**注：本论文仅供参考，不可直接提交。祝各位成功猎杀国奖！**

**B 题** **“拍照赚钱”的任务定价**

**摘 要**

在信息化时代，“拍照赚钱”已成为移动互联网下的一种自助式服务模式，通过用户下载 APP，注册成为 APP 的会员，然后从 APP 上领取需要拍照的任务，赚

* APP 对任务所标定的酬金。本文主要研究了用户任务执行情况变量与 gps 纬度， gps 经度，任务标价之间的函数关系问题。在充分理解题意及合理假设的基础上，利用 MATLAB，SPSS，Excel 软件对附件大数据进行预处理，对问题 1 用数据拟合的方法解决，对问题 3 用模糊数学的方法解决。

针对问题 1（任务定价），本文运用统计学中的数理统计方法首先建立了多元回归模型 I。利用 Excel 软件对附件给出的数据先剔除后利用分层抽样得到样本 400 个，以样本排序后的任务定价均值为聚点进行数据的筛选。画出任务定价与任务执行情况散点图，初步观察，依据最小二乘法原理，借助 SPSS 软件进行回归分析，求出回归系数，得到多元回归方程。但 R2 的值为 0.447，拟合效果不佳。于是我们建立了插值与数据拟合模型 II。根据任务定价与任务执行情况的关系式，将任务定价作为因变量，求解与任务执行情况的关系。利用 MATLAB 软件进行拟合得到 R2 的值为 0.74，拟合效果较好。同时，本文从样本中随机抽取了 3 组数据（每组 10 个采样），对模型 II 理论结果进行数据的检验。最终，运用对比思想，支持使用模型 II 来解决问题 1，从而得出定价规律。

针对问题２（定价方案），利用问题一得到的结论，通过插值与数据拟合模型II，即可完成问题三的定价问题。在模型分析里，我们不断地对上述模型的系数进行修正，相关程度始终保持不变，故该模型**灵敏度良好**，并且利用多种模型进行**鲁棒性分析**，相关程度变化较小，从而验证了模型的合理性、科学性和实用性。并进一步肯定了新方案的优越性。

针对问题 3（会员位置 GPS），利用逐步回归法在模型 I 的基础上首先建立多元逐步回归模型Ⅲ。因为模型 I 中两个自变量与ＧＰＳ的拟合度不高，所以我们首先引入对被解释变量卡顿时长占比贡献最大的解释变量 VMOS。借助 SPSS 软件求解回归系数，R2 的值为 0.695，拟合效果尚可。但由于缺乏对于非线性情况的讨论，为使模型具有实际可操作性，于是我们建立了模糊神经网络模型Ⅳ。由于附件总数据中任务执行情况为 0 的数据多，即有大量无用变量，故需对数据重新提取。利用 Excel 软件剔除附件一中任务执行情况为 0 以及附件二中预定任务限额１０以下的数据。画出任务执行情况与附件二中预定任务限额１０以上的对比图，发现图形重合度高。最后，支持使用模型Ⅳ来解决问题 3。

针对问题 4（新项目定价方案）利用问题三得到的结论，我们对自变量进行了正态性分布检验、交互相关性检验 ，确定了自变量之间的相互影响 , 且使用

**Spearman 秩相关系数**。利用任务定价与GPS关系得到积分方程，对其分别建立了**幂函数**曲线、**Logistic** 曲线增长模型，并基于Simulink动态仿真模型进行**多元非线性回归**，相关系数分别达到了 **0.7631**,**0.9592**。于是任务定价方案可以接受，该方案的实施效果较为理想。

最后，文章对模型的优缺点进行了客观的评价，以方便模型进一步的推广和使

用。

关键词：多元回归分析 插值与数据拟合 模糊网络模型 散点图 Spearman 系数 Logistic 曲线 鲁棒性分析;

§1.**问题的重述**

1.1 问题背景

“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

* 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。

1.2 问题提出

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。

1. 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。
2. 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？
3. 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。

§**2.问题的分析**

（一）问题 1.2 的分析

针对问题一：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。任务 GPS 经纬度与价格是用户是否下载使用 APP 存在着非常强的相关性。问题

1. 求执行任务情况与三个自变量（gps 纬度，gps 经度，价格）的函数关系式，属于分析自变量与因变量之间关系的数学问题，解决此类问题一般利用多元回归分析原理。

首先，题目附件给出的八百多个大数据，无法全部应用。这时，我们需要用正确方法进行数据预处理，剔除坏数据，筛选有用数据。其次，有时候在回归分析中，采用何种表达式只是一种猜测，该表达式是否满足还未可知，并且某些因子具有不可预测性，这些使得线性分析在某些情况下受到限制。

基于以上原因，我们可以首先建立一个多元回归的数学模型 I,然后建立一个插值与数据拟合的模型 II,利用 MATLAB 分析，SPSS 软件进行拟合，观测两个模型的 R²、sig 等统计量值，判断模型的拟合效果，通过对比，择优选择。

（二）问题 3 的分析

对于问题三同问题 1 类似，求解单个因变量与多个自变量的函数关系的问题。首先，我们要查看问题 1 的模型 I 的效果，考虑在此基础上逐步添加新变量，建立

回归模型。其次，对经纬度的小数位数保留不当的数据在接近九百个大数据中只占大概 4%，所以我们需要重新选取数据。我们在不了解价格与任务执行情况的具体关系前提下，可以利用模糊模型来求解。

基于以上原因，我们可以首先建立一个多元逐步回归的数学模型Ⅲ,然后建立一个模糊神经网络模型Ⅳ，比较最后的结果。

**（三）问题４的分析**

通过问问题一以及题三的分析可以以上述方法进一步的拟合问题四的新项目的任务定价方案，并通过拟合效果可以进一步的评价该方案的实施效果。分别采

* BP 算法和遗传算法优化的 BP 神经网络算法对ＧＰＳ和任务定价的的函数关系进行计算。首先对所用数据进行归一化处理，根据当前已有研究成果进行建模，最终可获得定价关于新项目ＧＰＳ的函数式并根据该关系式完成定价。

§**3.模型的假设**

为使解题过程得到简化，本文作出如下假设以规定模型的适用范围。1.假设文中所引用的文献和结论都正确可靠；

2.样本容量足够大，样本数据真实，能够反映具体情况；

3.样本数据中 GPS，任务标价，预定任务时间等反映其自身属性的数据波动范围较小，可取一个中间值固定下来呈现；

4.用户选择只与任务标价及 GPS 有关，无其他影响因素。

5.用户观看视频所用设备的性能指标相同；

6.用户对视频体验好坏的判断标准一样。

7.整个计时过程中无人为因素干扰；

|  |  |
| --- | --- |
|  | §4 名词解释与符号说明 |
|  |  |
| 符号 | 说明 |
|  |  |
| a1 | 任务标价与 gps 纬度之间对应的回归系数 |
| a2 | 任务标价与 gps 经度之间对应的回归系数 |
| a3 | 任务标价与任务执行情况之间对应的回归系数 |
| a4 | gps 纬度与任务执行情况之间对应的回归系数 |
| a5 | gps 经度比与任务执行情况之间对应的回归系数 |
| a6 | GPS 与任务预定限额对应的回归系数 |
|  | 任务标价 |
| v |  |

gps 纬度

V1

gps 经度

V2

|  |  |
| --- | --- |
| x4 | VMOS |
| e | 随机误差 |
|  |  |



|  |  |
| --- | --- |
| y | 任务执行情况 |
| ｚ | 预定任务限额 |
| ｇ | 信誉值 |
|  |  |

§5 模型建立与求解

第一部分：准备工作

(一）数据的处理

第一步，将附件一剔除经纬度小数点后位数不正确的数据后将剩余部分分为

1. 组（每组 2 个），利用 SPSS 软件分层抽样的方法每组抽取数据 400 个，总共抽取 4.5 万个样本；

第二步，利用 Average 函数求解样本任务标价的平均值，经计算，平均值为69.1107；

第三步，利用 Excel 排序功能对抽取样本的任务标价大小进行降序排列；第四步，前 18 条数据以及后 12 条数据与平均值差距过大，舍弃；

第五步，以剩余的数据的均值 67.8614 为聚点，以+1.5 为其邻域，删除邻域外的数据，至此，数据的预处理工作完成。

第五步，针对附件二用相同的方法，对 GPS 进行数据预处理，进行挑选并处理数据。

第二部分：解决问题 1 的两个模型

（一）模型 I（多元回归模型）

1.建模背景

查阅多元线性回归分析预测法概述**[1]**可知，多元回归，主要研究一个因变量与多个自变量之间的相关关系，亦称多元线性回归。它是反映一种现象或事物的数量依多种线性或数量的变动而相应地变动规律。建立多个变量之间线性或非线性数学模型数量关系式的统计方法。



设 y 为因变量，

时，则多元线性回归模型为：

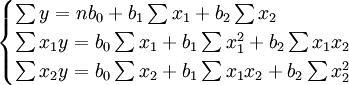


为自变量，并且自变量与因变量之间为线性关系

其中，b0 为常数项;为未知量的参数，即回归系数；e 为随机误差项。

多元性回归模型的[参数估计](http://wiki.mbalib.com/wiki/%E5%8F%82%E6%95%B0%E4%BC%B0%E8%AE%A1)，同一元线性回归方程一样，也是在要求误差平方

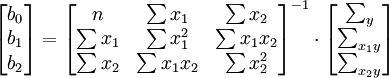
和()为最小的前提下，用[最小二乘法](http://wiki.mbalib.com/wiki/%E6%9C%80%E5%B0%8F%E4%BA%8C%E4%B9%98%E6%B3%95)求解参数。以二线性回归模型为例，求解回归参数的标准方程组为：



解此方程可求得 b0,b1,b2 的数值。亦可用下列矩阵法求得

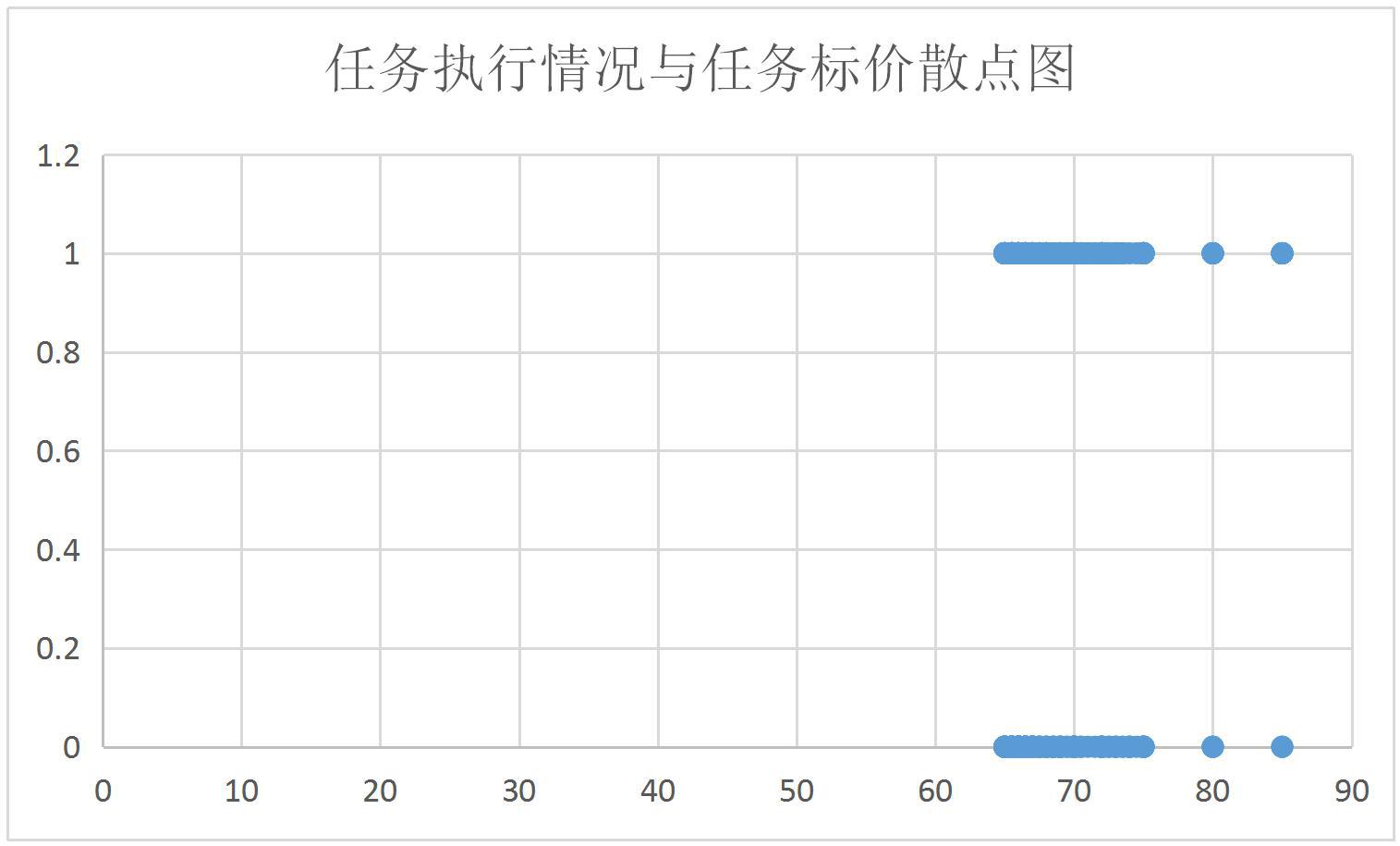


即

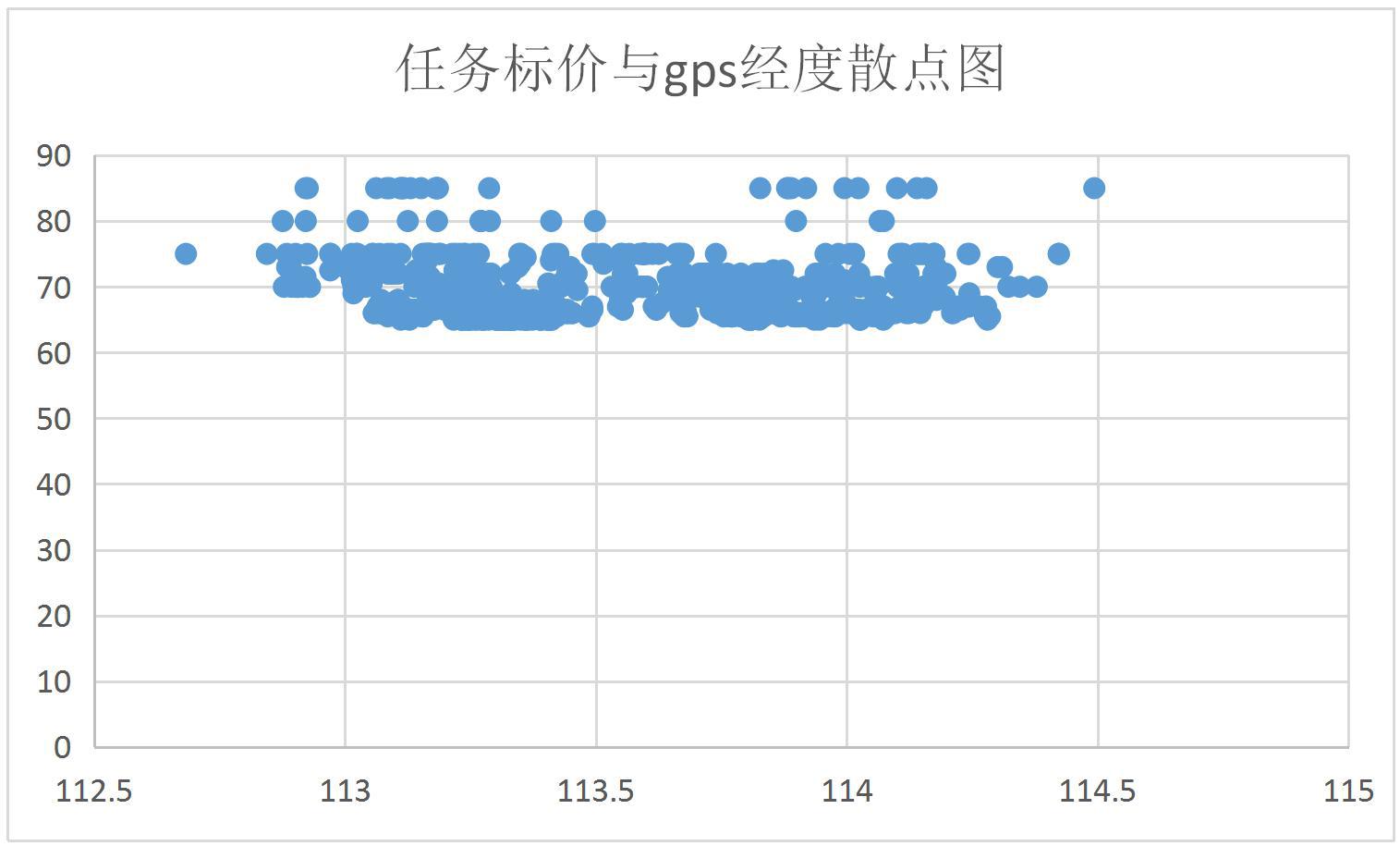


2.模型 I 的建立和求解

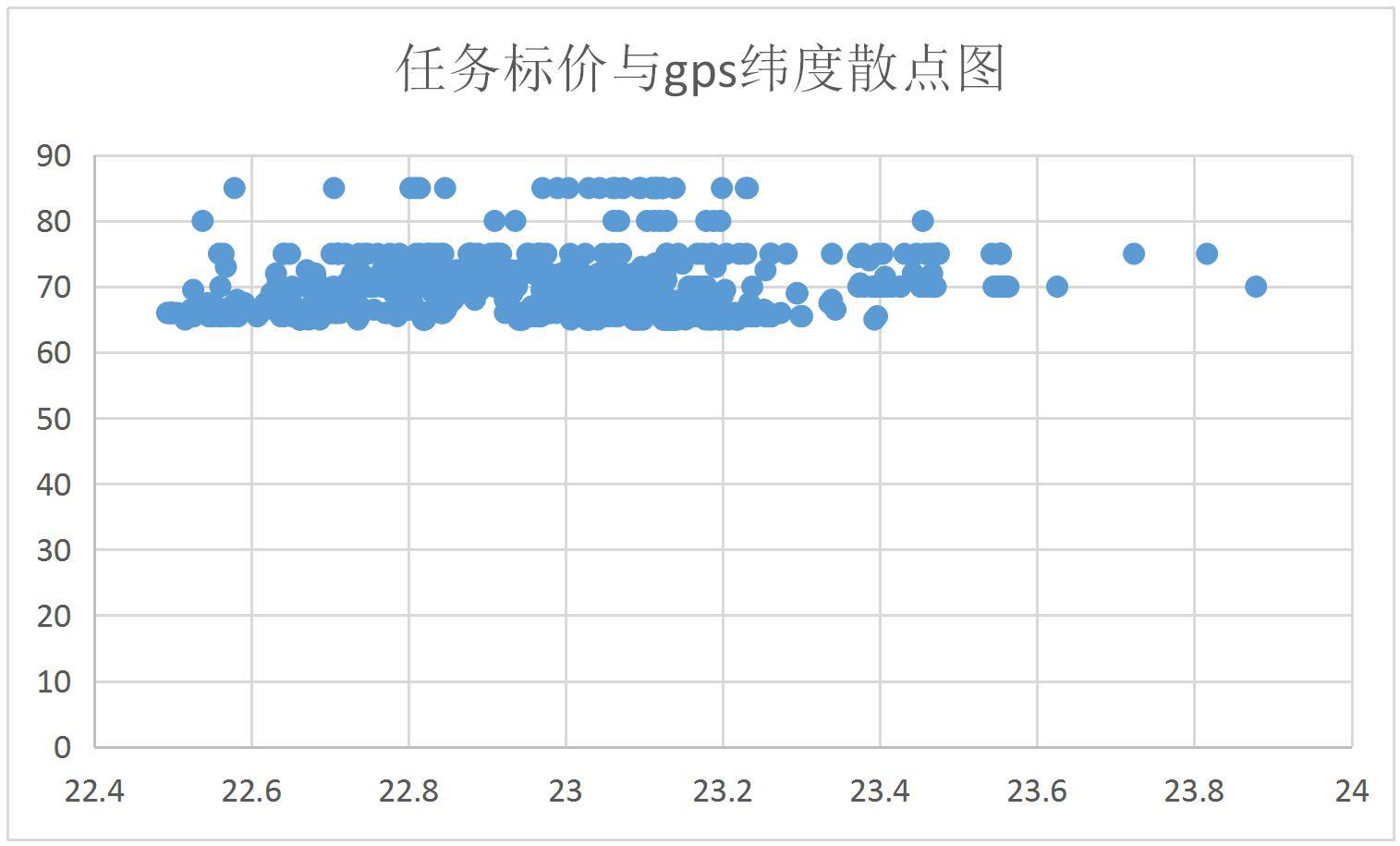
（1）针对任务标价与任务执行情况之间的函数关系，我们根据样本数据画出任务执行情况与 gps 纬度，gps 经度，任务标价的散点图。



**图 1-1**



**图 1-2**



* 1-3

(3)根据对图像的初步观察，我们假设初始缓冲时延与网络侧变量之间线性相关，建立如下回归方程：

y=a1x1+a2x2+a3x3+e

我们依据最小二乘法原理，借助 SPSS 软件对 a1，a2，a3 的值进行估计，可以得到以下图表：

表 5-2-1.Anovab



a.预测变量：（常量）纬度，经度，任务执行情况b.因变量：任务标价

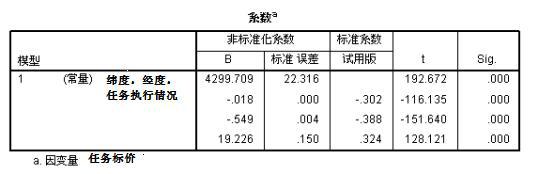


表 5-2-3.模型汇总



各个参数结果如下：

e = 4.729

a1 = −0.018

a2 = -0.549

a3 = 19.226

R2=0.447

所以任务定价与 gps 纬度，gps 经度，任务执行情况的函数关系式为：

y=-0.018x1-0.549x2+19.226x3+4.729

R2=0.447，拟合效果不佳，于是我们建立模型 II。

（二）模型 II（插值与数据拟合模型）

1.建模背景

参考西南石油大学插值与拟合 PPT**[2]**材料结论，在实际问题中常常要处理由实验或测量所得到的一些离散数据。插值与拟合方法就是要通过这些数据去确定某一类已知函数的参数或寻求某个近似函数，使所得到的近似函数与已知函数由较高的拟合精度。

如果要求的这个近似函数经过所已知的所有数据点，则称此类问题为插值问题。如果不要近似函数通过所有数据点，而是要求它能较好地反映数据变化规律的近似函数的方法称为数据拟合。

2.模型 II 的建立和求解

由上式我们得到任务定价与 gps 纬度，经度具有相关性，假设三者间的关系式为：

v=c1v12+c2v22+c3v1+c4v2+c5v1v2+g （2）

根据问题 1 的需要，用 SPSS 软件聚类分析，取得采样值。将采样所对应的特征值如附录表 1 所示。

利用 MATLAB 软件做线性最小二乘拟合，分别作多项 f(x)=d1xm+d2xm-1+...+d 拟合，可利用已有次序 a=ployfit(x,y,m) (其中 m 表示拟合多项式次数）；

多项式在 x 处的值 y 可以用一下命令计算：

y=polyval(d,x)

得到拟合结果如下：

c1=24.1000

c2=-1.0000

c3=9.8206

c4=2.0000

c5=-1.0000

g=2529.3522

R2=0.74

可得关系式如下：

v=24.100v12-v22+9.8206v1+2.0000v2-v1v2+2529.3522 (3)

3.模型 II 的数字模拟

从样本中随机抽取 4 组数据（每组 10 个采样）（见附录 2 表 2）对模型 II 理论结果进行数据的可靠性计量，其中一组的情况如下：



（三）问题 1 的两种数学模型的比较

模型 I 的拟合度小于 0.5，效果较差。模型 II 的拟合度大于 0.7，效果良好。于是对于问题 1 决定采用模型 II 解决。由以上可得出任务定价多分配于 65-75 之间，其 gps 纬度在 22.5，gps 经度在 113.9 左右，误差波动范围较小。由 SPSS 和 MATLAB 求解过程知，定价任务失败的原因在于没有考虑 GPS 位置（gps经度，gps 纬度）因素对定价任务的影响。

通过对模型一，模型二的分析，

v=24.100v12-v22+9.8206v1+2.0000v2-v1v2+2529.3522 (3)

我们可以得到定价规律，将 gps 纬度及经度带入到上式得出定价。

利用 Kendall 及 Spearman 相关性检验对 其 进行拟合，并对拟合参数进行微调观察相关系数变化程度以判断其灵敏性

|  |  |  |  |
| --- | --- | --- | --- |
|  | 原参数 | 变动百位数 | 变动个位数 |
|  |  |  |  |
| Kendall 检验-R2 | 0.7549 | 0.7547 | 0.7514 |
| Spearman 检验-R2 | 0.8981 | 0.8980 | 0.8963 |
| Pearson 检验- R2 | 0.8400 | 0.8104 | 0.8204 |

可看到，对模型的 Pearson 相关系数检验达到 0.8400，Kendall 的相关系数高达 0.7549，Spearman 的相关系数高达 0.8981，可以看出拟合效果较好,且对百位数和个位数进行微调模型的相关性变化极小，即模型对参数的灵敏度较低，鲁棒性高。

根据以上判断相比原方案将会大大的提高任务执行情况，充分的考虑了 gps 纬度，gps 经度的影响因素。可以和很好的满足用户需求，从而提高执行情况。

（二）模型Ⅲ（逐步回归模型）

1.建模背景

多元线性回归章节**[4]**指出，当自变量的个数很多时，不是所有的自变量都对因变量由显著影响，我们需要选择一些对因变量有显著影响的自变量应用多元回归的方法建立“最优”回归方程。每引入一个解释变量后都要进行 t 检验，当原来引入的解释变量由于后面解释变量的引入变得不再显著时，将其删除。经过逐步回归，使得最后保留在模型中的解释变量既是重要的，又没有严重多重共线性。

逐步回归分析相比较线性回归分析，不仅有作为线性回归分析本身建立的回归方程，描述自变量与因变量的数量关系的作用，还有它本身独具的优势：

a.剩余标准差较小，方程的稳定性较好。

B.每一步都做检验，保证了方程中所有自变量都是有显著性的。

2.模型Ⅲ的建立与求解

（1）因为模型 I 中三个自变量与任务定价的拟合度不高，所以我们首先引入对

被解释变量卡顿时长占比贡献最大的解释变量 VMOS。根据华为技术有限公司，

视频体验优化特性参数描述[EB/OL]和彭放，杨瑞琰，肖海军，何永明.的《数学

建模方法》可以确定 VMOS 与任务定价，GPS，任务预定限额三个量的关系式为：

VMOS=squality/5(p1+p2)\*(p1\*sloading+p2\*sstalling)

（2）引入新的自变量 VMOS，假设 z=a3v1+a4v2+a5v3+x4,利用 SPSS 软件对

其进行求解。画出 VMOS 与各个自变量之间的散点图如下：

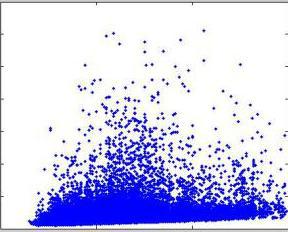


图 1 VMOS 与 V1 的散点图

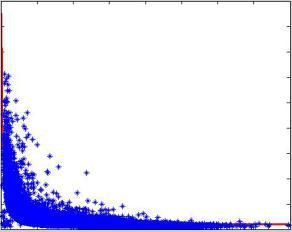


图 1 VMOS 与 V2/V3 的散点图由上面两个散点图可知，可假设逐步回归方程为：

z=a3v1+a4v2+a5v3+x4（4）

借助 SPSS 软件对 a3,a4,a5,x4 的值进行估计，我们得到下表：

表 5-3-1



表 5-3-2



对比表 5-3-1 和表 5-3-2 发现引入解释变量 VMOS 前后对于拟合结果的影响,引入后拟合度增加了 0.337。

表 5-3-3



各个参数结果如下：

e=1.410

a4=23.292

a5=-1.896

a6=1.100

c=-0.260

R2=0.695

所以卡顿占比与初始缓冲峰值速率，播放阶段平均下载速率，E2E RTT 以及 VMOS 之间的函数关系式为：

y=1.100V+23.292V1-1.896V2-0.260X4+1.410（5）

R2= 0.695，拟合效果尚可，但样本容性不高，缺乏对非线性情况的讨论，于是接下来建立模型Ⅳ。

（二）模型Ⅳ（模糊神经网络模型）

1.建模背景

参照模糊神经网络的应用研究**[6]，**模糊神经网络结合了神经网络系统和模

糊系统的长处，它在处理非线性，模糊性等问题上有很大的优势，在只能信息处

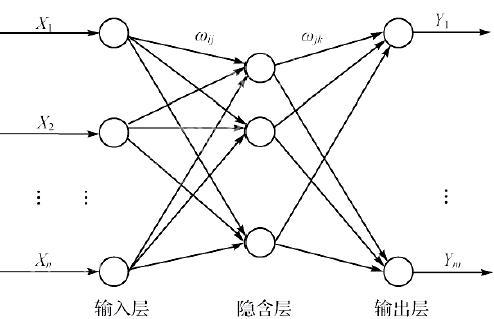
理方面有很大优势。

模糊集合的定义：

设 A 是论域 X 到区间[0,1]的一个映射，即 A:X→[0,1],x→A(x)

称 A 是 X 上的 Fuzzy 集，A(x)称为 Fuzzy 集的 A 隶属函数。

下图为模糊神经网络一般结构：

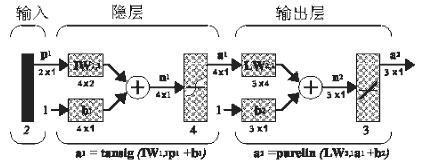


我们规定神经网络层数为 2 层，运用 MATLAB 神经网络工具箱求得：

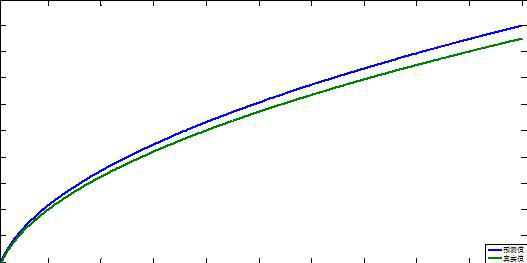
当 h-[h]≤0.2 时，f(h)=[h];

当 h-[h]≥0.2 时，f(h)=[h]+2;

一般 2 层神经网络的结构示意图为：



画出任务预定限额预测值与实际值折线对比：



观察可知，拟合度较高，该模型效果较好。

（三）问题 3 的两种数学模型的比较

模型Ⅲ的 R2 值为 0.695，拟合效果尚可。通过图形观察模型Ⅳ的拟合度大于0.7，效果良好。于是对于问题 2 决定采用模型Ⅳ解决。于是定价我们采用

y=1.100V+23.292V1-1.896V2-0.260X4+1.410（5）

进行定价。考虑了在实际情况下，多个任务可能因为位置比较集中，导致用户会争相选择的可能性。我们考虑是将这些任务联合在一起打包发布。从而进一步完善了ＡＰＰ拍照赚钱的方案，有利于提高收入，对问题二进一步完善从而更有可信度以及说服力。

**（四）定价问题及求解**

(1)数据预处理

7 / 26

数据归一化处理把所有数据都转化为 0,1 之间的数，其目的是取消各维数据

间数量级差别，避免因为输入输出数据数量级差别较大而造成预测误差较

大，方便后面数据的处理，并保证程序运行时收敛加快。我们将数据集按如下公

式进行归一化处理：

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *y*  |  |  | *y* | max |  |  | *y* | min | | *x* |  |  |  | *x* | min |  |  *y* |  |  |
|  |  |  |  |  |  |  | min |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | *x* | | | | | *x* | | | | | | | | |  |  |
|  |  |  |  |  |  | max | | | |  | min | | | | | | | |  |  |
| 其中 *y* min |  1, *y*max 1， *x* 为需要归一化的数据集。 | | | | | | | | | | | | | | | | | | |  |

(2)BP 神经网络初始化

在对数据进行归一化处理后，需要根据系统输入输出序列  *X* ,*Y*  确定网络输

入层节点数 *n* 、隐含层节点数 *l i* ，输出层节点数 *m* ，迭代次数 *e* ，学习速率 *s* ，目标误差 *g* 。考虑到本问题为2输入、1输出的非线性函数关系，因此建立2 *Ni* 1结构的神经网络，其中 *Ni* 为第 *i* 层隐含层节点数。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 隐藏层传输函数采用 tansig 函数： |  |  |  |  |
| *a*  | 2 | | 1 |  |
|  | 2*n* |  |
|  | 1 *e* | |  |  |

输出层传输函数采用 purelin 函数：

*a*  *n*

隐藏层神经元个数对 BP 神经预测精度有显著的影响，节点数太少，网

络不能很好地学习，需要增加训练次数，训练的精度也受影响；节点数太多，

训练时间增加，网络容易过拟合。我们参考如下公式来确定最适隐藏层神经元

个数

[8]。

*l*  *n* 1

*l*   *m*  *n*  *a*

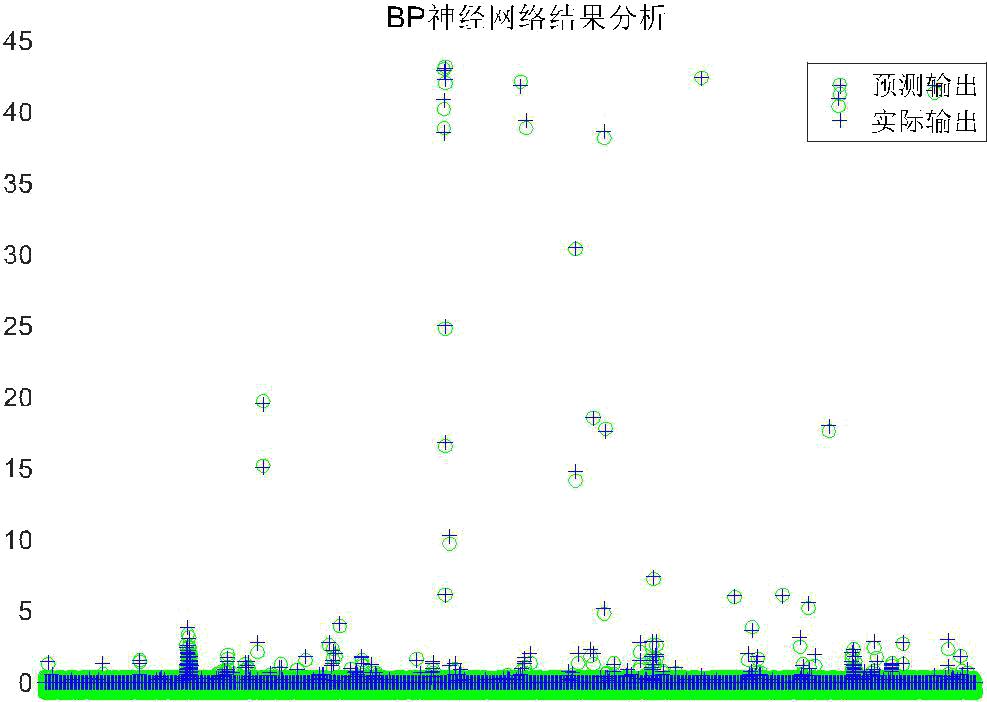


*l* log2 *n*

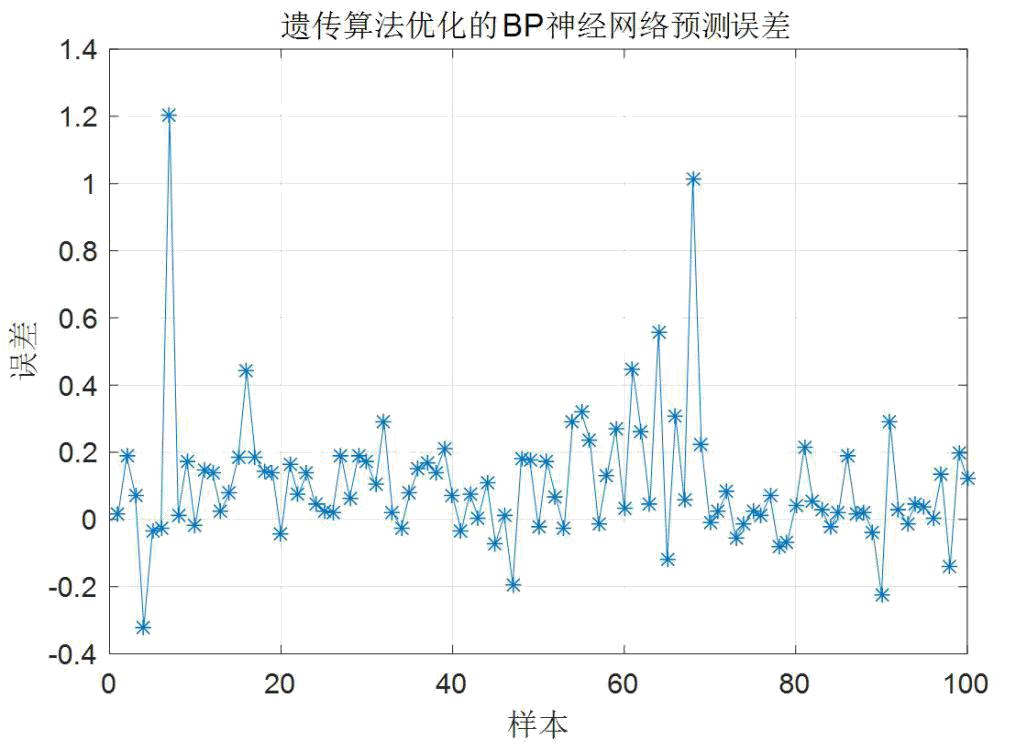
在实际问题中，隐含层节点数的选择首先是参考公式来确定节点数的大概范

围，然后用试凑法确定最佳的节点数。经过多次试验，我们选择 *N*1=8，*N*2=10，

*N*3=8此时BP神经网络达到了较高的精度。



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 每次迭代后种群平均适应度 | | |  |  |
| 迭代次数 | 1 | 2 | 3 | 4 | 5 |
|  |  |  |  |  |  |
| 平均适应度 | 24.9184 | 15.0291 | 12.9231 | 7.6431 | 7.0072 |
|  |  |  |  |  |  |
| 迭代次数 | 6 | 7 | 8 | 9 | 10 |
|  |  |  |  |  |  |
| 平均适应度 | 2.2965 | 1.5965 | 1.4870 | 1.4659 | 1.3948 |
|  |  |  |  |  |  |
| 迭代次数 | 11 | 12 | 13 | 14 | 15 |
|  |  |  |  |  |  |
| 平均适应度 | 1.2072 | 1.2041 | 1.1965 | 1.1870 | 1.1659 |
|  |  |  |  |  |  |
| 迭代次数 | 16 | 17 | 18 | 19 | 20 |
|  |  |  |  |  |  |
| 平均适应度 | 1.0048 | 1.0005 | 0.9834 | 0.8125 | 0.8583 |
|  |  |  |  |  |  |
| 迭代次数 | 21 | 22 | 23 | 24 | 25 |
|  |  |  |  |  |  |
| 平均适应度 | 0.8572 | 0.8365 | 0.7865 | 0.6870 | 0.4659 |
|  |  |  |  |  |  |
| 迭代次数 | 26 | 27 | 28 | 29 | 30 |
|  |  |  |  |  |  |
| 平均适应度 | 0.4248 | 0.4125 | 0.4079 | 0.3963 | 0.3960 |
|  |  |  |  |  |  |



从以上结果可以看出，训练得到的 BP 神经网络能较准确地反映ＧＰＳ

与任务定价的函数关系；于是在ＭＡＴＬＡＢ中即可模拟出定价与ｇｐｓ经纬度三者的一个对比。

根据上述模拟可得 v 与 v1 gps 纬度 以及 v2 gps 经度之间的关系:

V＝23.1982V1-4.6725V2+68.8410（６）

据（6）式可以得出附件三中 gps 经纬度位置与任务定价之间的关系式。据此可以给出附件三中的任务定价。该方案未能考虑在实际情况下，多个任务可能因为位置比较集中，导致用户会争相选择的可能性。只是模拟了 gps 经纬度和任务定价的函数关系式，考虑不够全面。定价方面存在一定的误差。但在拟

合数据的变化趋势以及散点图的拟合程度，使问题结果可视化，更加清晰、条理和直观。模型整体简单易懂但缺乏复杂性，对实际情况考虑欠佳。

§6 模型评价及建议

模型优点：

1.抽取样本对模型理论结果与实际值进行检验，增强了结果的可信度。

2.考虑了非线性回归的可能性，采用了模糊神经网络模型，确保了模型的可操作性，使算法更接近于实际应用。

3.每个问题从线性何非线性这两方面分别对给定的样本进行对比建模，避免了只采用一种模型所带来的片面性，结果说服力强。

4.在数据处理方面，先判断自变量间的交互影响性，减少变量所含信息重复导致的误差；对变量做出正态性检验、相关性检验，在建模之前确立了变量

间的 Spearman,Kendall 秩相关系数检验方法，增加了建模过程中相关系数的可信度；

5.在图像处理与显示方面，我们使用了 Matlab 与 SPSS 双重作图，拟合数据的变化趋势以及散点图的拟合程度，使问题结果可视化，更加清晰、条理和直观；

6.在算法应用上，运用了计算机动态仿真等前沿算法，对模型拟合程度很高。

模型缺点：

1.处理数据时采用分层抽样的方式，没有对全部数据进行处理，对实验结果由一定影响。

2.多元线性回归缺乏对非线性情况的讨论，由部分限制性。

3.没有考虑到预定任务开始时间，信誉度以及 GPS 之间是否相关关系。

4.随时间序列变化的函数没有进行降噪处理，离散数据不够平滑

5.对于异常的数据无法做到很好地筛去异常点

6.对于问题二的拟合程度还不够精确

模型改进和推广：

模糊神经网络汇集了神经网络和模糊理论的优点，集识别，信息处理为一身，具有自学习和联想功能，特别是对于问题的求解具有黑箱的特性。它不需要知道系统的精确数学模型，对处理不确定的非线性的系统是一种非常好的方法。

**7.参考文献**

[1]彭放，杨瑞琰，肖海军，何永明.《数学建模方法》.科学出版社，2012

年 2 月.

[2]堵秀凤，张剑，张宏民.《数学建模》.北京航空航天大学出版社，2011

年 8 月.

[3]华为技术有限公司，华为 U-vMOS 视频体验标准白皮书 V1.0,2015-09-15

[4]刘京娟，多元线性回归模型检验方法，湖南税务高等专科学校学报，第 5

期第 18 卷，2005

[5]华为技术有限公司，视频体验优化特性参数描述[EB/OL]，2016-06-30

[6]张凯兵,神经网络与数学建模模型及算法,湖北工业大学[7]卓金武,MATLAB 在数学建模中的应用，北京航空航天大学出版社[8]高新波，模糊聚类分析及其应用，西安电子科技大学出版社，2004-01-01

附录:

（表一）:求解数据

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 任务号码 | 任务 gps 纬度 | 任务 gps 经度 | 任务标价 | 任务执行情况 |
| A0775 | 23.13871675 | 113.111198 | 85 | 1 |
| A0776 | 22.98807885 | 114.0208579 | 66 | 1 |
| A0777 | 23.04326407 | 112.9260813 | 85 | 1 |
| A0778 | 23.061023 | 112.9221482 | 80 | 1 |
| A0779 | 22.94081175 | 113.9453226 | 72 | 1 |
| A0780 | 22.99587084 | 113.7621434 | 72 | 1 |
| A0781 | 22.91334142 | 113.8483693 | 72 | 1 |
| A0782 | 22.97333327 | 113.730314 | 67.5 | 1 |
| A0783 | 23.11276083 | 112.9299114 | 70 | 1 |
| A0784 | 23.12358718 | 113.1648355 | 68 | 1 |
| A0785 | 23.05913863 | 113.7431308 | 66 | 1 |
| A0786 | 23.06812838 | 113.0255363 | 80 | 1 |
| A0787 | 23.09724421 | 113.7523577 | 71 | 1 |
| A0788 | 23.19885366 | 113.1851605 | 85 | 1 |
| A0789 | 23.09348295 | 113.1826215 | 85 | 1 |
| A0790 | 23.23190692 | 113.0890874 | 85 | 1 |
| A0792 | 23.06233415 | 113.1056841 | 66.5 | 1 |
| A0793 | 23.10971032 | 113.1808405 | 85 | 1 |
| A0794 | 23.22923111 | 113.0827188 | 85 | 1 |
| A0795 | 23.11501909 | 112.9305637 | 70 | 1 |
| A0796 | 23.11526775 | 113.1169399 | 85 | 1 |
| A0797 | 22.80256864 | 113.2873979 | 85 | 1 |
| A0798 | 23.06007494 | 112.9216899 | 85 | 1 |
| A0799 | 23.15029122 | 113.1092646 | 67.5 | 1 |
| A0800 | 23.11065649 | 113.1309687 | 85 | 1 |
| A0801 | 23.0650595 | 113.0572779 | 72 | 1 |
| A0802 | 22.93595157 | 113.8990852 | 80 | 1 |
| A0716 | 22.74394534 | 114.1439999 | 75 | 1 |
| A0717 | 22.8387028 | 113.7066879 | 72 | 1 |
| A0718 | 22.70228568 | 114.1736681 | 75 | 1 |
| A0719 | 22.8291523 | 113.7347647 | 72 | 1 |
| A0720 | 23.02506148 | 114.1035973 | 75 | 1 |
| A0721 | 22.74829892 | 114.1391543 | 75 | 1 |
| A0722 | 22.87660886 | 113.675044 | 75 | 1 |
| A0723 | 22.73591184 | 114.2757904 | 65.5 | 1 |
| A0724 | 23.08248182 | 113.8756202 | 68.5 | 1 |
| A0623 | 22.97892268 | 113.956267 | 70 | 1 |
| A0624 | 22.83110817 | 113.7801718 | 69 | 1 |
| A0625 | 22.72831758 | 114.1967043 | 72 | 1 |
| A0626 | 22.78108249 | 114.0574014 | 70 | 1 |

A0627

A0628

A0629

A0630

A0631

A0632

A0633

A0634

A0635

A0636

A0637

A0638

A0639

A0640

A0641

A0642

A0643

A0644

A0645

A0646

A0647

A0648

A0649

A0650

A0651

A0652

A0653

A0654

A0655

A0656

|  |  |  |  |
| --- | --- | --- | --- |
| 22.79633118 | 114.0632326 | 70 | 1 |
| 22.82104527 | 113.8090357 | 65 | 1 |
| 23.04883737 | 113.8557656 | 70 | 1 |
| 22.81973576 | 113.8093302 | 65 | 1 |
| 22.92277014 | 113.6689502 | 66 | 1 |
| 22.8387028 | 113.7066879 | 68.5 | 1 |
| 23.03128773 | 113.8334312 | 67.5 | 1 |
| 22.99889137 | 113.6206429 | 66.5 | 1 |
| 22.7920915 | 114.1017881 | 70 | 1 |
| 22.96921119 | 113.7638371 | 71.5 | 1 |
| 22.76781337 | 114.0421691 | 70 | 1 |
| 23.05029788 | 113.7566159 | 65.5 | 1 |
| 22.8188764 | 113.6774155 | 65.5 | 1 |
| 22.88245435 | 114.0068175 | 75 | 1 |
| 22.7406896 | 114.1664449 | 70 | 1 |
| 22.96547337 | 114.1542792 | 75 | 1 |
| 22.69796842 | 114.1787142 | 68 | 1 |
| 22.96920507 | 114.1741551 | 75 | 1 |
| 23.03504106 | 113.8561107 | 69 | 1 |
| 23.02874532 | 113.8311519 | 67.5 | 1 |
| 23.10214598 | 113.8030771 | 69.5 | 1 |
| 23.0230261 | 113.8416605 | 68 | 1 |
| 22.84421485 | 114.1744379 | 70 | 1 |
| 23.06377156 | 113.834619 | 68.5 | 1 |
| 23.0284963 | 113.848423 | 68 | 1 |
| 22.84485611 | 114.1728089 | 70 | 1 |
| 23.02631551 | 113.8422567 | 67.5 | 1 |
| 22.85760935 | 113.0531252 | 70.5 | 1 |
| 22.98796007 | 113.0170859 | 69 | 1 |
| 22.84851008 | 113.1870299 | 70.5 | 1 |

（表二）：部分数据模拟验证

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  | 任务执行情 |
| 任务号码 任务 gps 纬度 任务 gps 经度 | | | 任务标价 | 况 |
| A0540 | 23.17934885 | 112.8994031 | 70 | 1 |
| A0541 | 23.17455266 | 113.0688513 | 75 | 1 |
| A0542 | 23.17677756 | 112.8784023 | 70 | 1 |
| A0543 | 23.16943397 | 112.9073857 | 70 | 1 |
| A0544 | 23.2615299 | 113.0247518 | 75 | 1 |
| A0545 | 23.16865186 | 112.8938126 | 70 | 1 |
| A0546 | 23.16321812 | 112.9148997 | 70 | 1 |
| A0547 | 23.07057951 | 112.8445408 | 75 | 1 |

A0548

A0549

A0550

A0551

A0552

A0553

A0554

A0555

A0556

A0557

A0558

A0559

A0560

A0561

A0562

A0563

A0564

A0565

A0566

A0567

A0568

A0569

A0570

A0571

A0572

A0573

A0574

A0576

A0577

A0578

A0579

A0580

|  |  |  |  |
| --- | --- | --- | --- |
| 22.90422292 | 112.9019826 | 75 | 1 |
| 23.0386685 | 113.2034172 | 70 | 1 |
| 23.20300498 | 113.1713931 | 69.5 | 1 |
| 23.17018893 | 112.9060636 | 70 | 1 |
| 22.87794292 | 113.2265578 | 70.5 | 1 |
| 23.18828575 | 113.2051584 | 68 | 1 |
| 23.1176526 | 113.1774208 | 66.5 | 1 |
| 23.14659839 | 113.1104485 | 67.5 | 1 |
| 23.12232763 | 113.1914852 | 67 | 1 |
| 23.16730584 | 112.902416 | 70 | 1 |
| 23.10435408 | 113.1991211 | 66.5 | 1 |
| 23.18585718 | 113.0927552 | 75 | 1 |
| 23.06677042 | 113.1421953 | 66.5 | 1 |
| 23.04403683 | 112.9214509 | 71.5 | 1 |
| 22.83661574 | 113.0000318 | 73 | 1 |
| 23.00707365 | 113.008964 | 73.5 | 1 |
| 22.95343786 | 112.9712874 | 75 | 1 |
| 23.07826539 | 113.0413813 | 70 | 1 |
| 23.11360001 | 113.0152889 | 73.5 | 1 |
| 23.0386685 | 113.2034172 | 70 | 1 |
| 23.0695963 | 113.156189 | 66 | 1 |
| 22.90311313 | 112.8923366 | 73 | 1 |
| 23.06369651 | 113.1355791 | 66.5 | 1 |
| 22.91055956 | 112.8841425 | 75 | 1 |
| 23.03863041 | 113.1704547 | 66.5 | 1 |
| 23.06573343 | 113.1321153 | 66.5 | 1 |
| 23.15666359 | 112.9049408 | 70 | 1 |
| 23.0723988 | 113.0896832 | 67.5 | 1 |
| 23.03791217 | 113.1563516 | 65.5 | 1 |
| 22.89624538 | 112.8883289 | 73.5 | 1 |
| 23.33555566 | 113.3075287 | 67.5 | 1 |
| 23.0723988 | 113.0896832 | 67.5 | 1 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 会员编号 | 会员位置(GPS) | | 预订任务限额 | 预订任务开始时间 | 信誉值 |
| B0158 | 23.022885 | 113.826933 | 26 | 6:57:00 | 183.2365 |
| B0086 | 22.979376 | 113.359597 | 15 | 6:57:00 | 479.3786 |
| B0087 | 22.617332 | 114.046699 | 12 | 7:18:00 | 466.0196 |
| B0088 | 23.110206 | 113.840149 | 36 | 6:45:00 | 464.8392 |
| B0089 | 23.467117 | 113.36579 | 19 | 6:33:00 | 463.2336 |
| B0090 | 22.661854 | 114.027580 | 12 | 7:00:00 | 462.5524 |
| B0091 | 22.805765 | 113.580688 | 58 | 6:30:00 | 456.8141 |
| B0092 | 22.573973 | 114.118394 | 14 | 7:06:00 | 443.4528 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| B0093 | 23.048773 | 113.320775 | 16 | 7:27:00 | 438.3089 |
| B0094 | 22.654465 | 114.042558 | 12 | 7:12:00 | 432.0017 |
| B0095 | 23.08834 113.811783 | | 42 | 6:48:00 | 430.3396 |
| B0096 | 23.042076 | 113.088167 | 36 | 6:39:00 | 425.5539 |
| B0097 | 22.715649 | 114.266454 | 34 | 6:36:00 | 419.474 |
| B0098 | 23.102272 | 113.314206 | 15 | 7:51:00 | 409.9957 |
| B0099 | 22.957762 | 113.885591 | 40 | 6:30:00 | 404.3689 |
| B0100 | 22.735681 | 114.276218 | 34 | 6:39:00 | 401.2549 |
| B0101 | 22.566656 | 114.048785 | 11 | 7:27:00 | 389.3617 |
| B0102 | 22.605696 | 114.052535 | 11 | 7:30:00 | 388.9843 |
| B0103 | 22.528087 | 114.053445 | 12 | 7:06:00 | 385.9114 |
| B0110 | 23.060244 | 113.228424 | 15 | 7:18:00 | 318.7701 |
| B0111 | 23.025509 | 113.136987 | 29 | 6:45:00 | 318.7701 |
| B0112 | 23.131169 | 113.297453 | 16 | 7:39:00 | 317.5724 |
| B0113 | 22.55917 113.964202 | | 16 | 7:21:00 | 317.3112 |
| B0114 | 22.571019 | 114.063028 | 10 | 7:42:00 | 316.9499 |
| B0129 | 22.669148 | 113.923368 | 11 | 7:00:00 | 240.7228 |
| B0130 | 22.697751 | 114.227313 | 22 | 6:45:00 | 239.112 |
| B0131 | 22.634345 | 114.079336 | 10 | 7:39:00 | 239.0776 |
| B0132 | 23.017262 | 113.334856 | 15 | 7:18:00 | 239.0776 |
| B0133 | 23.035953 | 113.163214 | 19 | 6:54:00 | 236.2993 |
| B0134 | 22.615165 | 114.168801 | 12 | 7:09:00 | 234.2849 |
| B0135 | 22.999192 | 114.343194 | 42 | 6:30:00 | 233.9414 |
| B0136 | 24.80413 113.605786 | | 23 | 6:36:00 | 229.1596 |
| B0137 | 23.107527 | 113.14625 | 22 | 6:51:00 | 225.3733 |
| B0138 | 22.670303 | 114.139487 | 11 | 7:24:00 | 225.2535 |

附录二：程序部分

幂函数

f=@(p,x)((1./x(2,:).\*(1483-p(1)/(p(2)+1).\*)

* + logic 模型的拟合 [beta,r,j]=nlinfit([x;y],z,f,[1 1 1 0 1]); [zpre,delta]=nlpredci(f,[x;y],beta,r,'jacobian',j); [b,bint,r,rint,stats]=regress(z2',z');
* VMOS 函数的 nlinfit 拟合

f=@(p,x)(4.33.\*(1/5\*(p(1)+p(2)).\*(p(1)\*x(1,:)+p(2)\*x(2,:)))) [beta,r,j]=nlinfit([SLoading';Sstalling'],VMOS',f,[0.5 0.5]); [VMO，Ｖ]=nlpredci(f,[SLoading';Sstalling'],beta,Ｖ１,Ｖ２,ｙ); [Ｖ,Ｖ１,Ｖ２,ｙ,ＶＭＯＳ]=regress(VMOS，Ｖ);

相关系数分析：

[corr(Ｖ１,V２,'ｖ３,'Kendall'),corr(ｖ１,ｖ２,ｖ３,'Spearman'); corr(ａ１,ａ２,ａ６,'Kendall'),corr(ａ１,ａ２,ａ６,'Spearman'); corr(ａ１,ａ３,ａ６,'Kendall'),corr(ａ１,ａ３,ａ６,'Spearman')];

[corr(Ｖ,V２,'ｖ３,'Kendall'),corr(ｖ,ｖ２,ｖ３,'Spearman'); corr(Ｖ,ａ２,ａ６,'Kendall'),corr(Ｖ,ａ２,ａ６,'Spearman'); corr(Ｖ,ａ３,ａ６,'Kendall'),corr(Ｖ,ａ３,ａ６,'Spearman')];

[corr(Ｖ,V１,'ｖ３,'Kendall'),corr(ｖ,ｖ１,ｖ３,'Spearman'); corr(Ｖ,ａ２,ａ１,'Kendall'),corr(Ｖ,ａ２,ａ１,'Spearman'); corr(Ｖ,ａ４,ａ５,'Kendall'),corr(Ｖ,ａ４,ａ５,'Spearman')];

核心代码及部分运算结果

General model:

cf\_(bofangsulv) = a./(bofangsulv+b)+c

Coefficients (with 95% confidence

bounds):

|  |  |  |
| --- | --- | --- |
| a = | ２３．２９ |  |
| ２ |  |
|  | ２４．１０ |  |
| b = | ０ |  |
| c = | １．８９６ |  |

sse: 1.405586496112479e+004 rsquare: 0.697034682629779

dfe: 89263

adjrsquare: 0.697027894479765 rmse: 0.396819539401220

附录：1. st\_ = [1000 -100 -100]; ft\_ = fittype('a./(chushisulv+b)+c' ,...

'dependent',{'chushishiyan'},'independent',{'chushisulv'},...

'coefficients',{'a', 'b', 'c'});

[cf\_,good]= fit(chushisulv,chushishiyan,ft\_ ,'Startpoint',st\_); h\_ = plot(cf\_,'fit',0.95);

legend off; % turn off legend from plot method call set(h\_(1),'Color',[1 0 0],...

'LineStyle','-', 'LineWidth',2,...

'Marker','none', 'MarkerSize',6); hold

on,plot(chushisulv,chushishiyan,'\*') st\_ = [1000 100 10];

ft\_ = fittype('a./(bofangsulv+b)+c' ,...

'dependent',{'kadunzhanbi'},'independent',{'bofangsulv'},...

'coefficients',{'a', 'b', 'c'});

[cf\_,good]= fit(bofangsulv,kadunzhanbi,ft\_ ,'Startpoint',st\_); h\_ = plot(cf\_,'fit',0.95);

legend off; % turn off legend from plot method call set(h\_(1),'Color',[1 0 0],...

'LineStyle','-', 'LineWidth',2,...

'Marker','none', 'MarkerSize',6); hold

on,plot(bofangsulv,kadunzhanbi,'.')

核心代码及部分运算结果

分段 1 General model:

cf\_(fenduankadunbi1) = a.\*fenduankadunbi1 +b Coefficients (with 95% confidence bounds):

|  |  |
| --- | --- |
| a = | -１．００００ |
| b = | １．４１００ |

sse: 0.003322134577609 rsquare: 0.999909795878552

dfe: 442

adjrsquare: 0.999909591796829 rmse: 0.002741558239086 分段 2

General model:

cf\_(fenduankadunbi2) = a.\*(fenduankadunbi2)+c Coefficients (with 95% confidence bounds):

|  |  |
| --- | --- |
| a = | ０．７４ |
|  | ０．２６ |
| c = | ０ |

sse: 1.836073756901905e-028 rsquare: 1

dfe: 1640 adjrsquare: 1

rmse: 3.345978435140646e-016

**BP 神经网络 MATLAB 程序源代码**

xlsfile='89266.xlsx';

[input1]=xlsread(xlsfile,1,'B1:C70000');

[output1]=xlsread(xlsfile,1,'E1:E70000');

[input2]=xlsread(xlsfile,1,'B70001:C89266');

[output2]=xlsread(xlsfile,1,'E70001:E89266');

[inputn,inputps]=mapminmax(input1');

[outputn,outputps]=mapminmax(output1');

inputn\_test=mapminmax('apply',input2',inputps);

net=newff(inputn,outputn,[8 10 8],{},'trainlm');

net.trainParam.epochs=100; net.trainParam.lr=0.1;

net.trainParam.goal=0.00001 net=train(net,inputn,outputn);

an=sim(net,inputn\_test);

BPoutput=mapminmax('reverse',an,outputps); figure(1)

plot(BPoutput,':og') hold on

plot(output2,'- \*');

legend('预测输出','实际输出');

title('BP 神经网络预测输出')

xlabel('样本')

ylabel('函数输出')

grid on;

error=BPoutput-output2';

figure(2)

plot(error,'-\*')

title('BP 神经网络预测误差')

xlabel('样本')

ylabel('误差')

grid on;