

doi: 10.3969/j.issn.1673-2340.2019.02.012

引文格式: 冯俊丰, 温晴岚, 胡丹青, 等. 基于互联网悬赏制众包模式的定价策略[J]. 南通大学学报(自然科学版), 2019, 18(2): 82-86.

## 基于互联网悬赏制众包模式的定价策略

冯俊丰, 温晴岚, 胡丹青, 施林丽, 金晶亮\*

(南通大学 理学院, 江苏 南通 226019)

**摘要:**互联网众包模式的出现提高了企业搜集信息的效率。基于需求导向定价理论和万有引力公式,选择任务的原始定价、任务与会员的距离以及每位会员的限购额作为企业制定平台任务价格的影响因素,研究悬赏制众包模式下任务定价的相关问题。首先,通过数据预处理剔除异常值,使用聚类分析将会员信息分成23类;其次,建立任务完成情况 Logistic 回归模型,发现平台任务完成情况与任务所处的纬度位置以及任务的原始定价有关;接着,选取每类数据的平均值作为研究对象,借助万有引力公式计算平台任务与会员之间的弹性因子,建立了任务定价模型;最后,通过模型计算得出当任务完成率不变时,企业所需的总成本要小于原始成本。

**关键词:**互联网悬赏制;众包模式;导向定价理论;聚类分析;Logistic 回归模型;万有引力公式

中图分类号:O221

文献标志码:A

文章编号:1673-2340(2019)02-0082-05

## Pricing Strategy Based on Crowdsourcing Model of Internet Reward System

FENG Junfeng, WENG Qinglan, HU Danqing, SHI Linli, JIN Jingliang\*

(School of Sciences, Nantong University, Nantong 226019, China)

**Abstract:** The emergence of the Internet crowdsourcing model has increased the efficiency of information-collecting. Based on the demand-oriented pricing theory and the universal gravitation formula, with the original price, the distance between the task and the member and the limit of each member's purchase taken as the influencing factors of the company's platform price, this paper aimed to study the problems of task pricing under the crowdsourcing mode. Firstly, the outliers were eliminated by data preprocessing, and the member information was divided into 23 categories by cluster analysis. Secondly, the Logistic regression model of task completion was established. It is found that the completion of the platform task is related to the latitude position of the task and the original pricing of the task. The average value of each type of data was used as the object. The gravity factor was used to determine the elasticity factor between the platform task and the member, and the task pricing model was established. The final model presents the total cost required by the enterprise when the task completion rate is constant, which is less than the original cost.

**Key words:** Internet reward system; crowdsourcing model; pricing strategy; cluster analysis; Logistic regression model; universal gravitation formula

收稿日期:2019-03-22

基金项目:国家自然科学基金项目(71603135);国家级大学生创新训练计划项目(201910304006Z)

第一作者简介:冯俊丰(1995—),男,硕士研究生。

\*通信联系人:金晶亮(1982—),男,副教授,博士,主要研究方向为能源系统规划。E-mail: chinajjl@ntu.edu.cn

Jeff Howe 在 2006 年首次提出“众包”这一概念, 他将众包定义为: 企业借助互联网, 将过去由内部员工完成的任务, 在众包平台上以自由自愿的形式, 交付给特定的大众群体完成<sup>[1]</sup>。悬赏制众包模式是众包模式中的一种, 其主要流程分为两个阶段: 一是任务发布阶段, 发包方在平台上填写具体任务需求, 再将任务赏金托管到平台上, 由平台进行发布, 等待用户领取任务; 二是任务反馈阶段, 当接包方领取任务后提交解决方案, 由发包方决定是否采纳方案, 同时通知平台将赏金发放给被采纳方案的接包方。该众包交易模式有如下特点: 平台全程托管全额赏金。当发包方企业公布任务需求后, 需要在众包平台上公布赏金, 并等待接包方领取任务。

目前, 国内外学者通过研究众包的定义、分类方式以及结构组合等多个方面, 丰富和发展了众包理论。在众包理论中, 制定合理的任务价格是企业所要考虑的重要问题之一。定价是否合理关系到任务能否准时完成, 部分学者在研究众包过程中提出不同的众包模型。Lerner 等人基于对程序开发类创意竞赛的研究, 发现在开放式的环境下, 相似任务的出价越高, 则该类型任务对用户的吸引力越低<sup>[2]</sup>。Dipalantino 等人基于威客处在风险中性的情况下, 得到任务发布者所收到任务解决方案的数量随着任务定价的增加呈对数增长, 同时指出任务发布者在考虑任务定价时, 需要设置所获得解决方案数的区间值<sup>[3]</sup>。孙茜等人采用多元回归最小二乘法和二元 Logistic 回归方法, 将网络众包创新平台环境分为 5 个维度进行分析, 研究表明: 众包平台的信任环境和群众的交流沟通行为对群众接受任务具有正面影响, 能够提升任务完成率<sup>[4]</sup>。刘晓钢基于服务定价理论以及在线逆向拍卖理论等相关理论的基础上, 通过实际数据分析任务属性以及市场竞争状况对任务发布者出价行为的影响, 得出任务属性以及市场竞争状况的差异将会导致任务发布者采取不同的出价策略<sup>[5]</sup>。

可以看出, 目前对任务定价的影响因素研究较多, 对定价模型建立的研究比较少。本文基于以往学者对任务定价的影响因素分析, 建立悬赏

制众包模式定价策略模型。首先, 通过选取广东省某众包平台的相关数据, 分析任务顺利完成的主要原因; 其次, 使用聚类分析对会员信息数据进行分类, 并通过计算均值, 得到会员与任务的中心位置; 接着, 建立会员限购额模型得到每个会员领取任务的限购额; 同时, 在需求导向定价理论的基础上引入万有引力定律, 建立任务与会员之间的弹性因子模型; 最后, 建立悬赏制众包模式定价策略模型, 通过计算获得的定价策略可以实现定价成本与任务完成率的和谐统一。

## 1 任务定价模型建立与求解

### 1.1 数据处理

原始数据包含的数据因子包括任务与会员的序号、任务经纬度、会员位置、限购额与信誉值、任务的原始定价以及任务的完成情况, 其中 0 表示任务没有完成, 1 表示任务已经完成<sup>[5]</sup>。根据任务的经纬度画出任务分布的散点图, 发现原始任务数据中存在异常值。根据箱线图法<sup>[6]</sup>剔除异常值后, 新的任务分布散点图如图 1 所示。图 1 中圆点表示任务已经完成, 三角形表示任务没有完成。

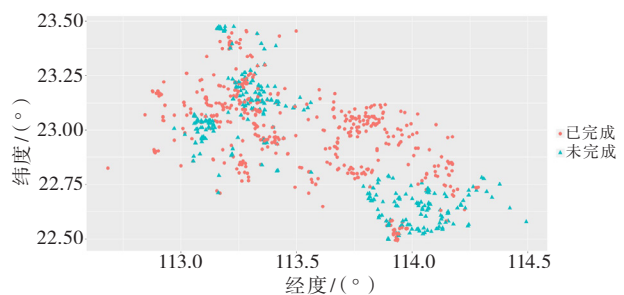


图 1 任务分布图

Fig.1 Task distribution

从图 1 中可以发现: 左上角地区任务大部分没有完成, 中间一块任务基本完成, 中间至右下角处任务完成度不高。结合任务经纬度数据及百度地图信息, 发现这些任务主要集中在广州、深圳、佛山及东莞等地。

表 1 为任务的相关信息。从表 1 中可以看出: 一共有 23 种价格类别; 价格定在 65.5, 66.0, 70.0 元的任务总量较多, 其中 65.5 元的任务多达 150 件, 但是它的任务完成率只有 51%; 除定价在 74.0, 74.5 元的外, 其他的任务完成率基本超过 45%。

表1 任务相关信息  
Tab.1 Task related information

任务价格/ 元	任务总量/ 个	任务完成量/ 个	任务完成率
65.0	65	35	0.54
65.5	150	76	0.51
66.0	103	46	0.45
66.5	63	35	0.56
67.0	38	18	0.47
67.5	23	17	0.74
68.0	30	25	0.83
68.5	11	6	0.55
69.0	19	12	0.63
69.5	8	6	0.75
70.0	84	65	0.77
70.5	11	9	0.82
71.0	4	3	0.75
71.5	5	4	0.80
72.0	60	42	0.70
72.5	9	7	0.78
73.0	10	7	0.70
73.5	5	3	0.60
74.0	5	1	0.20
74.5	2	1	0.20
75.0	72	53	0.74
80.0	13	9	0.69
85.0	27	24	0.89

### 1.2 基于 Logistic 回归的任务完成情况分析

为进一步探究任务完成情况与任务定价、任务所处经纬度的关系, 本文根据数据的性质, 以任务完成情况为因变量, 任务定价 $X_1$ 、任务所处经度 $X_2$ 、纬度 $X_3$ 为自变量, 建立 Logistic 回归模型<sup>[7]</sup>。模型结果见表 2。

表2 Logistic 回归系数表  
Tab.2 Logistic regression coefficient

变量	系数	标准差	p 值
Intercept	-92.531 1	39.186 9	0.018 2*
$X_1$	1.840 1	0.457 9	$5.85 \times 10^{-5}$ ***
$X_2$	0.385 6	0.274 3	0.159 8
$X_3$	0.101 5	0.019 9	$3.24 \times 10^{-7}$ ***

注: \* 表示比较显著; \*\*\* 表示非常显著。

从回归系数中可以看出: 在 0.05 的显著性水平下, 任务所处的经度对任务完成情况的影响并不显著,

无法拒绝参数为 0 的假设<sup>[8]</sup>。去除任务经度重新拟合, 模型结果见表 3。

表3 改进后模型系数表  
Tab.3 Improved model coefficient

变量	系数	标准差	p 值
Intercept	-38.573 5	7.661 8	$4.79 \times 10^{-7}$ ***
$X_1$	1.398 0	0.330 3	$2.30 \times 10^{-5}$ ***
$X_3$	0.101 2	0.019 8	$3.41 \times 10^{-7}$ ***

注: \*\*\* 表示非常显著。

新模型的每个回归系数均显著, 另外, 在 0.05 的显著性水平下, 通过卡方检验比较这两个模型, 结果显示两个模型的差异不显著( $p=0.159 6 > 0.05$ )。这表明两个预测变量的新模型和 3 个完整预测变量的模型拟合程度一样好。系数指数化后, 纬度项系数为 4.047, 价格项系数为 1.106 5。可以认为: 在保证任务价格不变的情况下, 每当任务所处的纬度增加  $1^\circ$  时, 任务完成情况的优势将乘以  $e^{1.398} = 4.047$ , 也就是考虑任务纬度因素的任务完成量是不考虑任务纬度因素的 4.047 倍。在保证任务纬度不变的情况下, 每当任务的价格增加 1 元时, 任务完成情况的优势将乘以  $e^{0.101 2} = 1.106 5$ 。因此, 随着任务位置纬度值和任务价格的增加, 任务完成情况优势将上升。Logistic 回归模型如下:

$$p = p(y=1|X_1, X_3) =$$

$$\frac{\exp(-38.57 + 1.397 \times X_1 + 0.101 \times X_3)}{1 + \exp(-38.57 + 1.397 \times X_1 + 0.101 \times X_3)},$$

$$p(y=0|X_1, X_3) = 1 - p(y=1|X_1, X_3)。 \quad (1)$$

根据表 3 的数据分析, 本文认为影响任务完成的因素为任务定价和任务的纬度, 其中任务定价越高、任务位置纬度越大, 任务完成率越高。

### 1.3 任务定价模型建立

首先, 引入需求导向定价理论, 这是本文模型的理论基础。在原始的需求导向定价理论<sup>[9]</sup>中, 商品定价 $P_{cost}$ 与原始定价 $X_1$ 成正比, 与商品弹性因子 $E_d$ 的绝对值成反比, 具体公式为

$$P_{cost} = \frac{1}{2} \times \left( 3 - \frac{1}{|E_d|} \right) \times X_1。 \quad (2)$$

接着, 通过如下 4 步建立任务定价模型:

第 1 步, 对数据进行聚类处理<sup>[10]</sup>。因为在数据中存在部分会员与会员之间距离很短, 距离可

以忽略不计, 所以, 可以用他们的平均值来代替这些会员的信息。根据表 1 对价格的分析, 任务共有 23 种原始定价, 所以, 本文把会员信息也分成 23 类, 并提取每一类的平均值。数据分类后, 平台上任务总量为 23 件, 任务定价总成本为 1 635 元。

第 2 步, 以分类处理后的数据为基础, 用会员任务限购额  $T$  作为因变量, 会员纬度  $Z_1$ 、会员经度  $Z_2$  及会员信誉值  $Z_3$  为自变量, 建立多目标会员预定限额线性回归模型<sup>[11]</sup>。由于会员数据单位不一致, 因此, 需要对数据进行标准化处理<sup>[12]</sup>。会员任务限购额模型为

$$T = 3.913 + 1.055Z_1 + 0.242Z_2 + 0.536Z_3, \quad (3)$$

模型拟合优度为 0.759, 表明模型可以解释 75.9% 的数据方差, 拟合程度很好。

第 3 步, 引入万有引力公式作为需求导向定价模型中的价格弹性因子模型。在现实自然中, 任意两个物体之间都是存在吸引力的, 而引力的大小跟这两个物体的质量乘积成正比, 跟两者距离的平方成反比。这是万有引力定律的内容<sup>[13]</sup>, 而在本文中所研究的任务与会员之间也是如此。任务与会员距离越远, 这两者吸引力越低; 任务的定价越高, 任务对会员的吸引力越强; 会员自身任务限购额越大, 会员领取的任务就越多, 任务对会员的吸引力也越高。模仿万有引力公式, 建立任务定价弹性因子模型, 结果如式 (4) 所示。

$$F = \frac{X_1 \times T}{r^2}, \quad (4)$$

其中:  $F$  表示任务对会员的吸引力;  $X_1$  表示任务原始定价;  $r$  表示任务与会员的距离;  $T$  表示会员自身的任务限购额。需要注意的是, 根据万有引力定律以及现实因素, 当不同的任务具有相同的定价时, 会员会根据距离的大小选择任务。

第 4 步, 建立任务定价多目标规划模型。模型的目标为任务定价成本及任务完成率。为了求解方便, 本文使用单纯形法, 将多目标规划模型转化为单目标规划模型<sup>[14-15]</sup>, 即用任务完成率代替任务定价成本。记  $t_i$  为第  $i$  件任务完成情况:

$$t_i = \begin{cases} 1, & \text{第 } i \text{ 件任务已完成} \\ 0, & \text{第 } i \text{ 件任务未完成} \end{cases} \quad i = 1, 2, \dots, 23. \quad (5)$$

根据前文分析及需求导向定价理论, 建立任务定价目标规划模型。模型为

$$\begin{aligned} \max P_{\text{total}} &= \sum_{i=1}^{23} P_{\text{cost}_i} t_i = \\ & \sum_{i=1}^{23} \left[ \frac{1}{2} \times \left( 3 - \frac{r^2}{|X_1 \times T|} \right) \times X_1 \right] t_i = \sum_{i=1}^{23} \left[ \frac{1}{2} \times \left( 3 - \frac{r^2}{|X_1 \times (3.913 + 1.055Z_1 + 0.242Z_2 + 0.536Z_3)|} \right) \times X_1 \right] t_i \\ \text{s.t. } & t_i(1 - t_i) = 0 \quad i = 1, 2, \dots, 23, \\ & 0 \leq \sum_{i=0}^{23} t_i \leq 23, \end{aligned} \quad (6)$$

其中  $P_{\text{cost}_i}$  为第  $i$  个任务的价格。最终, 将数据代入模型中, 得到的结果是在任务完成率不变的情况下, 企业发布任务定价总成本为 1 530.05 元, 比原来成本降低 105.05 元。任务定价如表 4 所示。

表 4 任务定价情况

Tab.4 Task pricing

任务ID	原定价/元	新定价/元	定价差/元
1	65.0	84.12	19.12
2	65.5	83.16	17.66
3	66.0	25.76	-40.24
4	66.5	26.31	-40.19
5	67.0	81.28	14.28
6	67.5	56.19	-11.31
7	68.0	79.12	11.12
8	68.5	97.85	29.35
9	69.0	22.51	-46.49
10	69.5	40.72	-28.78
11	70.0	16.45	-53.55
12	70.5	85.47	14.97
13	71.0	99.15	28.15
14	71.5	85.14	13.64
15	72.0	93.20	21.20
16	72.5	61.08	-11.42
17	73.0	108.85	35.85
18	73.5	75.15	1.65
19	74.0	70.96	-3.04
20	74.5	48.86	-25.64
21	75.0	69.90	-5.10
22	85.0	106.34	26.34
23	80.0	12.48	-72.52

从表 4 可以发现: 依靠新的定价方案, 有 12 件任务的价格得到提升, 占总任务数量的 52.17% ;

有 11 件任务的价格得到降低, 占总任务数量的 47.82%; 价格变化幅度最大的是 23 号任务, 任务定价降低 72.52 元; 变化幅度最小的是 18 号任务, 任务定价提高了 1.65 元。在任务完成率保持不变时, 按照原始方案, 任务定价总成本为 1 635 元, 而在新方案下, 定价总成本为 1 530.05 元, 比原来成本降低 105.05 元, 这表明新定价方案是好的。

## 2 结论

本文主要研究悬赏制众包模式的定价策略, 通过归纳总结以往学者的研究, 选择任务原始定价、任务与会员的距离及会员的任务限购额作为企业在制定任务定价时要考虑的影响因素。

在模型建立与求解中, 首先, 对搜集到的数据进行预处理, 剔除异常值; 通过建立 Logistic 回归模型, 分析得出任务完成的主要原因包括任务定价及任务所处的纬度位置, 但与任务所处的经度位置无关; 其次, 通过聚类分析, 将会员信息数据按照任务数据的数量聚成 23 类, 以降低计算量; 接着, 基于需求导向定价理论和万有引力公式, 计算任务与会员之间的吸引力并把它作为需求导向定价理论中的弹性因子; 最后, 建立新的任务定价方案。具体而言, 在该方案中, 根据悬赏制众包模式的性质, 企业在制定任务价格时除了要考虑以往任务的价格, 还需考虑任务与会员之间的距离以及每位会员自身的任务限购额。最终在新方案下, 任务总成本比原始方案降低了 105.05 元。

### 参考文献:

[1] 胡静思. 基于双边市场理论的众包平台定价策略研究[D]. 武汉: 华中师范大学, 2017.  
 [2] LERNER J, TIROLE J. Some simple economics of open source[J]. The Journal of Industrial Economics, 2003, 50(2): 197-234.  
 [3] DIPALANTINO D, VOJNOVIC M. Crowdsourcing and all-

pay auctions[C]//Proceedings of the tenth ACM conference on Electronic commerce - EC '09, Stanford, California, USA, July 6-10, 2009. New York: ACM Press, 2009.  
 [4] 孙茜, 刘海波, 杨绪勇, 等. 创新众包平台对接包方中标率的影响机制研究[J]. 科学学研究, 2016, 34(2): 279-287.  
 [5] 全国大学生数学建模竞赛[EB/OL]. (2017-09-14) [2018-10-18]. [http://www.mcm.edu.cn/html\\_cn/node/460baf68ab0ed0e1e557a0c79b1c4648.html](http://www.mcm.edu.cn/html_cn/node/460baf68ab0ed0e1e557a0c79b1c4648.html).  
 [6] 孙向东, 刘拥军, 陈雯雯, 等. 箱线图法在动物卫生数据异常值检验中的运用[J]. 中国动物检疫, 2010, 27(7): 66-68.  
 [7] 房钦钦, 赵为华. Logistic 分布位置-尺度参数联合回归建模及其 Score 检验[J]. 南通大学学报(自然科学版), 2017, 16(1): 81-86.  
 [8] KABACOFF R I. R 语言实战[M]. 2 版. 王小宁, 刘擷芯, 黄俊文, 译. 北京: 人民邮电出版社, 2016: 284-289.  
 [9] 刘晓钢. 众包中任务发布者出价行为的影响因素研究[D]. 重庆: 重庆大学, 2012.  
 [10] 王锦, 郝晋珉, 陈爱琪, 等. 基于模糊聚类分析的天津市生态环境承载力评价[J]. 中国农业大学学报, 2018, 23(10): 112-121.  
 [11] 韩丽娜, 李红, 韩改宁. 基于多元线性回归和多目标规划模型的决策问题研究[J]. 自动化与仪器仪表, 2016(9): 177-178.  
 [12] 顾安朋, 徐国智, 林潮彬, 等. 营销客户地址数据标准化应用分析与研究[J]. 科技与创新, 2018(16): 142-144.  
 [13] 李艳琼, 李冬冬, 王喆, 等. 万有引力近邻的多视角分类学习[J/OL]. 计算机工程与应用[2018-11-29]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20181101.0950.019.html>.  
 [14] 周轩伟. 较多约束多目标规划的最优性条件[J]. 应用数学, 2016, 29(4): 902-909.  
 [15] 孟香惠, 施保昌, 胡新生. 线性规划单纯形法的动态灵敏度分析及其应用[J]. 应用数学, 2018, 31(3): 697-703.

(责任编辑: 张燕)