

Research on Task Pricing of Mobile Internet Self-Service Labor Crowdsourcing

Rui Feng¹, Xiaoying Fu², Rongxun Wang³, Jiaqi Zhao¹

¹Shandong University of Science and Technology, Qingdao Shandong

²College of Mathematics and System Science, Shandong University of Science and Technology, Qingdao Shandong

³College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao Shandong

Email: 964413522@qq.com

Received: Jun. 17th, 2018; accepted: Jul. 5th, 2018; published: Jul. 12th, 2018

Abstract

The mobile Internet self-service labor crowdsourcing platform provides enterprises with various business inspections and information collections, effectively guarantees the authenticity of survey data and shortens the investigation cycle. This paper considers the task pricing of labor crowdsourcing as a multi-objective optimization problem. The objective function is the task completion rate and survey cost. Therefore, according to the result of the optimization, a new task pricing plan is obtained. The result shows that the new scheme has a beneficial effect on the improvement of the completion rate. The data covered in this article come from the 2017 China Undergraduate Mathematical Contest in Modeling (CUMCM2017B).

Keywords

Person Correlation Analysis, Double Goal Optimization, Reverse Analysis, Task Packaging

移动互联网自助式劳务众包任务定价问题的研究

冯 睿¹, 付晓莹², 王荣勋³, 赵佳琪¹

¹山东科技大学, 山东 青岛

²山东科技大学, 数学与系统科学学院, 山东 青岛

³山东科技大学, 计算机科学与工程学院, 山东 青岛

Email: 964413522@qq.com

收稿日期: 2018年6月17日; 录用日期: 2018年7月5日; 发布日期: 2018年7月12日

摘 要

移动互联网自助式劳务众包平台, 为企业提供各种商业检查和信息搜集, 有效保证了调查数据的真实性,

文章引用: 冯睿, 付晓莹, 王荣勋, 赵佳琪. 移动互联网自助式劳务众包任务定价问题的研究[J]. 应用数学进展, 2018, 7(7): 766-775. DOI: 10.12677/aam.2018.77092

缩短了调查的周期。本文将劳务众包的任务定价问题考虑为一个多目标优化问题，目标函数是任务完成率和调查成本。从而，根据优化的结果得出了新的任务定价方案，结果表明新方案对完成率的提高产生了有益的效果。本文所涉及数据来自于2017年全国大学生数学建模竞赛B题(CUMCM2017B)。

关键词

Person相关性分析, 双目标优化, 逆向分析, 任务打包

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 前言

“拍照赚钱”是移动互联网下的一种自助式服务。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, 赚取 APP 对任务所标定的酬金。这种基于移动互联网的自助式劳务众包平台, 为企业提供各种商业检查和信息搜集, 相比传统的市场调查方式可以大大节省调查成本, 而且有效地保证了调查数据真实性, 缩短了调查的周期。因此, APP 成为该平台运行的核心, 而 APP 中的任务定价又是其核心要素。刘晓钢[1]认为任务的出价策略受到任务自身属性以及市场竞争性状况的影响。而在众包模式与交易机制匹配方面, 选择型众包可与悬赏制、招标制进行匹配, 整合型众包可与雇佣制、计件制进行匹配[2]。同时, 李勇军[3]认为人的性格会决定其做事的态度, 并且一个人已完成的任务信息也能为任务分配提供一定的参考依据, 所以将接包方的非技术能力(性格能力与在线信誉)与任务需求的相关性考虑在内。

会员的信誉状况、任务点附近的会员分布情况都在很大程度上增加了任务完成率的不确定性, 因此很难合理且快速地给出一套定价方案, 因此根据任务点和会员的分布, 给出完成率尽可能高、总成本尽可能低的定价方案非常具有挑战性。

问题的提出

本文获取的会员信息数据, 包含会员的位置、信誉值、参考其信誉给出的任务开始预订时间和预订限额。原则上会员信誉越高, 越优先开始挑选任务, 其配额也就越大; 获得的项目任务数据, 包含任务的位置信息。本文研究并尝试解决了以下问题:

问题一: 研究各项任务的定价规律, 分析任务未完成的原因;

问题二: 设计新的任务定价方案, 并与原定价方案进行比较, 达到优化原方案的目的;

问题三: 考虑将部分任务联合在一起打包发布, 进而修改问题二中给出的新的定价方案, 考察打包发布对目标函数的优化效果;

2. 问题分析

2.1. 问题一分析(如图 1 所示)

针对任务的定价规律, 考虑到若参考任务周围会员的信誉值进行定价, 会降低定价的效率、增加成本, 所以先不考虑信誉值这项数据信息。通过分析数据, 发现可能影响问题一中任务定价规律的因素有图示四种因素。进而, 处理数据, 考察这些因素与定价的相关性, 如果相关性显著, 则可认为该因素影响定价, 最终得到定价规律。

针对任务未完成的原因, 本文将会员分布及任务分布进行可视化处理, 获得任务与会员的分布图, 对任务未完成的原因进行了定性分析。

2.2. 问题二分析(如图 2 所示)

本文认为, 此定价方案的本质是一个双目标优化问题。目标函数是完成率和调查成本。众包平台的优点是能够大大节省调查成本, 而缺点是如果定价不合理, 有的任务就会无人问津, 导致商品检查的失败。所以说, 如果采用众包的方式来进行调查, 首先需要保证任务的完成率, 其次才是降低调查成本, 即完成率要先于调查成本来考虑。

此优化问题的约束条件是任务的执行情况和标价。

2.3. 问题三分析

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, 一种考虑是将这些任务联合

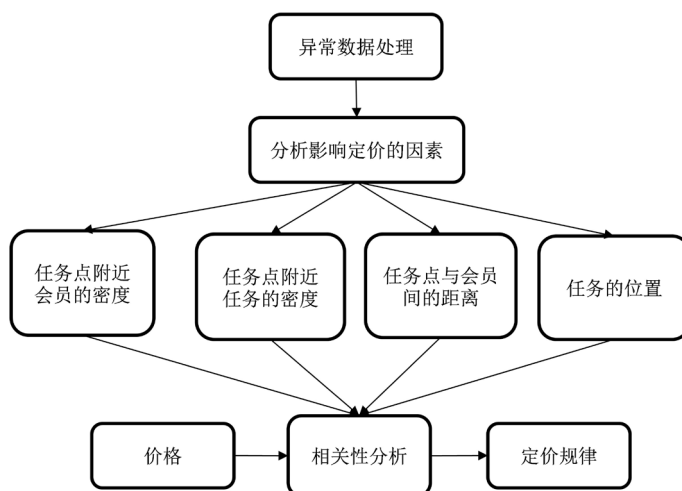


Figure 1. Study map of pricing rules

图 1. 研究定价规律的思维导图

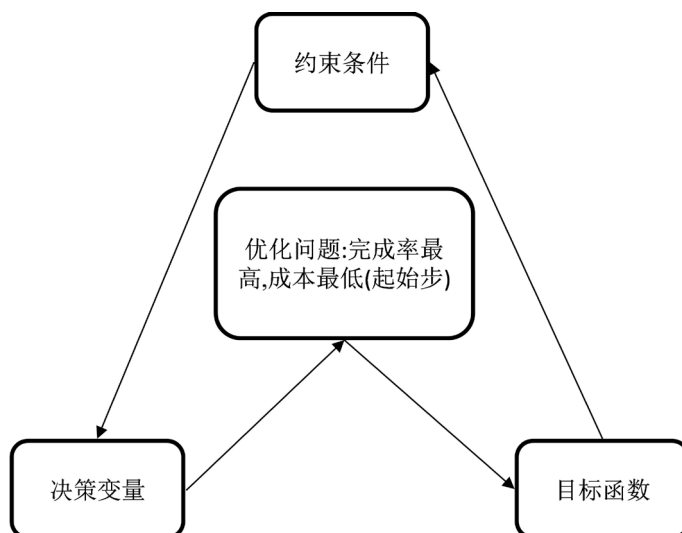


Figure 2. Mind map of optimization problems in the new pricing scheme

图 2. 新定价方案中优化问题的思维导图

在一起打包发布。因为，任务打包以前，如果会员分多次完成所选任务，时间和路程花费较高。任务归拢打包以后，会员可一次出行完成多个任务，成本会明显下降，并且因为一次可完成多个任务，会员的积极性也会显著提高，在现有信誉值的基础上，提高完成任务的可能性。

在这个问题中本文主要考虑因素如表 1 所示。

3. 模型的建立与应用

考察“任务点附近会员的密度”、“任务点附近任务的密度”、“任务点距用户的最近距离”等因素与定价的相关性，如果相关性显著，则可认为该因素影响定价，得到对应的定价规律。图 3 为任务分布图。

3.1. 影响定价因素的模型(问题一)

3.1.1. 影响定价的因素

因素一：任务点附近会员的密度

$$\rho_i^v = \frac{N_v(r)}{\pi r^2},$$

其中， $N_v(r)$ 表示以任务点为中心、 r 为半径的圆所圈定的会员的数目。

因素二：任务点附近任务的密度

$$\rho_i^t = \frac{N_t(r)}{\pi r^2},$$

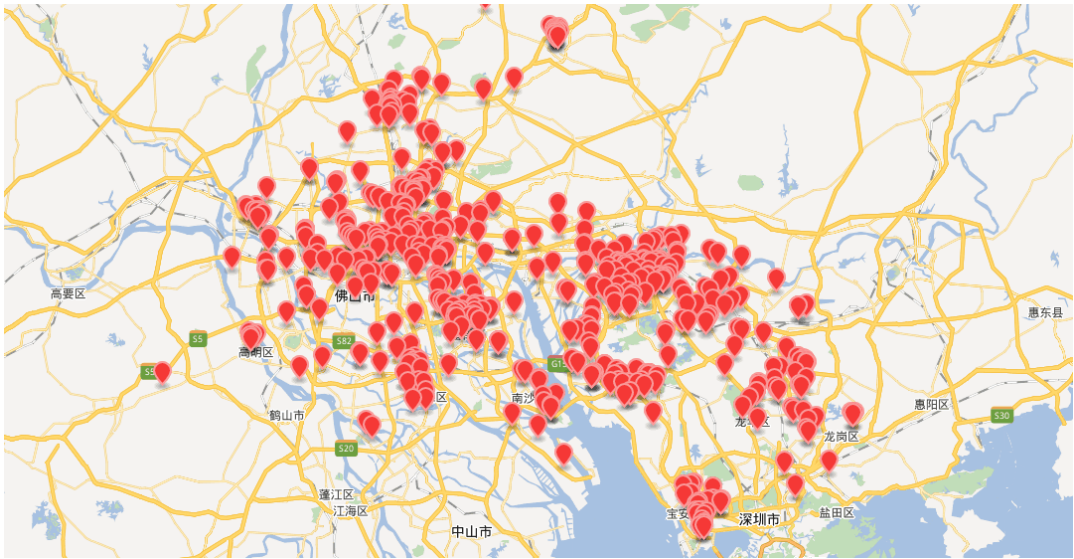


Figure 3. Task map
图 3. 任务分布图

Table 1. The main factor of influence
表 1. 主要影响因素

影响因素	中心任务点的位置
	临近任务点的位置
	周围会员的位置和信誉值(开始时间)

其中, $N_i(r)$ 表示以任务点为中心、 r 为半径的圆所圈定的任务的数目。

因素三：任务点与最近的 n 个会员间的平均距离

$$d_i^n = \frac{\min_n(d_i)}{n},$$

其中, $\min_n(d_i)$ 表示与第 i 个任务点相距最近的 n 个项目之间距离的总和。在对该因素进行计算时, 本文令 $n=10,15,20,25,50$ 分情况讨论, 试图寻找一个使相关性最显著的 n 值。

3.1.2. 相关性分析与价格规律的阐述

本文使用 SPSS 对上述各因素的计算结果与任务标价分别进行相关性分析, 部分相关性较显著的结果如表 2 所示。

根据 SPSS 的相关性分析的结果可以看出 25 个距任务最近会员的平均距离、50 个距任务最近会员的平均距离、任务密度、用户密度与任务标价的皮尔孙相关系数均大于 0.4, 属于中度相关, 可以认为对任务价格产生影响, 且任务密度呈负相关, 说明任务密集的地区任务价格低而任务间距远的地区定价高, 其他三个指标与定价呈正相关, 说明会员少的地区价格高。任务的经纬度、10 个距任务最近会员的平均距离、15 个距任务最近会员的平均距离与任务价格的相关系数低, 可以认为与定价无关。

本文选取相关系数较高的三个指标: 50 个距任务最近会员的平均距离、任务密度和用户密度, 并以任务定价为因变量, 使用 Matlab 进行逐步回归分析, 剔除不显著的变量, 建立回归模型[4]:

$$y = 0.3446x_1 - 2.603x_2 - 1.2074x_3 + 69.0211$$

其中, x_1, x_2, x_3 为选取的三个指标值, y 为任务价格。得到显著性为 $3.11e^{-49}$, 接近于 0, 说明该结果可信度高, 可认为该模型即为原始定价规律。

3.1.3. 任务未完成原因的分析

本文运用 Matlab 将位置数据和执行情况融入到图 4 中。

通过图 4 的任务及用户分布图, 可以看出完成情况存在明显的区域性, 各区域导致任务未完成的原因不同。通过对典型区域进行分析, 可以发现, 有些地区人口集中但未完成任务居多, 且多为低价任务, 如(a) (b)区, 说明该地区任务定价过低, 导致完成率低; 有些地区任务价格高, 但仍未被完成, 如(c)区,

Table 2. Partial results of correlation analysis

表 2. 相关性分析的部分结果

相关		
变数		任务标价
50 个最近会员的平均距离	皮尔森(Pearson)相关	0.435**
	显著性(双尾)	6.26E-40
	N	835
任务密度	皮尔森(Pearson)相关	-0.412**
	显著性(双尾)	1.77E-35
	N	835
用户密度	皮尔森(Pearson)相关	-0.423**
	显著性(双尾)	1.39E-37
	N	835

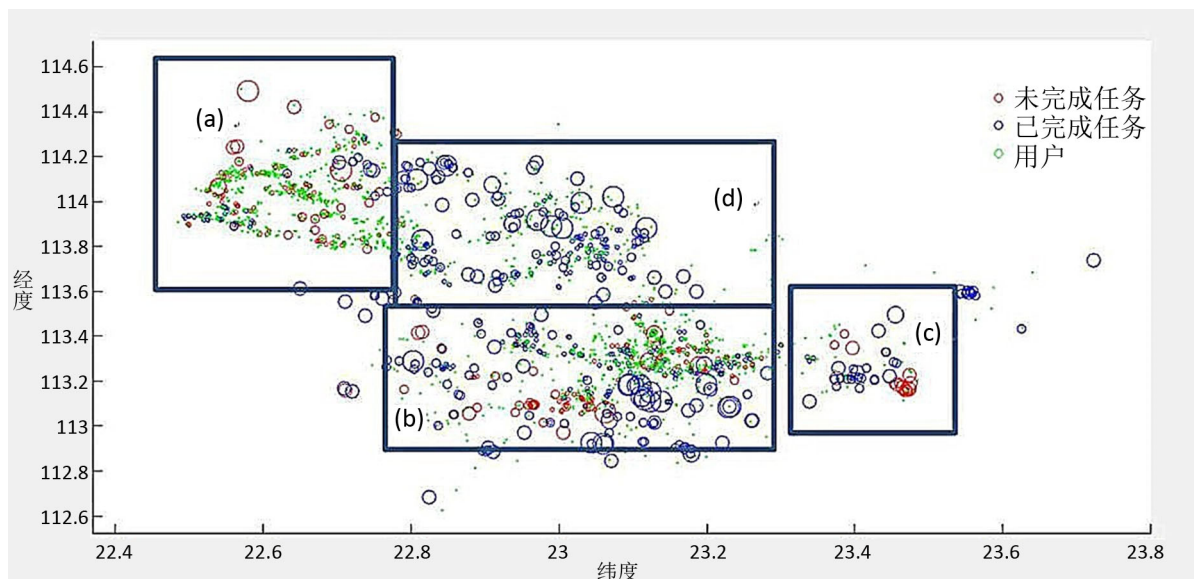


Figure 4. Task and user profile (the higher the task pricing, the larger the corresponding circle)

图 4. 任务及用户分布图(任务定价越高, 对应的圆圈越大)

原因为该区域人口过少。有些地区人口量适中且任务完成率高, 如(d)区。

3.2. 优化模型(问题二)

本文进行了逆向推导, 先来考虑任务的执行情况, 执行情况只有完成与未完成两种, 显然这其中的判别条件变得非常重要。本文以每个任务的利润作为这个判别指标, 当该任务的利润大于预期利润, 则表明这件任务的执行情况较为乐观, 反之较为悲观。利润的定义是最初标价减去会员执行该任务的成本, 而该成本则可以认为是任务与会员之间的距离乘以每公里的执行任务成本。从而, 任务的执行情况这一约束条件便转化为由决策变量(即各任务)所决定的函数不等关系。

对于另外一个约束条件(任务的标价), 本文根据不同任务与会员之间距离的不同, 对任务进行分类。容易推得, 距离越大, 则说明该任务周围的会员越少, 这个任务越不易被完成, 固然, 本文应该对这类任务进行涨价, 以提高任务整体的完成率, 反之则应该降价。

3.2.1. 目标函数

$$\begin{cases} \max H = \frac{\sum D_i}{\sum i} \\ \min Q = \sum (P_i \cdot D_i) \end{cases},$$

其中, H 表示完成率, Q 表示调查的总成本, D_i 表示第 i 个任务的执行情况(取值 0 或 1), P_i 表示第 i 个任务的价格。

3.2.2. 约束条件

目标函数中的未知量有第 i 个任务的执行情况和第 i 个任务的价格, 现确定其表示形式。

任务的执行情况 D_i

$$D_i = \begin{cases} 0, & w_i < \bar{w} \\ 1, & w_i > \bar{w} \end{cases},$$

其中, w_i 为第 i 个任务的利润, $\bar{w}$ 为会员对利润的期望值与其信誉等级的乘积。当任务所带来利润高于此值, 这个任务将被执行; 反之, 将不被执行。

会员信誉等级如表 3 所示。

第 i 个任务的利润 w_i

$$w_i = P_i^0 - l_i,$$

其中, P_i^0 为初始的任务标价, l_i 为执行第 i 个任务的成本。它表示会员的实际所得为任务标价减去执行该任务所需的成本。

执行第 i 个任务的成本

$$l_i = d_i^n \cdot c,$$

其中, d_i^n 表示任务点与最近的 n 个会员间的平均距离, c 表示单位距离所需成本。

任务标价 P_i

在第一问中, 本文发现任务标价 P_i 和任务点与最近会员间距离 d_i^n 呈正相关。所以, 本文考虑将任务标价 P_i 与任务点与最近会员间距离 d_i^n 进行关联, 即

$$P_i = k \cdot d_i^n.$$

通过对约束条件中的参数进行调整, 以达到对目标函数优化的目的。

3.2.3. 结果分析(如图 5 所示)

原定价方案完成任务总数为 522, 采用问题二定价方案, 完成任务总数为 665, 增幅约为 27.395%, 表明新方案对完成率的提高效果明显。

3.3. 任务打包发布模型(问题三)

3.3.1. 模型阐述

在这个问题中本文主要考虑的因素如表 4 所示。

通过对以上因素并结合相关文献综合考虑, 本文选取如下打包规则:

- ✧ 临近任务点距离中心任务点 850 米以内的允许打包, 否则不允许打包;
- ✧ 单个联合打包的任务点, 其包含的任务量必须低于周围 75% 的会员的配额;
- ✧ 每增加一个临近任务, 总定价将下调 2%。

按照以上规定, 本文进行了第一次打包处理。因为信誉值高的会员可以优先选择任务点, 所以本文

Table 3. Member credit rating

表 3. 会员信誉等级

信誉等级	信誉值的范围	会员数
1	大于 140	188
2	(43,140]	186
3	(20,43]	175
4	(19,20]	711 (包含新会员)
5	(1.993,19]	170
6	(1.0,1.993]	217
7	[0,1.0]	230

任务号码	50个最近会员的平均距离	用户密度	任务gps 纬度	任务gps经度	任务标价	改进定价	任务执行情况	改进执行情况
A0526	3. 033955499	0. 598726115	23. 06060865	113. 7374003	66. 5	63. 2	1	1
A0527	3. 912323233	0. 496815287	23. 01783954	113. 8233686	69	70. 7	1	1
A0528	3. 059619949	0. 636942675	23. 05828484	113. 8147964	67	67	0	0
A0529	4. 05591369	0. 382165605	22. 96936454	113. 8966138	69. 5	69. 5	1	1
A0530	3. 302824599	0. 649681529	23. 01465206	113. 7680676	65. 5	65. 5	0	0
A0531	4. 679828637	0. 382165605	22. 99522565	113. 8000233	71	72. 8	0	1
A0532	3. 50574431	0. 611464968	23. 05866525	113. 8321688	68	69. 7	0	1
A0533	4. 472715675	0. 343949045	22. 98744806	113. 725147	68	69. 7	0	1
A0534	4. 364033287	0. 331210191	22. 99313225	113. 7323716	67	68. 7	0	1
A0535	3. 015301696	0. 649681529	23. 02871297	113. 8045789	66	66	0	0
A0536	4. 713697203	0. 369426752	22. 97417485	113. 7345883	67. 5	69. 2	0	1
A0537	4. 822138975	0. 356687898	22. 94132785	113. 6767951	66	66	0	0
A0538	8. 19435138	0. 140127389	22. 84339687	113. 2557574	66	67. 6	0	1
A0539	3. 041871412	0. 636942675	23. 06167985	113. 8102199	67	67	0	0
A0540	8. 282280976	0. 178343949	23. 17934885	112. 8994031	70	71. 8	0	1
A0541	5. 756425266	0. 178343949	23. 17455266	113. 0688513	75	78. 8	0	1
A0542	8. 683325435	0. 152866242	23. 17677756	112. 8784023	70	71. 8	0	1
A0543	8. 035436672	0. 178343949	23. 16943397	112. 9073857	70	71. 8	0	1
A0544	7. 848880422	0. 063694268	23. 2615299	113. 0247518	75	78. 8	0	1
A0545	8. 287580078	0. 178343949	23. 16865186	112. 8938126	70	71. 8	0	1
A0546	7. 8807534	0. 178343949	23. 16321812	112. 9148997	70	71. 8	0	1
A0547	11. 02449988	0. 050955414	23. 07057951	112. 8445408	75	78. 8	1	1
A0548	11. 28011251	0. 089171975	22. 90422292	112. 9019826	75	78. 8	1	1
A0549	3. 572967294	0. 598726115	23. 0386685	113. 2034172	70	70	0	0
A0550	3. 327956584	0. 828025478	23. 20300498	113. 1713931	69. 5	69. 5	0	0
A0551	8. 061820578	0. 178343949	23. 17018893	112. 9060636	70	71. 8	0	1
A0552	7. 714617283	0. 152866242	22. 87794292	113. 2265578	70. 5	74	0	1
A0553	2. 344134403	1. 006369427	23. 18828575	113. 2051584	68	64. 6	0	0
A0554	2. 996120202	1. 108280255	23. 1176526	113. 1774208	66. 5	63. 2	0	0
A0555	4. 763496405	0. 25477707	23. 14659839	113. 1104485	67. 5	69. 2	0	1
A0556	2. 70078896	1. 477707006	23. 12232763	113. 1914852	67	67	0	0
A0557	8. 118496751	0. 178343949	23. 16730584	112. 902416	70	71. 8	0	1

Figure 5. Partial result example
图 5. 部分结果示例

Table 4. Main considerations
表 4. 主要考虑因素

影响因素	中心任务点的位置
	临近任务点的位置
	周围会员的位置和信誉值(开始时间)

在进行任务分配时，模拟联合打包任务点范围内，会员按信誉值依次选择任务点。

本文认为只做出以上假设和规定，还不够完善，可以按照图 6 的方案做进一步改进：

有一种任务，其周围会员分布稀疏，或信誉值整体较差，而任务又在较偏远的位置，本文称这种任务为“冷门任务”。此种情况下，可以将此“冷门任务”打包到临近的一个联合打包任务点。同时制定如下规定：

- ✧ “冷门任务”离联合打包任务点距离不能超过 2500 米；
- ✧ 每个联合打包任务点，只能绑定一个“冷门任务”；
- ✧ “冷门任务”离联合打包任务点中其他任务中心位置每增加 500 米，将“冷门任务”定价提高 2%。

虽然增加了小部分成本，但对“冷门任务”的完成，做出了保障，可有效改善任务总的完成情况。

打包发布对最终任务完成情况的影响如下：打包模型可以在降低整体定价的前提下，保障任务的完成情况，并且在优化处理之后，可以很好地解决“冷门任务”无人问津的情况。并且大胆做出估计，打

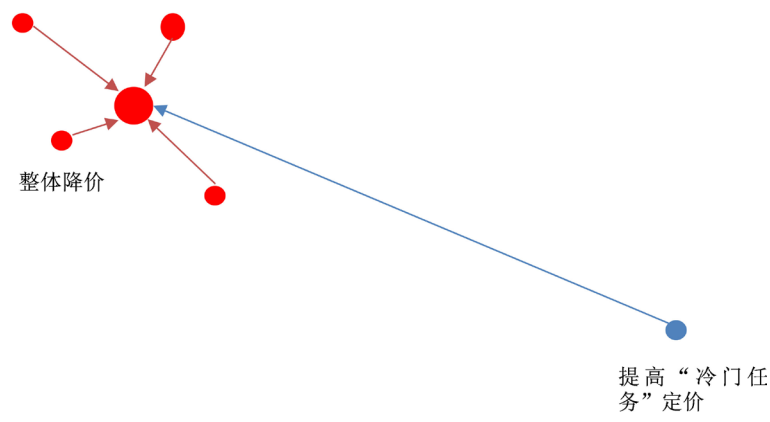


Figure 6. Improved packaging solution
图 6. 改进后的打包方案

任务号码	任务gps 纬度	任务gps经度	任务标价	改进定价	打包后价格	打包组别	原任务执行情况	改进执行情况	打包执行情况
A0308	23. 45165523	113. 2845092	70	70	63	6	1	1	1
A0311	23. 55840651	113. 5895144	70	73. 5	66. 2	11	1	1	1
A0393	23. 446661	113. 2201755	75	75	67. 5	6	1	1	1
A0395	23. 47430929	113. 163008	75	71. 2	64. 1	9	0	0	1
A0397	23. 44184274	113. 3286502	72	73. 8	66. 4	5	0	1	1
A0398	23. 37803295	113. 2569259	75	71. 2	64. 1	1	0	0	1
A0399	23. 4422023	113. 3300279	72	73. 8	66. 4	6	1	1	1
A0400	23. 40329889	113. 2398514	75	67. 5	60. 8	4	0	0	0
A0401	23. 43115969	113. 4254505	75	78. 8	70. 9	5	1	1	1
A0407	23. 46444973	113. 1728925	75	71. 2	70. 9	7	1	1	1
A0418	23. 54196985	113. 5979552	75	78. 8	73. 8	9	0	1	1
A0424	23. 55366671	113. 5951132	75	78. 8	70. 9	10	0	1	1
A0426	23. 45457434	113. 4983044	80	82	73. 8	6	0	1	1
A0429	23. 5532428	113. 5953837	75	78. 8	70. 9	10	0	1	1
A0431	23. 55380843	113. 5953328	75	78. 8	70. 9	11	1	1	1
A0435	23. 45803366	113. 1869153	75	71. 2	64. 1	7	0	0	0
A0436	23. 3754398	113. 2134638	75	71. 2	64. 1	1	1	1	1
A0437	23. 47484287	113. 1908172	75	67. 5	60. 8	9	1	1	1
A0438	23. 47290189	113. 1700943	75	67. 5	60. 8	8	1	1	1
A0439	23. 46841629	113. 1611801	75	71. 2	64. 1	8	0	0	1
A0457	23. 46633279	113. 159492	72	68. 4	61. 6	8	0	0	1

Figure 7. Partially packaged result example
图 7. 部分打包结果示例

包模型可以促进中低信誉的会员的积极性，可以进一步提高任务总体完成情况。

3.3.2. 结果示例及分析(如图 7 所示)

取北纬 23.35°N~24.56°N，东经 112.98°E~113.60°E 区域，做打包处理进行实验。最后任务完成率比原定价方案结果高出 37.5%，比问题二优化方案高出 15.79%，说明打包方案切实可行，总任务完成情况超过 80%。

4. 结语

本文研究了众包任务的定价规律，发现任务标价与 50 个最近会员的平均距离和用户密度这两个因素的相关性最强，继而逐步回归分析，剔除不显著的变量，建立出回归模型，将定性的价格规律量化。继而，基于多目标优化的思想提出了新的定价方案，以提高任务完成率并降低调查成本。最后，考虑了将临近任务联合打包发布。任务归拢打包以后，会员可一次出行完成多个任务，成本会明显下降，并且因为一次可完成多个任务，会员的积极性也会显著提高，在现有信誉值的基础上，提高了完成任务的可能性。

参考文献

- [1] 刘晓钢. 众包中任务发布者出价行为的影响因素研究[D]: [硕士学位论文]. 重庆: 重庆大学, 2012.
- [2] 李忆, 姜丹丹, 王付雪. 众包式知识交易模式与运行机制匹配研究[J]. 科技进步与对策, 2013(7): 127-130.
- [3] 李勇军, 郭基凤, 缙西梅. 软件“众包”任务分配方法[J]. 计算机系统应用, 2015, 24(2): 1-6.
- [4] 董大校. 基于 MATLAB 的多元非线性回归模型[J]. 云南师范大学学报, 2009(3): 45-48.

知网检索的两种方式:

- 1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2324-7991, 即可查询
- 2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: aam@hanspub.org