

# 基于多元线性回归与聚类分析的网络任务定价模型

## 摘要

随着互联网经济的高速发展，用户通过平台发布和接受任务的新型“雇佣劳务关系”正逐渐受到年轻人的青睐，而任务定价作为这种新型关系中最重要的一环，其合理性也就成了吸引双方的决定性因素，因此平台若能通过用户数据分析建立合理的定价模型向用户给出合理的推荐价格，势必能使其在强大竞争环境下站稳脚跟。本文根据所给数据及问题分别建立了四个对应模型：原始定价模型、修正定价模型、打包定价模型及预测定价模型。

原始定价模型建立在对根据附件数据的分析上，分析发现任务定价主要与三个变量有关：供求距离  $D_i$ 、消偏信誉值  $E_i$  及总预订任务限额  $Q_i$  相关，据此建立多元线性回归模型进行拟合计算任务点的预期定价，结果得到定价函数：

$$W_i = 2.622D_i - 0.019E_i - 0.072Q_i + 71.599$$

用得到的预期价格与实际价格进行比较并对未完成任务进行 K-means 聚类分析将未完成原因分为三类：价格过低，无人员分布及地理位置相对偏僻。

修正定价模型建立在根据不同原因对三类未完成任务进行定价改进的基础上。对于价格过低的任务，筛选只使用已完成任务数据进行多元拟合，以其得到的定价函数对其进行定价修正；对于无人员分布的任务，用同样无人员分布但已完成的离其距离最短的任务的定价作为其修正定价；对于地理位置偏僻的任务，引入常量地理位置因子  $C$  对其价格修正，其值为各任务实际价格与用第一类修正定价函数得到的预期定价的差值的均值，计算得出  $C = 5.5$ 。

打包定价模型以每个任务点为中心，将半径  $r = 1$  范围内所有任务装入同一任务包，再根据任务序号剔除不同任务包内的重复任务，再用原始定价模型中的函数进行拟合，对该模型的评价采用与实际价格的  $\pm 1.05$  倍作为合理范围，计算合理任务的比例作为优劣度  $\alpha$  进行评价，发现其大于 0.8，故认为具有可信度。

预测定价模型同样采用多元线性模型，通过分析前三种模型的分析发现采用打包定价模型的定价函数最优，用其预测新任务价格，预测结果的评价用离其最近的已结束任务价格仿照打包定价模型进行优劣度  $\alpha$  分析，发现具有可信度。

**关键词：** 任务定价 多元线性回归 K-means 聚类 任务点评价

## 一、 问题背景与重述

“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。

题目中给出了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）；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）；一个新的检查项目任务数据，只有任务的位置信息。让我们完成下面的问题：

1. 研究项目的任务定价规律，分析任务未完成的原因。
2. 为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

### 1. 问题一分析

针对问题一，题目要求我们研究项目的定价规律和分析任务未完成的原因。对于前面一个问题，首先我们通过分析，把会员看做一种商品，又根据经济学知识可知商品的价格主要受市场的供给需求关系影响，所以我们推断项目的定价规律由拍照任务的供给和需求决定。我们把会员接的任务数当做是供给方，拍照任务数当做需求方，当拍照任务较少而会员较多时，会出现供过于求，则价格较低，制定的价格会较低；当拍照任务较多而会员较少时，会出现

供不应求，则价格较高，制定的价格会较高。接着，我们结合日常生活的经验，选取变量作为任务定价的影响因素。我们推断一个任务点旁边的的会员数量和会员离此的距离长短都会影响任务的定价，因此我们取一个任务点方圆  $1\text{ km}$  的会员与它的距离求和除以  $n^{k_1}$  ( $k_1$  为特征值) 称为供求距离，作为一个变量。然后，我们分析得到一个任务点周围会员的信誉值也会影响任务的定价。类似于供求距离的由来，我们把一个任务点周围方圆  $1\text{ km}$  的会员信誉值依次搜索出来求和除以  $n^{k_2}$  ( $k_2$  为特征值) 称为消偏信誉值作为第二个变量。之后，我们取一个任务点方圆  $1\text{ km}$  的会员预定任务限额求和称为总任务限额为第三个变量，我们认为会员任务限额也会影响市场的供求关系。我们用 MATLAB 计算出三个变量的数据，然后结合任务定价的数据进行函数拟合找出项目的定价规律。

对于后面的问题，分析未完成的原因。我们先筛选出未完成的任务的数据，找出它们的经纬度位置、定价和其周围的会员信息，然后对它们进行分析，找出它们失败的原因。

## 2. 问题二分析

针对问题二，设计新的定价方案并和原方案比较。我们从问题一得出的未完成原因出发设计一套新的方案使得定价更合理。问题一得出的未完成原因分三种情况，周围有会员但定价过低、定价不低但有效识别范围内没有会员和定价不低且有效识别内有会员但受地理位置影响而任务没有被完成。我们针对这三种原因分别制定不同的定价方案以此来合理定价未完成的任务。在第一种原因下，我们认为可以用被完成的任务点的数据进行回归分析，拟合出一条直线并写出表达式并将其作为新的定价方案；在第二种原因下，这些未完成的任务点周围是没有会员的，因此没有可利用的数据进行分析并完成定价，然后我们联想到那些周围没有会员却完成任务的点，并进行比对，最后我们将离未完成的任务最近的没有会员却完成的任务点的定价作为它们新的定价；在第三种原因下，我们将定价结合了日常实际出现的情况，我们要想办法合理提高定价来消除任务点位置偏僻带来的影响，我们认为应该在第一问得出的任务定价函数上加一个常数，而这个常数由任务点地理位置带来的影响决定。

### 3. 问题三分析

针对问题三，考虑到实际情况出现打包的方案，让我们修改前面的定价模型。首先，我们认为任务打包分布可以避免会员过多进行恶性竞争的现象，而且打包可以提高任务完成的效率。然后我们对任务点的分布进行分析寻找打包的方法，我们认为把以一个任务点为圆心， $r$ 为半径的一个圆范围内的所有任务看作是一个任务包，已经被归纳入任务包的不进入下一次搜索。接着我们对任务包进行定价，从原始任务点出发对任务包定价，一个任务包里面的任务定价相同。最后，我们要分析打包对任务完成情况的影响，我们对附件一中的未完成数据进行重新定价然后与修正预期价格模型得到的定价进行比对，进行优劣度检验，如果优劣度 $\alpha > 0.8$ ，则认为打包对任务完成情况具有积极意义。

### 4. 问题四分析

问题四考虑对新任务定价，首先在问题一、二、三所建立的定价模型中挑选以对新任务进行预测定价，再对预测结果进行评价。评价标准采用离新任务最近的已结束任务点的已知定价作为参考中心价格 $P_n$ ，再取5%的误差定下合理定价范围，用预测结果进行比较，计算优劣度 $\alpha$ ，若其大于0.8则认为该模型定价具有一定的合理性。反之则认为模型定价效果相当较差。

## 三、问题假设

1. 假设任务是否完成与任务难度无关，只取决于是否有会员前往完成。
2. 假设进行任务不存在失败的情况，也就是说只要有会员前往完成任务就被能完成。
3. 假设任务点之间的距离、任务点与会员的距离都是绝对距离，不考虑街道等因素对距离带来的影响。
4. 假设会员对任务点的选择只考虑任务定价、距离的大小和地理位置是否方便，会员对任务的选择都不带有主观厌恶与喜好情感。
5. 假设所以会员都是理性人。
6. 假设数据都是来自于实际生活中完整准确的数据。

## 四、符号说明

符号	符号说明
$D_i$	第 <i>i</i> 个任务点的供求距离
$d_{ij}$	第 <i>i</i> 个任务点与其周围第 <i>j</i> 个会员的距离
$E_i$	第 <i>i</i> 个任务点的削偏信誉值
$e_{ij}$	第 <i>i</i> 个任务点周围第 <i>j</i> 个会员的信誉值
$Q_i$	第 <i>i</i> 个任务点的总任务限额
$q_{ij}$	第 <i>i</i> 个任务点周围第 <i>j</i> 个会员的任务限额
$W_i$	第 <i>i</i> 个任务点的预期定价
$W_i'$	第 <i>i</i> 个任务点的修正预期定价
$P_i$	第 <i>i</i> 个任务的真实定价
$d_{mn}$	第 <i>m</i> 个任务点与第 <i>n</i> 个任务点的距离
$C$	地理位置因子
$P_i'$	重新定价

## 五、模型的建立与求解

### 5.1 问题一定价模型的建立与求解

#### 5.1.1 原始定价模型的建立

我们把会员看作是一件商品，任务的定价看作是商品的价格，商品的价格受市场的供求关系影响，即任务的定价受市场上会员的供求关系影响，按照经济学原理任务的定价为会员供给等于需求时的价格。对于这个问题，我们从市场的供给与需求角度出发，建立一个新的定价模型。我们选取了供求距离、消偏信誉值和消偏任务限额三个变量作为任务定价的因素。

##### (1) 供求距离的定义与求值

我们把任务点方圆  $1\text{ km}$  以内的会员与其的距离  $d_{ij}$  求和除以  $n^{k_1}$  定义为任务供求距离  $D_i$ 。计算公式如下：

$$D_i = \sum_{j=1}^n d_{ij} / n^{k_1} \quad (5-1)$$

$D_i$  为第*i*个任务点的供求距离； $d_{ij}$  为第*i*个任务点周围  $r\text{ km}$  内的第*j*个会员

与任务点的距离模型中的距离都为绝对距离,如果考虑具体街道的话距离很难计算;  $k_1$  为特征值。由于附件一、二给出的任务及会员位置为经纬度分布,故在计算距离时需要用经纬度转化求距离公式<sup>[1]</sup>, 设 A 点(纬度  $\varphi_1$ , 经度  $\lambda_1$ ), B 点(纬度  $\varphi_2$ , 经度  $\lambda_2$ ), 两点的距离为:

$$d_{12} = 111.199 \left[ (\varphi_1 - \varphi_2)^2 + (\lambda_1 - \lambda_2)^2 \cos^2 \left( \frac{\varphi_1 + \varphi_2}{2} \right) \right]^{0.5} \quad (5-2)$$

平均距离一样但会员分布不同这样的对于任务定价有不同影响,如果供求距离单纯是一个平均数,那么不论会员分布如何只要平均距离是一样的会员信息对任务定价的影响是相同的,为消除这种情况对模型的影响,我们引入特征值,把求和的值除以  $n^{k_1}$ 。供求距离的公式将会员信息对任务定价的影响考虑进去,是从供给侧出发,会员与任务点的距离和任务点周围的会员数都会对供求距离的大小产生影响。

#### (2) 消偏信誉值的定义与求值

我们把任务点方圆 1 km 以内的会员信誉值加总求和除以  $n^{k_2}$  定义为消偏信誉值  $E_i$ 。计算公式如下:

$$E_i = \sum_{j=1}^n e_{ij} / n^{k_2} \quad (5-3)$$

$E_i$  为第  $i$  个任务点的消偏信誉值,  $e_{ij}$  为第  $i$  个任务点周围的第  $j$  个会员的信誉值,  $k_2$  为特征值。我们判断会员的信誉值会影响会员的供求关系而影响价格,并且我们认为会员的信誉值是通过影响会员预定任务开始时间进而影响市场的供给关系从而影响价格。我们利用 SPSS 对会员信誉值与会员预定任务开始时间进行相关回归分析,发现它们存在显著负相关关系,也就是说信誉值越高的会员通常他们开始预定任务的时间越早。回归分析结果如表 1-1 所示。

表 1-1 会员信誉值与预定任务开始时间回归分析结果

	平方和	自由度	平均值平方	F	显著性	标准化系数
回归	0.041	1	0.041	92.295	0	$-2 \times 10^{-6}$
残差	0.828	1875	0			
总计	0.868	1876				

当显著值  $P < 0.5$  则认为存在显著性相关，从结果  $P = 0$ ，可以看出会员信誉值与预定任务时间存在显著性负相关。

### (3) 总任务限额的定义与求值

我们把任务点方圆  $1\text{ km}$  以内的会员任务限额加总求和定义为总任务限额  $Q_i$ 。计算公式如下：

$$Q_i = \sum_{j=1}^n q_{ij} \quad (5-4)$$

$Q_i$  为第  $i$  个任务点的总任务限额， $q_{ij}$  为第  $i$  个任务点周围的第  $j$  个会员的任务限额。我们推断会员任务限额会影响市场的供求关系进而影响任务定价。

### (4) 多元线性回归模型的建立

根据上面的影响因素  $D_i$ 、 $E_i$  与  $Q_i$  和题目给出的项目各个任务的预期定价  $W_i$ ，建立多元线性回归模型，利用这些得到的数据拟合函数。多元线性回归模型如下所示：

$$W_i = \beta_1 D_i + \beta_2 E_i + \beta_3 Q_i + \lambda \quad (5-5)$$

$\beta_1$ 、 $\beta_2$ 、 $\beta_3$  为回归系数， $\lambda$  为常数。我们利用 SPSS 求出回归系数和常数的，回归系数和常数的值要使得函数值与真实值尽量全部一样，也就是要使得尽可能多的点落在直线上，而不在直线上的点均匀分布在两侧。我们使用  $\sum (P_i - W_i)^2$  来衡量模型的偏差，也就是说，我们回归模型的结果要尽量将  $\sum (P_i - W_i)^2$  控制的足够小。

## 5.1.2 定价模型的求解

我们将题目给出的数据进行分析计算，我们利用 MATLAB 计算出各任务点三个变量的值，程序代码见附录一，完整结果见支撑材料。这里我们给出部分的结果。

表 1-2 变量部分结果

任务序号	完成情况	任务定价	供求距离	消偏信誉值	总配额
1	0	66	0	0	0
2	0	65.5	2.158	35.803	9
3	1	65.5	1.089	8939.362	6
4	0	75	0	0	0
5	0	65.5	0.413	7.552	4
6	0	75	0.247	0.194	1
7	1	65.5	1.198	21.609	5
8	0	65.5	0.868	96.954	0
9	0	66	0.094	10.676	5
10	1	66	1.280	103.498	7

在计算各任务点变量值时，我们多次改变特征值，发现特征值  $k_1=0.6$ ， $k_2=0.4$  时拟合效果更好更直观，所以我们取特征值  $k_1$  为 0.6， $k_2$  为 0.4。接着，我们将数据进行多元线性回归，拟合出一条直线并求出任务定价的函数表达式。

我们利用 SPSS 进行多元线性回归，我们得到的多元直线函数表达式为

$$W_i = 2.622D_i - 0.019 E_i - 0.072Q_i + 71.599 \quad (5-6)$$

$D_i$  为第  $i$  个任务点的供求距离， $E_i$  为第  $i$  个任务点的消偏信誉值， $Q_i$  为第  $i$  个任务点的总任务限额， $W_i$  为任务预期定价。

回归结果如下表所示：

表 1-3 多元线性回归显著性分析结果

	平方和	自由度	平均值平方	F	显著性 sig
回归	1554.081	3	518.027	30.854	0
残差	11752.641	700	16.789		
总计	13306.722	703			

当显著性 sig 值小于 0.05 时，我们认为线性回归模型通过检验。这里的显



著值为 0 则通过回归模型通过显著性检验。

表 1-4 各系数及其显著性分析结果

	$\lambda$	$\beta_1$	$\beta_2$	$\beta_3$
系数	71.599	2.622	-0.019	-0.072
T	196.825	8.041	-3.281	-2.659
显著性	0	0	0.001	0.008

表 1-4 是三个变量与常数的系数值和 T 显著性检验的结果。当显著值小于 0.05 时则认为该变量与项目定价存在显著性相关。这里的结果显示，供求距离与消偏信誉值、总任务限额都存在显著性相关，并且我们得到他们的系数值。

至此，我们得出任务定价的规律。任务的定价与任务周边会员数量和离其的距离有关，另外周边会员的信誉值也会影响任务的定价，我们分析出会员任务的限额也会影响任务的定价。我们还得到任务定价与各变量之间满足以下关系： $W_i = 2.622D_i - 0.019E_i - 0.072Q_i + 71.599$ 。任务周边会员数量和离其的距离通过影响  $D_i$  从而影响定价；周边会员的信誉值影响  $E_i$  从而影响定价；任务限额的影响通过  $Q_i$  影响定价。在这个模型中，我们对周边的定义为任务点周围方圆 1 km。

## 5.2 基于 K-means 聚类模型分析任务未完成原因

### 5.2.1 未完成原因的分类

由于假设不考虑任务本身难度，故任务未完成只能是没有会员接单导致任务匹配超时自动视为未完成。将附件一中所有未完成任务的各指标数据代入原始定价模型的定价函数  $W_i$  求解出所有任务的预期定价，再将其与实际定价比较，发现三种结果：

(1)  $P_i < W_i$ ，即预期价格小于实际价格。由于将会员看作“商品”，故其价格会围绕供求关系上下波动，若任务定价不符合真实情况下市场定价规律，即预期价格大于实际价格，则任务报酬未达到会员预期，会导致无人愿意接受任务，导致任务无法完成。故从该种结果中的出任务未完成的原因之一，即价格过低。

(2)  $W_i = \lambda$ ，即供求距离为 0，以任务点为中心半径  $r$  范围内无会员分布，即

任务点不属于任一会员可接受接单范围，如表 1-2 所示。故得出另一导致任务未完成情的原因，即无人员分布。

(3)  $P_i > W_i$ ，即预期价格大于实际价格。该种情况相较前两种更不易观察，任务既满足合理定价规律，又存在人员分布。故将该种情况下的未完成原因可能与地形，交通便利程度等因素有关，故需再将该情况下所有数据单独进行进一步分析。

### 5.2.2 未知原因的 K-means 聚类探索

#### (1) K-means 聚类方法原理<sup>[2]</sup>

K-means 聚类的基本思想是初始随机生成  $m$  个聚类中心，然后按照最近原则把样本点分到各个类别，然后计算每个类别所以点位置的平均值，作为新的聚类中心，不断迭代下去，直到聚类中心的移动距离小于某个值。

将  $P_i < W_i$  的所有未完成任务汇总进行 K-means 聚类，具体步骤如下：

1. 随机生成  $m$  个聚类中心

2. 计算每个样本点到聚类中心的距离，将每个点聚类到离该点最近的类别中去。我们采用的距离是欧几里得距离，对于两个数据点  $(x_1, y_1)$  和  $(x_2, y_2)$ ，计算公式为

$$Euclid = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (5-7)$$

3. 利用得到的类别集合计算每个类别中所以点的平均值，作为新的簇中心对于一个类别集合  $A = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ ，类别  $A$  的新中心为  $(\bar{x}, \bar{y})$ ， $\bar{x} = (x_1 + x_2 + \dots + x_n) / n$ ， $\bar{y} = (y_1 + y_2 + \dots + y_n) / n$ 。

4. 反复进行 2, 3 步，直到聚类中心不再进行大范围移动。

改变  $m$  的值进行多次测试，发现当  $m$  为 3 时效果最佳，如图 1-1 所示，其各类的聚类中心如图 1-2 所示。

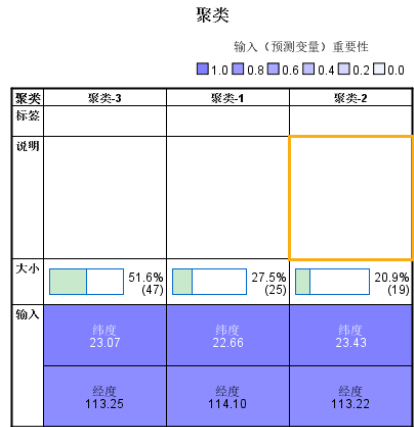
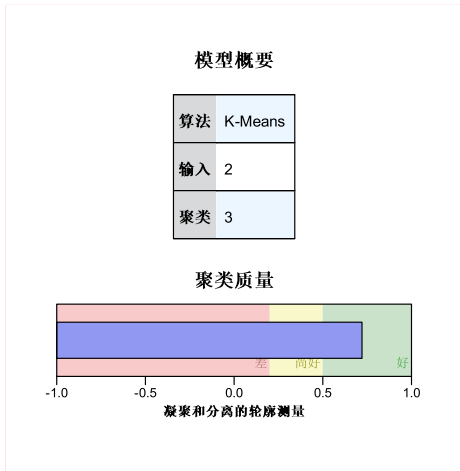


图 1-1 聚类数为 3 时的聚类效果图

图 1-2 各类聚类中心经纬度

每类所包含的部分任务序号如表 1-5 所示，完整聚类情况见支撑材料。三个类别的聚类中心如图 1-2 所示

表 1-5 部分聚类结果

任务序号	纬度	经度	聚类结果
437	23.47484	113.1908	类-2
167	23.23746	113.2043	类-3
436	23.37544	113.2135	类-2
224	23.39484	113.2215	类-2
240	23.39484	113.2215	类-2
658	22.91391	113.2241	类-3
107	22.67052	113.9236	类-1
102	22.65235	113.9299	类-1

考虑到未知原因可能与地理位置等因素有关，故在地图上搜索定位将各类聚类中心分析其周边地理分布，定位结果的平面位置图如图 1-3、1-4、1-5 所示。



图 1-3 类 1 中心

图 1-4 类 2 中心

图 1-5 类 3 中心

定位结果：

- (1)类 1 中心：清平高速与机荷高速附近。
- (2)类 2 中心：理工职业技术学校附近。
- (3)类 3 中心：珠江江岸附件。

由于 K-means 聚类的依据为离中心的距离，故各类点都应靠近其中心。分析定位结果可知，类 1 聚集在高速公路，位置较为偏僻；类 2 虽聚集在技校周边，但该技校自身位置就已经较为偏僻；类 3 任务点聚集在在芳村大道南附近，即珠江边的沿江路旁，也是相对偏僻的地方。所以第三个导致任务未完成的原因可以推测为地理位置，地理位置偏僻，人口密度相对较低的任务点越难找到合适的接单会员，导致任务越难完成。

### 5.2.3 未完成原因分类结论

根据对预期价格和实际定价的比较，以及对聚类模型的结果分析，可以将任务未完成的原因分为三类：

(1)价格过低。根据原始任务的实际定价低于利用原始定价模型计算得出的会员的预期定价，会员选择放弃接受未达期望报酬的任务，导致任务没有其接受者，从而其结果为未完成。我们共搜索出 186 个未完成任务点，具体见支撑材料

(2)无人员分布。任务发布点的有效识别范围内无会员分布，导致任务没有其接受者，从而其结果为未完成。我们搜索出了 36 个这种原因下的未完成任务点，具体见支撑材料。

(3)地理位置偏僻。任务发布点的有效识别范围内有会员分布，但任务发布点的地理位置位于相对偏僻，人口密度低的地理位置如高速公路、机场、村庄、郊区、城外沿海或沿江路等，而任务定价与地理位置相对优越的发布点采用同一方案，即同一定价函数，虽实际定价大于该方案下的预期定价，但未考虑到任务可接受者对偏僻位置的价格补偿，即会员真实的预期定价大于原始定价模型下的预期定价。导致实际定价小于会员真实的预期价格，从而其结果为未完成。我们搜索出了 91 个未完成任务点，具体见支撑材料。

## 5.3 修正模型的建立与求解

在问题一中我们分析了任务未完成的原因，归纳出三种原因：周围有会员但

定价过低、定价不低但有效识别范围内没有会员和定价不低且有效识别内有会员但受地理位置影响而任务没有被完成。我们认为应该针对这三种情况，分别建立新的定价方案使得这些任务未完成的点定价更合理。

### 5.3.1 第一种原因下未完成任务点的新定价模型的建立与求解

我们得出的第一种原因是周围不缺乏会员但定价过低使得任务未完成。我们使用已完成的任务点数据进行多元线性回归，利用直线拟合，得到函数表达式，再将未完成任务的变量数据代入函数表达式得出新的任务定价。也就是说，利用未完成的任务点拟合出的直线的函数表达式为第一种原因下未完成任务的新定价模型。类似于第一问的模型建立，我们选用了三个变量作为任务定价的影响因素，分别是供求距离、消偏信誉值和总任务限额。

由问题中的模型可知三个变量的定义与求值，供求距离  $D_i = \sum_{j=1}^n d_{ij} / n^{k_1}$ ，消偏信誉值  $E_i = \sum_{j=1}^n e_{ij} / n^{k_2}$ ，总任务限额  $Q_i = \sum_{j=1}^n q_{ij}$ 。多元线性回归模型如下所示：

$$W_i' = \beta_1' D_i + \beta_2' E_i + \beta_3' Q_i + \lambda' \quad (5-8)$$

$\beta_1'$ 、 $\beta_2'$ 、 $\beta_3'$  为回归系数， $\lambda'$  为常数。我们利用 MATLAB 计算出每个任务的三个变量值，再利用 SPSS 求出回归系数和常数的，回归系数和常数的值要使得函数值与真实值尽量全部一样，也就是-要使得尽可能多的点落在直线上，不在线上的点均匀分布在两侧。我们使用  $\sum (P_i - W_i')^2$  来衡量模型的偏差，也就是说，我们回归模型的结果要使得  $\sum (P_i - W_i')^2$  尽量小。

我们也是先求出各个任务已完成的点的三个变量数值，我们用 MATLAB 计算出结果，程序代码和算法与第一问相同见附录一。具体数据结果见支撑材料，这里我们给出部分数据。

表 3-1 已完成任务点的变量部分结果

任务序号	供求距离	消偏信誉值	总任务限额	任务定价
3	1.089	69.585	6.0	65.5
7	1.198	28.837	5.0	65.5
10	1.280	20.152	7.0	66.0

13	0.711	15.569	10.0	65.5
14	1.286	13.738	0.0	66.0
15	0.720	12.831	5.0	66.0
16	1.016	12.707	40.0	66.0
17	1.734	11.896	0.0	65.5
19	0.929	10.727	10.0	65.5
20	0.587	9.645	7.0	66.0

在计算变量值时，特征值的取值与第一问相同  $k_1 = 0.6$ ,  $k_2 = 0.4$ ，接着，我们将数据进行多元线性回归，拟合出一条直线并求出任务定价的函数表达式。我们利用 SPSS 将三个变量与任务定价进行多元线性回归并进行拟合，得到新定价规律：

$$W_i' = 3.425D_i - 0.138E_i - 0.02Q_i + 72.417 \quad (5-9)$$

回归结果如下表所示：

表 3-2 多元线性回归显著性分析结果

	平方和	自由度	平均值平方	F	显著性 sig
回归	2245.337	3	748.446	39.365	0
残差	9848.736	518	19.013		
总计	12094.073	521			

当显著性值 sig 小于 0.05 时，我们可以认为自变量与因变量之间存在显著线性关系。结果显示显著性值 sig 为 0，则可认定回归方程通过显著性检验。

表 3-3 各系数及其显著性分析结果

	$\beta_1'$	$\beta_2'$	$\beta_3'$	$\lambda'$
系数值	3.425	-0.138	-0.02	72.417
T	8.763	-2.674	-2.858	232.420
显著性	0	0.008	0.004	0

结果显示，每个系数的显著性值都小于 0.05，则可认为三个变量都与任务的定价存在显著性相关，通过显著性检验。

我们得到这种原因下未完成任务点的新定价方案：

$$W_i' = 3.425D_i - 0.138E_i - 0.02Q_i + 72.417$$

接着,我们利用新的定价方案,对这种情况下未完成任务点进行修正定价,这里我们给出部分任务点的新定价我们称为修正预期价格,完整结果见支撑材料。

表 3-4 第一种原因下的部分未完成任务修正定价结果

任务序号	真实定价 $P_i$	预期价格 $W_i$	修正预期价格 $W_i'$
5	65.5	69.2	69.3
8	65.5	68.4	68.4
9	66.0	70.5	70.8
11	65.5	66.6	64.8
12	65.5	66.1	65.0
18	66.0	70.2	70.4
23	65.5	68.9	68.9

由结果我们可以看出任务定价都有提高,那么经过重新定价以后这些任务有很大的可能性会被完成。

### 5.3.2 第二种原因下未完成任务点的新定价模型的建立与求解

我们得到的第二种原因是定价不低但有效识别范围内没有会员,这些任务点是属于周围没有会员没有被完成的类型,这种情况下需要提高定价吸引会员前来完成。在模型建立的时候我们首先考虑从问题一得到的定价函数进行改进提高任务定价,但是这些任务点周围是没有会员的也就是说这些任务点是不具备定价函数的变量数据,因此这种想法被推翻。接着,我们将这种类型的任务点与周围没有会员却被完成的类型任务点联想在一起,认为这些未完成任务点的定价与周围没有会员却被完成的类型任务点的定价有关。

我们建立新的模型给出第二种原因下未完成任务的定价,我们认为这些定价等于离它最近的已完成任务点的定价,但是这些任务点必须满足方圆  $1\text{ km}$  没有会员却被完成这个条件我们称为目标任务点。我们就得到这种原因下未被完成的任务的新定价模型:

$$\begin{cases} W'_m = P_n \\ D_n = 0 \\ D_m = 0 \\ \min d_{mn} \end{cases} \quad (5-10)$$

$W'_m$  为这种原因下第  $m$  个任务点的新预期定价它是未完成的； $P_n$  为这个未完第  $n$  个任务点的真实定价它是已完成的； $D_m$  和  $D_n$  分别是任务点  $m$  和  $n$  的供求距离, 它们都为 0 说明它们方圆  $1km$  以内没有会员。我们利用 MATLAB 搜索每一个这种原因下未完成任务点附件满足模型限制条件的已完成任务点, 这个已完成任务点的真实定价就是未完成任务点的新定价, 我们利用 MATLAB 对目标任务点进行搜索, 程序代码和算法见附录二我们得到未完成任务点的新定价, 这里我们给出部分结果, 完整结果见支撑材料。部分结果见下表所示。

表 3-5 第二种原因下的部分未完成任务修正定价结果

任务序号	任务定价 $P_i$	目标任务点序号	重新定价 $P'_i$
1	66.0	786	80
4	75.0	786	80
25	67.5	745	75
29	66.0	786	80
30	68.0	777	85
56	73.0	777	85
58	69.0	777	85

### 5.3.3 第三种原因下未完成任务点的修正定价模型建立与求解

第三种原因是任务点的定价不低但地理位置偏僻, 因此我们认为应该在原先的定价上进行位置偏僻的价格补偿, 也就是在原基础上提高定价。我们在问题一得到的函数上面加一个常数地理位置因子  $C$ 。在这个模型中我们是这样计算  $C$  的:

$$C = \sum |W'_i - P_i| / n_3 \quad (5-11)$$

$W'_i$  是将这种原因下未完成的任务点的数据代入 5-7 公式中求出的定价,  $P_i$



为任务的真实定价，地理位置因子  $C$  是两个定价差值的绝对值的平均数。然后我们就能得到这种情况下的任务修正定价模型：

$$W_i' = \beta_1' D_i + \beta_2' E_i + \beta_3' Q_i + \lambda' + C \quad (5-12)$$

$\beta_1''$ 、 $\beta_2''$ 、 $\beta_3''$  为回归系数， $\lambda'$  为常数。新任务定价只是在问题一中函数的基础上加上一个地理位置因子。我们计算出地理位置因子为 5.5，这样们就可以给出任务修正定价模型：

$$W_i' = 3.425D_i - 0.138E_i - 0.02Q_i + 72.417 + 5.5 \quad (5-13)$$

公式可以简化为：

$$W_i' = 3.425D_i - 0.138E_i - 0.02Q_i + 77.917 \quad (5-14)$$

利用新的定价模型我们可以对这种原因下的未完成任务重新定价，我们计算出结果，完整结果见支撑材料，这里我们给出部分结果如下表所示：

表 3-6 第三种原因下的部分未完成任务修正定价结果

任务序号	真实定价 $P_i$	修正预期定价 $W_i'$	重新定价 $P_i'$
2	65.5	61.3	66.8
6	75.0	70.4	75.9
35	66.5	62.1	67.6
37	70.0	70.2	75.7
38	75.0	70.2	75.7
40	66.0	63.5	69.0
44	69.0	64.3	69.8

### 5.3.4 问题二新定价模型总结

考虑到已完成的任务中存在由三类原因导致的不合理定价，故修正定价模型也从这三类原因出发分别进行修正。

(1) 第一类修正：从附件一中筛选已完成任务的相关数据仿照原始定价模型再次进行多元线性回归拟合，从而对问题一中第一类未完成原因下的任务重新计算其修正价格，以此作为这些任务的修正价格。

(2) 第二类修正：对于由于所选范围内无会员分布的未完成任务的定价修正，

采用离其最近的相同条件下的已完成项目的定价作为修正价格。

(3) 第三类修正：对于属于第三类因地理位置影响导致未完成的任务，引入地理位置因子对其进行修正，定义为用实际价格与用第一类修正模型下的拟合函数所得的预期价格的差值的期望值。

## 5.4 打包定价模型的建立与求解

### 5.4.1 任务装包模型

问题三要求将任务打包发布，其目的是为了防止出现因任务位置集中导致用户争相选择的情况。将任务积累打包再同时发布，可以增大一次发布任务的数量，缓解会员“供大于求”的情况。故将任务装包的关键因素为各任务之间的绝对距离（直线距离，不考虑实际情况下的连通路程）。

故任务装包的基本思想为：按任务序号升序搜索，以某一半径  $r$  搜索其范围内所有其它任务，将同属同一范围的任务装入同一任务包，再进行发布，按入包的先后顺序剔除重复任务，使每个任务包中的任务各不相同。因为原始定价模型中采用半径为 1，任务距离过大可能导致匹配会员时出现较大误差，故也令  $r = 1$ ，即判断任务是否装入同一任务包的依据为：

$$d_{mn} \leq 1 \quad (5-15)$$

其中  $m, n$  表示任务序号， $d_{mn}$  表示第  $m$  个任务与第  $n$  个任务间的绝对距离。

按此方法进行任务装包的最终结果为共装 217 包，装包情况如表 3-1 所示，完整装包情况见支撑材料。具体算法代码见附录三。

表 4-1 部分任务装包情况

任务包序号	包内任务序号
1	29、362、444
2	364、4446
3	360
4	22、23、32、462
5	52、61、68

### 5.4.2 任务包整体多元线性定价模型

任务装包后在某一时间发布，即包内所有任务同时发布，由于同一任务包内任务位置集中，故采用整体法将一个任务包看作一个“大任务”。故每个包有其自己的定价，且包内所有任务价格都等于任务包的预期定价。由修正定价模型知任务的预期价格四因素有关：供求距离、消偏信誉值、总配额和地理位置。任务打包发布时同一包内任务发布时间相同，一定程度上消除了最早预订任务时间对定价的影响，但仍与上述四因素相关，故任务包定价参照原始定价模型进行多元线性回归拟合，设任务包的预期价格为：

$$W_i'' = \beta_1'' D_i + \beta_2'' E_i + \beta_3'' Q_i + \lambda'' \quad (5-16)$$

拟合所用的任务包的实际价格  $P_n$  采用包内各任务实际价格的平均值作近似处理， $D_i$  为供求距离， $E_i$  为消偏信誉值， $Q_i$  为总任务限额，利用 SPSS 进行多元线性拟合结果如表 3-2, 表 3-3 所示：

表 4-2 任务包定价模型显著性分析结果

	平方和	自由度	平均值平方	F	显著性 sig
回归	1138.25	3	426.336	35.368	0
残差	8867.54	214	14.235		
总计	10005.79	217			

表 4-3 任务包定价函数系数检验结果

	$\beta_1''$	$\beta_2''$	$\beta_3''$	$\lambda''$
系数值	2.689	-0.115	-0.04	71.568
T	8.549	-2.588	-2.901	197.012
显著性	0	0.012	0.008	0

我们得到任务包定价模型：

$$W_i'' = 2.689 D_i - 0.115 E_i - 0.04 Q_i + 71.568 \quad (5-17)$$

#### 5.4.3 打包定价模型的优劣度检验

利用拟合出的定价函数对附件一的未完成任务进行再定价，将其与修正定价模型中的对未完成任务的修正定价进行比较检验打包定价模型的优劣。规定若

$$-1.05 * W_i' < W_i'' < 1.05 * W_i' \quad (5-18)$$

则认为打包定价合理。具体操作如下：

- (1) 找到未完成任务所属任务包。
- (2) 将该任务的修正定价  $w_i'$  与所属任务包定价  $w_i''$  比较。
- (3) 若满足上述公式则认为该打包定价模型能使该未完成任务被完成。

规定用优劣度  $\alpha$  作为打包定价模型的合理程度， $\alpha$  的计算公式：

$$\alpha = \frac{\text{属于合理定价范围内的任务总数}}{\text{已知总任务数}} \quad (5-19)$$

部分比较结果如表 3-4 所示，完整对照表见支撑材料。

表 4-5 部分价格对照

任务序号 \ 价格	修正价格	打包价格	是否属于合理范围
5	69.3	70.2	是
8	68.4	68.9	是
1	66	64.8	是
4	75	72.1	是
2	61.3	65.7	否
6	70.4	72	是

表 4-5 分别选取了修正定价模型中的三类未完成原因下的任务各 2 个进行展示。将所有未完成任务进行价格比照，在 (公式 5-16) 的参照标准下发现  $\alpha = 84.22\%$  的定价属于合理范围。故可得出打包定价对任务的完成情况有积极影响，增大任务被完成的概率。

## 5.5 价格预测评价模型

### 5.5.1 价格预测模型的确立

由问题一、二、三的分析建模得到了原始定价模型、修正定价模型，以及打包定价模型。将三种定价模型进行对比可知，修正定价模型是在原始定价模型上的改进，但由于修正定价模型需要考虑每个任务点的地理位置因素，还需分类讨论，较为繁琐，而由打包定价的优劣度评价知打包定价模型的优劣度  $\alpha > 80\%$ ，故可认为打包定价模型的合理程度相对较高，且其不需分情况及考虑地理位置因素，故在新任务的预测中采用打包定价模型，即 (5-14)，将附件三数据代入得各新

任务预期价格，选取属于不同任务包的任务进行展示，如表 4-1 所示。

表 4-1 新任务部分

任务序号 \ 价格	所属任务包	供求距离	消偏信誉值	总任务限额	预测价格
C0001	1	0.0922	10.7238	260	68.5
C0023	2	1.2859	2606.9992	4478	70
C1030	3	0	0	0	71.6
C1339	4	0.7678	26.5362	1200	72.5
C1553	5	1.0113	55.6259	2000	75.1
C1637	6	0.6752	42.9977	1500	74.3

### 5.5.2 价格预测模型的最短距任务价格评价

虽然打包定价模型有一定的合理性，但其与修正定价模型还存在一定误差，加上拟合的误差，故还需进行价格预测模型的评价。对预测定价的合理性评价采用以最短距任务价格确定合理范围进行评价，具体步骤如下：

(1) 将每个新任务  $m$  与所有已结束项目分别计算绝对距离，记为  $d_{mn}$ 。具体计算方案见附录三。

(2) 每个任务点都取其  $\min d_{mn}$  所对应的已结束项目的实际价格  $P_n$  作为该新任务点的参考中心价格。利用 MATLAB 计算的部分结果如表 4-2 所示，具体代码见附录四，整结果见支撑材料

表 4-2 部分新任务的参考中心价格

任务号码	最短距已结束任务点序号	参考价格
C0001	A0442	69
C0002	A0077	65.5
C0003	A0047	67
C0004	A0077	65.5

C0005	A0442	69
-------	-------	----

- (3) 参照打包定价模型的评价方法，取±5%确定合理定价范围。
- (4) 判断记录由价格预测函数计算得到的预测价格是否属于合理定价范围。
- (5) 计算优劣度 $\alpha$ 的值判断其是否大于0.8，若满足则可认为模型具有一定的合理性。部分比较结果如表4-3所示。

表 4-3 部分评价结果

任务号码	最短距已结束任务点序号	参考价格	预测价格	是否合理
C0001	A0442	69	70.2	是
C0002	A0077	65.5	67.5	是
C0003	A0047	67	65.5	是
C0004	A0077	65.5	66	是
C0005	A0442	69	67.5	是

统计发现新任务总数为2066，合理定价任务数为1672，计算得优劣度

$$\alpha = \frac{\text{合理定价任务数}}{\text{总任务数}} = \frac{1672}{2066} \approx 0.8087, \text{ 大于 } 0.8, \text{ 于是得出评价结果: 打包定价}$$

模型对新任务定价效果较好，有一定的可信度。

## 六、模型的优缺点与改进

### 6.1 模型的优点

(1) 我们建立的任务定价模型都是从任务与会员的供求关系出发，研究影响任务定价的因素是以影响市场供求关系为切入点。该模型与微观经济学知识紧密联系。

(2) 在求解供求距离与消偏信誉值时，我们引入特征值。因为平均距离一样但会员分布不同这样的对于任务定价有不同影响，如果供求距离单纯是一个平均数，那么不论会员分布如何只要平均距离是一样的会员信息对任务定价的影响是相同的，而特征值得引入可以消除这种影响。

### 6.2 模型的缺点

- (1) 在考虑任务定价的影响变量时，我们可能忽略了一些重要因素变量。
- (2) 在建立模型时，我们没有过多的考虑现实中的客观情况对完成任务的影响。我们还假设了每个会员都是理性人，这对于分析实际生活情况会造成很大影响。
- (3) 在建立回归模型时，我们只进行了线性回归和直线拟合，没有尝试更多

的拟合方式

### 6.3 模型的改进

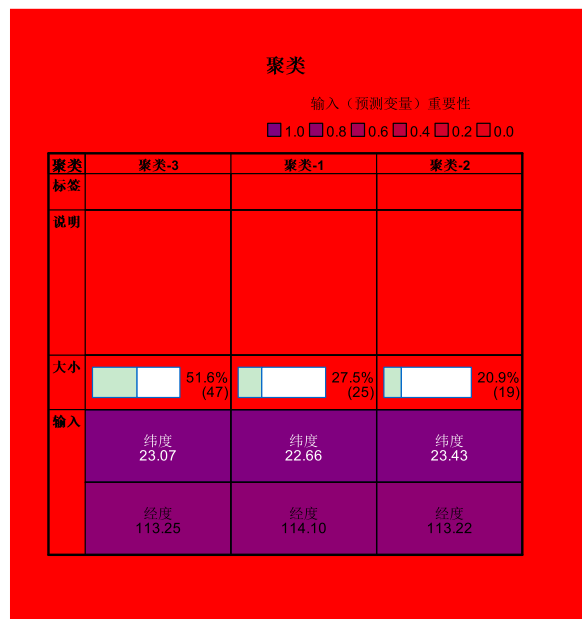
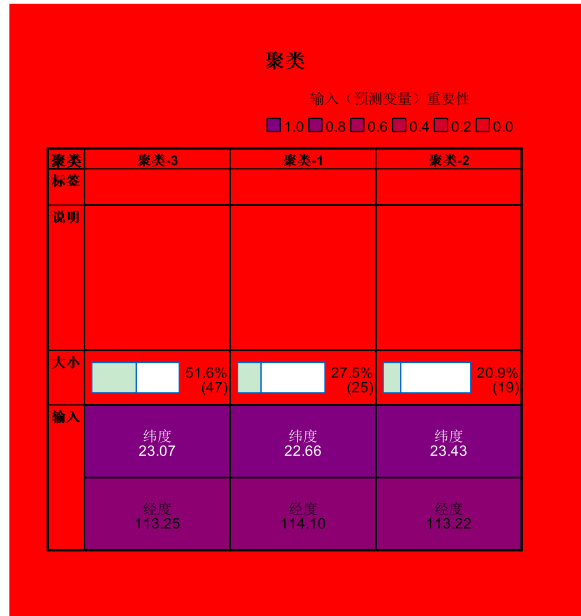
(1) 我们可以获得更多的数据以建立更多的变量，来使得定价模型更合理。

(2) 在分析会员与任务的供求关系时，可以结合行为经济学来对会员完成任务的行为做更多的分析，或许能够在供求关系的研究上有重大突破。

(3) 在进行函数拟合时，可以其他更多的函数去进行拟合。

## 七、参考文献

- [1] 国家地震局地球物理研究所.近震分析[M].地震出版社，1977.
- [2] 张文彤,钟云飞.数据分析与挖掘实战案例[M].北京：清华大学出版社,2013:308.
- [3] 李治问,金娥，熊强.用户质量对平台定价策略的影响研究[A].江苏大学,2017.
- [4] 吴瑞杰，孙鹏，孙昱.分布式任务计划动态调整模型及算法[A].空军工程大学信息与导航学院，2017.



附录一

```
function credit1=xinyu(a,c,e)
credit1=zeros(835,1);
n=1;
sum=0;
k1=0.4;
for i=1:835
    for j=1:1877
```



```

        if 111.199*sqrt((a(i,1)-c(j,1))*(a(i,1)-c(j,1))+a(i,2)-c(j,2))*(a(i,2)-
c(j,2))*cos((a(i,1)+c(j,1))/2)*cos((a(i,1)+c(j,1))/2)) <= 1
            sum=sum+e(j,1);
            n=n+1;
        end
    end
    if n==1
        credit1(i,1)=sum/(n^k1);
    else
        credit1(i,1)=sum/((n-1)^k1);
    end
    n=1;
    sum=0;
end
for i=1:835
    if credit1(i,1)>500
        credit1(i,1)=credit1(i,1)/5
    end
end
end

```

```

xlswrite('C:\Users\123\Desktop\信誉值.xls',credit1)
end

```

```

function credit2=pe2(b)
n=1;
sum=0;
c=xlswrite('C:\Users\123\Desktop\附件一：已结束项目任务数据 2.xls');
d=xlswrite('C:\Users\123\Desktop\会员位置.xls');
e=textread('C:\Users\123\Desktop\配额.txt');
task2=zeros(217,1);
k1=0.4;
for i=1:835
    for j=1:835
        sum=e(i,1);
        if b(j,i)~=0
            sum=sum+e(j,1);
            n=n+1;
        end
    end
    credit2(i,1)=sum/(n^k1);
    n=1;
end
xlswrite('C:\Users\123\Desktop\总配额.xls',task2)
end

```

```

function distance=jl(a,c)
distance=zeros(835,1);
n=1;
sum=0;
k1=0.6;
for i=1:835
    for j=1:1877
        if 111.199*sqrt((a(i,1)-c(j,1))*(a(i,1)-c(j,1))+(a(i,2)-c(j,2))*(a(i,2)-
c(j,2))*cos((a(i,1)+c(j,1))/2)*cos((a(i,1)+c(j,1))/2)) <= 1
            sum=sum+111.199*sqrt((a(i,1)-c(j,1))*(a(i,1)-c(j,1))+(a(i,2)-c(j,2))*(a(i,2)-
c(j,2))*cos((a(i,1)+c(j,1))/2)*cos((a(i,1)+c(j,1))/2));
            n=n+1;
        end
    end
    if n==1
        distance(i,1)=sum/(n^k1);
    else
        distance(i,1)=sum/((n-1)^k1);
    end
    n=1;
    sum=0;
end
xlswrite('C:\Users\123\Desktop\供求距离.xls',distance)
end

```

## 附录二

```

function [n,p]=nobodyprice()
a=textread('C:\Users\123\Desktop\无人且未完成.txt');
b=textread('C:\Users\123\Desktop\无人且完成.txt');
m=1000000;
n=zeros(36,1);
p=zeros(36,1)
for i=1:36
    for j=1:95
        if 111.199*sqrt((a(i,3)-b(j,3))*(a(i,3)-b(j,3))+(a(i,4)-b(j,4))*(a(i,4)-
b(j,4))*cos((a(i,3)+b(j,3))/2)*cos((a(i,3)+b(j,3))/2))) <= m
            m=111.199*sqrt((a(i,3)-b(j,3))*(a(i,3)-b(j,3))+(a(i,4)-b(j,4))*(a(i,4)-
b(j,4))*cos((a(i,3)+b(j,3))/2)*cos((a(i,3)+b(j,3))/2)));
            p(i,1)=b(j,1);
            n(i,1)=b(j,2);
        end
    end
end
end

```

end

### 附录三

```
function task2=pe2(b)
n=1;
sum=0;
c=xlsread('C:\Users\123\Desktop\附件一：已结束项目任务数据 2.xls');
d=textread('C:\Users\123\Desktop\会员位置.txt');
e=textread('C:\Users\123\Desktop\配额.txt');
task2=zeros(217,1);
k1=0.4;
count=0;
q=1;
for i=1:835
    for j=1:835
        if b(j,i)~=0
            for m=1:1877
                if 111.199*sqrt((c(j,1)-d(m,1))*(c(j,1)-d(m,1))+(c(j,2)-d(j,2))*(c(j,2)-
d(m,2))*cos((c(j,1)+d(m,1))/2)*cos((c(j,1)+d(m,1))/2)) <= 1
                    sum=sum+e(m,1);
                    n=n+1;
                end
            end
        end
    end
    if sum~=0
        task2(q,1)=sum;
        q=q+1;
    end
    n=1;
end
xlswrite('C:\Users\123\Desktop\第三问配额.xls',task2);
end
```

```
function distance2=jl2(b)
n=1;
sum=0;
c=xlsread('C:\Users\123\Desktop\附件一：已结束项目任务数据 2.xls');
d=xlsread('C:\Users\123\Desktop\会员位置.xls');
distance2=zeros(217,1);
k1=0.6;
for i=1:217
    for m=1:1877
```

```

        if 111.199*sqrt((c(i,1)-d(m,1))*(c(i,1)-d(m,1))+(c(i,2)-d(m,2))*(c(i,2)-
d(m,2))*cos((c(i,1)+d(m,1))/2)*cos((c(i,1)+d(m,1))/2)) <= 1
            sum=sum+111.199*sqrt((c(i,1)-d(m,1))*(c(i,1)-d(m,1))+(c(i,2)-d(m,2))*(c(i,2)-
d(m,2))*cos((c(i,1)+d(m,1))/2)*cos((c(i,1)+d(m,1))/2));
            n=n+1;
        end
    end
    for j=1:835
        if b(j,i)~=0
            for m=1:1877
                if 111.199*sqrt((c(i,1)-d(m,1))*(c(i,1)-d(m,1))+(c(i,2)-d(m,2))*(c(i,2)-
d(m,2))*cos((c(i,1)+d(m,1))/2)*cos((c(i,1)+d(m,1))/2)) <= 1
                    sum=sum+111.199*sqrt((c(i,1)-d(m,1))*(c(i,1)-d(m,1))+(c(i,2)-
d(m,2))*(c(i,2)-d(m,2))*cos((c(i,1)+d(m,1))/2)*cos((c(i,1)+d(m,1))/2));
                    n=n+1;
                end
            end
        end
    end
    end
    distance2(i,1)=sum/(n^k1);
    n=1;
    sum=0;
end
xlswrite('C:\Users\123\Desktop\第三问供求距离.xls',distance2);
end

```

```

function credit2=xinyu2(b)
n=1;
sum=0;
c=xlswrite('C:\Users\123\Desktop\附件一：已结束项目任务数据 2.xls');
d=xlswrite('C:\Users\123\Desktop\会员位置.xls');
e=textread('C:\Users\123\Desktop\信誉值.txt');
credit2=zeros(217,1);
k1=1.2;
q=1;
count=0;
p=zeros(217,1)
for i=1:835
    for j=1:835
        if b(j,i)~=0
            count=count+1;
            for m=1:1877
                if 111.199*sqrt((c(j,1)-d(m,1))*(c(j,1)-d(m,1))+(c(j,2)-d(j,2))*(c(j,2)-
d(m,2))*cos((c(j,1)+d(m,1))/2)*cos((c(j,1)+d(m,1))/2)) <= 1

```

```

                sum=sum+e(m,1);
                n=n+1;
            end
        end
    end
end
if sum~=0
    credit2(q,1)=sum/(n^k1);
    q=q+1;
end
n=1;
end
xlswrite('C:\Users\123\Desktop\第三问信誉值.xls',credit2)
end

```

```

function a=pro3(d)
a=zeros(835,835);
a=d;
for i=1:835
    for j=1:835
        if (a(j,i)<1 && a(j,i)~=0)
            a(j,i)=1;
        else
            a(j,i)=0;
        end
    end
end
for i=2:835
    for m=1:i-1
        for j=i+1:835
            if (a(j,m)==1 && a(j,i)==1)
                a(j,i)=0;
            end
        end
    end
end
end
end

```

## 附录四

```

function p=duibi()
p=zeros(2066,2);
b=xlswrite('C:\Users\123\Desktop\附件一：已结束项目任务数据 2.xls');
a=xlswrite('C:\Users\123\Desktop\附件三：新项目任务数据.xls');

```

```

j=1;
for i=1:2066
    m=111.199*sqrt((a(i,1)-b(j,1))*(a(i,1)-b(j,1))+a(i,2)-b(j,2))*(a(i,2)-
b(j,2))*cos((a(i,1)+b(j,1))/2)*cos((a(i,1)+b(j,1))/2));
    for j=2:835
        if 111.199*sqrt((a(i,1)-b(j,1))*(a(i,1)-b(j,1))+a(i,2)-b(j,2))*(a(i,2)-
b(j,2))*cos((a(i,1)+b(j,1))/2)*cos((a(i,1)+b(j,1))/2)) < m
            m=111.199*sqrt((a(i,1)-b(j,1))*(a(i,1)-b(j,1))+a(i,2)-b(j,2))*(a(i,2)-
b(j,2))*cos((a(i,1)+b(j,1))/2)*cos((a(i,1)+b(j,1))/2));
            p(i,1)=j;
            p(i,2)=b(j,3);
        end
    end
end
j=1;
end
xlswrite('C:\Users\123\Desktop\第四问任务匹配价格.xls',p);
end

```