

“拍照赚钱”的任务定价

2019年8月27日

摘要

“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。定价是一个至关重要的环节。本文要解决的问题是研究任务的定价规律，设计新的定价方案，评价新的方案的可行性。本文采用回归分析求解，先对数据进行了可视化，之后建立了基于整体数据的回归模型，找到了定价的影响因素，并基于可视化的二维平面图找到了未完成的原因。并用已完成的数据作为训练集，训练了一个更加精确的回归模型，并用未完成的数据进行测试，说明了其合理性。之后，采取打包的方法，将邻近的任务打包成一个整体，重新建立新的回归模型，得到了更有效的模型。然后，用新的模型给未定价的数据进行定价。最后，实现了对于这个较为简易模型的优缺点进行了分析。

关键词: 定价 回归分析 拍照赚钱

1 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。每个会员的信誉值不同，预定任务开始时间不同，预定任务限额也不同。定价是一个至关重要的环节，如果定价过低，有些任务就会无人问津，导致任务无法完成，造成不必要的损失。

附件一给出任务的位置（经纬度）以及完成情况，“1”代表完成，“0”代表未完成；附件二给出了此APP的会员信息，包括：会员所处位置、任务限额、任务开始时间以及信誉度。附件三给出的是新的任务位置，而没有完成情况。

要解决的问题如下：

问题一：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。

问题二：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。

问题三：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。有一种方案是将这些任务联合在一起打包发布，形成“集群”效应。在这种方案下，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？

问题四：对附件三中的新项目给出新的任务定价方案，并评价该方案的可行性。

2 模型假设

1. 假设任务和会员的地理位置不随着时间发生变化。
2. 假设会员的数量、预定任务限额、预定任务开始时间、信誉度不发生变化。

3. 假设所有任务的性质相同，即任务的执行情况只受其地理位置和周边会员情况的影响，不受其难易程度的影响。

3 符号说明

- x_1 临近区域内的任务的数量
- x_2 临近区域内的会员预定任务限额之和
- x_3 临近区域内会员信誉之和
- y 任务定价
- M 会员总数
- N 任务总数

4 模型的建立与求解

4.1 问题一的模型建立与求解

通过对数据的分析，与地图对比经纬度，可以发现该区域是我国广东省广州、深圳、东莞、佛山四城市。

4.1.1 任务定价的规律

对每一个任务，以其经纬度为基准，在其经度0.05内、纬度0.05内的区域称为其临近区域。计算每个任务的临近区域内的任务的数量 x_1 、临近区域内的会员预定任务限额之和 x_2 和临近区域内会员信誉之和 x_3

任务定价为 y ，利用matlab进行回归分析，表达式为：

$$y = 72.4265 - 0.1067x_1 - 0.0034x_2$$

在题中给的区域内， y 是 x_1 和 x_2 的减函数，这就说明任务的密度越大，会员的预定任务限额密度越大，任务定价越低。 y 与 x_3 无关，因此在此不予考虑。

对于 y 和 x_1 的关系，可以这样来理解：任务较密集时，会员可供选择的任务较多，因此价格较低；任务较稀疏时，会员可供选择的任务较少，因此价格较高。

对于 y 和 x_2 的关系，可以这样来理解：人力较密集时，会员竞争做任务比较激烈，因此价格较低；人力较稀疏时，为吸引周边不多的人力，因此价格较高。

因此得到任务定价规律：任务的密度和会员的预定任务限额密度越大，任务的定价越低，即任务定价与任务分布和会员分布有关。

4.1.2 任务未完成的原因

分别以纬度、经度为横坐标，任务定价为纵坐标，建立二维图像模型，结果如下。

注：

- (1) 黄色点代表完成，蓝色点代表未完成，红色点代表会员的预定任务限额。

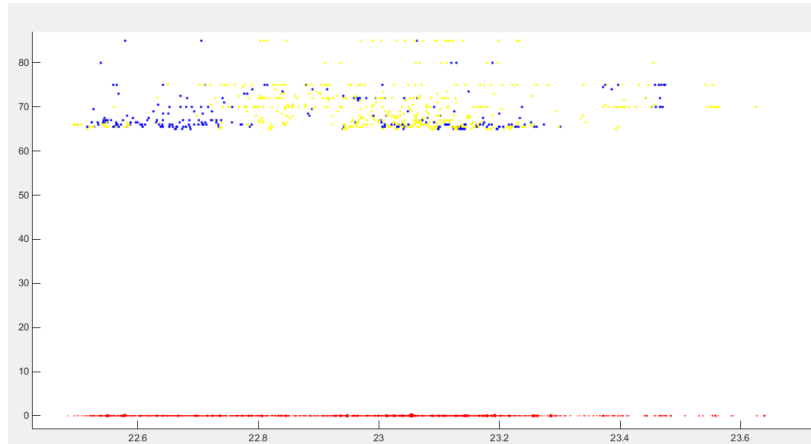


Figure 1: 纬度分析图

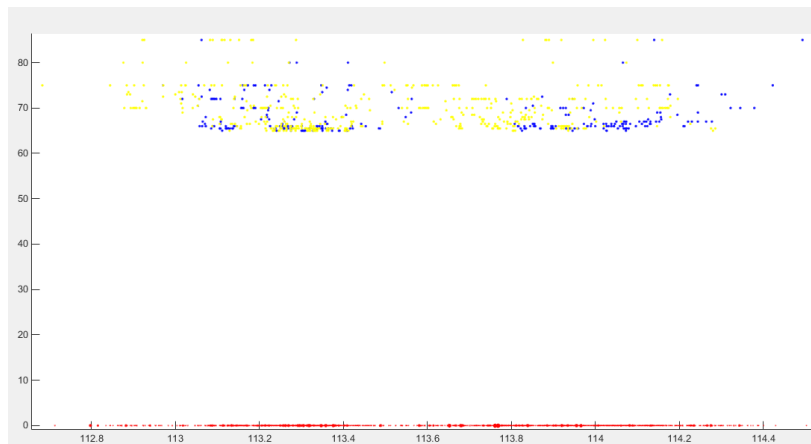


Figure 2: 经度分析图

(2) 红色点越深，代表该处人力越密集，反之代表人力越稀疏。

(3) 纵坐标为任务定价，其中会员的预定任务限额的任务定价为0。

同时以经度为横坐标，纬度为纵坐标，建立二维图像模型。

从这三张图中，可以总结出任务未完成的原因：

(1) 从“经纬度”图中可以看出，未完成任务在空间上具有聚集现象，其中深圳、佛山地区未完成任务较多，东莞地区未完成任务较少，得到第一个结论是在某些城市定价不合理，对会员的吸引力不够，从而造成任务未完成。

(2) 从“经度分析”和“纬度分析”两张图中可以看出，在会员人数多、预定任务限额高的地方，完成任务较多，未完成任务较少，反之，在会员人数少、预定任务限额低的地方，未完成任务较多，得到第二个结论是任务未完成与会员预定任务限额较低有关。

(3) 从“经度分析”和“纬度分析”两张图中可以看出，在一些预定任务限额尚可的地方，由于任务定价较低，导致对会员的吸引力不够，得到第三个结论是价格定价过低导致对会员的吸引力不够，从而造成任务未完成。

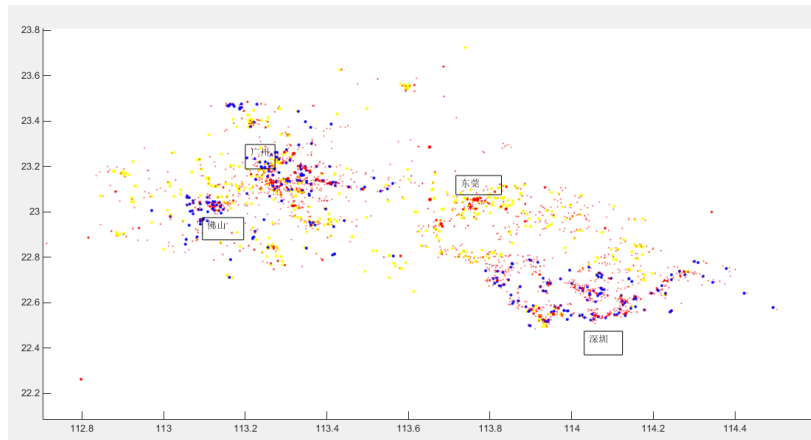


Figure 3: 经度纬度图

4.2 问题二的模型建立与求解

4.2.1 新的任务定价方案

使用完成的任务为训练集，建立新的回归分析模型。

对每一个任务，以其经纬度为基准，在其经度0.05内、纬度0.05内的区域称为其临近区域。计算每个任务的近区域内的任务的数量 x_1 、临近区域内的会员预定任务限额之和 x_2 和临近区域内会员信誉之和 x_3

任务定价为 y ，利用matlab进行回归分析，表达式为

$$y = 73.3072 - 0.1429x_1 - 0.0048x_2$$

，得到其结果，如图所示。（图中并不完全）

任务号码	任务gps 纬度	任务gps经度	任务标价	任务执行情况	新的任务标价
A0001	22.56614225	113.9808368	66	0	68.7685
A0002	22.68620526	113.9405252	65.5	0	71.4008
A0004	22.56484081	114.2445711	75	0	73.0118
A0005	22.55888775	113.9507227	65.5	0	67.5877
A0006	22.55899906	114.2413174	75	0	73.0118
A0008	22.56277351	113.9565735	65.5	0	67.9178
A0009	22.50001192	113.8956606	66	0	71.749
A0011	22.52486369	113.9308596	65.5	0	69.3844
A0012	22.519087	113.9358436	65.5	0	69.4959
A0018	22.4981901	113.8984817	66	0	71.773
A0023	22.51578568	113.9420557	65.5	0	69.4826
A0024	22.54220963	114.0196493	66.5	0	70.5866
A0025	22.54458961	113.9986601	67.5	0	69.1562
A0029	22.56470734	113.9820093	66	0	68.6448
A0030	22.58193448	114.1563953	68	0	69.2917

Figure 4: 对未完成的新的定价

4.2.2 与原方案的比较

原方案和新方案各个任务价格的散点图如图所示。

可以看到，新方案散点图比原方案整体上有所提高。对未完成的任务使用该模型，可以得到在该模型基础上的定价接下来进行计算，计算得到新方案散点图定价总和为58251.69，原方案为57707.50，整体上有明显提高。

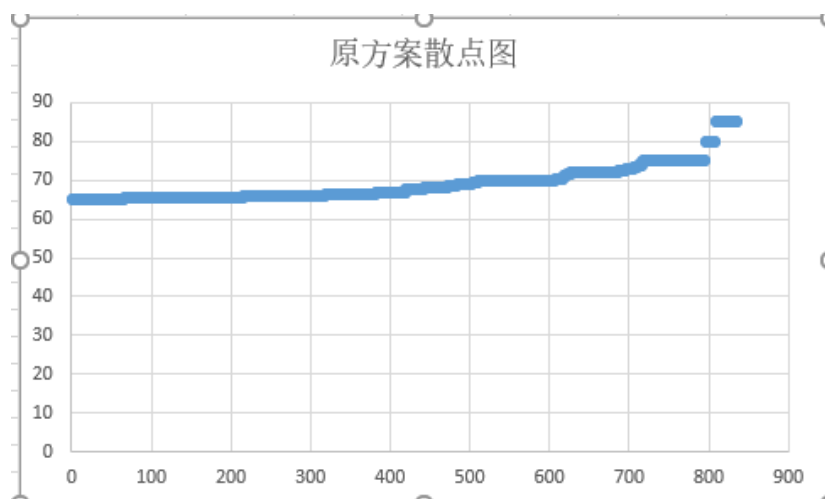


Figure 5: 原方案散点图

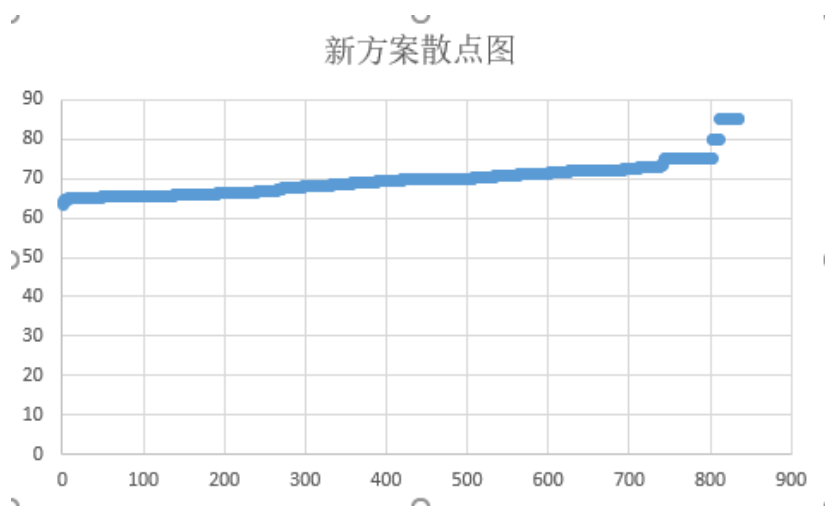


Figure 6: 新方案散点图

由上面的分析知，定价偏低是任务未完成的一个重要因素，而新的方案提高了定价，使得任务的完成率提高。

4.3 问题三的模型建立与求解

由于多个任务位置比较集中，可以考虑打包发布。打包发布的原则是：

- (1) 最多3个任务为一组进行打包。
- (2) 以经纬度为横纵坐标，两点之间的欧氏距离 $\sqrt{(a_1 - a_2)^2 + (b_1 - b_2)^2}$ 小于0.1。

仍然以附件一中完成的数据为训练集，建立回归模型，变量 y, x_1, x_2 的意义与上面相同。将距离较近的任务进行打包，得到回归方程为

$$y = 185.8624 + 1.8669x_1 - 0.0061$$

对所有任务使用该模型，对未完成的任务进行打包，在该模型基础上打的各个包的定价如下所示。（不完全）

202.3174	194.4039	196.8262	191.274	191.4448
191.744	189.0167	194.1904	219.8942	207.0513
187.5768	189.7978	190.8653	216.2272	187.6622
207.5643	192.7813	195.8011	221.8343	203.2504
207.729	191.5427	196.6615	222.6273	203.2504
205.8862	196.2589	196.6859	194.7394	205.1112
204.0559	191.4939	189.2485	207.2288	207.7107
202.25	193.9525	203.2144	210.4377	205.16
204.2511	195.124	206.7893	206.6798	214.6049
205.8191	190.0845	203.7085	211.414	199.6569
205.8438	191.0483	203.2388	206.1735	207.4237
191.5183	193.8549	192.6529	206.4236	207.7653
209.2543	190.4078	204.0626	194.7394	201.8596
194.349	194.0379	203.3849	197.5768	205.5141
187.6439	195.8987	190.4871	197.7961	196.0393
196.6859	192.4028	192.5736	192.8908	201.5851
191.7989	195.7462	194.4344	193.3117	212.8472
191.2865	188.7422	198.4125	199.8033	187.4853
194.2941	196.2647	200.4804	207.2471	206.1055
194.9166	201.4753	187.6622	201.3652	207.8812

Figure 7: 打包后新的定价

计算打包后定价总和为63175.87，比未打包的有明显提高。

4.4 问题四的模型建立与求解

使用问题二建立的（未打包）回归模型：

$$y = 73.3072 - 0.1429x_1 - 0.0048x_2$$

将附件三的数据导入Matlab中，变量 y, x_1, x_2 的意义与上面相同，计算对应的值，得到其结果，如图所示。（图中并不完全）

由于任务数量较多，导致响应的定价偏低。此时采用打包方法较为合适。

5 模型的分析

5.1 模型缺点

- 模型的精确性不够强。
- 打包的方式过于机械化。
- 模型没有考虑会员起始的时间。

任务号码	任务GPS纬度	任务GPS经度	定价
C0001	22.73004117	114.2408795	64.6996
C0002	22.72704287	114.2996199	64.6782
C0003	22.70131065	114.2336007	64.1631
C0004	22.73235925	114.2866672	55.5906
C0005	22.71839144	114.2575495	51.1559
C0006	22.75392493	114.3819253	72.9542
C0007	22.72404221	114.2721836	54.3466
C0008	22.71937803	114.2732478	55.2136
C0009	22.73028254	114.2304955	66.9812
C0010	22.7187968	114.267027	54.4895
C0011	22.65746229	114.3476957	70.3063
C0012	22.71611614	114.2474716	53.4131
C0013	22.72986374	114.2939012	59.6254
C0014	22.746174	114.285869	59.6974
C0015	22.7333086	114.2943071	59.1967
C0016	22.74551648	114.2744633	57.6403
C0017	22.72282036	114.2654813	53.9467
C0018	22.72492675	114.2731306	53.9179
C0019	22.74551648	114.2744633	57.6403
C0020	22.75335785	114.285655	64.8381

Figure 8: 定价

5.2 模型优点

- 模型可以比较快速的求解。
- 模型可操作性强，易于理解。

5.3 模型改进

- 可以考虑建立高次的回归模型，从而提高模型的精确性。

References

- [1] 全国大学生数学建模竞赛组委会, 高教社杯全国大学生数学建模竞赛论文格式规范, 北京, 2009。
- [2] 韩中庚, 数学建模竞赛获奖论文精选与点评, 北京: 科学出版社, 2007。

附录

Matalb程序由以下几部分组成:

- initialize.m (初始化导入数据)
- model_whole.m (初始回归模型)
- huatu.m (画经纬度与定价的图)
- model_yes.m (用完成的数据训练回归模型)
- train_no.m (将回归模型应用于未完成的数据)
- dabao_hao.m (用完成的数据训练打包模型)
- dabao.m (将打包模型应用于所有数据)
- q4ini.m (导入附件三)
- q4by2.m (将回归模型应用于附件三)

附录1 model_whole.m

```
initialize;
a=longtitude;
b=latitude;
counttask=zeros(835,1);
countlabor=zeros(835,1);
countxy=zeros(835,1);
for i=1:835
    for j=1:835
        if a(j,1)-0.05<=a(i,1) && a(i,1)<a(j,1)+0.05 && b(j,1)-0.05<=b(i,1) && b(i,1)<b(j,1)+0.05
            counttask(i,1)=counttask(i,1)+1;
        end
    end
end

for i=1:835
    for j=1:1877
        if vip_longtitude(j,1)-0.05<=a(i,1) && a(i,1)<vip_longtitude(j,1)+0.05 && vip_latitude(j,1)-0.05<=b(i,1) && b(i,1)<vip_latitude(j,1)+0.05
            countlabor(i,1)=countlabor(i,1)+capacity(j,1);
        end
    end
end

for i=1:835
```



```

        for j=1:1877
            if vip_longitude(j,1)-0.05<=a(i,1) && a(i,1)<vip_longitude(j,1)+0.05 && vip_
                countxy(i,1)=countxy(i,1)+xinyu(j,1);
            end
        end
    end
end
y=price;
x1=counttask;
x2=countlabor;
x3=countxy;
Y=y;
X=[ones(length(y'),1),x1,x2,x3];
b=regress(Y,X)

```

附录2 huatu.m

```

initialize;

% scatter(113.26498600,23.12909397,'*r');
% hold on;
% scatter(113.7519901,23.02097619,'*r');
% hold on;
% scatter(113.1220032,23.0218445,'*r');
% hold on;
% scatter(114.0579807,22.54277481,'*r');
% hold on;

% scatter(longtitude_yes,latitude_yes,price_yes,'.y');
% hold on;
% scatter(longtitude_no,latitude_no,price_no,'.b');
% hold on;
% scatter(vip_longitude,vip_latitude,capacity,'.r');

% scatter(vip_latitude,vip_price,capacity,'.r');
% hold on;
% scatter(latitude_no,price_no,'.b');
% hold on;
% scatter(latitude_yes,price_yes,'.y');

% scatter(vip_longitude,vip_price,capacity,'.r');

```

```

% hold on;
% scatter(longtitude_no,price_no,'.b');
% hold on;
% scatter(longtitude_yes,price_yes,'.y');

% scatter3(latitude_yes,longtitude_yes,price_yes,'.y');
% hold on;
% scatter3(latitude_no,longtitude_no,price_no,'.b');
% hold on;
% scatter3(vip_latitude,vip_longtitude,vip_price,'.r');

```

附录3 model_yes.m

```

initialize;
counttask=zeros(522,1);
countlabor=zeros(522,1);
countxy=zeros(522,1);
a=longtitude_yes;
b=latitude_yes;
for i=1:522
    for j=1:522
        if a(j,1)-0.05<=a(i,1) && a(i,1)<a(j,1)+0.05 && b(j,1)-0.05<=b(i,1) && b(i,1)<b(j,1)+0.05
            counttask(i,1)=counttask(i,1)+1;
        end
    end
end

for i=1:522
    for j=1:1877
        if vip_longtitude(j,1)-0.05<=a(i,1) && a(i,1)<vip_longtitude(j,1)+0.05 && vip_latitude(j,1)-0.05<=b(i,1) && b(i,1)<vip_latitude(j,1)+0.05
            countlabor(i,1)=countlabor(i,1)+capacity(j,1);
        end
    end
end

for i=1:522
    for j=1:1877
        if vip_longtitude(j,1)-0.05<=a(i,1) && a(i,1)<vip_longtitude(j,1)+0.05 && vip_latitude(j,1)-0.05<=b(i,1) && b(i,1)<vip_latitude(j,1)+0.05
            countxy(i,1)=countxy(i,1)+xinyu(j,1);
        end
    end
end

```

```

end
y=price_yes;
x1=counttask;
x2=countlabor;
x3=countxy;
Y=y;
X=[ones(length(y'),1),x1,x2,x3];
b=regress(Y,X)

```

附录4 train_no.m

```

initialize;
counttask=zeros(313,1);
countlabor=zeros(313,1);
a=longtitude_no;
b=latitude_no;
for i=1:313
    for j=1:313
        if a(j,1)-0.05<=a(i,1) && a(i,1)<a(j,1)+0.05 && b(j,1)-0.05<=b(i,1) && b(i,1)<b(j,1)+0.05
            counttask(i,1)=counttask(i,1)+1;
        end
    end
end

for i=1:313
    for j=1:1877
        if vip_longtitude(j,1)-0.05<=a(i,1) && a(i,1)<vip_longtitude(j,1)+0.05 && vip_latitude(j,1)-0.05<=b(i,1) && b(i,1)<vip_latitude(j,1)+0.05
            countlabor(i,1)=countlabor(i,1)+capacity(j,1);
        end
    end
end
x1=counttask;
x2=countlabor;
y=73.3072-0.1429*x1-0.0048*x2;

```

附录5 dabao_hao.m

```

initialize;
flag=zeros(522,1);
k=0;
for i=1:522
    if flag(i,1) == 0
        pacnum=1;
        wei(1)=latitude_yes(i,1);
        jing(1)=longitude_yes(i,1);
        qian(1)=price_yes(i,1);
        for j=(i+1):522
            if sqrt((latitude_yes(i,1)-latitude_yes(j,1))^2+(longitude_yes(i,1)-longitude
                pacnum=pacnum+1;
                flag(j,1)=1;
                wei(pacnum)=latitude_yes(j,1);
                jing(pacnum)=longitude_yes(j,1);
                qian(pacnum)=price_yes(j,1);
            end
        end
        k=k+1;
        if pacnum == 1
            new_wei(k)=latitude_yes(i,1);
            new_jing(k)=longitude_yes(i,1);
            new_qian(k)=price_yes(i,1);
        else if pacnum == 2
            new_wei(k)=(wei(1)+wei(2))/2;
            new_jing(k)=(jing(1)+jing(2))/2;
            new_qian(k)=qian(1)+qian(2);
        else if pacnum == 3
            new_wei(k)=(wei(1)+wei(2)+wei(3))/3;
            new_jing(k)=(jing(1)+jing(2)+jing(3))/3;
            new_qian(k)=qian(1)+qian(2)+qian(3);
        end
    end
end
end
new_wei=new_wei';
new_jing=new_jing';
new_qian=new_qian';
counttask=zeros(201,1);

```

```

countlabor=zeros(201,1);
a=new_jing;
b=new_wei;
for i=1:201
    for j=1:201
        if a(j,1)-0.05<=a(i,1) && a(i,1)<a(j,1)+0.05 && b(j,1)-0.05<=b(i,1) && b(i,1)<
            counttask(i,1)=counttask(i,1)+1;
        end
    end
end

for i=1:201
    for j=1:1877
        if vip_longitude(j,1)-0.05<=a(i,1) && a(i,1)<vip_longitude(j,1)+0.05 && vip_
            countlabor(i,1)=countlabor(i,1)+capacity(j,1);
        end
    end
end

y=new_qian;
x1=counttask;
x2=countlabor;
Y=y;
X=[ones(length(y'),1),x1,x2];
b=regress(Y,X)

```

附录6 dabao.m

```

initialize;
flag=zeros(835,1);
k=0;
for i=1:835
    if flag(i,1) == 0
        pacnum=1;
        wei(1)=latitude(i,1);
        jing(1)=longitude(i,1);
        qian(1)=price(i,1);
    end
end

```

```

for j=(i+1):835
    if sqrt((latitude(i,1)-latitude(j,1))^2+(longtitude(i,1)-longtitude(j,1))^2) < 0.05
        pacnum=pacnum+1;
        flag(j,1)=1;
        wei(pacnum)=latitude(j,1);
        jing(pacnum)=longtitude(j,1);
        qian(pacnum)=price(j,1);
    end
end
k=k+1;
if pacnum == 1
    new_wei(k)=latitude(i,1);
    new_jing(k)=longtitude(i,1);
    new_qian(k)=price(i,1);
else if pacnum == 2
    new_wei(k)=(wei(1)+wei(2))/2;
    new_jing(k)=(jing(1)+jing(2))/2;
    new_qian(k)=qian(1)+qian(2);
else if pacnum == 3
    new_wei(k)=(wei(1)+wei(2)+wei(3))/3;
    new_jing(k)=(jing(1)+jing(2)+jing(3))/3;
    new_qian(k)=qian(1)+qian(2)+qian(3);
end
end
end
new_wei=new_wei';
new_jing=new_jing';
new_qian=new_qian';
counttask=zeros(316,1);
countlabor=zeros(316,1);
a=new_jing;
b=new_wei;
for i=1:316
    for j=1:316
        if a(j,1)-0.05<=a(i,1) && a(i,1)<a(j,1)+0.05 && b(j,1)-0.05<=b(i,1) && b(i,1)<b(j,1)+0.05
            counttask(i,1)=counttask(i,1)+1;
        end
    end
end
end

```

```

end

for i=1:316
    for j=1:1877
        if vip_longitude(j,1)-0.05<=a(i,1) && a(i,1)<vip_longitude(j,1)+0.05 && vip_
            countlabor(i,1)=countlabor(i,1)+capacity(j,1);
        end
    end
end

x1=counttask;
x2=countlabor;
y=185.8624+1.8669*x1-0.0061*x2;

```

附录7 q4by2.m

```

initialize;
q4ini;
counttask=zeros(2066,1);
countlabor=zeros(2066,1);
a=q4jing;
b=q4wei;
for i=1:2066
    for j=1:2066
        if a(j,1)-0.02<=a(i,1) && a(i,1)<a(j,1)+0.02 && b(j,1)-0.02<=b(i,1) && b(i,1)<
            counttask(i,1)=counttask(i,1)+1;
        end
    end
end

for i=1:2066
    for j=1:1877
        if vip_longitude(j,1)-0.02<=a(i,1) && a(i,1)<vip_longitude(j,1)+0.02 && vip_
            countlabor(i,1)=countlabor(i,1)+capacity(j,1);
        end
    end
end

x1=counttask;
x2=countlabor;

```

$$y=73.3072-0.1429*x_1-0.0048*x_2;$$