

Research on Mission Pricing Scheme under Crowdsourcing Business Model

Xiaoying Fu¹, Rui Feng²

¹College of Mathematics and System Science, Shandong University of Science and Technology, Qingdao Shandong

²Shandong University of Science and Technology, Qingdao Shandong
Email: 1013670821@qq.com

Received: Apr. 3rd, 2018; accepted: Apr. 17th, 2018; published: Apr. 24th, 2018

Abstract

This paper studies the self-service crowdsourcing platform of mobile Internet, and establishes the task pricing model based on multivariate linear regression model. Compared with the original plan, the completion rate of the improved model reached 67.305%, which was significantly higher than 62.395% of the original plan. And the overall relative reduction of mission expenditure, optimization effect is better. According to the analysis of the task "package release", we carried out the simulation based on the principle of support vector machine (SVM). On the basis of the second problem, the rules of packing are added, and the rules of simulation are changed, thus introducing the idea of greedy algorithm to establish the packing mechanism to improve the pricing scheme established before. After solving, in the improved model, the task completion rate increased to 72.302%, and the overall task is completed better. This article involves the data from the National Undergraduate Mathematical Contest 2017 Modeling B questions.

Keywords

Multiple Linear Regression, Support Vector Machine, Greedy Algorithm, Simulation, Punishment Mechanism

众包商业模式下的任务定价方案研究

付晓莹¹, 冯睿²

¹数学与系统科学学院, 山东科技大学, 山东 青岛

²山东科技大学, 山东 青岛
Email: 1013670821@qq.com

收稿日期: 2018年4月3日; 录用日期: 2018年4月17日; 发布日期: 2018年4月24日

摘要

本文对移动互联网的自助式劳务众包平台进行研究,建立了基于多元线性回归模型下的任务定价模型。与原方案进行对比,改进后模型的任务完成率达到67.305%,相对于原方案的62.395%有了明显提高,并且任务相对支出整体上降低,优化效果较好。进一步,根据对任务“打包发布”方式的分析,进行了基于支持向量机样本分类原理下的仿真模拟,在问题二的基础上添加打包规则、更改仿真规则,并且引入贪心算法的思想建立打包机制,改进之前建立的定价方案。经求解,得出改进后的模型中,任务完成率提高到72.302%,整体任务完成情况较好。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。

关键词

多元线性回归, 支持向量机, 贪心算法, 仿真模拟, 惩罚机制

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. 引言

“拍照赚钱”是移动互联网下的一种自助式服务模式。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(比如上超市去检查某种商品的上架情况),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。APP 作为平台运行的核心,其中的任务定价是核心要素。假如定价不合理,有的任务就会无人问津,导致商品检查失败。

研究现状:随着互联网普及率的逐渐提高以及 Web2.0 时代网络虚拟产品交易的兴起,众包已经成为一种新型的电子商务模式并占据着越来越重要的市场地位。众包是一种开放式创新,其成功与否与任务发布者的出价密切相关,研究任务的出价策略对于任务发布者获取满意方案、降低成本(时间成本与经济成本)具有重要意义。众包的本质是服务产品在线交易行为,大多数学者基于服务定价理论和在线逆向拍卖理论为指导,通过实证的方法研究众包网站上任务发布者制定的任务最终悬赏金额与任务属性及市场竞争的关系,从而得出成功的出价策略。

问题的产生

基于组委会给出相关地区数据,本文研究了如下问题:

- 1) 研究某已知项目的任务定价规律,分析任务未完成的原因。
- 2) 设计新的任务定价方案,并和原方案进行比较。
- 3) 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在一起打包发布。在这种考虑下,如何修改前面的定价模型,对最终的任务完成情况又有什么影响?

2. 问题分析

本文选取的影响任务定价的因素:任务难度、会员-任务数量比、会员距离任务地点的远近程度、

地区差异、会员信誉均值、会员预定任务限额(图 1~3)。

3. 模型的建立

3.1. 任务定价影响因素体系

1) 任务难度

同一个项目中任务难度相当, 因此样本底价相同, 用常数值 P_0 表示。

2) 会员 - 任务数量比

$$\rho = \frac{N_i}{M_i}, \quad i = 1, 2, 3, 4, \quad (1)$$

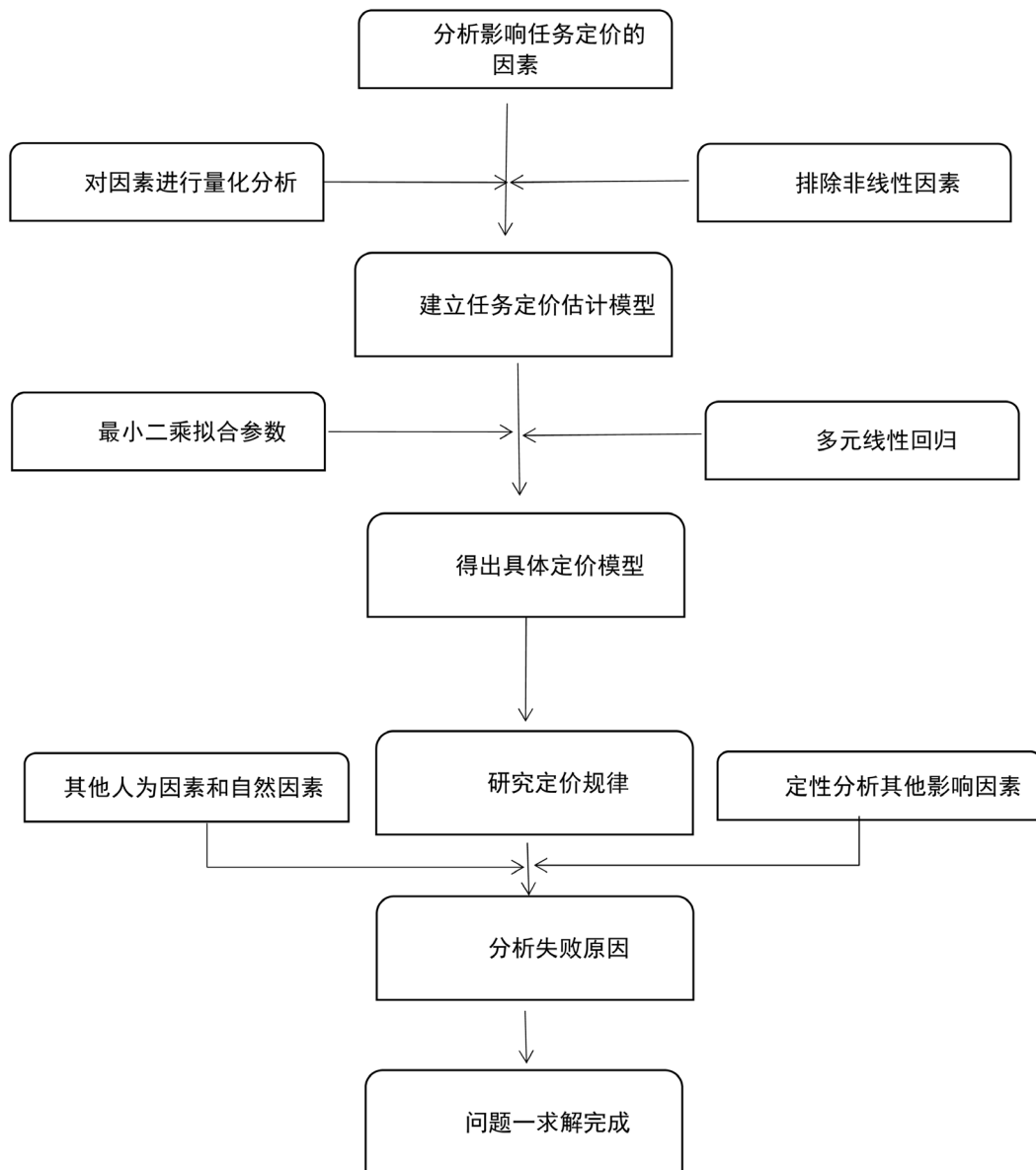


Figure 1. Flow chart of problem 1
图 1. 问题一模型求解流程图

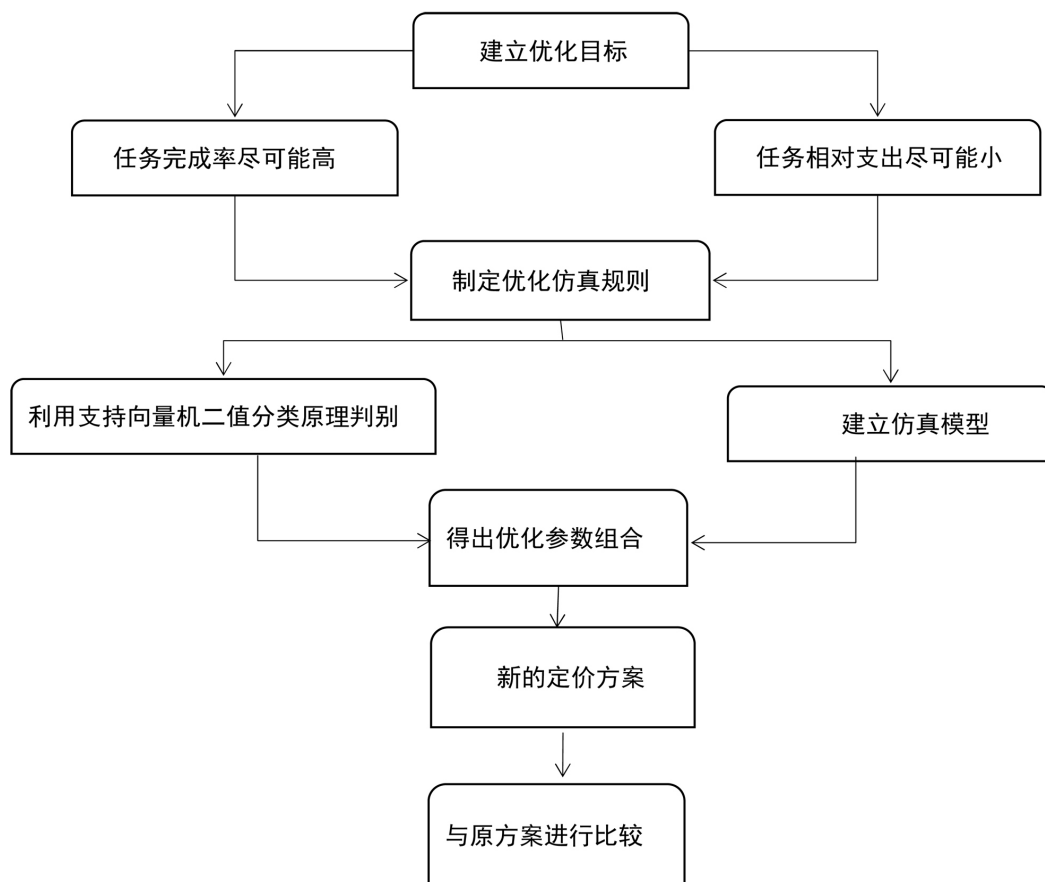


Figure 2. Flow chart of problem 2
图 2. 问题二模型求解流程图

其中, N_i 代表一定区域内会员分布的数量, M_i 代表一定区域内任务分布的数量, i 表示本文划分的若干个小区域。在此, 引入经济学中的需求函数进行该部分的定价 P_ρ 表示, B 表示 P_ρ 关于 ρ 的敏感程度, A/B 表示本文所考虑的 P_ρ 最大值:

$$P_\rho = \frac{A - \sqrt{\rho}}{B}. \quad (2)$$

3) 会员距离任务地点的远近程度

数据只给出了会员与人物之间的经纬度坐标, 在此本文利用 Google 地图提供的方法[1], 进行距离换算, 计算精度与谷歌地图精度相似, 误差在 0.2 米以下, 具体运算方式如下:

$$S = 2 \arcsin \sqrt{\sin^2 \frac{a}{2} + \cos(Lat1) \times \cos(Lat2) \times \sin^2 \frac{b}{2}} \times 6378.137, \quad (3)$$

其中, $Lat1$ $Lung1$ 表示 A 点经纬度, $Lat2$ $Lung2$ 表示 B 点经纬度, $a = Lat1 - Lat2$ 为两点纬度之差, $b = Lung1 - Lung2$ 表示两点经度之差, 6378.137 为地球半径, 单位为千米。

“距离定价”定义为距离总和的均值, 即:

$$d = \frac{\sum_{j=1}^n d_j}{n}, \quad j=1, 2, \dots, n, \quad (4)$$

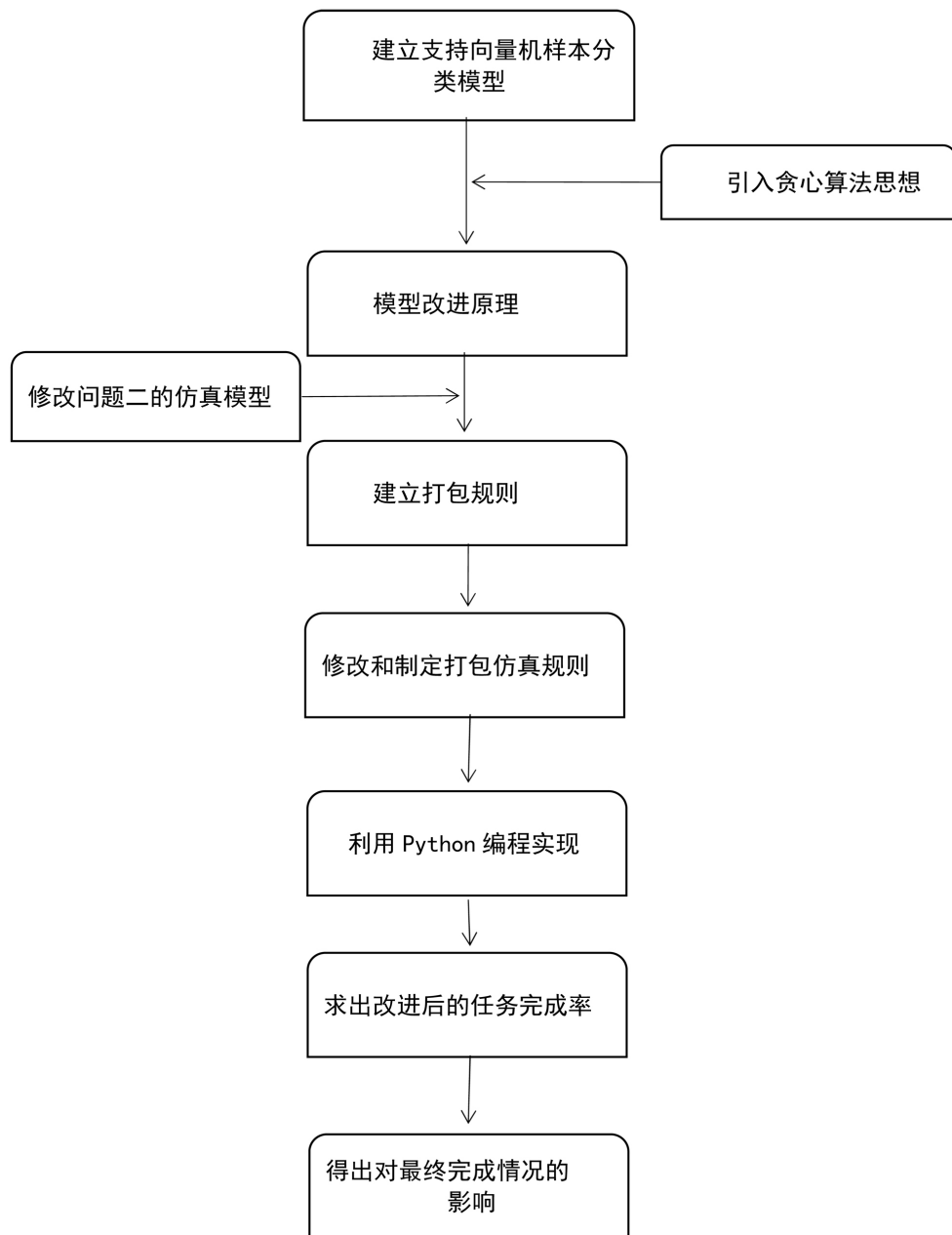


Figure 3. Flow chart of problem 3
图 3. 问题三模型求解流程图

其中, d_j 表示一个任务周围 5 km 内的会员距离, 本文利用 Java 求解 835 个任务位置 5 km 以内的会员个数和距离均值。

4) 区域差异

四个城市之间的地区差异体现在多方面, 故导致的任务定价不同, 本文用误差项 $\varepsilon(o)$ 表示。其中, 误差项并不只是一个单纯的常数值, 它包含了可能存在的浮动因素, 浮动因素会使定价模型出现动态非线性的特点, 为此本文不考虑非线性模型, 因此将其考虑进误差项中, 将定价模型转换成线性模型。

$$\varepsilon(p) = \text{动态因素引起的任动态定价差异} \tag{5}$$

5) 会员信誉均值

$$\bar{R} = \frac{\sum_{j=1}^n R_j}{n}, \quad j=1, 2, \dots, n, \quad (6)$$

6) 会员的预定任务限额

本文用一定区域内所有会员的预定任务限额表示, 即

$$Q = \sum_{j=1}^n Q_j, \quad j=1, 2, \dots, n. \quad (7)$$

3.2. 任务定价估计模型[2]

基于上述分析, 本文建立以下众包平台任务定价模型, 并且该模型可以给出单个的任务定价:

$$price = \alpha_1 \cdot P_p + \alpha_2 \cdot d + \alpha_3 \cdot \bar{R} + \alpha_4 \cdot Q + P_0 + \varepsilon(p). \quad (8)$$

利用 Matlab 将量化后的四类数据与定价代入到模型中, 根据 0~1 分类进行多元线性回归, 最小二乘拟合出完成与未完成样本的参数估计值, 得到参数结果如表 1 所示。

综上可知, 在划定区域内(任务点周围 5 km 以内), 通过参数 α_1, α_2 可以看出会员 - 任务的数量比 ρ 和会员到该任务地点的距离因素 d 是影响任务定价 $price$ 的主要因素, 并且通过分析对比可知, 这两个因素也显著影响任务完成率。因此, 任务的定价是影响任务完成的主要因素。

4. 优化与仿真模型

4.1. 确定优化目标

本文拟通过对 α_1, α_2 的优化, 达到以下两个优化目标, 作为问题二模型改进的方向。

- 1) 对于任务发布者来说, 尽可能提高任务完成率。
- 2) 对于任务发布者来说, 尽可能减少任务的相对支出(支出与任务完成率的比值)。

4.2. 对任务是否完成进行判别: 支持向量机二值分类原理及应用

支持向量机是针对二值分类问题提出的, 选取样本, 建立约束条件和超平面约束条件, 然后进行分类, 从而建立一个支持向量数目相对较少的最优分类平面, 从训练样本中找到一个支持向量, 建构出最好的分类超平面, 是本文的主要目标。

本文将训练样本集 $(x_i, y_i), i=1, \dots, l$ 分成两个类别, 标记为正 ($y_i=1$) 的属于第一类 $x_i \in R^n$, 标记为负 ($y_i=-1$) 的属于第二类 $x_i \in R^n$, 如果本文建构的决策函数能够将被测试数据分开, 那么学习目标就达成了。其中, $n=6$, 六个维度依次代表: 任务难度、会员 - 任务数量比、会员到任务地点的距离、会员信誉均值、地区差异、会员预定任务限额。

Table 1. Parameter fit values for completed and uncompleted task pricing models

表 1. 完成与未完成任务定价模型参数拟合值

完成情况	参数 α_1	参数 α_2	参数 α_3	参数 α_4
所有任务(835 个)	1.7329	0.0523	-0.001	-0.0102
完成(522 个)	3.4719	0.0964	-0.001	-0.009
未完成(313 个)	0.3128	-0.7022	-0.001	-0.0082
相对变动率	91.02%	-87.13%	0%	8.9%

由于题目中给出的任务完成情况是由 0 和 1 两种表示(1 表示完成, 0 表示未完成)。本文在优化第一个目标, 即任务完成率时, 就需要根据支持向量机的二值分类特点进行任务是否完成的判断, 从而得出任务完成情况。

4.3. 仿真模型的建立

首先, 本文确定仿真模型的规则如下:

Rule 1: 每个会员只会选择其直线距离 5 km 以内的任务。

Rule 2: 每个会员会选择可选范围内给出定价最高的任务。

Rule 3: 每个会员按照预定任务开始时间进行任务选择。

Rule 4: 根据题意, 每个会员获得任务数量为:

$$\frac{\text{该会员的预定任务限额}}{\text{该会员周围5 km内所有会员的任务限额总和}} \times 5 \text{ km内所有任务数} \quad (9)$$

Rule 5: 对重要参数 α_1, α_2 在其 0.8~1.2 倍范围内进行枚举搜索。

Rule 6: 不考虑可能的自然因素和人为因素影响。

在满足规则的前提下, 本文对 α_1, α_2 在一定范围内进行枚举, 增量范围可以在参数 0.8~1.2 倍之间。对其进行网格化处理, 使用启发式算法搜索最优参数, 使得任务完成率最大, 并且此时的价格即为最优价格。

4.4. 求解

本文利用 Java 实现定价模型优化的仿真模拟。发现 α_2 对优化后的模型影响不大, 任务完成率和单位任务支出的数据变动几乎可以忽略不计。所以最终的优化模型结果中, 忽略了 α_2 的参与作用, 认为其仍然是原来的值固定不变, 只通过优化 α_1 的值来改进模型。 α_1 部分枚举结果如表 2 所示。

本文将优化后的最优 α_1 值代入原模型中, 得出新的定价模型, 即为改进后的方案。部分运行结果如表 3 所示。

4.5. 与原模型的对比及结果分析

通过两个优化目标确定 α_1 最优值, 并且与原始参数的对比如表 4。其中, 通过对比可以看出, 优化后的模型任务完成率比原模型增加了 5% 以上, 支出比原来有所上升, 但是相对支出是降低的, 由原来的 58,411.54 减少为 58,258.81。因此, 本文认为仿真模拟后得出的模型优化效果较好。

Table 2. Partial model solution results

表 2. 部分模型求解结果

α_1	完成率	支出
4.2	0.671856287	39,093.35706
4.18	0.671856287	39,094.3591
4.16	0.671856287	39,095.36113
4.14	0.671856287	39,096.36316
4.12	0.671856287	39,097.3652
4.1	0.673053892	39,162.86288
4.08	0.673053892	39,163.89389
.....

Table 3. Part of new pricing schemes for problem 2
表 3. 部分问题二新的定价方案

编号	价格	信誉均值	配额总和	ρ	平均距离	完成度	新模型定价
A0001	66	1614.966	421	6.32	3	0	71.03466
A0002	65.5	143.5646	137	1.44	2.23	0	62.08756
A0003	65.5	1344.131	546	6.32	3.16	1	68.2859
A0004	75	0.3657	2	6.32	2.73	0	61.36891
A0005	65.5	991.9538	666	6.32	3.25	0	68.07084
A0006	75	0.3636	1	6.32	0.4	0	62.91792
A0007	65.5	1320.498	456	6.32	3.12	1	61.79373
A0008	65.5	1059.654	612	6.32	3.32	0	67.45145
A0009	66	64.38424	102	6.32	4.07	0	63.27501
.....

Table 4. Parameter values before and after optimization
表 4. 优化前后参数值

	α_1	完成率	支出	相对支出
优化后参数值	3.16	0.673	39,211.32	58,258.81
原参数值	3.47	0.624	36,446	58,411.54

5. 任务打包模型：基于贪心算法的打包规则

本文主要是增加了打包的思想：多个任务可能因为位置比较集中，导致用户会争相选择，本文考虑将这些任务联合在一起打包发布。因此，本文在问题二的基础上添加打包规则、更改仿真规则。

使用贪心算法求解问题时，总是做出在当前看来是最好的选择。即不从整体最优上加以考虑，他所做出的是在某种意义上的局部最优解。结合支持向量机的样本分类功能，本文采用贪心算法的打包流程如图 4。

5.1. 仿真规则的改进

根据问题三的实际情况本文在问题二的基础上增加改进了以下仿真规则：

- 1) 限定最多三个任务被打包组合在一个包内；
- 2) 视打包后整体为一个任务，该任务的位置取包内任务距离中心；
- 3) 每个会员可接受任务的范围发生改变，该范围与包内任务之间的距离呈递减关系，可表示为：

$$\bar{d} = 5 - b \cdot \Delta d^a, \quad (10)$$

其中， a, b 均为大于零的常数； Δd 表示包内任务之间的距离，当 $\Delta d = 0$ 时可接受范围为之前的 5 km。

5.2. 改进后模型的求解

本文利用 Java 实现改进后定价模型的仿真模拟，部分任务的定价如表 5 所示。

5.3. 模型改进后的效果分析

模型改进后与改进前的任务完成率具体比较如表 6 所示。

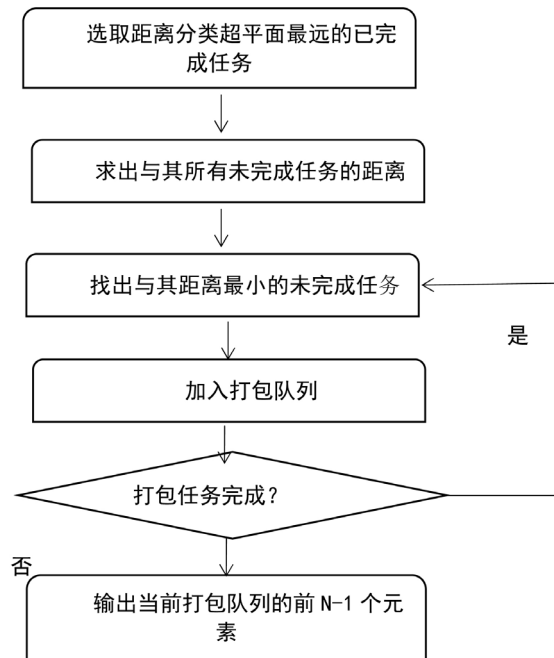


Figure 4. Flowchart of the greedy algorithm
图 4. 贪心算法流程图

Table 5. Partially packaged task pricing
表 5. 部分打包后的任务定价

打包后的任务编号	定价
A0151 A0141 A0148	196.5
A0150 A0136	198.5
A0339 A0176	146.5
A0337 A0316 A0314	131.5
.....

Table 6. Task completion rates before and after model improvement
表 6. 模型改进前后的任务完成率

	任务完成率
模型改进前	67.305%
模型改进后	72.302%

由结果可知, 改进后的任务完成情况较好, 任务完成有所提高。

参考文献

- [1] 谷歌地图计算两个坐标的距离函数[EB/OL]. 博客.
<http://www.cnblogs.com/yony/archive/2012/07/31/2616253.html>
- [2] 李训, 林川, 胡明. 我国地域关系对审计定价影响的实证研究[J]. 财经论丛(浙江财经大学学报), 2013(2): 93-100.

附录

2017 年全国大学生数学建模竞赛 B 题 全部数据见:

http://www.mcm.edu.cn/html_cn/node/460baf68ab0ed0e1e557a0c79b1c4648.html

因篇幅原因以下只给出部分数据展示:

附件一: 已结束项目任务数据

任务号码	任务 gps 纬度	任务 gps 经度	任务标价	任务执行情况
A0001	22.56614225	113.9808368	66	0
A0002	22.68620526	113.9405252	65.5	0
A0003	22.57651183	113.957198	65.5	1
A0004	22.56484081	114.2445711	75	0
A0005	22.55888775	113.9507227	65.5	0
A0006	22.55899906	114.2413174	75	0
A0007	22.54900371	113.9722597	65.5	1
A0008	22.56277351	113.9565735	65.5	0

附件二: 会员信息数据

会员编号	会员位置(GPS)	预订任务限额	预订任务开始时间	信誉值
B0001	22.947097 113.679983	114	6:30:00	67,997.3868
B0002	22.577792 113.966524	163	6:30:00	37,926.5416
B0003	23.192458 113.347272	139	6:30:00	27,953.0363
B0004	23.255965 113.31875	98	6:30:00	25,085.6986
B0005	33.65205 116.97047	66	6:30:00	20,919.0667

附件三: 新项目任务数据

任务号码	任务 GPS 纬度	任务 GPS 经度
C0001	22.73004117	114.2408795
C0002	22.72704287	114.2996199
C0003	22.70131065	114.2336007
C0004	22.73235925	114.2866672
C0005	22.71839144	114.2575495
C0006	22.75392493	114.3819253

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：2160-7311，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：mm@hanspub.org