^p |x_k - y_k|^2 \right]^{\frac{1}{2}}。$$

通过枚举法确定聚类数, 选定具体聚类数的准则为保证每个任务包中的任务数小于 15 个。选定包内任务数不超过 15 个的原因是任务数过多必然会延长任务的完成时间, 以及降低最终完成任务的数目。

通过上述 Q 型聚类分析, 能够解决所有任务按照地理位置该如何打包, 以及每个包内含有多少任务数的问题。

(2) 在任务联合打包后修改定价

设任务联合打包后包内的任务数为 m , 包内的任务数为 m 越多, 任务完成的时间必然越长, 故为提升任务的完成效率, 在任务联合打包的情况下新的任务标

价必然与 m 成反比关系，故在本问题中 m 影响新的定价 c'' 。

在上述讨论后制定的新的标价 c'' 是关于经度坐标 x 、维度坐标 y 、各任务的
销售范围内会员数 n 、包内的任务数为 m 的函数，设为：

$$c'' = g(x, y, n, s, m)。$$

下面研究新的标价 c'' 的具体函数表达式：

①设 m 对 c'' 影响权重的表达式为： $a_1 m^2 + a_2 m + a_3$ ，即单调性中途发生改变。由于 c'' 与 m 成反比关系， $m > 0$ ，因此可对问题二的定价模型进行优化，
将 m 表达式代入问题二的模型中，作为新的定价 c'' 的因变量之一。。

②又在问题二的定价模型中，调价率 g 中的弹性变量前的符号为负，为满足
所以将 $a_1 m^2 + a_2 m + a_3$ 放在需求价格弹性绝对值 $|E|$ 所在的分数分子位置。

③将 $a_1 m^2 + a_2 m + a_3$ 代入问题二的定价模型

$$\begin{cases} c' = \frac{1}{2} \left(3 - \frac{1}{|E|} \right) c \\ |E| = s + \frac{1}{n} \end{cases},$$

确定新的定价 c'' 的具体函数表达式：

$$c'' = \frac{1}{2} \left(3 - \frac{a_1 m^2 + a_2 m + a_3}{s + \frac{1}{n}} \right) c。$$

5.3.2 模型求解

(1) 对位置较集中的任务联合打包

采用 Q 型聚类分析对位置较集中的任务联合打包，利用欧几里得距离计算
两个任务点之间的位置关系，枚举法确定聚类数，选定具体聚类数的准则为保证
每个任务包中的任务数小于 15 个，解决所有任务按照地理位置该如何打包，以
及每个包内含有多少任务数 m 的问题。

(2) 以任务发布平台的酬金成本 p 为规划目标，进行非线性规划，求出最化
的情况下的最优系数 a_1 、 a_2 、 a_3 和 p 的最优解。

酬金成本 p 的表达式为：

$$p = \sum_{i=1}^N c_i,$$

上式中 c_i 表示各任务的标价， N 表示总的任务数。

非线性目标规划为：

$$\begin{aligned} & \min p \\ \text{s.t.} & \begin{cases} a_1 < 0 \\ -\frac{b}{2a} > 0 \\ a_1 + a_2 + a_3 = 1 \end{cases} \end{aligned}$$

或

$$\min p$$

$$s. t. \begin{cases} a_1 > 0 \\ b^2 - 4ac \leq 0 \\ a_1 + a_2 + a_3 = 1 \end{cases}$$

将上述求得最优化系数值代入：

$$c'' = \frac{1}{2} \left(3 - \frac{a_1 m^2 + a_2 m + a_3}{s + \frac{1}{n}} \right) c。$$

得到新的定价 c'' 的具体函数表达式。

(3) 在任务联合打包发布的情况下，用新的定价方案 c'' 计算出各任务的具体定价。以任务打包后所属包内（同一类）里任务的所有任务意愿半径 r 内会员数 n 的平均值代替原来个体的 n ，同理以任务打包后所属包内（同一类）里任务的所有任务意愿半径 r 内会员平均信誉值 s 的平均值代替原来个体的 s ，和打包后各任务的具体定价一起代入问题二中训练好的支持向量机，对该情况下的任务完成情况进行预测，计算出在任务联合打包发布的情况下的任务完成率 μ_3 和任务发布平台的酬金成本 p ，并于问题二的数据进行比较，分析任务联合打包发布对任务完成情况的影响。

5.3.3 模型检验与分析

根据 AIC 准则，即赤池信息量准则是由赤池弘次在研究信息论特别是在解决时间序列定阶问题时提出来的^[14]，求出本题 Q 型聚类的 AIC 值，评判聚类模型的效果。评判原则为，聚类模型的规模的确定应权衡模型适用性和紧凑性，以使 AIC 值最小为好。

5.3.4 问题答案

(1) 非线性规划求解结果为：

①最优化系数为： $a_1 = 0.336$ 、 $a_2 = -1.832$ 、 $a_3 = 2.496$ ；

②最优化任务发布平台的酬金成本 $p = 75326.93$ （元）

(2) 新的定价模型表达式为：

$$c'' = \frac{1}{2} \left(3 - \frac{0.336m^2 - 1.832m + 2.496}{s + \frac{1}{n}} \right) c。$$

(3) 问题三的定价模型支持向量机预测完成率 $\mu_3 = 88.50\%$ ，又计算得问题二的酬金成本 p 为 84818.66 元，完成率 $\mu_2 = 93.53\%$ 。分析可知与问题二相比，问题三模型使得酬金成本下降了 11.19%，与此同时完成率仅下降了 5.03%，模型结果较为理想。

(4) 检验聚类效果，得到本题聚类分析 AIC 最小值为 0，最佳分类数 $k = 451$ 。又结合实际情况，将聚类数定为 500，模型适用性和紧凑性较为合理。

5.3.5 问题总结

通过建立了基于 Q 型聚类分析和非线性规划的定价优化模型，解决了在位置较集中的任务联合打包发布的情况下修改定价模型并分析对任务完成情况影响的问题。最后用 AIC 准则对已建模型进行了检验。求得合理的问题答案为：任务完成率 $\mu_3 = 88.50\%$ ，任务完成情况较为理想。本节的创新之处在于使用 Q 型聚类分析，利用枚举法根据实际情况控制聚类数，规避了密度聚类法带来的链

式反应。并考虑到降低任务发布者的酬金成本等因素,使用规划,优化定价模型。

5.4 问题四的模型建立和求解

通过利用问题二中基于支持向量机的完成率评判模型,解决确定附件三中新项目的任务定价方案的问题;通过建立基于修改圈近似算法的旅行商模型、基于模糊层次分析法和熵值法的综合评价模型,解决基于修改圈近似算法的旅行商模型、基于模糊层次分析法和熵值法的综合评价模型问题。将采用修改圈算法求解遍历所有任务点的最短路程。最后用相对比较赋权法对已建的综合评价模型进行检验。给予各指标相同的增量,比较各指标对评价值影响大小进行灵敏度分析。这样就能求得合理的问题答案。

5.4.1 模型建立

(1) 确定新项目的定价方案

问题一、二、三实质上存在3种定价方案,故为附件三中的新项目选定定价方案,需比较新的项目中问题一、二、三各自的定价方案的**任务完成率 μ_1 、 μ_2 、 μ_3** ,**选择任务完成率较高的定价方案。**

(2) 建立基于修改圈近似算法的旅行商模型、基于模糊层次分析法和熵值法的综合评价模型

定理 5.4.1.1: 在已知任务发布平台的酬金成本 p 、任务数最大的包内完成所有任务的最短路程 L 、任务的完成率 μ 和时间完成度 T 这4项指标值的情况下,建立基于模糊层次分析法和熵值法的综合评价模型,计算表达式如下:

$$y = 0.1723p + 0.2086L + 0.3457\mu + 0.2734T$$

即可计算评价值,判定新的任务在**该定价方案下的实施效果**。

证明:

1) 指标选取

为评价方案的实施效果,从任务的成本损耗和任务的执行满意度两个方面,选取了4个指标,分别为任务发布平台的酬金成本 p 、任务数最大的包内完成所有任务的最短路程 L 、任务的完成率 μ 和时间完成度 T 。

下面详细介绍各指标的定义及计算方法:

①**任务发布平台的酬金成本 p**

任务发布平台的酬金成本 p 定义为众包平台预期所有任务均完成所需支付给用户的酬金,计算公式如下:

$$p = \sum_{i=1}^N c_i,$$

上式中 c_i 表示各任务的标价, N 表示总的任务数。

若用 c_{max} 表示新的项目中任务标价的最大值,则酬金成本 p 的标准化公式如下:

$$p' = \frac{p}{Nc_{max}}。$$

②**任务数最大的包内完成所有任务的最短路程 L**

最短路程 L 定义为在任务数最大的包内用户完成所有任务,即遍历所有任务点的最短路程。在任务数最大的包内各任务点之间的距离已经确定的情况下,可**利用修改圈近似算法**,建立**基于修改圈近似算法的旅行商模型**。

用 d_{max} 表示任务数最大的包内两个任务点间距离最大值，则最短路程 L 如下：

$$L' = \frac{L}{d_{max}m_2}。$$

③任务的完成率 μ

任务的完成率 μ 计算公式如下：

$$\mu = \frac{a}{b}。$$

上式中， a 为已完成的任务数， b 为未完成的任务数。

④时间完成度 T

经调查，“拍照赚钱”类任务的任务完成期限一般为 24 小时。即 t_i 表示任务 i 是否完成，取值如下：

$$t_i = \begin{cases} 1, & \text{任务在 24 小时内完成} \\ 0, & \text{任务未在 24 小时内完成} \end{cases}，$$

一般当 m_2 大于 10 时， $t_i = 0$ ；当 m_2 取值为 1 至 10 内的整数时， $t_i = 1$ 。

则 24 小时内任务的完成度 T 计算公式如下：

$$T = \frac{\sum_{i=1}^N t_i}{N}。$$

2) 基于模糊层次分析法和熵权法的综合评价模型

由于本问题的 4 个指标是从任务的成本损耗和任务的执行满意度两个方面选取的，这 4 个指标通过任务的成本损耗和任务的执行满意度两个方面影响定价方案的实施效果，故应采用层次分析法建立综合评价模型。但传统层次分析法存在以下缺点^[5]：1)判断矩阵的一致性检验困难重重、难以操作。2)一致性的衡量标准是人为设定的，客观科学性差。3)人脑常规思维模式中的一致性明显异于判断矩阵一致性，易引起思维错误。

为解决以上问题，本文采用模糊层次分析法和熵值法相结合的方法建立综合评价模型。

(1) 模糊层次分析法计算权重

模糊层次分析法与传统的层析分析法的不同点^[5]仅存在：1) 构造模糊一致判断矩阵。2) 在确定不同元素相对重要性的权重上，模糊层次分析法由模糊判断矩阵求解。

引起模糊层次分析法与传统的层析分析法的不同点的本质原因为模糊层次分析法构造模糊一致判断矩阵。而构造模糊一致判断矩阵最重要的是舍弃传统的 1-9 标度方法，而采用模糊层次标度法。^[15]本文采用模糊层次标度法^[15]，见下表：

表 5.4.1.1 本文采用的模糊层次标度法含义表

标度	含义
0.5	两方案对某属性同样重要
0.7	两方案对某属性，一方案比另方案稍微重要
0.7930	两方案对某属性，一方案比另方案明显重要
0.8542	两方案对某属性，一方案比另方案强烈重要
0.9	两方案对某属性，一方案比另方案极端重要
0.6262, 0.7524, 0.8262, 0.8786	表示相邻两标度之间折衷时的标度
上列标度互补	方案 A_i 对方案 A_j 的标度为 r_{ij} ，反之为 $1-r_{ij}$

本文建立模糊层次分析法的层次划分如下：

- 1) 目标层：定价方案的实施效果。
 - 2) 准则层：任务的成本损耗、任务的执行满意度。
 - 3) 方案层：任务发布平台的佣金成本 p 、任务数最大的包内完成所有任务的最短路程 L 、任务的完成率 μ 、时间完成度 T 。
- 模糊层次分析法的其余步骤与模糊层次分析法一致，在此不做赘述。

由上述步骤，计算模糊层次分析法的权重 w_j ， $j = 1, 2, 3, 4$ 。

(2) 熵值法计算熵值及差异性系数

1) 归一化处理指标数值，在指标选取中已经介绍 4 个指标归一化处理的方式。

2) 计算指标的熵值

计算熵值 H_j ，计算公式^[16]如下：

$$H_j = -\frac{1}{\ln n} \sum_{i=1}^n a_{ij} \ln a_{ij}$$

上式中， n 为评价的对象数， a_{ij} 为标准化处理后第 i 个评价对象的 j 指标值。

3) 计算差异性系数

计算差异性系数 g_j ，计算公式^[17]如下：

$$g_j = 1 - H_j, \quad j = 1, 2, 3, 4。$$

(3) 计算组合权重

组合权重 ω_j ，计算公式^[17]如下：

$$\omega_j = \frac{w_j g_j}{\sum_{i=1}^n w_j g_j}$$

5.4.2 模型求解

(1) 确定新项目的定价方案

比较新的项目中问题一、二、三各自的定价方案的任务完成率 μ_1 、 μ_2 、 μ_3 ，选择任务完成率较高的定价方案。

(2) 评价改定价方案实施效果

1) 计算模糊层次分析法的权重

①求解模糊判断矩阵

$$A = \begin{bmatrix} 0.5 & 0.207 \\ 0.793 & 0.5 \end{bmatrix}。$$

②计算组合向量

$$B_1 = \begin{bmatrix} 0.5 & 0.7 & 0.8542 & 0.8542 \\ 0.3 & 0.5 & 0.7930 & 0.7930 \\ 0.1458 & 0.2070 & 0.5 & 0.7930 \\ 0.1458 & 0.2070 & 0.2070 & 0.5 \end{bmatrix}，$$

$$B_2 = \begin{bmatrix} 0.5 & 0.3 & 0.1458 & 0.1458 \\ 0.7 & 0.5 & 0.207 & 0.207 \\ 0.8542 & 0.7930 & 0.5 & 0.7 \\ 0.8542 & 0.7930 & 0.3 & 0.5 \end{bmatrix}。$$

- ③按公式计算组合向量
- ④按公式确定个指标的权重
- 2) 熵值法计算差异性系数
- 3) 计算组合权重

5.4.3 模型检验与分析

建立基于相对比较赋权法的综合评价模型，计算出新的评价值，与本题建立的综合评价模型的评价值对比，进行模型检验。

相对比较赋权法的过程如下：将所有的评价指标分别按行和列排列，构成一个正方形的表；再根据三级比例标度对任意两个指标的相对重要关系进行分析，并将评分值记入表中相应的位置；将各个指标评分值按行求和，得到各个指标的评分总和；最后做归一化处理，求得指标的权重系数。

三级比例标度两两相对比比较评分的分值为，则标度值及其含义如下：

$$q_{ij} = \begin{cases} 1 & x_i \text{ 比 } x_j \text{ 重要时} \\ 0.5 & x_i \text{ 比 } x_j \text{ 同样重要时} \\ 0 & x_i \text{ 比 } x_j \text{ 不重要时} \end{cases}$$

则评分构成的矩阵 $Q = (q_{ij})_{4 \times 4}$ ，指标 x_i 的权重 ω'_i ：

$$\omega'_i = \frac{\sum_{j=1}^4 q_{ij}}{\sum_{i=1}^4 \sum_{j=1}^4 q_{ij}}$$

使用该方法确定指标权重时，任意两个指标之间的相对重要程度要有可比性。这种可比性在主观判断评分时，应满足比较的传递性，即 x_1 比 x_2 重要， x_2 比 x_3 重要，则 x_1 比 x_3 重要。

这里，我们根据实际情况，综合考虑任务发布者和会员的立场，优先改善任务完成情况，得到评分矩阵：

$$Q = \begin{bmatrix} 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 \\ 1 & 0.5 & 1 \\ 0.5 & 0.5 & 0.5 \end{bmatrix}$$

通过公式 $Q\omega' = \lambda_{\max}\omega'$ ，计算得到权重 $\omega' = [\omega'_1, \omega'_2, \omega'_3, \omega'_4]^T$ ，则综合评价模型计算表达式为：

$$y' = \omega'_1 x_1 + \omega'_2 x_2 + \omega'_3 x_3 + \omega'_4 x_4$$

若 y' 与 y 的值相差不大，则说明模型具有合理性。

5.4.4 灵敏度分析

在求得综合评价表达式后，分别给予 4 个指标相同的增量，通过对各指标依次改变后对应评价值的改变量有大到小进行排序，比较各指标对评价值影响大小进行灵敏度分析。在相同增量的情况下，对应评价值的改变量越大，则该指标对应评价值的影响越大。

5.4.5 问题答案

(1) 新项目的定价方案

比较新的项目中问题一、二、三各自的定价方案的任务完成率

$$\mu_1 = 23.23\%, \mu_2 = 78.27\%, \mu_3 = 79.82\%,$$

得到 μ_1 、 μ_2 、 μ_3 的大小关系如下：

$$\mu_1 < \mu_2 < \mu_3$$

因此选择任务完成率较高的问题三中基于 Q 型聚类分析和非线性规划的优化定价方案。

(2) 评价定价方案实施效果

1) 计算组合权重

$$\omega_1 = 0.1723, \omega_2 = 0.2086, \omega_3 = 0.3457, \omega_4 = 0.2734。$$

2) 综合评价表达式

$$Z = 0.1723x_1 + 0.2086x_2 + 0.3457x_3 + 0.2734x_4$$

3) 评价值

评价值：

$$Z = 0.7038$$

根据此评价模型得到的最终评价值可知：此模型实施效果尚可，不是特别完美，但是相较于前面的两种定价方案模型有所改进。

(3) 模型检验结果

模型检验求得的评价值：

$$Z' = 0.7767$$

新求得的评价值与原评价值的差值大小在可允许的范围内，故模型整体上通过了检验。

(4) 灵敏度分析结果

分别给予 4 个指标相同的增量 0.1，得到对应评价值的增量见下表：

表 5.4.5.1 4 个指标分别增加 0.1 评价值的增量

指标增加 0.1	酬金成本	任务数最大的包内完成任务的最短路程	任务的完成率	时间完成度
评价值的增量	0.7210	0.7246	0.7384	0.7311

由上表可知，4 个指标中每个指标增加 0.1，任务的完成率指标评价值的增量最大，故 4 个指标中，任务的完成率对评价值影响最大。

5.4.6 问题总结

本节通过利用了问题二中基于支持向量机的完成率评判模型，解决了确定附件三中新项目的任务定价方案的问题；通过建立了基于修改圈近似算法的旅行商模型、基于模糊层次分析法和熵值法的综合评价模型，解决了评价定价方案实施效果的问题。采用了修改圈算法求解遍历所有任务点的最短路程。最后用相对比较赋权法对已建的综合评价模型进行检验。给予各指标相同的增量，比较各指标对评价值影响大小进行了灵敏度分析。问题的合理答案：(1) 选用问题三中基于 Q 型聚类分析和非线性规划的优化定价方案。(2) 评价值为 0.7038，根据此评价模型得到的最终评价值可知：此模型实施效果尚可，不是特别完美，但是相较于前面的两种定价方案模型有所改进。(3) 4 个指标中，任务的完成率对评价值影

响最大。本节的创新之处在于采用了模糊层次分析法和熵值法的组合赋权法，不仅适合于层次鲜明的指标评价体系，而且避免了主观赋权法中权值主观随意性的问题。方法新颖独特，实用性强。

六、评价与改进

6.1 模型一的评价

优点：采用多元二项式回归，避免了出现拟合精度低、模型适用性弱等缺点。并在分析任务未完成原因是引入意愿半径思想，简化问题，建模思维极富创造性。

缺点：在研究定价规律时，仅考虑了经纬度，即地理经济差异性，未考虑任务密度等其他因素的影响。

改进：加入任务密度，所属城市社会经济等因素，丰富定价模型指标后在进行回归分析。

6.2 模型二的评价

优点：引入弹性函数的概念，具体确定了如何从会员数、会员的平均信誉值两个方面具体计算定价方案

缺点：未引入与任务完成时间相关的变量，模型的全面性还有待提高。

改进：加入时间变量，调整定价模型。

6.3 模型三的评价

优点：使用 Q 型聚类分析，利用枚举法根据实际情况控制聚类数，规避了密度聚类法带来的链式反应。并考虑到降低任务发布者的酬金成本等因素，使用规划，优化定价模型。

缺点：求取任务联合打包后包内的任务数计算任务量较大，不简便。

6.4 模型四的评价

优点：综合考虑了任务打包后会员工作量，引入最短路径指标，并创新引入了时间完成度的概念，使得综合评价模型丰满，指标丰富。

缺点：由于数据较少，熵权法确定权重是，客观性有所欠缺。

改进：映入附件一的数据计算相关指标后，与附件三的指标共同作为数据，用熵权法确定权重。

6.5 模型的推广

本题的定价模型不仅仅可以考虑应用于基于移动互联网的自助式劳务众包平台的任务定价问题，还可以应用于交通出行收费、外卖配送价格的制定、货物工作同配送收入分配等众多众包行业的一系列涉及最短路径与动态定价的定价问题，

参考文献

- [1] 廖什的市场区位理论，
<https://baike.baidu.com/item/%E5%BB%96%E4%BB%80%E7%9A%84%E5%B8%82%E5%9C%BA%E5%8C%BA%E4%BD%8D%E7%90%86%E8%A E%BA/12748922?fr=aladdin>，2016-03-10
- [2] 约翰·W.马林斯，奥维尔·C.小沃克，营销管理:战略决策制定方法（第6版）[M]，清华大学出版社，北京，2008
- [3] 沈琼，马士基中美集装箱航线的现状分析与改进方案[D]，复旦大学，2010

- [4] 余烨, 张昌民, 李少华, 朱锐, 杜家元, 王莉, 多元统计分析在地质学中的应用——以惠州凹陷 M 层物源分析为例[J], 地质科学, 2014, 49(01): 191-201
- [5] 张吉军, 模糊系统与数学[J], 2000, 14(02): 80-85
- [6] 缪其浩, 情报工作如何从“众包模式”获益?[J], 竞争情报, 2014(1): 1-1
- [7] J Howe, The rise of crowdsourcing[J], Wired magaine, 2006, 14(6): 1-4
- [8] 插值,
<https://baike.baidu.com/item/%E6%8F%92%E5%80%BC/1196063?fr=aladdin>, 2016-08-30
- [9] 司守奎, 孙兆亮, 数学建模算法与应用[M], 国防工业出版社, 北京, 2015
- [10] 吕占华, 产品与服务混合定价策略探讨[J], 中国科技博览, 2011(11): 58-58
- [11] 何蕾, 钟书华, 企业加速器服务的四种定价方法[J], 统计与决策, 2011(14): 175-178
- [12] 姜雪, 卢文喜, 杨青春, 赵海卿, 应用支持向量机评价土壤环境质量[J], 中国环境科学, 2014, 34(5): 1229-1235
- [13] 焦卫东, 林树森, 整体改进的基于支持向量机的故障诊断方法[J], 仪器仪表学报, 2015, 36(8): 1861-1870
- [14] Chen P C. A non-line-sight mitigation algorithm in location estimation [C], IEEE Wireless Communications and Networking Conference, New Orleans , USA, 1999: 316 - 320
- [15] 兰继斌, 徐扬, 霍良安, 刘家忠, 模糊层次分析法权重研[J], 系统工程理论与实践, 2006, 26(9):107-112
- [16] 杨宇, 多指标综合评价中赋权方法评析[J], 统计与决策, 2006(13): 17-19
- [17] 陈海峰, 基于 AHP-熵值法的税源风险评估研究[D], 厦门大学, 2014

附录

附录一 问题一的模型算法求解程序

```
%%%%%%%% 根据附件一数据画三维插值拟合图 %%%%%%%%%%
scatter3(x1,y1,z1)%散点图
figure
[X,Y,Z]=griddata(x1,y1,z1,linspace(min(x1),max(x1)),linspace(min(y1),max(y1)),v4'
);%二元三维非等距插值，v4 是算法
pcolor(X,Y,Z);shading interp%伪彩色图，使彩色平滑，与热力图类似（无等高线）
figure,contourf(X,Y,Z) %绘制等高线图
xlabel('任务纬度','Fontname','宋体','FontSize',12);
ylabel('任务经度','Fontname','宋体','FontSize',12);
figure,surf(X,Y,Z);%绘制三维曲面图
%%%%%%%% 多元二项式回归分析 %%%%%%%%%%
x = ones(835,1);
xx1 = [x,y1,x1];
yy1 = z1;
[b,bint,r,rint,stats] = regress(yy1,xx1); %最小二乘法线性回归
b,bint,stats,rcoplot(r,rint) %画出残差及其置信区间图
xxx1 = [y1,x1];
rstool(xxx1,yy1,'purequadratic') %纯二次回归法
% “Exprot” 输出 beta（回归系数）、rmse（剩余标准差）、residuals（残差）
beta,rmse
rstool(xxx1,yy1,'linear') %线性回归
beta1,rmse1
rstool(xxx1,yy1,'interaction') %交叉回归
beta2,rmse2
rstool(xxx1,yy1,'quadratic') %完全二次回归
beta3,rmse3
rmsee = [rmse,rmse1,rmse2,rmse3] %对比四种回归方法的残差，取最小者

%%%%%%%% 回归优度检验 %%%%%%%%%%
y1y1 = 0;x1y1 = 0;x1x1 = 0;
fori = 1:1:835
y1y1 = y1y1 + y1(i)*y1(i);
x1y1 = x1y1 + x1(i)*y1(i);
x1x1 = x1x1 + x1(i)*x1(i);
end
zz1
=
beta3(1)+beta3(2)*y1+beta3(3)*x1+beta3(4)*y1y1+beta3(5)*x1y1+beta3(6)*x1x1;
%Q-Q 图检验回归优度
xbar = mean(zz1);s = std(zz1); %求均值和标准差
sa = sort(zz1); %按照从小到大排序
n = length(zz1);pi = ([1:n]-1/2)/n;
```



```

yi = norminv(pi,xbar,s)'; %计算对应的 yi 值
plot(yi,sa, '.') %描点画 Q-Q 图
normplot(zz1) %正态概率图
probplot('norm',zz1) %P-P 图
%%% 由任意两点的经纬度求两点间最短距离 %%%
%d = 6371.004*acos(sin(yyyy1)*sin(yyyy2) +
cos(yyyy1)*cos(yyyy2)*cos(xxxx1-xxxx2))*3.1415926/180
%xxxx1 : 1 点经度 yyyy1 : 1 点纬度
%xxxx2 : 2 点经度 yyyy2 : 2 点纬度
%%%%% 计算意愿半径 R 内的三项指标 %%%%
%导入数据省略
%计算平均价格 cc1 和 cc2
cc1 = mean(z11)
cc2 = mean(z10)
%计算意愿半径内平均会员数 nn1、nn2 和平均信誉值 ss1、ss2
r = 3; %意愿半径
%已完成任务：
n1 = zeros(522,1);
s1 = zeros(522,1);
sss1 = zeros(522,1);
fori = 1:1:522
for j = 1:1:1877
d = 6371.004*acos(sin(y11(i))*sin(y2(j)) +
cos(y11(i))*cos(y2(j))*cos(x11(i)-x2(j)))*3.1415926/180;
if d<r
n1(i) = n1(i) + 1;
sss1(i) = sss1(i) + g2(j);
end
end
if n1(i)>0
s1(i) = sss1(i)/n1(i);
end
end
nn1 = mean(n1) %已完成任务的平均会员数
ss1 = mean(s1) %已完成任务的平均信誉值
%未完成任务：
n2 = zeros(522,1);
s2 = zeros(522,1);
sss2 = zeros(522,1);

fori = 1:1:313
for j = 1:1:1877
d = 6371.004*acos(sin(y10(i))*sin(y2(j)) +
cos(y10(i))*cos(y2(j))*cos(x10(i)-x2(j)))*3.1415926/180;

```

```

if d<r
n2(i) = n2(i) + 1;
sss2(i) = sss2(i) + g2(j);
end
end
if n2(i)>0
s2(i) = sss2(i)/n2(i);
end
end
nn2 = mean(n2) % 未完成任务的平均会员数
ss2 = mean(s2) % 未完成任务的平均信誉值

```

附录二 问题二的模型算法求解程序

```

%%%%%% 计算意愿半径 R 内所有任务的三项指标 %%%%%%%%%
x1 = xlsread('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题一\附件一：
已结束项目任务数据 - 副本.xls','t_tasklaunch','D2:D836');
y1 = xlsread('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题一\附件一：
已结束项目任务数据 - 副本.xls','t_tasklaunch','E2:E836');
z1 = xlsread('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题一\附件一：
已结束项目任务数据 - 副本.xls','t_tasklaunch','H2:H836');
x2 = xlsread('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题一\附件二：
会员信息数据 - 副本.xlsx','Sheet1','C2:C1878');
y2 = xlsread('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题一\附件二：
会员信息数据 - 副本.xlsx','Sheet1','D2:D1878');
g2 = xlsread('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题一\附件二：
会员信息数据 - 副本.xlsx','Sheet1','G2:G1878')
%计算平均价格 cc
cc = mean(z1);
%cc =69.1108;
%计算意愿半径内平均会员数 nn 和平均信誉值 ss
r = 3; %意愿半径
n = zeros(835,1);
s = zeros(835,1);
sss = zeros(835,1);
s = zeros(835,1);
cx = zeros(835,1);
for i = 1:1:835
for j = 1:1:1877
d = 6371.004*acos(sin(y1(i))*sin(y2(j))
cos(y1(i))*cos(y2(j))*cos(x1(i)-x2(j)))*3.1415926/180;
if d<r
n(i) = n(i) + 1;
sss(i) = sss(i) + g2(j);
end
end
end

```

```

if n(i)>0
s(i) = sss(i)/n(i);
    %e(i) = s(i)/n(i);    %备用
e(i) = s(i) + 1/n(i);
end
if n(i)==0
e(i) = 1;
end
end
nn = mean(n); %已完成任务的平均会员数
ss = mean(s); %已完成任务的平均信誉值
%nn = 22.2599;ss = 232.9349;
%%%%% 计算新定价模型下所有任务新价格  %%%%%
for i = 1:1:835
cx(i) = 0.5*(3-1/e(i))*z1(i);
end
status = xlswrite('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题二\附件
一任务新定价.xls',cx,'t_tasklaunch','H2:H836') %将新定价导出并返回导出结果成
功与否
%%%%% 支持向量机训练预测新定价模型成果  %%%%%
%导入数据
%x1 = xlsread('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题一\附件一：
已结束项目任务数据 - 副本.xls','t_tasklaunch','D2:D836');
%y1 = xlsread('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题一\附件一：
已结束项目任务数据 - 副本.xls','t_tasklaunch','E2:E836');
%z1 = xlsread('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题一\附件一：
已结束项目任务数据 - 副本.xls','t_tasklaunch','H2:H836');
h1 = xlsread('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题一\附件一：
已结束项目任务数据 - 副本.xls','t_tasklaunch','I2:I836');
z2 = xlsread('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题二\附件一
任务新定价.xls','t_tasklaunch','H2:H836');
b0=[z1,n,s]'; dd0=[z2,n,s]'; %提取已分类和待分类的数据
[b,ps]=mapstd(b0); %已分类数据的标准化
dd=mapstd('apply',dd0,ps); %待分类数据的标准化
group=h1; %已知样本点的类别标号
s66=svmtrain(b',group) %训练支持向量机分离器
%s77 = svmtrain(b',group,'Method','SMO','Kernel_Function','quadratic' %使用序列
最小化方法训练支持向量机的分类器
sv_index=s66.SupportVectorIndices; %返回支持向量的标号
beta=s66.Alpha; %返回分类函数的权系数
bb=s66.Bias; %返回分类函数的常数项 %-0.0570
mean_and_std_trans=s66.ScaleData; %第1行返回的是已知样本点均值向量的相反
数，第2行返回的是标准差向量的倒数
check=svmclassify(s66,b'); %验证已知样本点

```

```

status1 = xlswrite('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题二\附件一任务新定价.xls',check,'t_tasklaunch','L2:L836') %将预测结果导出并返回导出结果成功与否
err_rate=1-sum(group==check)/length(group) %计算已知样本点的错判率 %0.3509
solution=svmclassify(s66,dd'); %对待判样本点进行分类
y = solution;
status2 = xlswrite('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题二\附件一任务新定价.xls',y,'t_tasklaunch','I2:I836') %将预测结果导出并返回导出结果成功与否
%%%%% 计算完成率 %%%%
yita1 = sum(h1)/835 %0.6251 原数据完成率
yita3 = sum(check)/835 %0.6287 验证数据完成率
yita2 = sum(y)/835 %0.8707 新定价模型完成率
%yita2 = sum(y)/835 %0.9353 备用新定价模型完成率

```

附录三 问题三模型算法求解程序

```

%%%%% Q 型聚类法将同类任务打包 %%%%
%导入数据代码省略
y=pdist(a,'cityblock'); %求 a 的两两行向量间的绝对值距离
yc=squareform(y); %变换成距离方阵
z=linkage(y); %产生等级聚类树
T=cluster(z,'maxclust',500); %把对象划分成 500 类
jz = zeros(500,13); %构造分类数据矩阵
for i=1:500
    tm=find(T==i); %求第 i 类的对象
    tm=reshape(tm,1,length(tm)); %变成行向量
    ll = size(tm);
    jz(i,1) = i;
    jz(i,2) = ll(2);
    lll = 2 + ll(2);
    jz(i,3:lll) = tm(1:ll(2));
    fprintf('第%d 类的有%d 个任务,分别是: %s\n',i,ll(2),int2str(tm)); %显示分类结果
end
status1 = xlswrite('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题三\一项目任务数据.xls',jz,'聚类结果')
numm = xlsread('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题三\一项目任务数据.xls','聚类结果','B1:B500');
duix = xlsread('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题三\一项目任务数据.xls','聚类结果','C1:O500');
mm = xlsread('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题三\一项目任务数据.xls','数据备份','A2:A836');
m2 = zeros(835,1);
S = zeros(835,1);
SS = zeros(835,1);

```

```

o = zeros(1,13);

for k = 1:1:835    %求解 m2 和 SS
    for i = 1:1:500
        for j = 1:1:13
            if duix(i,j)==mm(k)
                m2(k) = numm(i);
                o = duix(i,:);
                if o(j)>0
                    S(k) = ss(o(j)) + S(k);
                end
                SS(k) = S(k)/m2(k);
            end
        end
    end
end
status2 = xlswrite('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题三\一
项目任务数据.xls',m2,'数据备份','C2:C836')
%%%% 线性规划计算价格相关系数最优值 %%%%
mm2 = mean(m2); %2.724550898
%目标函数.m 文件
function f = fun1(x)
    mm2 = 2.724550898;
    f = mm2.^2*x(1) + mm2*x(2) + x(3);
end
%非线性约束函数.m 文件
function [g,h] = fun2(x)
    g = -4*x(1)*x(3) + x(2).^2;
    h = [];
end

%[x,y] = fmincon('fun1',rand(3,1),[1,0;0,-1;0,0],[0;0],[1,1,1],1)
[x,y] = fmincon('fun1',rand(3,1),[-1,0,0],0,[1,1,1],1,[],[],'fun2')
%%%% 通过函数计算最终价格 %%%%
FF = zeros(835,1);
for i = 1:1:835
    if m2(i)>0
        FF(i) = 0.5*(3-(x(1)*m2(i).^2+x(2)*m2(i)+x(3))/(SS(i)+1/nn(i)))*c0(i);
    else
        FF(i) = c0(i);
    end
end
P = sum(FF) %价格总定量 75326.9337107756
cx3 = FF;

```

```
%支持向量机模拟计算问题三打包模型定价的完成率
b0=[z1,nn,ss]'; dd0=[cx3,nn,SS]'; %提取已分类和待分类的数据
[b,ps]=mapstd(b0); %已分类数据的标准化
dd=mapstd('apply',dd0,ps); %待分类数据的标准化
group=h1; %已知样本点的类别标号
s66=svmtrain(b',group) %训练支持向量机分离器
solution=svmclassify(s66,dd'); %对待判样本点进行分类
yita = sum(solution)/835 %验证数据完成率 0.8850
```

附录四 问题四的模型算法求解程序

```
%%%%% 4.2 四个评价指标及其标准化+熵权法评价 %%%%
%%%%% 计算任务发布者总定价及标准化 %%%%
x1 = xlsread('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题四\附件三：
新项目任务数据 - 副本.xls','t_tasklaunch','B2:B2067');
y1 = xlsread('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题四\附件三：
新项目任务数据 - 副本.xls','t_tasklaunch','C2:C2067');
mm = xlsread('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题四\附件三：
新项目任务数据 - 副本.xls','任务编号','A1:A2066');
```

```
P = 211531.538;
cmax = max(c3); %110.5797751
N = 2066;
PP = P/(cmax*N); %标准化数据 0.9259
```

```
%%%%% 计算最短路径及标准化 %%%%
%dis
[696,698,703,708,740,745,754,758,790,809,832,851,854,910,911,916,941,945,955,96
5,975,1783,1785,1792,1798,1827,1851,1864,1881,1882,1895,1925,1926,1979,1991,1
993,1995,2009,2012,2021,2036,2044,2054];
dis
[518,522,537,572,573,585,610,613,614,615,621,624,627,629,632,645,648,663,664,67
4,1598,1628,1643,1667,1668,1681,1699,1702,1705,1716,1717,1720,1721,1725,1727,
1735,1751];
distan=zeros(37,37);
x2 = x1; y2 = y1;
```

```
for i = 1:1:37
    for j = 1:1:37
        i1 = dis(i); j1 = dis(j);
        if x1(i1)==x2(i1)&y1==y2(j1)
            distan(i,j) = 0;
        else
            distan(i,j) = 6371.004*acos(sin(y1(i1))*sin(y2(j1))
+ cos(y1(i1))*cos(y2(j1))*cos(x1(i1)-x2(j1)))*3.1415926/180;
        end
    end
end
```



```

end
end
status = xlswrite('C:\Users\Administrator\Desktop\2017 数学建模\数据\问题四\附件
三：新项目任务数据 - 副本.xls',distan,'距离')

a = distan;
L=size(a,1);
c=[1:37 1]; %选取初始圈
[circle,long]=modifycircle(a,L,c) %调用下面修改圈的子函数
maxx = max(distan); %0.2138
LL = long/(maxx(1) * 37) %去量纲标准化最短路径
%%%%% 完成率 %%%%%
yita1 = 0.2323;
yita2 = 0.7827;
yita3 = 0.7982;
%%%%% 计算 T 时间完成度 %%%%%
t = zeros(2066,1);
for i = 1:1:2066
    if m2(i)>10
        t(i) = 0;
    else
        t(i) = 1;
    end
end
end

T = sum(t)/2066; %0.9351
%%%%%%%%%%%%%
%%%%%%%%%%%%% 层次分析熵权法评价 %%%%%%%%%%%%%%
%四个评价指标值： 0.9259 0.0607 0.7982 0.9351

%求矩阵最大特征值对应的权向量 w
A=[1 1/5
    5 1];%输入一个二维矩阵（对角阵）
X=size(A);
n=X(1,1);%n 表示矩阵 A 的行数
m=X(1,2);%m 表示矩阵 A 的列数
W=A;%定义一个矩阵并赋初值
M=sum(A);%对矩阵 A 列求和
for j=1:m;
    for i=1:n;
        W(i,j)=A(i,j)/M(1,j);
    end
end
end
Y=sum(W,2);

```

```

w=Y;
for i=1:n
    w(i,1)=Y(i,1)/sum(Y); end
N=A*w;
B=N;
for j=1:n
    B(j,1)=N(j,1)/w(j,1);
end
w
%w = [0.1667 0.8333]';
%再分别求 B1 和 B2 的权向量
B1 = [0.5 0.7 0.8542 0.8542
0.3 0.5 0.7930 0.7930
0.1458 0.2070 0.5 0.7930
0.1458 0.2070 0.2070 0.5];
%w1 = [0.3863 0.2978 0.1860 0.1300]';
B2 = [0.5 0.3 0.1458 0.1458
0.7 0.5 0.207 0.207
0.8542 0.7930 0.5 0.7
0.8542 0.7930 0.3 0.5];
%w2 = [0.1295 0.1908 0.3776 0.3021]';
W = [w1,w2]*w
x = [0.9259 0.0607 0.7982 0.9351];
y = 0;
for i = 1:4
    y = y + W(i)*x(i);
end
y %0.7038

```

