“拍照赚钱”定价方案的研究

摘要

“拍照赚钱”已经成为移动互联网下的一种新兴自助式服务模式。普通用户可以通过完成 APP 上的任务赚取报酬，而企业也可通过 APP 更廉价快速地完成市场调查。

本文运用经济学模型、回归分析等方法对以这种新兴模式为背景提出的四个问题进行了分析，提出了更高效的定价方案。

对于问题一，我们通过可视化方法得出在现有定价方案中对任务价格影响较大的三个因素并利用这三个因素以构造的方式给出了一个现有定价方案的近似方程。最后我们通过回归分析的方法证明了给出方程的有效性，并从地域差异等原方案忽视的角度对部分任务未完成的原因进行了分析。

对于问题二，我们利用 logistic 模型及回归分析法，给出了一个计算任务被完成几率的公式。然后，我们利用模型中的假设对要求解的问题进行了化简，即对一小足够小区域内的所有任务采取相同定价。借助这个条件，我们得出给定任务的成功概率可用只以该任务定价为变量的函数表示。我们考虑成功概率总和及成本总和，给出目标函数，通过对这个函数的分析，找到适当定价方案。

对于问题三，我们用网络流模型来解释人员流动的过程，证明了该网络流模型中的最小费用流是人员选择任务时的配额分布情况。为了改善资源分配，以达到提高任务完成率的目的，我们以距离为主要参数将任务打包发布，来缓解中心地区的压力。最后使用 dinic 算法求解最小费用流来拟合任务获得的期望配额，提高概率模型的准确率。

对于问题四，在第三问的求解中，我们发现每个任务可接受的配额密度近似于区域会员配额密度与任务密度之比，为方便模型求解，我们就用其代替任务可接受的配额密度。同时发现第二问中的模型仍需完善。之后我们进行了 logistic 回归分析，解出各因素的参数，再借助层次分析法，估计打包后新增的变量的参数并进行调整。得到任务完成概率与任务定价的函数关。最后给出定价方案。

最后，在完成对模型的检验分析后，我们给出了关于模型优缺点的客观评价，并指出了改进的方法。

关键字： 定价因素　logistic 回归　打包　目标规划

一、问题重述

1.1 问题背景

“拍照赚钱”作为互联网下的一种新的自助式服务模式，其 APP 的会员用户可以通过完成相应拍照任务赚钱一定的酬金。这种新兴的市场调查方式不仅大大节省了调查成本，而且有效的保证了调查数据的真实性，缩短了调查的周期。因此，如何对 APP 中的任务进行定价就成了这种新方式的核心问题。

1.2 待解决问题

1. 根据对已经完成的一个项目定价规律的分析，得出某些任务未完成的原因。
2. 对上述已完成的项目重新定价，并与原方案进行比较。
3. 考虑实际情况下，对 1 中已完成的项目的部分任务打包发布，并讨论这种发布形式对任务完成情况的影响。
4. 对一个新项目给出定价方案，并评价该方案的实施效果。

二、问题分析

2.1 问题一的分析

我们通过构造测量网格，对价格，用户密度，用户信誉参数进行插值，然后转化为高度图来分析数据的空间分布特征, 最后得到了如下图 1 至图 4。其中，点的明暗表示了点的数值大小，颜色越亮，数值越大；反之越小。

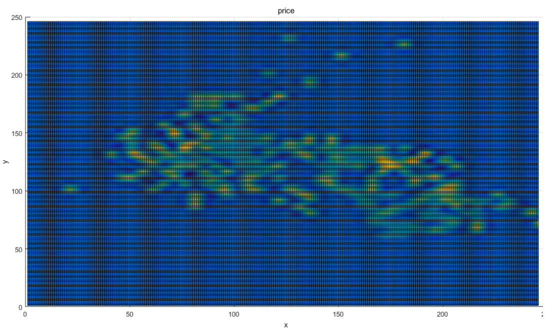


图 1　任务价格空间分布

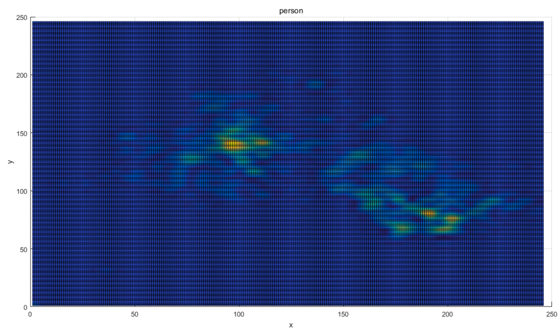
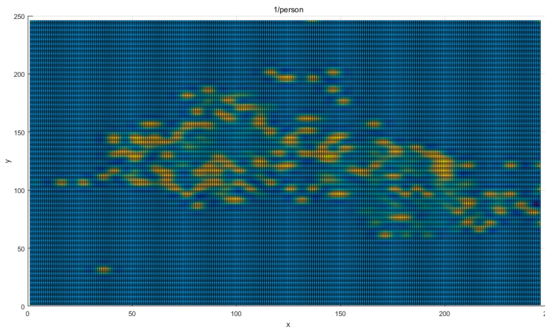


图 2　会员密度空间分布



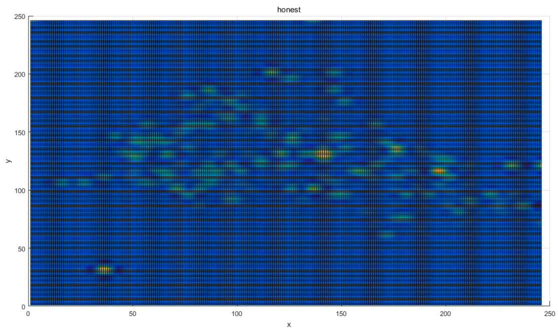


图 3　会员密度倒数空间分布　　　　**图 4　信誉值空间分布**

通过图 1 与图 2 的对比可知，在图 1 中颜色较亮的区域，在图 2 中颜色较暗，因此我们将用户密度取倒数得到图 3，可以清晰的见到图 3 与图 1 的色彩分布一致，由此我们得出定价与会员密度的倒数成正相关。通过图 1 与图 4 的对比，我们可以看到，图 4 中与图 1 中的明亮分布并无明显关系。因此得出结论，在已给的定价规律中，与其他因素相比，用户信誉值对价格确定的影响可以被忽略。

同时，我们根据任务的经纬度信息，将其完成情况映射到实际地图上。从而得到如下图 5：

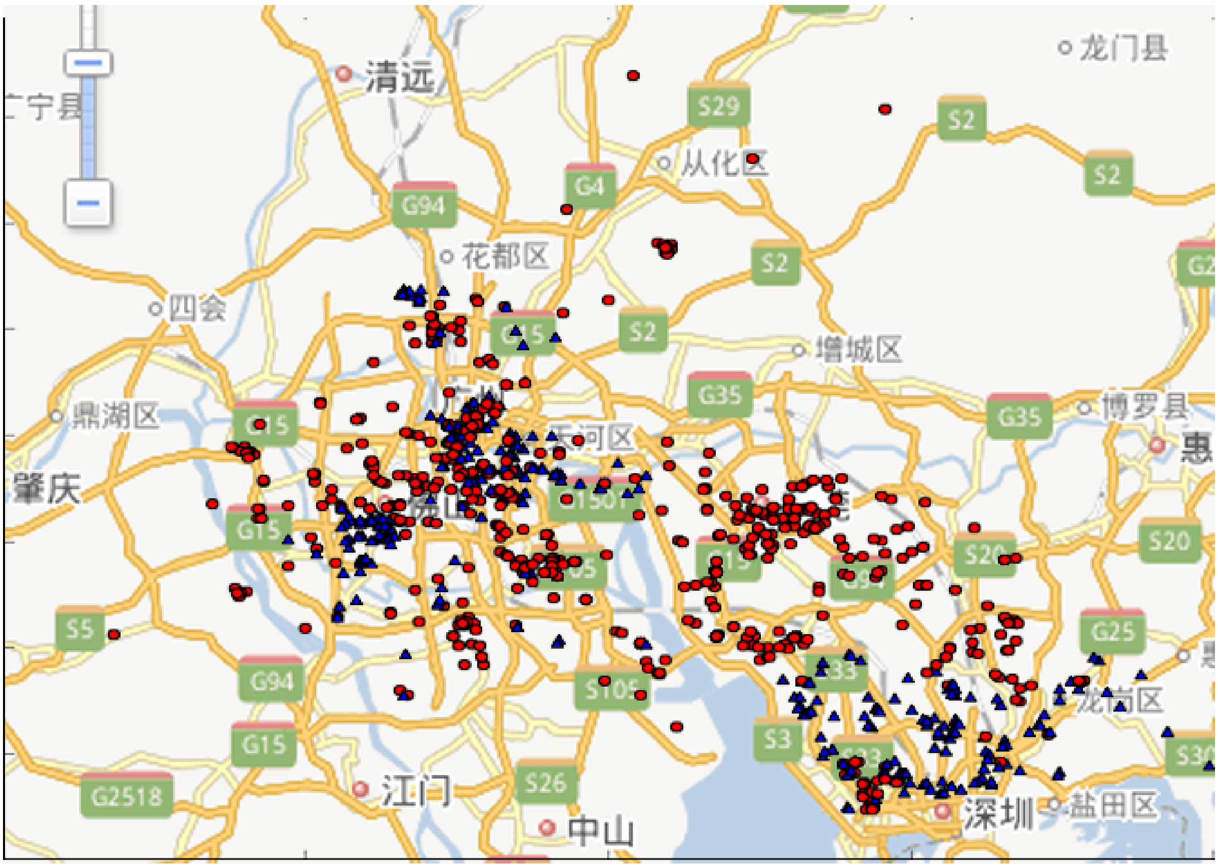


图 5　注：其中红色圆点代表已完成任务，蓝色三角形代表未完成任务。

从图 5 中可以看到，这些任务主要集中在三个区域，因此，我们从地图上对应区

域确定了三个城市：东莞、深圳和广州佛山都市群（从第三产业发展情况分析，将广州佛山看作一个整体是合理的 [1]，后文用广州代替广州佛山都市群），并规定除去极少数特殊点，其余点均属于且仅属于三个城市中的一个。我们通过将给定的任务点及会员位置的经纬度与电子地图中该经纬度的地理位置信息，将这些任务按城市分类，得到了附录 B 中表 1 6。在分类过程中，我们只损失了少数点，因此，我们的分类方法是有效的。在完成对数据的分类后，我们可以进一步从图 1 的色彩明暗看出：靠近城市边缘的任务定价较高；远离城市边缘的任务定价较低。因此得出结论：任务的定价与其在城市中的位置有关。

注：对于游离于三个城市外的少数点，因其过于分散，不具有代表性，因此我们将其舍去。后文也是如此。剩余点请见附录 B 表 7.

2.2　问题二的分析

根据对问题一的分析，可知在现有定价方案下，由于未能考虑任务周边会员信誉值的分布情况或信誉值分布情况和地域差异，导致一些区域的未完成任务数量远高于其他区域。因此，在我们的定价方案中，需要考虑这两个因素，从而提高任务的完成率。同时，原定价模型中的个参数不一定取到最优情形，所以我们转而考虑任务定价与完成概率之间的关系模型，并希望能通过平衡完成概率与支付成本来确定新的定价方案。

2.3　问题三的分析

在现有任务分配规则下，会员更多会选择几个距离相近的任务完成，已在尽量短的时间内可以完成更多任务。这就导致会员资源过度集中在一些任务周围，而那些较分散的任务没有人去做。由图 5 可知这种现象在深圳地区最为明显。用打包的方式解决这个问题，最关键的是使会员在任务集中地区只有很少的几个选项，并且通过降低这些打包任务的定价来让这些任务的吸引力降低；同时，可以将打包降价得到的金额加到那些较分散的任务上，增加对会员的吸引。对于打包方式，为了不使一个包的任务因任务繁重而导致没有会员选择，因此每个包所含任务数量应给出一个上界。同时，为了防止资源的浪费，只将距离近的任务打包。

2.4　问题四的分析

通过上述网络流模型求解，发现每个任务可接受的配额密度近似于区域会员配额密度与任务密度之比，为方便模型求解，我们就用其代替任务可接受的配额密度。同时我们发现在第二问的模型中，我们考虑会员密度而不考虑会员配额密度并不太准确，所以我们先修改第二问的模型，在进行 logistic 回归分析，解出各因素的参数，再借助

层次分析法，估计打包后新增的变量的参数并调整。最后得出任务完成概率与任务定价的函数关系，类似于第二问方法，通过求解目标函数极值点，得出定价方案。

三、模型的假设

- 所有任务对于已经到达任务地点的会员在完成时难易程度相同，不会影响价格的确定和会员的选择。
- 可用一个会员的信誉值来表示其活跃程度的相对关系。
- 对于一充分小区域，该区域的数据可用该区域的相应密度值代表。
- 任务可接受的配额密度可用区域会员配额密度与任务密度之比代替。

四、符号说明

符号	意义
V	任务定价（元）
N	任务总数（个）
P	任务完成情况（P=0 或 1）
d	任务位置距市中心的距离
ρ_m	会员密度（人/平方千米）
ρ_t	任务密度（个/平方千米)
ρ_p	任务配额密度（个/平方千米)
ρ_n	任务量密度（个/平方千米)
ρ_r	会员信誉值密度

注：所有有关密度的量，其计算方式均是现将地图上按经纬度划分为等面积的小区域，然后求出区域内所有相关值的和，在除以划分后每一区域的面积。（其中，对于信誉值的处理是将其先加上一后取自然对数）

五、模型建立与求解

5.1 问题一

5.1.1 定价规律

从对问题一的分析中我们可以得到如下关系：

$$\mathrm{V} \propto \rho_m$$

$$V \propto d$$

同时，我们进行了多次实验，通过控制变量法，发现任务的定价还与任务在一个区域内的密度有关。

在问题假设基础上，借鉴第三方在线购物平台中信誉与竞争因素对价格和销量的影响 citeone，我们建立各因素对价格的影响模型，采用如下线性回归模型对原定价进行拟合：

$$V = \alpha_{10} + \alpha_{11} * \ln \rho_m + \alpha_{12} * \ln \rho_t + \alpha_{13} * \ln d^2 \tag{1}$$

其中 α_{10}、α_{11}、α_{12}、α_{13} 为未知系数通过回归分析可得 α_{10}= 72.3735、α_{11}=-1.4810、α_{12}=-0.9774、α_{13}=12.7459，检验数 p= 1.3673e-63，因此我们的拟合方法是合理的。

$$V = 72.3735 + -1.4810 * \ln \rho_m + -0.9774 * \ln \rho_t + 12.7459 * \ln d^2 \tag{2}$$

5.1.2 任务未完成原因分析

首先，我们可以从图 5 中可以看到，蓝色三角形主要集中在两个区域，也就是未完成任务主要集中在两个城市：广州和深圳；而红色圆点在三个城市均有分布，也就是三个城市中的任务均有人完成。其中，广州市的已完成任务和未完成任务混杂在一起，而东莞市只有已完成任务，深圳市的已完成任务分布在城市边缘。因此，这些任务的完成情况存在着明显的城市差异。通过查阅当地的最新统计年鉴 [3][4][5]，我们得到了三个城市的 2015 年的人均可支配收入，将其和我们计算得到的三个城市任务的平均价格（见附表 B）汇总到下表：

表 2 三个城市数据比较（保留一位小数）

城市	2015 年人均可支配收入	任务定价的均值
广州	46734.6	69.1
深圳	44633.3	67.0
东莞	39793	70.3

从表 1 及图 5 可以得出，位于深圳的任务大部分未完成的原因是其居民的收入较高，定价较低的任务对其缺乏吸引力；而东莞的情况恰好相反，当地居民的收入在三个城市里是最低的，但当地的任务平均定价却是三个城市中最高的，因此所有任务均

被完成；对于广州，则可对比图 1 和图 5，发现定价较高区域完成情况较好。因此，我们认为导致一些任务未完成的原因之一是运营方在制定价格时未考虑到三个城市居民的收入差异，无法对高收入地区用户产生吸引。

而导致部分任务未完成的另一因素则是运营方在定价时未能考虑到任务周边会员的信誉值，从而可能会出现在任务的一定范围内的用户普遍信誉值不高，但运营方并未因此提升任务定价，来起到激励这些信誉值不高的会员或距离较远的高信誉值会员来完成任务的作用。

5.2 问题二

依据对问题的分析，我们可以把对定价 V 的影响因素限制在：任务距市中心距离 d，会员密度 ρ_m，任务密度 ρ_t，会员信誉值密度 ρ_r. 由此，我们可以有如下公式：

$$V = f(\rho_t, \rho_m, \rho_r, d) \tag{3}$$

其中 $f(\rho_t, \rho_m, \rho_r, d)$ 为一未知函数。

由模型假设中的第三个，我们对充分小区域用其区域内量的均值来代表区域内所有点性质，从而对一充分小区域，任务距市中心距离 d，会员密度 ρ_m，任务密度 ρ_t，会员信誉值密度 ρ_r 均相同，因此可以得出：在一充分小区域内所有任务定价相同。

同时，对于问题的完成情况 P，可以将其影响因素分为会员因素和任务因素两大类。首先，会员的信誉值会直接影响到任务的完成情况，在这里，我们用一个区域内会员信誉的均值 ρ_r 来代表该区域内所有会员的信誉情况；其次，由于任务的密度 ρ_t，任务周边会员密度 ρ_m 会影响到会员是否能抢到任务，从而影响任务的完成，并且，任务的定价 V 和任务距市中心的距离 d 会影响到会员是否愿意抢下此任务。综上，我们可以找到 P 及各因素直接的关系，又因其为 0-1 规划，所以我们采取 logistic 模型，有下面公式：

$$P = \frac{1}{1+e^x} \tag{4}$$

$$-x = \alpha_{20} + \alpha_{21} * \ln\rho_t + \alpha_{22} * \ln\rho_m + \alpha_{23} * \ln\rho_r + \alpha_{24} * \ln d^2 + \alpha_{25} * \ln\frac{V}{\bar{V}} + \alpha_{26} * \ln V \tag{5}$$

其中 α_{20}、α_{21}、α_{22}、α_{23}、α_{24}、α_{25}、α_{26} 为未知系数

经过回归分析可得：$\alpha_{20} = -45.2614$，$\alpha_{21} = 0.6729$，$\alpha_{22} = -0.8100$，$\alpha_{23} = 0.0823$，$\alpha_{24} = -0.2063$，$\alpha_{25} = -12.7427$，$\alpha_{26} = 11.0179$

则公式为：

$$-x=-45.2614+0.6729*\ln\rho_t-0.8100*\ln\rho_m+0.0823*\ln\rho_r-0.2063*\ln d^2+-12.7427*\ln\frac{V}{\bar{V}}+11.0179*\ln V \tag{6}$$

由模型假设的第三个，对于一充分小区域有：$\frac{V}{\bar{V}}=1$, 即 $\ln\frac{V}{\bar{V}}=0$; 且 $b=\alpha_{20}+\alpha_{21}*\ln\rho_t+\alpha_{22}*\ln\rho_m+\alpha_{23}*\ln\rho_r+\alpha_{24}*\ln d^2$ 为一常数。因此 x 可化简得：

$$P(V)=\frac{1}{1+e^{b+\alpha_5*\ln V}} \tag{7}$$

我们先按照已经分好的三个区域分别对此方程进行了回归分析，但用已知完成情况进行检验时，发现有较大误差。通过分析，发现造成误差的主要原因是分区域进行回归时，样本空间过于集中。因此，我们不得不将三个区域看成一个整体来分析。

5.2.1 方法一

对于上述 P(V)，我们希望找到一个 λ, 满足:P(V)>λ 时，可近似认为任务完成;P(V)<λ 时。可近似认为任务未完成。通过附录 A 中代码 1.2.2 及 1.2.3 可以画出如下图像：

同时，我们可以得到当 λ =0.4844 时，P(V) 可以有最大值。利用 λ 反解出 V，再带入 P(V) 公式，可得到对于当前 λ 值，完成率的期望为 $P_0=0.6866$，大于已给方案的 0.6252。因此，我们可以认为用 λ =0.4844 给出的方案是合理的。具体每个任务的定价请见附录 B 表 7 中方案一。

5.2.2 方法二

通过公式 (6)，我们可以算得所有任务的价格总期望为

$$E(V)=\sum_{i=1}^{N}P_i*T_i \tag{8}$$

其中 $T_i=\ln V_i$.

同样，我们可以得到所有任务完成率的总期望为：

$$E(P)=\sum_{i=1}^{N}k*P_i \tag{9}$$

由此，我们可以定义目标函数 [6]:

$$I=\sum_{i=1}^{N}*k*P_i-\sum_{i}*P_i*T_i \tag{10}$$

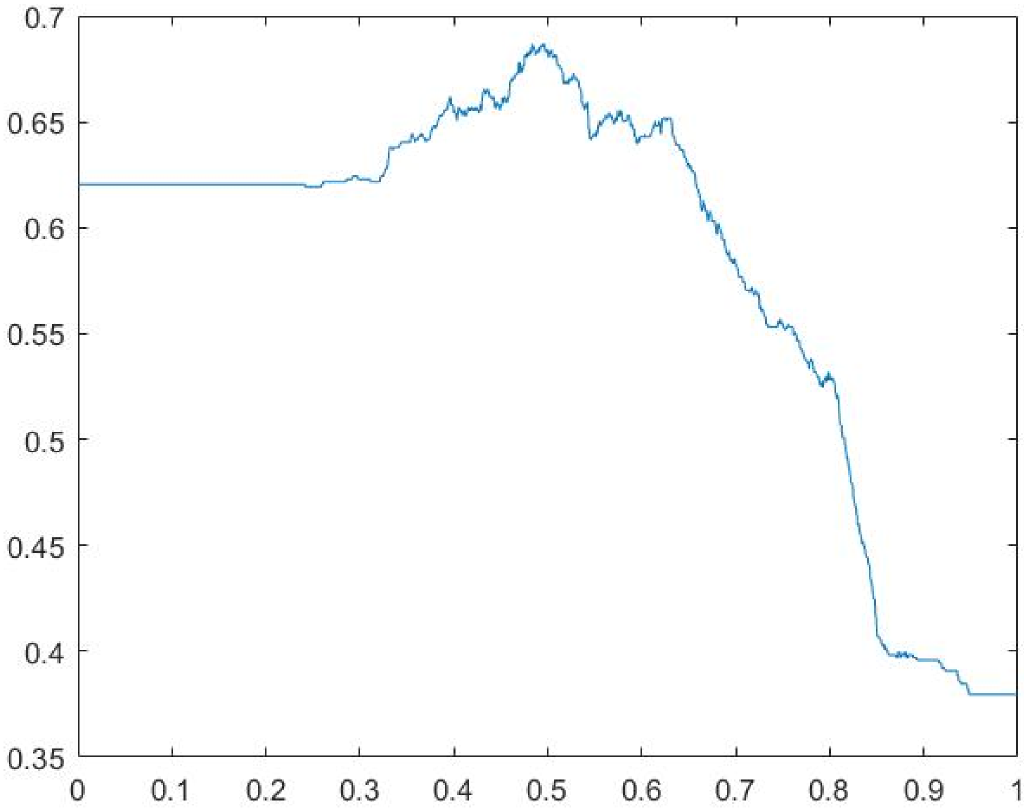


图 6　λ-P 关系图

当 I 取最大时对应的 T_i 即可对应求得 V_i 为对个任务的定价。对其进行化简可得：

$$I = \sum_{i}^{N} \left(\frac{k - T_i}{1 + e^{b+\alpha_5 * T_i}}\right) \tag{11}$$

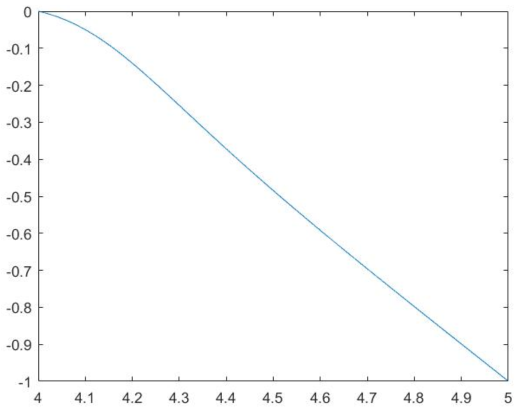
记 $I_i = \frac{k-T_i}{1+e^{b+\alpha_5 * T_i}}$, 易见当每个 I_i 取最大值时，I 为最大值。故，不妨考虑 I_i 的最值，让其对 V 求导可得：

$$I_i{}' = \frac{e^{b+\alpha_5 * T_i} * (\alpha_5 * (k - T_i) + 1) + 1}{(e^{b+\alpha_5 * T_i})^2} \tag{12}$$

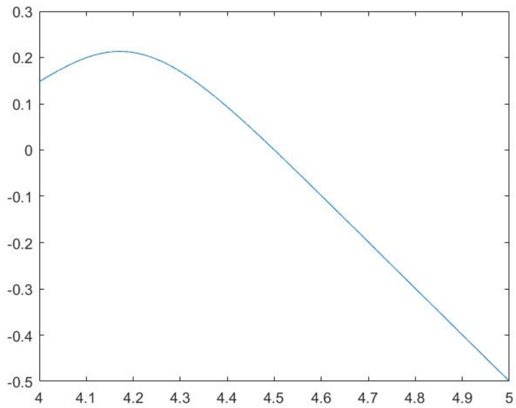
令 $I_i{}' = 0$。通过附录 A 中代码 1.2.4，对于参数 k 取不同值时，可画出 lnV-I_i 的函数图像如下：

当 $k = 4$ 时，函数单调，不符合我们对目标函数的预期；当 $k = 5.5$ 时，得出的任务定价有大量值大于 85，如果采取此方案，会极大地增大成本，不是一种可行方案。因此，基于上述分析，我们采用 $k = 4.5$ 和 $k = 5$ 两个值。(得出的具体定价情况请见附录 B 表 7 方案二及方案三）

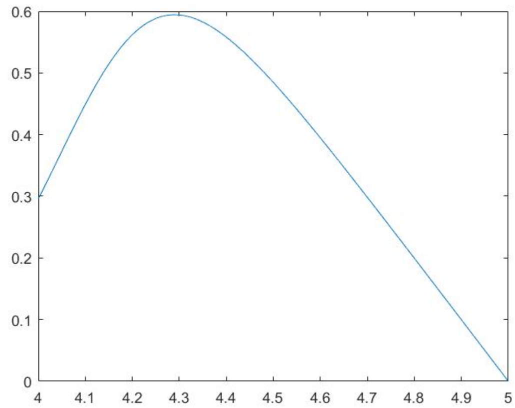
两种新定价方案与已实行的方案的对比如下：



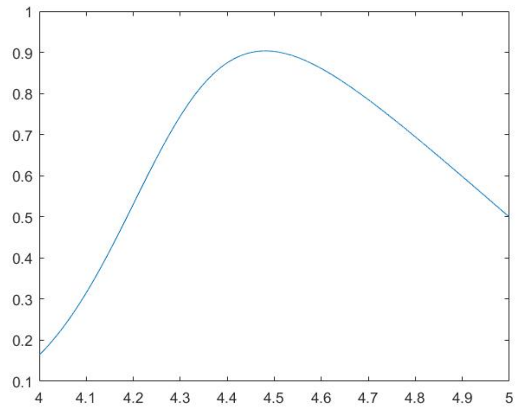
(a) $k = 4$



(b) $k = 4.5$



(c) $k = 5$



(d) $k = 5.5$

图 7　参数 k 取不同值时，lnV-I_i 的函数图像

表 3　三种方案的对比

方案	任务完成率总期望	任务价格总和期望
原方案	500.7452	36446
$k = 4.5$	475.7758	31994
$k = 5$	658.8843	50766

可以从表中看出 $k = 4.5$ 时，成本较原方案降低，但完成率期望却也随之降低；$k = 5$ 时则恰好相反。所以，我们的这两种方案可以适应不同的企业，供其根据需求进行选择。

5.3 问题三

考虑到用户争相选择位置集中的任务会使很多人往一个地方去导致资源浪费，将集中的任务进行适当打包来使得会员去向分散，这样促使更多的任务被完成。可用下图表示：

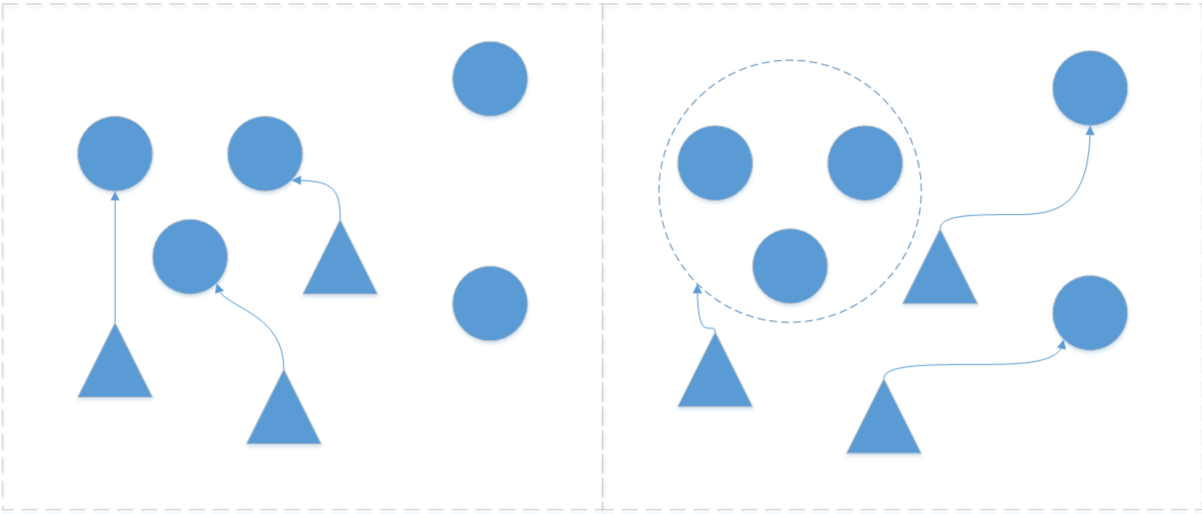


图 8

为了实现这个人员流动“推挤”的过程，我们使用网络流模型来模拟这个过程。我们建立以下网络流模型：

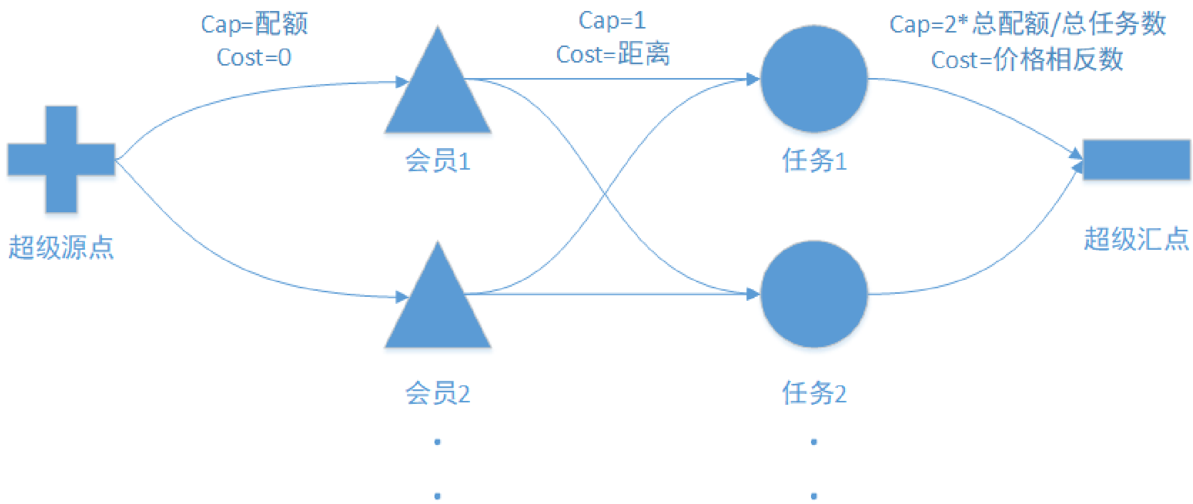


图 9

并给出如下定义：

任务的周边配额密度 = 任务实际流量/容量

一条路径的费用 = 流量 * 单位费用 = 流量 *(距离 -任务价格)

显然，任务的打包与任务量直接相关。同时，对于一个会员来说，他会想办法使这条路径的费用达到最低，这符合最大化利益动机。而对于网络来说，需要流量的最

大化，也就算是使任务周边配额密度和最大化。因为任务周边配额密度越大表示越多的人想做这个任务，也就意味着任务的完成概率越大，这符合企业的动机，所以网络的任务完成度恰好符合它的最小费用最大流的流量。

基于以上分析，我们就可给出打包算法：

首先，我们用并查集的方法维护任务包，对任一任务点枚举其它所有点，将距其小于 D 的任务或任务包合并，但合并后包内任务数量不超过包容量上限 K. 打完包后需要重新对任务进行定价，得到相应的最小费用最大流。并用得到的新的周边配额密度为每个点计算完成概率。

由附录 A 中 1.2.5 我们可以得到打包后的方案（见附录 B 表 8）打包后的任务分布为

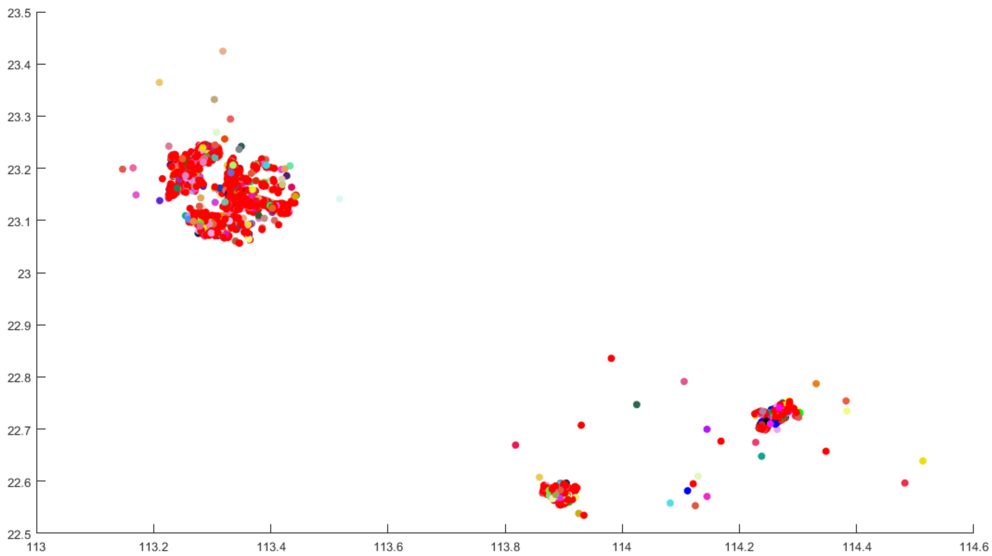


图 10　打包后任务分布

由此，我们可以知道，对于前面的模型，需要将会员密度参量改为会员配额密度。经过修改，我们的模型将能更好的对任务的完成概率进行预测，从而得出更优的定价方案。

5.4　问题四

我们先修正问题二的模型（因打包前任务量可与任务密度看做同一因素，暂不添加此变量）：

进行 logistic 回归处理：$\alpha_0 = -45.2614$ $\alpha_1 = 0.6729$ $\alpha_2 = -0.8100$ $\alpha_3 = 0.0823$ $\alpha_4 = -0.2063$ $\alpha_5 = -12.7427$ $\alpha_6 = 11.0179$

p 值为：4.94209361568777e-07，7.86331458411476e-05，1.13405964726879e-09，0.444914297067 0.00115895972348319，0.000334884346677496，2.01948502994176e-07。

我们发现信誉值影响几乎可忽略不计且其 p 值较大，于是我们在模型中忽略信誉值。再进行 logistic 回归处理: $\alpha_0 = -44.8824$ $\alpha_1 = 0.6645$ $\alpha_2 = -0.7773$ $\alpha_3 = 0$ $\alpha_4 = -0.2042$ $\alpha_5 = -12.6592$ $\alpha_6 = 10.9407$

p 值为：5.31741669333309e-07，9.07298885599064e-05，5.80421111008928e-10，NaN，0.00127839964487700，0.000346363035161265，2.12259951974742e-07。

最后，考虑打包后情形，我们加上任务量因素给出新的模型：

$$P = \frac{1}{1+e^x} \tag{13}$$

$$-x = \alpha_0 + \alpha_1 * \ln \rho_t + \alpha_2 * \ln \rho_m + \alpha_4 * \ln d^2 + \alpha_5 * \ln \frac{V}{\overline{V}} + \alpha_6 * \ln V + \alpha_7 * \ln \rho_n \tag{14}$$

这里，我们认为除任务量因素外，其余因素的参数直接继承打包前的模型参数，即：$\alpha_0 = -44.8824$ $\alpha_1 = 0.6645$ $\alpha_2 = -0.7773$ $\alpha_4 = -0.2042$ $\alpha_5 = -12.6592$ $\alpha_6 = 10.9407$ 对于任务量参数 6，我们考虑层次分析法 [7]，任务密度的重要程度略高于任务量密度的重要程度（2,1/2），而距离因素（人口流动性）的影响又稍弱于任务量（1/3,3），于是我们大致估计任务量密度的参数为 0.3。

我们先计算出 $b = \alpha_{20} + \alpha_{21} * \ln \rho_t + \alpha_{22} * \ln \rho_m + \alpha_{23} * \ln \rho_r + \alpha_{24} * \ln d^2$，发现其基本稳定在 31 左右，这样带入 $P = \frac{1}{1+e^x}$，p 值几乎恒等于 1. 于是我们再调整参数 7，当其取 7.7 左右时，P 值波动较为合适。得到新模型：$-x = -44.8824 + 0.6645 * \ln \rho_t - 0.7773 * \ln \rho_m - 0.2042 * \ln d^2 + -12.6592 * \ln \frac{V}{\overline{V}} + 7.7 * \ln V + 0.3 * \ln \rho_n$

利用和第二问一样的规划：

$$\max_V I = \sum_i^N \left(\frac{k - T_i}{1 + e^{b + \alpha_5 * T_i}}\right) \tag{15}$$

取适当 k 值, 有如下两幅图：

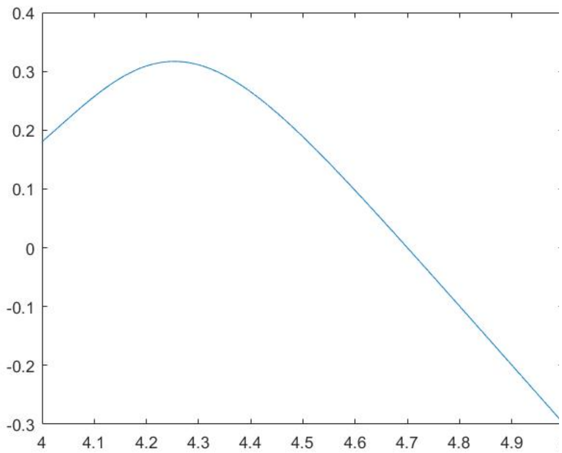


图 11　k=4.7

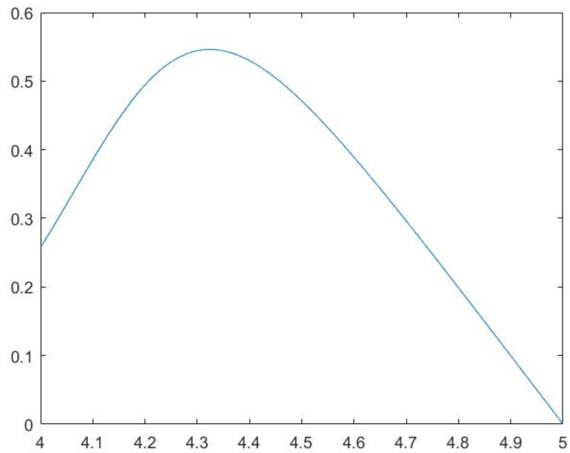


图 12　k=5

表 4　两种方案的对比

方案	任务完成率总期望	任务价格总和期望
$k = 4.7$	536.0083	37516
$k = 5$	629.3241	48285

最后利用同第二问相同的求解方式，得出我们的两个任务定价方式 (参见附录 B 表 9 及表 10)。同时得出两个方案的成功率期望及成本期望。从表中可以得知，当成本小幅上升时，可以带来成功率的大幅上涨。因此，我们的方案是有效的。在任务显著变多的情况下，我们的成本期望和上次任务的成本期望相比并没有显著涨幅，同时也保证了较高的成功概率。

六、模型的检验

首先，我们考虑问题二方法一中，当取最优的 λ 值时，拟合度不足七成。所以我们检验当 λ 出现波动时，对定价结果的干扰。通过观察下图 λ 从 0.45 变化到 0.5 时某一随机任务此方法最优定价的变化，我们可以发现两者之间呈线性关系，但变化率并不显著，定价差距控制在两元以内。因此，此方法在一定程度上可行。

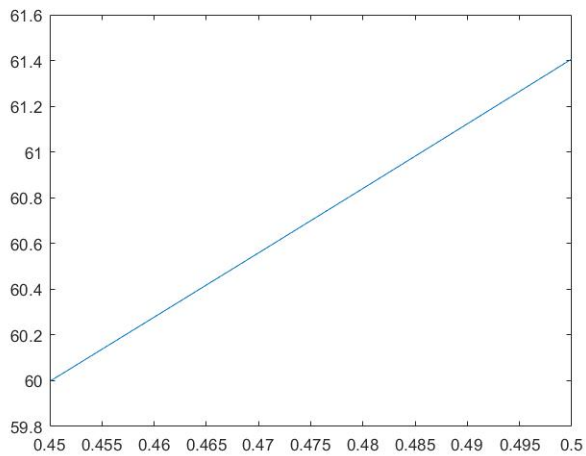


图 13　λ 检验

同样，我们需要检验方法二目标函数中权重 k 的影响。下图为任务完成率总期望及任务价格总和期望随 k 值变化的图像。从中可知，两者随 k 值变化的趋势几乎相同。也就是说 k 值的稳定性不高，所以若有继续工作的可能，可以考虑 k 更精确的取值。

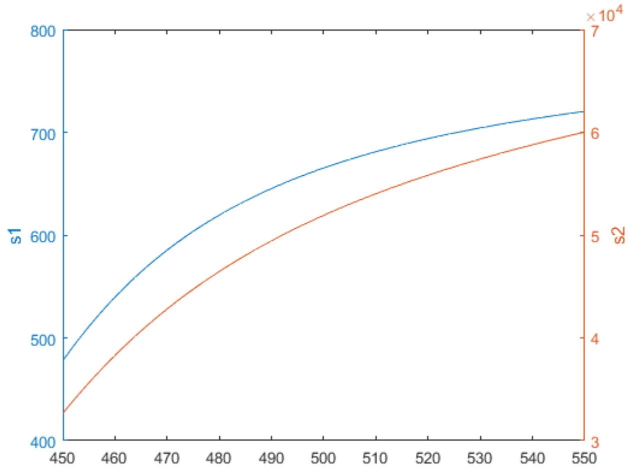


图 14　k 的检验

最后我们检验问题四模型中两个参数 α_7 及 k 的影响。分别作出 $\alpha_7 = 0.25$□$0.3, 0.35$ 时，任务完成率总期望及任务价格总和期望随 k 值变化的图像。

从以下三幅图的对比中可以发现当 $alpha_7 = 0.3$ 时，与其它两种情形相比，任务完成率总期望及任务价格总和期望变化趋势有差异，使得能较好地确定极值，但差异在可接受的范围内。使得在极值附近不会出现较大波动，因此模型有合理性。

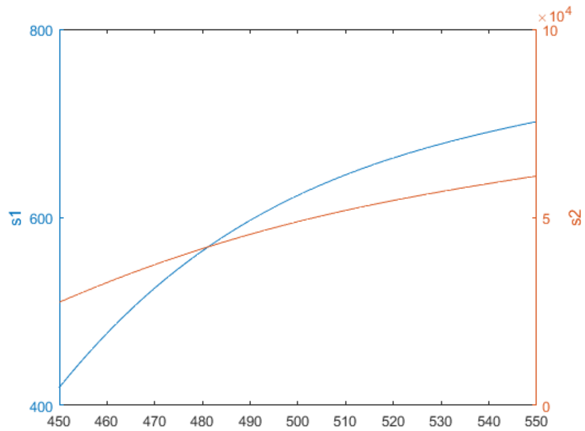


图 15　$\alpha_7 = 0.25$

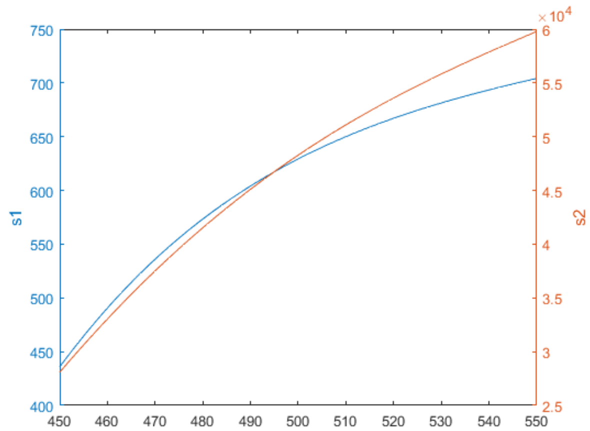


图 16　$\alpha_7 = 0.3$

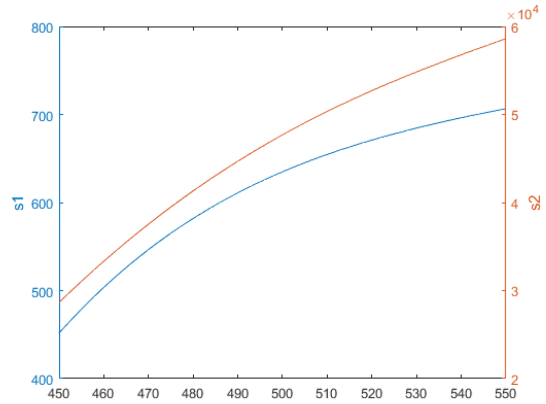


图 17　$\alpha_7 = 0.35$

七、模型优缺点及改进方案

7.1　缺点

1. 忽略了任务难易程度，每个会员是否完成选定任务的随机因素等因素。2. 考虑到有地域因素影响，却因各地域样本过于特殊，而在模型中放弃考虑此因素。3. 模型大部分由数据分析得到，而缺少理论推导。

7.2　优点

1. 给出的方案有效提高了任务的完成率。2. 整个模型在算法上较为简易。

7.3　改进方案

1. 可在模型中加入地域经济差异因素，此因素可从统计局统计年鉴中获取。比如：人均可支配收入、恩格尔系数。2. 对于已确定有影响的因素，可通过文献材料及相关学科知识分析其对模型影响的本质原因。

八、参考文献

[1] 陈鸿宇, 郭超. “广佛都市圈”的形成和发展动因分析——对广州、佛山产业结构变动的实证研究 [J]. 广东经济,2006,(01):19-23. [2017-09-17].

[2]] 徐敏. 第三方在线购物平台中信誉对价格和销量的影响研究 [D]. 哈尔滨工业大学,2016

[3] 广州统计局. 统计年鉴 2016[J], 中国统计出版社,2016,4,.

[4] 深圳统计局. 深圳 2016 年统计年鉴 [J], 中国统计出版社,2016,3,.

[5] 东莞统计局. 2016 年东莞统计年鉴 [J], 中国统计出版社,2016,5,.

[6] 黄湘民, 刘大成, 周阳方. 国外物流成本研究前沿及进展——一个文献综述 [J]. 商业研究,2006,(23):203-209. [2017-09-17]. DOI：10.13902/j.cnki.syyj.2006.23.057

[7] 董君. 层次分析法权重计算方法分析及其应用研究 [J/OL]. 科技资讯,2015,13(29):218+220. (2016-03-08)[2017-09-17]. http://kns.cnki.net/kcms/detail/11.5042.n.20160308.1549.125.htmlDOI：10.16661/j.cnki.1672-3791.2015.29.218

附录 A　我的源程序

1.1　Python 代码

1.1.1　区域划分

```
#coding: utf-8

bt=open("bound.txt","r")
bd={"dg":[],"gz":[],"sz":[]}
for lines in bt.readlines():
  line=lines.replace("\n","")
  if line=="dongguan":
    lis=bd["dg"]
    continue
  elif line=="guangzhou":
    lis=bd["gz"]
    continue
  elif line=="shenzhen":
    lis=bd["sz"]
    continue
  elif line=="":continue

  y,x=map(float,line.split()[0].split(","))
  lis.append((x,y))

def vsub(A,B):
  return (A[0]-B[0],A[1]-B[1])

def vadd(A,B):
  return (A[0]+B[0],A[1]+B[1])

def Cross(A,B):
  return A[0]*B[1]-A[1]*B[0]

def Seg_pi_Seg(a1,a2,b1,b2):
  c1=Cross(vsub(a2,a1),vsub(b1,a1))
  c2=Cross(vsub(a2,a1),vsub(b2,a1))
  c3=Cross(vsub(b2,b1),vsub(a1,b1))
  c4=Cross(vsub(b2,b1),vsub(a2,b1))
  return c1*c2<0 and c3*c4<0

def Point_in_poly(a,city):
```

```
  cnt=0
  lis=bd[city]
  l=len(lis)
  for i in range(0,l):
    if Seg_pi_Seg(a,(a[0]+100,a[1]),lis[i],lis[(i+1)%l]):
      cnt+=1
      # plt.plot([lis[i][0],lis[(i+1)%l][0]],[lis[i][1],lis[(i+1)%l][1]],"r
                                                   -")
  # print(cnt)
  # plt.plot([a[0],a[0]+2],[a[1],a[1]],"b-")
  if(cnt%2==1):return True
  else: return False

def in_city(a,city):
  return Point_in_poly(a,city)

if __name__=="__main__":

  from matplotlib import pyplot as plt
  import random
  for cn in bd.keys():
    lis=bd[cn]
    plt.plot([xx[0] for xx in lis],[xx[1] for xx in lis],"b-")
  # plt.plot([xx[0] for xx in lis],[xx[1] for xx in lis],"b-")
  # for i in range(0,1000):
    # y=random.random()*2+22
    # x=random.random()*2+112.5
    # plt.plot(x,y,"ro" if in_city((x,y),"dg") else "bo")

  # plt.show()
```

1.1.2 对于用来做公式（1）回归分析的矩阵

```
#coding : utf-8

import numpy as np
from skimage import io,feature
from matplotlib import pyplot as plt
import citysplit
import math

def xyc(x,y):
  if(isinstance(x,str)):
```

```
    x=float(x)
    y=float(y)
  return  (min(49,max(0,int((x-112.5)*25))),min(49,max(0,int((y-22)*25))))

#------------------1/person
f=open("fuckb2.csv","r")

ma_honr=np.zeros((50,50),dtype="d")
person_cnt=np.zeros((50,50))
for line in f.readlines():

  lis=line.replace(" ,",",").replace(" ",",").replace("\n","").split(",")
  id,x,y,cap,tim,hon=lis[0],lis[1],lis[2],lis[3],lis[4],lis[5]

  x=float(x)
  y=float(y)
  x,y=y,x
  if(x<100 or y>30):continue
  ix,iy=xyc(x,y)
  ma_honr[iy,ix]+=math.log(float(hon)+1)
  person_cnt[iy,ix]+=1

for i in range(50):
  for j in range(50):
    ma_honr[i,j]=0 if person_cnt[i,j]==0 else ma_honr[i,j]/person_cnt[i,j]
f.close()
#-----------------------------------------price
f=open("fuckb1.csv","r")
ma_price=np.zeros((50,50),dtype='d')
task_cnt=np.zeros((50,50),dtype='d')
for line in f.readlines():
  id,x,y,price,comp=line.replace("\n","").split(",")
  x,y=y,x
  plt.scatter(x, y, c='r'if comp=='1' else 'b', marker='o' if comp=='1'
                                      else '^',linewidths=None)



  ix,iy=xyc(x,y)
  ma_price[iy,ix]+=float(price)
  task_cnt[iy,ix]+=1
for i in range(50):
  for j in range(50):
    ma_price[i,j]=0 if task_cnt[i,j]==0 else ma_price[i,j]/task_cnt[i,j]
```

```
f.close()

#------------------------------------------task

def get_d(x,y):
  def distance(A,B):
    return math.sqrt((A[0]-B[0])*(A[0]-B[0])+(A[1]-B[1])*(A[1]-B[1]))
  center={"dg":(113.758231,23.026998),"sz":(114.064803,22.549054),"gz":(113
                                      .270714,23.13552)}
  for ct in center.keys():
    if(citysplit.in_city((x,y),ct)):
      return distance((x,y),center[ct])
  print("not in city")
  return 0

fo=open("regdata_price.csv","w")
f=open("fuckbl.csv","r")
for line in f.readlines():
  id,x,y,price,comp=line.replace("\n","").split(",")
  x,y=y,x
  x,y=float(x),float(y)
  ix,iy=xyc(x,y)
  d=get_d(x,y)
  fo.write(','.join(map(str,[
  1,math.log(max(1,person_cnt[iy,ix])),
  math.log(task_cnt[iy,ix]),
  math.log(d**2+1)]))+","+price.replace("\n","")+"\n")

fo.close()
plt.show()
```

1.2 Matlab 代码

1.2.1 对于公式 (1) 的回归分析

```
tab=csvread('regdata_price.csv');
y=tab(:,5);
X=tab(:,1:4);
[b,bint,r,rint,stats] = regress(y,X);
b
stats
```

1.2.2 P(V) 的代码实现

```
function [ pro ] = prob( a,X,b,y)
sum=0;
for i=1:length(X)
    c=1/(1+exp([1 X(i,:)]*b));
    if (c>a && y(i)==1)||(c<a && y(i)==0)
        sum=sum+1;
    end

end
pro=sum/length(X);
```

1.2.3 求解 λ

```
tab=csvread('regdata_comp.csv');
y=tab(:,8);
X=tab(:,2:7);
[b,DEV,STATS] = glmfit(X,[y ones(length(y),1)],'binomial', 'link', 'logit
b=-b;
s=0;ss=0;
for j=1:10000
    c(j)=prob(j/10000,X,b,y);
    if c(j)>s
        s=c(j);
        ss=j;
    end
end
plot(0.0001:0.0001:1,c)
s
ss/1000
```

1.2.4 对于给定参数 k，求解 $I_i' = 0$

```
s1=0;s2=0;
t=4:0.0001:5;
k=input('请输入k:␣');
p=1./(1+exp(d(2)+b(7).*t));
```

```
f=k.*p-t.*p;
plot(t,f)

for i=1:length(d)

        f=@(t)(exp(d(i)+b(7).*t).*(b(7)*(k-t)+1)+1);
        [m, n]=fzero(f,4);
        v(i)=exp(m);
        p=1/(1+exp(d(i)+b(7)*m));
        f=k*p-m*p;
        s1=s1+p;
        s2=s2+p*v(i);
end
s1
s2
% csvwrite('ans224.5.csv',v)
```

1.2.5 任务打包

```
[data,id]=task2mat('data_task_id.csv');%经度[ 纬度价格编号 ][]
n=size(data,1);
global fa;
fa=int32(1:n);
global siz;
siz=ones(1,n);
maxsize=4;

for i=1:n-1
    for j=i+1:n
        d=get_distance(data(i,1:2),data(j,1:2));
        if (d<=0.5 && fnd(i)~=fnd(j) && siz(fnd(i))+siz(fnd(j))<maxsize)
            join(j,i);
        end
    end
end
tm=unique(fa);
colo=zeros(n,3);
```

```
for i=1:n
    fnd(i);
    colo(i,:)=get_hashed_color(fnd(i));
end
scatter(data(:,1),data(:,2),[],colo,'filled');

fid=fopen('dabao_id_old.csv','w');
for i=1:n
    fprintf(fid,'%s,%d,%d\n',char(id{1,1}(i)),i,fnd(i));
end
fclose(fid);
function top=fnd(x)
    global fa;
    top=x;
    while(top~=fa(top))
        top=fa(top);
    end
    while(x~=top)
        tmp=x;
        x=fa(x);
        fa(tmp)=top;
    end
end

function join(x,y)
    global fa;
    global siz;

    fx=fnd(x); fy=fnd(y);
    if(fx~=fy)
        fa(fx)=fy;
        siz(fy)=siz(fy)+siz(fx);
    end
end

function c=get_hashed_color(x)
```

```
    xx=x^4+1564563423;
    R=mod(xx,256)/256;
    xx=xx/256;
    G=mod(xx,256)/256;
    xx=xx/256;
    B=mod(xx,256)/256;
    c=[R,G,B];
end
```

1.2.6 第四问

```
clear all
clc

b=[-44.8824,0.6645,-0.7773,-0.2042,-12.6592,0.3,7.7]';
b=-b;
mat=csvread('data_matrix_new.csv');
d=mat*b(1:6);
s1=0;s2=0;
t=4:0.0001:5;
k=input('请输入k:␣');
p=1./(1+exp(d(2)+b(7)*t));
f=k.*p-t.*p;
plot(t,f)

for i=1:length(d)

        f=@(t)(exp(d(i)+b(7).*t).*(b(7).*(k-t)+1)+1);
        [m, n]=fzero(f,4);
        v(i)=exp(m);
        p=1/(1+exp(d(i)+b(7)*m));
        f=k*p-m*p;
        s1=s1+p;
        s2=s2+p*v(i);
end
s1
s2
```

```
%csvwrite('ans22 5.csv',v)
```

1.2.7 第四问

```
clear all
tab=csvread('regdata_comp.csv');
y=tab(:,8);
X=tab(:,2:7);
[b,DEV,STATS] = glmfit(X,[y ones(length(y),1)],'binomial', 'link', 'logit
b=-b;

j=unidrnd(801);
d=[ones(length(X),1) X(:,1:4)]*b(1:5);
for i=1:500
p=0.45+i/10000;
t(i)=p;
v(i)=exp((log(1/p-1)-d(j))./b(7));
end
plot(t,v)
```

附录 B 数据

表 1. 东莞市会员表：Member(1)（东莞）.csv

表 2. 广州市会员表：Member(2)（广州）.csv

表 3. 深圳市会员表：Member(3)（深圳）.csv

表 4. 东莞市任务表：Task(1)（东莞）.csv

表 5. 广州市任务表：Task(2)（广州）.csv

表 6. 深圳市任务表：Task(3)（深圳）.csv

表 7. 给出的三种定价方案：任务定价表.csv

表 8. 打包后的任务：打包后方案.csv

表 9. 第四问中给出的一种定价方案：k=4.7 时定价方案.csv

表 10. 第四问中给出的另一种定价方案：k=5 定价方案.csv

注：所有表均包含在支撑材料里。