

“拍照赚钱”任务分配与定价评估模型

唐 寅, 高云聪, 陈灿宇

指导教师:

(中国科学院大学本科部, 北京 100049)

编者按: 本文仿照库伦定律, 建立了任务对会员的吸引力因子与任务价格、会员信誉度和会员到任务点距离之间的关系, 并定义了任务的完成率, 然后通过模拟计算得到了定价模型效果的评价. 作者对问题的理解比较到位, 抓住了本问题建模的关键因素——任务对会员的吸引力, 建模思路有新意. 需要指出的是, 本文关于任务完成概率的定义有不合理之处, 导致按此定义得到的完成率偏低. 另外, 模拟计算还可以做得更细致一些.

摘 要: “拍照赚钱”是移动互联网下的一种自助式服务模式, 其中任务的分配是一个重要的环节. 在我们的研究中, 我们对于每个定价方案给出了任务分配的机制, 并利用自己建立的机制对它们进行了评估. 我们引入了吸引力因子, 用两种方法刻画任务点对会员的吸引力, 并分别给出了不同的任务分配的机制. 之后, 我们还引入了模型的评估机制, 定义了完成率、用户总盈利、平台总收益等评价指标. 我们还考虑了对任务进行打包发布的情形. 我们先对任务点进行了聚类, 将每一类中的任务点联合在一起发布. 参照经济学中的搭售模型, 我们估算出了打包过程中定价的上下界. 我们还对吸引力因子的定义进行了推广, 给出了联合发布情形中的表达式. 但经过模拟, 这种做法对完成率几乎没有影响, 但是影响到用户总盈利、平台总收益等.

关键词: 众包; 搭售; 聚类分析; 模拟

分类号: AMS(2000) 97M40

中图分类号: O29

文献标识码: A

1 问题重述 (略)

2 符号约定

我们事先约定一下后面需要用的一些符号. 设所有任务点为 Y_1, Y_2, \dots , 所有会员记为 X_1, X_2, \dots . p_i 为任务点 Y_i 的价格, r_{ij} 为任务点 Y_i 与会员 X_j 之间的距离, m_i 为原始的会员信誉值. 其他需要用到的变量会在模型中定义.

3 模型假设

因所给会员信誉值差异过大, 我们在保持原始序关系的前提下, 将会员信誉值控制在 0.2–0.8 范围内, 其转换关系式为

$$\varphi(m_i) = 0.067930997 \log_{10} m_i + 0.471723987. \quad (1)$$

在后文中, 和信誉值有关的计算, 我们都用 $\varphi(m_i)$ 来表示.

其他假设都是平凡的, 在此不作赘述.

4 定价方案下任务分配与定价评估模型的建立

4.1 吸引力因子的引入

我们可以引入“吸引力因子”，来评判任务点对会员的吸引作用，我们不妨假定，吸引作用越强，会员越有可能去做该任务。吸引力因子决定了会员对任务的选择方式，即任务的分配问题。我们用 F_{ij} 表示任务点 Y_i 对会员 X_j 的吸引力因子。

4.2 综合分析模型

定性判断可知，所定义的吸引力一定是与任务点的定价正相关的，任务点的价格越高，会员自然越希望去完成这个任务。会员的信誉度也和吸引力正相关，从运营方的角度来说，派信誉高的会员去做任务，可以提高完成率。类似分析可以发现，吸引力与会员到任务点的距离应该是负相关的，同样的价格，会员自然希望去做更近的任务。但经过前面的分析，完成度的差异说明，各个地区对任务的完成度有很大差异，我们把这些差异带来的影响统一为一个参数——困难因子，来刻画综合区位因素、经济发展水平、信息不对称导致信息接收不及时等原因造成的客观条件上的吸引力削弱。

我们联想到了库伦定律的表达式

$$F = \frac{kq_1q_2}{r^2}. \quad (2)$$

可以把任务点类比成负电荷，会员类成正电荷，它们之间的吸引力因子可以用类似的公式给出，但是在这里我们作了一些修改，公式如下

$$F_{ij} = \frac{k_{ij}p_i\varphi(m_j)}{r_{ij}^\alpha}, \quad (3)$$

其中 α 为参数， k_{ij} 即为刚刚定义的困难因子，来体现客观条件对任务点 Y_i 对会员 X_j 的吸引力的影响。对于不同的地区， k_{ij} 可以不是常数，其他影响因素都可以吸收到 k_{ij} 中。

我们发现，吸引力因子的定义是合理的。首先，吸引力因子应该与任务点的价格 p_i 成正比，因为价格越高，对用户的吸引力当然越大。其次，吸引力因子与会员到任务点的距离 r_{ij} 应该成反比，因为会员自然会选择离自己更近的任务点，也就是说，更近的任务点对会员的吸引力会大一些。另外，吸引力因子与会员的信誉度 $\varphi(m_j)$ 也应该成正比。其中的原因比较复杂，需要结合任务分配方案来考虑。对于同样的距离和任务点距离，在分配的时候当然尽量让信誉度高的会员来做，相应地，可以认为其他条件相同，会员的信誉度越高，那么任务对会员的吸引力越大，因此吸引力因子可以认为与会员信誉度正相关。

我们的目标是完成会员与任务的完全匹配。基于这个目标，我们可以给出如下的算法：

S1：对所有会员进行预定任务开始时间从早到晚的排序，时间相同者，按照信誉度由高到低排序；

S2：开始时间最早的会员中信誉度最高的 X_1 首先挑选出对其吸引力因子最大的 N_1 个任务，其中任务个数 N_1 不超过会员 X_1 预定任务的限额；

S3：在所有任务的集合中删除 N_1 个已选任务，避免后面的会员重复选择；

S4：排序第二的会员 X_2 挑选 N_2 个任务，其中任务个数 N_2 不超过会员 X_2 预定任务的限额；

S5：不断循环往复，直到每个会员都分配到了任务(包括新会员)，且任务已经分配完成。

4.3 单位长度期望模型

在这个模型中，我们沿用吸引力因子这个概念。

另一种直观的想法是，会员选择任务点的标准应该是“拿更多的钱，走更少的路”。也就是说，通过单位距离所能够得到的钱对吸引力因子产生了很大的影响。然而，还有经济水平、区位等因素的存在，我们可以仿照前一个模型，把其他因素都吸收到一个参数 k_j 中。

我们得到如下公式

$$F_{ij} = k_j \frac{\sum_{i=1}^{N_j} p_{ik}}{|l_j|}, \quad (4)$$

其中 l_j 表示从 X_j 会员点出发的一条路径, 经过 N_j 个任务点 i_1, i_2, \dots, i_{N_j} . 设 $i = (i_1, i_2, \dots, i_{N_j})$. 此时, 单位长度会员盈利的期望即为

$$k_j \frac{\sum_{i=1}^{N_j} p_{ik}}{|l_j|},$$

其中 $|l_j|$ 表示路径 l_j 的长度, k_j 则为此路径的困难程度. 我们就将这种模型的吸引力因子 F_{ij} 定义为单位长度会员盈利的期望值.

关于会员任务分配的机制, 我们给出如下算法:

S1: 对所有会员进行预定任务开始时间从早到晚的排序, 时间相同者, 按照信誉度由高到低排序;

S2: 开始时间最早的会员中信誉度最高的 X_1 首先挑选出对其吸引力因子最大的 N_1 个任务, 其中任务个数 N_1 不超过会员 X_1 预定任务的限额;

2a): 该会员以其坐标为基准, 遍历 n 个任务点, 分别计算每个点对自己的吸引力因子, 取吸引力最大的任务点为 l_1 路径的第一个点 i_1 ;

2b): 以 i_1 为基准, 遍历剩下的任务点, 分别计算到每个点的吸引力, 取吸引力最大的点为 l_1 路径的下一个备选点 (记为 i_0);

2c): 若将备选点 i_0 加到 l_1 上后, 新的路径的吸引力因子不小于原来路径的吸引力因子, 则将备选点加入到路径 l_1 上, 返回 2b). 否则终止循环, 得到路径 l_1 , 设 l_1 上的点为 $i_1, i_2, \dots, i'_{N_j}$, 共 N'_j 个点;

S3: 设会员 X_1 分配的限额为 N'_1 , X_1 得到使 F_{ij} 最大的 N_1 个任务 i_1, i_2, \dots, i_{N_j} , 其中 $N_1 = \min\{N'_1, N''_1\}$. 将这 N_1 个任务从剩余任务中删除;

S4: 重复以上操作, 给第二个会员分配任务, 直至任务分配完成为止.

4.4 模型的评估与反馈

为了评估这个算法, 我们需要引入一些变量. 定义 p_1, p_2, \dots, p_N 的取值集合为 $\{Q_1, Q_2, \dots, Q_r\}$, 其中 $Q_1 < Q_2 < \dots < Q_r$. 定义 C_i 为选择了价格为 Q_i 的任务点的所有会员的完成率, 可用公式定义为

$$C_i = k_j \frac{\sum_{k=1}^{\lambda_i} \eta_{ik}}{\lambda_i} \times 100\%, \quad (5)$$

其中 λ_i 为选择价格 Q_i 的会员点数量, $i_1, i_2, \dots, i_{\lambda_i}$ 即为选择 Q_i 的会员编号, η_{ik} 表示会员 i_k 的完成概率, 其值由会员 i_k 的信誉 m_{ik} 确定. 我们用如下公式定义

$$\eta_i = \frac{\varphi(m_i)}{\max_i \varphi(m_i)}. \quad (6)$$

对于每一个定价方案 (p_1, p_2, \dots, p_n) , 我们的算法能够给出一个对各个价格取值的任务完成率的组合 (C_1, C_2, \dots, C_r) , 这相当于一个从所有定价方案组成的集合 P 到所有价格取值的任务完成率组合组成的集合 C 的一个映射, 即

$$f: P \rightarrow C, \quad (p_1, p_2, \dots, p_n) \rightarrow (C_1, C_2, \dots, C_r). \quad (7)$$

定义 n_i 为价格为 Q_i 的任务的数量, 这里 $i = 1, 2, \dots, r$. 自然我们可以用每个价格取值完成率的加权平均定义总完成率 C , 即

$$C = \frac{\sum_{k=1}^r n_i C_i}{\sum_{k=1}^r n_i}. \quad (8)$$

我们设商家的总报酬为常数 S . 通过以上的定义, 我们可以求出会员的总盈利 S_1 和平台的总收益 S_2 , 这里

$$S_1 = \sum_{k=1}^r Q_i n_i C_i, \quad S_2 = S - \sum_{k=1}^r Q_i n_i. \quad (9)$$

下面阐述我们的评估机制. 我们的评价指标包含以下四个方面:

- 1) 总完成率 C . 总完成率越高, 对应着会员的收益越多, 同时, 运营商也能更好地完成商家的任务, 形成“双赢”的局面. 反之, 造成的就是“双亏”;
- 2) $C-Q$ 关系图. 对于每一种价格, 任务的完成率都不应该过低, 这一指标可以用 $C-Q$ 关系图的稳定性来判断, 如图1所示;

案例中价格的 $C-Q$ 关系图

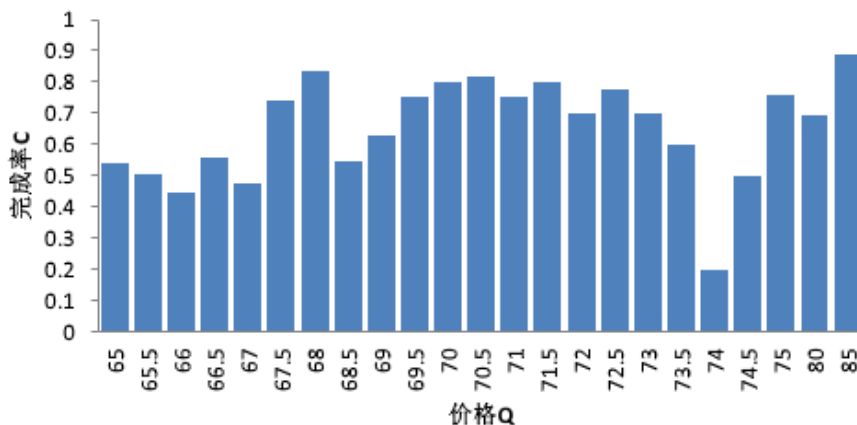


图1: 一个已完成案例中价格的 $C-Q$ 关系图

- 3) 会员的总盈利 S_1 和平台的总收益 S_2 . 显然, 这是非常重要的. 但是, 根据上述公式, 在任务点价格确定的基础上, S_1 和 S_2 的关系主要取决于任务的完成率 C_i ;

- 4) 模拟验证新模型优于原模型. 验证方法如下: 设原来的定价方案为 (p_1, p_2, \dots, p_n) , 且

$$f(p_1, p_2, \dots, p_n) = (C_1, C_2, \dots, C_r).$$

构造新的一组定价方案 $(p'_1, p'_2, \dots, p'_n)$, 满足 $p'_i < p_i$, $i = 1, 2, \dots, n$. 设

$$f(p'_1, p'_2, \dots, p'_n) = (C'_1, C'_2, \dots, C'_r),$$

模拟之后有 $C'_i < C_i$, $i = 1, 2, \dots, r$, 则可以说明新模型优于原模型.

我们还引入了一些反馈机制, 主要包括会员的评价机制和任务的评价机制. 会员评价机制方面, 需要考虑会员的信誉、预定的开始时间、预定任务的变化; 任务评价机制则包括商家与运营方的议价, 即考虑价格的增减.

5 任务打包发布的模型建立

5.1 想法的来源

将一些任务联合在一起发布, 相当于经济学中的“搭售”. 多个集中任务进行搭售, 有利于提升完成率, 降低平台成本, 双方共赢.

对于任务点 i , 我们考虑其周围的一个 δ 邻域 $U_\delta(i)$ (为了考虑这个邻域, 我们定义度量 $d(i, j)$ 就表示第 i 个任务点和第 j 个任务点之间的距离). 可以把这个邻域内的任务放在一起搭售, 定价为 P_i .

5.2 基于综合分析模型的搭售

和之前的模型中类似地, 我们可以定义

$$F_{U_\delta(i)j} = \frac{k_{ij}P_i\varphi(m_j)}{\max_{k \in U_\delta(i)}\{r_{kj}\}^\alpha}. \quad (10)$$

取 δ 很小, 那么这个小邻域的困难程度基本相同, 我们就可以近似用 k_{ij} 来计算. 而所需要走的距离可近似认为是小邻域内最远的点的距离, 即 $\max_{k \in U_\delta(i)}\{r_{kj}\}$.

5.3 基于单位长度期望模型的搭售

在这里, 我们用另一种方式定义 $F_{U_\delta(i)j}$. 令

$$F_{U_\delta(i)j} = \frac{k_{ij}P_i}{r_{ij} + \epsilon_i}, \quad (11)$$

其中 k_{ij} , P_i , r_{ij} 的定义和前文中提到的完全相同, ϵ_i 为走遍 $U_\delta(i)$ 中所有的任务点所需的最短路径长度. 关于 ϵ_i 的求解, 可以使用哈密顿图算法.

5.4 搭售过程中定价的估计

我们很容易得到 P_i 的取值范围. 我们知道搭售存在着一个价格关系, 即搭售后的价格大于每一个元素单独的价格, 而小于每个元素价格的总和. 也就是说

$$p_k < P_i < \sum_{k \in U_\delta(i)} p_k, \forall k \in U_\delta(i). \quad (12)$$

要使搭售更加受会员推崇, 那就是要求搭售对会员的吸引力因子比单个的吸引力因子要高, 即

$$F_{U_\delta(i)j} > F_{ij} \quad (13)$$

恒成立. 其中 $F_{U_\delta(i)j}$ 为搭售后在 $U_\delta(i)$ 邻域内的点对会员 j 的吸引力因子. 由于这种情况中估算定价是不依赖路径的, 我们采用之前综合分析模型里面的吸引力因子的定义式. 将定义式代入可得

$$\frac{k_{ij}P_i\varphi(m_j)}{\max_{k \in U_\delta(i)}\{r_{kj}\}^\alpha} > \frac{k_{ij}p_i\varphi(m_j)}{r_{ij}^\alpha}. \quad (14)$$

由于 $k \in U_\delta(i)$, 由三角不等式可得

$$r_{kj} \leq r_{ki} + r_{ij} < \delta + r_{ij}. \quad (15)$$

因此

$$\frac{P_i}{(r_{ij} + \delta)^\alpha} > \frac{p_i}{r_{ij}^\alpha}, \quad (16)$$

即

$$P_i > p_i \frac{(r_{ij} + \delta)^\alpha}{r_{ij}^\alpha}. \quad (17)$$

综合可得

$$p_i \frac{(r_{ij} + \delta)^\alpha}{r_{ij}^\alpha} < P_i < \sum_{k \in U_\delta(i)} p_k. \quad (18)$$

5.5 算法模拟

运用聚类分析的方法, 我们先在平面上随机取出若干个点作为圆心, 然后考虑这些点的 δ 邻域. 通过框入那些 δ 邻域内的点, 可以将任务点整合成一个个小包, 把那些邻域中的任务点当成一个任务点处理. 然后再利用综合分析模型中的算法进行任务的分配, 这样不仅更加有利于任务的分配, 且有利于整体完成率和平台利润率的提升.

我们对这种方案用综合分析模型中的算法进行了模拟, 但是通过模拟后对总完成率 C 的计算, 我们发现, 任务的联合发布对总完成率几乎没有影响. 但是, 这种影响会体现在 S_1, S_2 上, 由于联合发布任务的价格会低于单独发布每个任务的价格之和, 因此, 公司方面的成本降低, 盈利增加, 而会员拿到的总收入会减少.

参考文献:

- [1] 赵兴龙. 基于 K-means-遗传算法的众包配送网络优化研究[D]. 北京: 北京交通大学, 2016
Zhao X L. Network optimization of crowdsourcing distribution based on K-means-genetic algorithm[D]. Beijing: Beijing Jiaotong University, 2016
- [2] 孙信昕. 众包环境下的任务分配技术研究[D]. 扬州: 扬州大学, 2016
Sun X X. Research on task allocation technology in crowdsourcing[D]. Yangzhou: Yangzhou University, 2016
- [3] 韩雅雯. K-means 聚类算法的改进及其在信息检索系统中的应用[D]. 云南: 云南大学, 2016
Han Y W. Improvement of K-means clustering algorithm and its application in information retrieval system[D]. Yunnan: Yunnan University, 2016

Task-distribution and Price-evaluation Models in “Money-making Via Shots”

TANG Yin, GAO Yun-cong, CHEN Can-yu

Advisor:

(Undergraduate, University of Chinese Academy of Sciences, Beijing 100049)

Abstract: “Money-making via shots” is a self-help service mode based on mobile internet, where assigning tasks is crucial to successful transaction. The present study designed two distribution systems for each pricing method and evaluated its effectiveness. Firstly, we successfully introduced an attraction factor, which describes the degree of attractiveness in two ways. We also provided two different distribution systems of tasks. Secondly, we introduced an assessing rubric of these systems, using completion rate, users’ total profit and the platform’s total revenue as evaluation indexes. The case of packaging tasks was also considered. By clustering task points, tasks in a category could be publicized together. By referring to the tie-in sale model in economics, the bound of pricing in packing process could be estimated. The attraction factor in our systems could also be generalized to the case of packaging tasks. By simulation of the systems, packaging tasks had little impact on completion rate, but it could affect users’ total profit and the platform’s total revenue.

Keywords: crowdsourcing; tie-in sale; cluster analysis; simulation