

基于影响域的新型众包定价算法模型构建

周春樵¹, 肖昌昊², 刘 扬³, 姚安琪², 黄君扬³

(1.上海理工大学计算中心, 上海 200093

2.上海理工大学能源与动力工程学院, 上海 200093;

3.上海理工大学光电信息与计算机工程学院, 上海 200093)

摘 要:“众包”已成为时下新兴的一种基于互联网进行信息检查和搜集的商业模式,其成功率取决于诸多因素的影响,其中最大的影响因素为任务发布者的出价。针对此问题,本文提出了一种基于“影响域”的新型众包定价策略,该策略以经济学中的供求关系模型为建模方法,利用任务与劳动者的地理位置分布规律动态定价,同时,对新数据与原始数据进行相似性分析,通过机器学习模拟任务的完成概率,从而评价定价策略的优劣。本文以“拍照赚钱”自助式服务模式作为研究样本;在利用影响域定价模型重新定价后,经济效用较原始方案增长80.65%,效果良好。

关键词:众包;影响域;供求关系;标准化欧氏距离;机器学习

中图分类号: TP301 **文献标识码:** A

Research on New Pricing Strategy Based on Domain-of-Influence

ZHOU Chunqiao¹,XIAO Changhao², LIU Yang³,YAO Anqi²,HUANG Junyang³

(1.Computer Center,University of Shanghai for Science & Technology,Shanghai 200093,China;

2.School of Energy and Power Engineering,University of Shanghai for Science and Technology,Shanghai 200093,China;

3.School of Optical-Electrical and Computer Engineering,University of Shanghai for Science and Technology,Shanghai 200093,China)

Abstract:Crowdsourcing has become a new business model nowadays.It not only makes human knowledge and wisdom improved and disseminated infinitely,but also creates amazing social wealth.However,the success rate of crowdsourcing missions depends on a number of factors,among which the most important one is the bid given by mission publishers.In this paper,a new pricing strategy based on domain-of-influence is proposed,which uses the geographical distribution of missions and the employees to price dynamically,then readjusts the size of the affected domain for an iterative calculation until the pricing result is stable.In addition,this paper establishes a mathematical model to simulate the probability of completion of a mission,which is used to test the merits of the pricing strategy based on domain-of-influence.This paper takes the self-service model of *Photographing for Money* as the study subject and the financial rewards have increased by 80.65% after repricing by means of domain-of-influence compared with the previous pricing method.

Keywords:crowdsourcing;domain-of-influence;supply-demand relationship;standardized Euclidean distance;machine learning

1 引言(Introduction)

2006年6月份的《连线》杂志中,记者Jeff Howe^[1]在《众包的崛起》一文中首次提出了“众包”的概念。众包,指公司或机构把工作任务通过网络外包给非特定的大众,是“网络大众”与“外包”的合成词汇。它弥补了公司或机构自身的资源缺陷问题,提高了企业的工作效率,同时给完成任务的网络大众带来一定的经济收入,可谓双赢的经营模式。时

下,已有学者对众包做出了详尽的经济学解析^[2-4],也对任务定价的影响因素做了具体研究^[5]。众包任务的定价策略是该模式中举足轻重的部分——过低的定价可能会导致任务无法顺利完成;过高的定价则为企业带来沉重的负担。本文提出的基于影响域的众包定价策略,通过在每个任务点的周围划定了“影响域”,统计该任务点附近的任务密度和劳动力密度,根据供求关系模型^[6,7]确定该任务点的定价。其后,计算

样本数据之间的标准化欧氏距离^[8]，并以此作为机器学习^[9,10]的训练集，模拟新定价下的任务完成概率，计算经济效用，从而检验该定价模型的优劣。本文在建立“拍照赚钱”自助式服务模式数学模型基础上，对所提供的数据进行模型验证，收效良好。

2 影响域定价策略(Pricing strategy of domain-of-influence)

2.1 影响域

“影响域”为每个任务点周围划定的一个圆形区域，如图1所示。假设该地区共有 n 个任务，则每个任务都对应一个不同的定价 p_i 和影响域半径 r_i ($i=1,2,3,\dots,n$)。

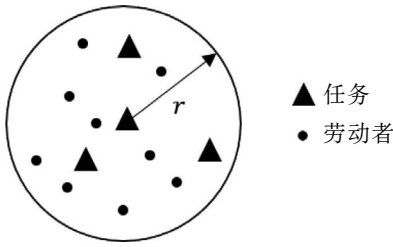


图1 影响域概念示意图

Fig.1 Schematic diagram of domain-of-influence

定价更高的任务会对距离更远的劳动者产生吸引力，因此任务的影响域半径 r_i 应与其定价 p_i 成正相关。规定 d_i 和 d_e 分别为包含其他任务点、劳动者到该任务点地理距离的数值矩阵，并规定影响域半径 r_i 满足：

$$d_i = \max(\min(d_t), \min(d_e)) \quad (1)$$

$$\bar{d} = \frac{1}{n} \sum_{i=0}^n d_i \quad (2)$$

$$\bar{p} = \frac{1}{n} \sum_{i=0}^n p_i \quad (3)$$

$$r_i = \bar{d} \left(1 + \frac{p_i}{\bar{p}} \right) \quad (4)$$

2.2 定价原理及方案

图1圆形区域内的三角形数量反映了“任务密度”，圆点数量则反映了“劳动力密度”。在忽略地区经济、交通等其他影响因素的情况下，供求关系决定了任务点的合理定价。由实际情况可知：(1)影响域内的各任务之间存在竞争关系。在影响域内的劳动力数量不变的情况下，市场的均衡任务价格会随任务数量的增加而上升；反之，则会因任务数量的减少而下降。(2)影响域内的劳动力之间亦存在竞争关系，在影响域内任务数量不变的情况下，市场的均衡任务价格随着劳动力数量的增加而上涨；反之，则会因劳动力数量的减少而下降。

因而，本文定义任务密度 ρ_t ，类比市场需求(demand)；

定义劳动力密度 ρ_e ，类比于市场供给(supply)。具体的市场运行机制可分别参考经济模型：需求变动对均衡的影响、供给变动对均衡的影响。

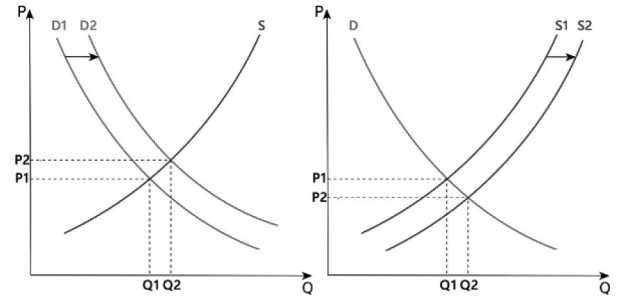


图2 供求曲线

Fig.2 Supply and demand curve

继而引入参与定价的重要参数 β ，表示单位劳动力可领取/完成的任务数量，满足：

$$\beta = \frac{p_i}{\rho_e} \quad (5)$$

$$p \propto \beta \quad (6)$$

研究区域内每个任务都对应一个不同的 β_i ($i=1,2,3,\dots,n$)。在该地区计划投入总额 p_{sum} 恒定不变的前提下，基于分配的思想为每个任务定价，配比则由每个任务对应的 β_i 决定：

$$\beta'_i = \beta_i / \sum_{i=0}^n \beta_i \quad (7)$$

$$p'_i = \beta'_i p_{sum} = \beta'_i \sum_{i=0}^n p_i \quad (8)$$

经模型分析，上述分配方案无法保障劳动者完成任务的基本回报，计算出的任务定价极差较大。因此设定基础价格 p_{std} 对上述公式进行修正：

$$p_{std} = \alpha p_{sum} = \alpha \sum_{i=0}^n p_i \quad (9)$$

$$p'_i = \beta'_i (p_{sum} - p_{std}) + \frac{p_{std}}{n} \quad (10)$$

其中，基础价格比例参数 $\alpha \in [0,1]$ ，不同 α 值对应定价的经济效用可能会有所不同，最终可通过比较选取使经济效用达到最高的 α 值。

2.3 算法流程

影响域定价策略需要通过迭代计算得到稳定的定价。如果已有先前的实践数据，可以利用各个任务的原定价作为迭代初值；如果缺乏经验数据，则可以用计划投入总额除以任务数量得到的平均任务定价作为迭代初值。

利用初始定价确定影响域半径，继而划定每个任务对应的影响域，并统计任务密度和劳动力密度，确定参数 β' ，再根据 β' 向量为所有任务动态分配价格。重复上述步骤，直至所有任务的定价收敛至恒定值。算法流程见图3。

3 模型的验证方案(Model verification scheme)

3.1 模型经济效用

本文定义了经济效用 U 来评价定价方案的优劣:

$$U = \frac{\bar{f}}{p_{\text{sum}}} = \bar{f} / \sum_{i=0}^n p_i \quad (11)$$

其中, $\bar{f}$ 为该地区任务的平均完成率。

当企业对外包任务的投入金额越少且回报越高时, 经济效用越高, 定价方案更优。如果企业已有其他定价策略, 可以通过比较两种策略的经济效用, 从而选择较优方案。

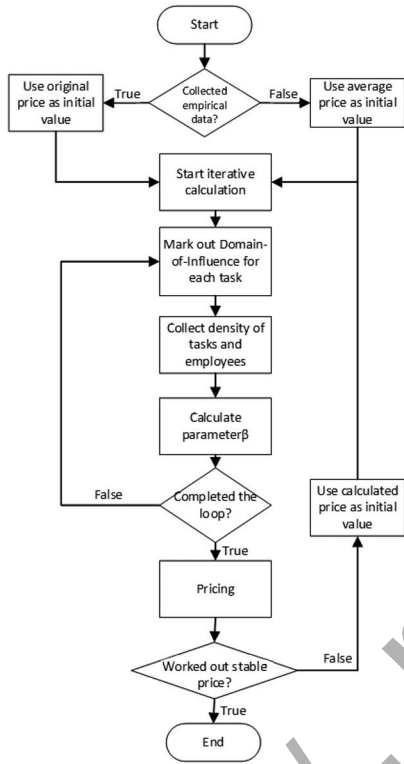


图3 定价算法流程图

Fig.3 Pricing algorithm flow chart

3.2 模拟完成率的原理及方案

上述公式中, 任务的平均完成率 $\bar{f}$ 尚为一个未知量。对于已经投入市场检验过的定价方案, 已完成的任务数量 t_f^0 和未完成的任务数量 t_u^0 都是已知的, 此时:

$$\bar{f} = \bar{f}^0 = \frac{t_f^0}{t_f^0 + t_u^0} \quad (12)$$

而对于利用影响域定价策略计算出的任务价格, 由于尚未投入市场实际检验, 因此 t_f 与 t_u 未知, 需要对任务完成率进行模拟。

首先, 调查已经过市场检验的定价方案, 把 q 个任务完成情况 f_j^0 ($j=1,2,3,\dots,q$) 二值化后存储在向量 F^0 中, 对应任务的定价 p_j^0 ($j=1,2,3,\dots,q$) 存储在向量 P^0 中, 其他可能会对任务完成率造成影响的 m 个参数 $x_{i,k}^0$ ($j=1,2,3,\dots,q; k=1,2,3,\dots,m$) 存储在 $q \times m$

维矩阵 X^0 中(若无其他影响因素或影响因素无法量化, m 亦可取0)。其中, 任务完成情况的二值化方法为:

$$f_j^0 = \begin{cases} 1, & \text{finished} \\ 0, & \text{unfinished} \end{cases} \quad (13)$$

同理, 将新方案下的 n 个任务的价格 p_i 和其他可能对任务完成率造成影响的参数 $x_{i,k}$ ($j=1,2,3,\dots,n; k=1,2,3,\dots,m$) 分别存储在向量 P 和矩阵 X 中。

其次, 计算新方案与原方案每个任务之间的相似度。因为标准化欧式距离可以去除各维度数据的量纲和不同维度之间数据的关联性, 因此采用标准化欧式距离 $D_{i,j}$ 来衡量相似度水平:

$$D_{i,j} = \sqrt{\left(\frac{p_i - p_j^0}{s_p}\right)^2 + \sum_{k=1}^m \left(\frac{x_{i,k} - x_{j,k}^0}{s_k}\right)^2} \quad (14)$$

其中, s_p 为向量 P 中所有元素和 p_i 的标准差, s_k 为矩阵 X^0 第 k 列所有元素和 $x_{i,k}$ 的标准差。如果没有其他影响因素或影响因素无法量化, 即 $m=0$ 时, 上述公式可以简化为:

$$D_{i,j} = \frac{p_i - p_j^0}{s_p} \quad (15)$$

由于标准化欧式距离 $D_{i,j}$ 与任务相似程度成负相关, 与常规逻辑相反, 故定义任务相似度 $S_{i,j}$ 满足:

$$S_{i,j} = D_{i,j}^{-\omega_1} e^{-\omega_2 D_{i,j}^{\omega_3}} \quad (16)$$

其中, ω_1 、 ω_2 和 ω_3 为自然数且不同时为0, 可使得 $S_{i,j}$ 收敛。因为:

- ① $D_{i,j}^{-\omega_1} > 0$, 且 $\lim_{\omega_1 \rightarrow \infty} D_{i,j}^{-\omega_1}$ 收敛;
- ② $e^{-\omega_2 D_{i,j}^{\omega_3}} > 0$, 且 $\lim_{\omega_2 \rightarrow \infty} e^{-\omega_2 D_{i,j}^{\omega_3}}$ 收敛。

由收敛级数的性质可知, 两个绝对收敛的正项级数之积仍然绝对收敛, 易证式(16)成立。

最后, 对相似度 $S_{i,j}$ 归一化处理得到 $S_{i,j}^*$, 并以之为权重, 对 f_j^0 各元素加权求和得到新方案下各任务的预期完成概率 f_i , 进而求出预期的任务平均完成率 $\bar{f}$:

$$S_{i,j}^* = S_{i,j} / \sum_{j=0}^q S_{i,j} \quad (17)$$

$$f_i = \sum_{j=0}^q S_{i,j}^* f_j^0 \quad (18)$$

$$\bar{f} = \frac{1}{n} \sum_{i=0}^n f_i \quad (19)$$

3.3 算法流程

在相似度 $S_{i,j}$ 的计算过程中, 参数 ω_1 、 ω_2 和 ω_3 的选取需要由算法循环实现。有关模拟完成率的完整算法流程如图4所示。

初始状态时, $\omega_1 = \omega_2 = 0$, $\omega_3 = 1$ 。将标准化欧式距离 $D_{i,j}$ 代入公式中计算当前参数下的任务近似度 $S_{i,j}$ 。不断调整 ω_1 、 ω_2 和

ω_3 的大小, 当 $S_{i,j}$ 趋于稳定值时跳出循环, 利用稳定的 $S_{i,j}$ 对向量 F^0 中各元素加权求和, 即可求得任务平均完成率 $\bar{f}$ 。

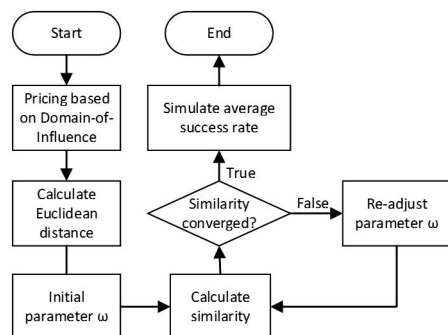


图4 模拟完成率算法流程图

Fig.4 Flow chart of simulation completion rate algorithm

4 模型的实践与分析(The practice and analysis of the model)

本文在建立了“拍照赚钱”自助式服务模式数学模型基础上, 选取了模型中的数据来支撑本文提出的定价策略。原始数据中包含了835个任务点的数据, 以及1878个劳动者的数据, 选取其中位于广东省深圳市南山区的161个任务数据, 对其利用影响域定价策略重新定价。不同基础价格参数 α 下新定价方案经济效益的增长率如图5所示。

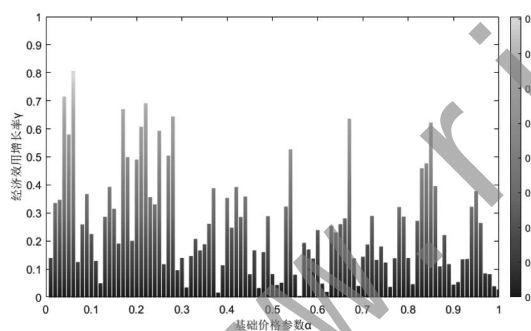


图5 $\alpha - \gamma$ 对应关系图

Fig.5 $\alpha - \gamma$ correspondence diagram

由图5可知, 如果选取合适的基础价格参数 α , 经济效益的增长率最大可以达到80.65%。这体现了影响域定价策略的良好效果。

5 结论(Conclusion)

影响域定价策略为线下众包任务的定价提供了新思路。它以市场供求关系作为研究基础, 将任务和劳动力的地理位置分布作为考虑因素, 并按照分配的思想为任务定价。其优点在于, 尽可能地使众包任务市场供求均衡状态, 因此任务发布者和任务执行者处于对等地位, 双方的关系相对和谐稳定。此外, 以经济效益作为定价方案优劣的评判标准, 充分

满足了任务发布者的需求; 设立基础价格系数 α , 切实提高了任务执行者的收益。其缺点在于, 目前缺乏大量的实践经验, 只能通过机器学习大致模拟出任务的完成概率, 很难具有说服力。未来将对影响域半径的选取做进一步调查研究, 完善影响域定价策略。

参考文献(References)

- [1] Steve Fleetwood.Do labour supply and demand curves exist[J]. Cambridge Journal Of Economics,2014,38(5):1087-1113.
- [2] Al-Roomi,M.Cloud computing pricing models:A survey[J].International Journal of Grid and Distributed Computing,2013,6(5):93-106.
- [3] Jeff Howe.Crowdsourcing:why the power of the crowd is driving the future of business[M].Beijing:CITIC Press,2009:6-9.
- [4] 李桂林,陈晓云.关于聚类分析中相似度的讨论[J].计算机工程与应用,2004,40(31):64-82.
- [5] 张利斌,钟复平,涂慧.众包问题研究综述[J].科技进步与对策,2012(6):154-160.
- [6] 高鸿业.西方经济学[M].北京:中国人民大学出版社,2011:20-24.
- [7] 刘瑞元.加权欧氏距离及其应用[J].数理统计与管理,2002(5):17-19.
- [8] 吕岩威,李平.一种加权主成分距离的聚类分析方法[J].统计研究,2016,33(11):102-108.
- [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] 冯剑红,李国良,冯建华.众包技术研究综述[J].计算机学报,2015(9):1713-1726.

作者简介:

周春樵(1988-), 男, 硕士, 工程师.研究领域: 协同计算, 计算机技术.

肖昌昊(1997-), 男, 本科生.研究领域: 动力机械工程, 计算机技术.

刘 扬(1996-), 男, 本科生.研究领域: 光学工程, 虚拟化技术.

姚安琪(1997-), 女, 本科生.研究领域: 制冷与空调工程, 计算机技术.

黄君扬(1996-), 男, 本科生.研究领域: 软件工程.