

基于可信工人路径规划的声望检验算法

张小欣, 刘佳旭, 蔡文广, 刘超, 张贺尧

(辽宁工程技术大学软件学院, 辽宁葫芦岛 125105)

摘要:针对工人自选择任务模式下无法确定路径规划算法工人可信度的问题,提出了基于可信工人路径规划的声望检验算法。首先,根据欧氏距离与区域限制对任务进行 k 均值聚类,缩小可信工人候选任务集,从而降低路径规划算法的时间复杂度;其次,提出了利用权重 k 对可信工人路径优化的算法,该算法综合考虑可检验工人数量、距离成本和截止时间计算任务权值,基于任务权值最大化寻找最佳插入位置来建立和更新可信工人的执行路径;最后,通过采用抽查检验的方式对可信工人与普通工人提交数据对比的结果来建立工人声望模型,该模型使用了可信性、不可信性和不确定性三个参数的期望值来描述工人的可靠性。在模拟数据集和gMission真实数据集上对声望检验算法进行实验,结果表明提出的声望检验算法数据采集质量提升13%,路径规划算法在保证抽查质量的基础上,可信工人的旅行成本降低11%。

关键词:空间众包;聚类算法;路径规划;声望系统;工人自选择任务模式

中图分类号:TP301.6

文献标识码:A

文章编号:1673-629X(2024)06-0016-07

doi:10.20165/j.cnki.ISSN1673-629X.2024.0078

Reputation Check Algorithm Based on Mobile Security Agent Route Planning

ZHANG Xiao-xin, LIU Jia-xu, CAI Wen-guang, LIU Chao, ZHANG He-yao

(School of Software, Liaoning Technical University, Huludao 125105, China)

Abstract:To address the problem of not being able to determine the worker validation of route planning algorithms in the worker self-selection task model, a reputation check algorithm based on mobile security agents is proposed. Firstly, k -mean clustering of tasks based on Euclidean distance with region restriction narrows the set of mobile security agent candidate tasks, which reduces the time complexity of the route planning algorithm. Secondly, an algorithm for optimizing the route of mobile security agents by using the weight k is proposed. The algorithm comprehensively considers the number of checkable workers, distance cost and deadline to calculate the task weight, and finds the best insertion position based on the maximization of task weight to establish and update the execution route of mobile security agents. Finally, the worker reputation model is established by comparing the results of the data submitted by the mobile security agents and the ordinary workers by means of random inspection. The model uses the expectations of the three parameters of believability, disbelievability and uncertainty to describe the reliability of the workers. Experiments are conducted on the synthetic dataset and the gMission real dataset for the reputation check algorithm, which show that the data collection quality of the reputation check algorithm proposed is improved by 13%, and the route planning algorithm reduces the travel cost of mobile security agents by 11% on the basis of ensuring the quality of sampling.

Key words: spatial crowdsourcing; clustering algorithm; route planning; reputation system; worker self-selection task model

0 引言

近年来,随着移动互联网的迅速发展,美团、滴滴打车等空间众包应用改变了人们传统的衣食住行模式。相对于这些流行的服务器分配任务模式^[1](Server Assigned Task, SAT)下的空间众包应用,工人自选择任务模式^[2](Worker Selected Task, WST)应用

也不断涌现。但由于受限于WST模式自身的局限性以及缺少有效声望系统的支撑,导致数据的采集质量难以得到保障,无法像SAT空间众包应用一样能够快速得到推广和普及。因此,为WST模式建立有效的声望系统,对于空间众包领域的研究具有非常重要的意义。

收稿日期:2023-09-14

修回日期:2024-01-16

基金项目:辽宁省教育青年项目(LJ2019QL022)

作者简介:张小欣(1998-),女,硕士研究生,CCF会员(L7220G),研究方向为空间众包、数据挖掘;通信作者:刘佳旭(1979-),男,硕导,博士,研究方向为空间众包、空间-时间数据库、网络安全。

目前,大部分声望系统的研究都集中在众包领域。Bhattacharjee 等^[3]提出了 QnQ (Quantity and Quality) 声望模型,根据工人的声望得分来区别不同的工人类别,如诚实、自私或恶意。Huang 等^[4]提出了一种基于线性判别分析的随机森林众包参与者声望评价模型 (Random Forest based on Linear Discriminant Analysis, LDA-RF)。Zhang 等^[5]通过模型质量参数计算信誉值来评估工人的可靠性。Sun 等^[6]介绍了利用区块链技术进行分布式信任和声望建模。Restuccia 等^[7]定义了一组危害现有应用程序可靠性的攻击并提出了一个可扩展且安全的基于信任的模型 FIDES。Yu 等^[8]提出了一种新的声望系统,该系统使用 Gompertz 函数计算声望分数以反映贡献数据的可信度。上述研究局限于传统的众包领域,其研究方法主要依赖于问卷调查、抽检和专家手动测试等方式来获得数据,用于对比工人结果的标准答案通常是预先确定的,从而平台可以将工人的结果与标准答案进行对比,以快速、低成本地建立工人声望模型。然而,空间众包领域并没有预先确定的标准答案,这导致对于工人提交的时空数据进行收集和验证变得更加复杂。鉴于工人可能提供虚假信息,平台可能需要派遣人员前往任务地点采集数据,以验证这些数据的真实性。

路径规划是一个任务分配的子问题^[9],任务分配以平台收益最大化、旅行成本最小化以及工人等待时间最小化的目标为主,采用的方法包括贪心^[10]、流算法^[11]和分支定界^[10-11]等。任务分配问题主要关注分配的实时性,根据对实时性要求的不同可以分为离线算法^[11-12]和在线算法^[10],而路径规划问题关注的是实体之间的关联性。根据关联关系可以分为单工人多任务的路径规划^[12]和多工人多任务的路径规划^[11,13]。Alotaibi 等^[14]开发了一种基于信任感知的多目标元启发式优化的安全聚类与路由规划技术。Bouatouche 等^[15]使用基于禁忌抓取的方法,在尊重某些约束的情况下,通过最大化工人的奖励,找到一条通往工人的最佳路线。Liu 等^[16]在车联网场景下创新性地引入竞价机制,提出了一种新的动态路径规划方法。上述研究中都缺乏对声望系统的关注,默认工人完成任务时是可信的,因此导致数据采集质量差、用户满意度低的问题。

针对上述问题,该文提出了一种基于可信工人路径规划的声望检验算法 (Mobile Security Agent - Reputation Checking, MSA - RC), 引入了可信工人 (Mobile Security Agent, MSA), MSA 是可信工人的统称,可以由固定的专职人员、高声望水平的众包工人或者无人机充当,MSA 与普通众包工人一起执行任务,根据任务提交的数据结果对比来判断工人的可靠性。

为了降低算法成本,提出了可信工人的路径优化的算法 (Cluster Division Route Planning, CDRP)。该算法综合考虑可检验工人数量、距离成本和截止时间计算任务权值,按照任务与可信工人的欧氏距离进行聚类。然后,在每个簇内,对任务的权值进行排序,并根据插值收益最大化的原则为工人生成规划路径。

1 问题定义

定义 1(工人):形式为 $\langle w, l_w \rangle$ 的工人 w 是移动设备的载体,自愿执行空间任务。其中 l_w 代表工人在 2D 区域内所处的位置。

定义 2(可信工人(MSA)):形式为 $\langle m, l_m, g_m \rangle$ 的可信工人 m 默认是完全可信的,可以是固定的专职人员、高声望水平的众包工人或者无人机等,是用于验证工人提交数据质量的移动设备的载体,即 MSA 和工人一起完成任务,从而工人的可靠性通过比较结果获得,与位置 l_m 和空间区域 g_m 相关联。

定义 3(空间任务):形式为 $\langle s, l_s, d_s, g_s, k \rangle$ 的空间任务 s , 其中 l_s 代表任务在 2D 区域内所处的位置, g_s 代表任务可以被完成的区域限制即以任务 s 为圆心指定长度为半径的圆形区域, d_s 代表截止时间, k 代表任务的优先级,与周围工人数量、距离成本和截止时间相关。

定义 4(声望水平 ω):向量 $\omega = [b, d, u]$ 满足 $b + d + u = 1$, 其中 $\{b, d, u\} \in [0, 1]^3$, b, d, u 分别对应可信、不可信和不确定因素。工人 w_i 的声望水平初值为:

$$\rho_i^0 = [r_0, 1 - r_0, 0] \quad (1)$$

其中, r_0 为工人的可信度, $1 - r_0$ 为不可信度。

该文提出的算法成本主要集中在 MSA 上,所以需要为 MSA 的路径进行规划,规划的目标是路径收益最大化,针对这一目标引入了该文的问题,具体定义如下:

定义 5(路径收益最大化问题, MRBT):给定实例集 $I_t = \langle M, S \rangle$, MRBT 的解决方案是为全部 MSA 找到一个路径规划 $P = \{R_{m_1}, R_{m_2}, \dots, R_{m_n}\}$ 使收益最大化,具体可以表示为:

$$\max \{V_{R_1} + V_{R_2} + \dots + V_{R_n}\} \quad (2)$$

其中, $V_R = \sum_{s \in R} k_s$, 具体地 V_R 代表路径 R 中任务产生的收益, k_s 代表任务 s 的权值。所有 MSA 的路径都需要满足每个任务的空间约束和截止时间约束,即:

$$\begin{cases} \text{set}(R_{m_i}) \in g_{m_i}, 1 \leq i \leq |M_t| \\ a_s \leq d_s, s \in \bigcup_{i=1}^p \text{set}(R_{m_i}) \end{cases} \quad (3)$$

MRBT 的目的是为 MSA 找到一条路径规划 P 使

任务收益最大化,满足:

- (1) 任务周围可验证工人的数量最大化;
- (2) 任务的平均旅行成本之和最小化;
- (3) 任务的截止时间最小化。

MRBT 是 NP 难的,对于单个工人与单个任务之间的最大任务调度问题^[14] (Maximum Task Scheduling, MTS) 已经被证明是 NP 难的,文中的多个工人与多个任务之间的最大任务调度问题也是 NP 难的。

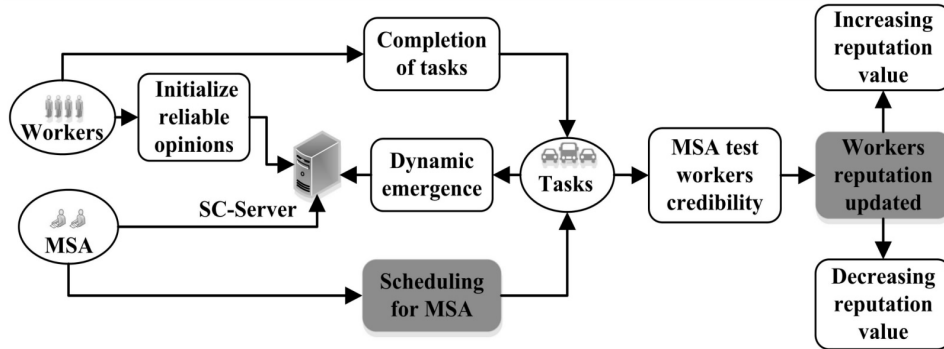


图 1 算法框架

2.1 路径规划模块

路径规划模块主要是为每个 MSA 规划一条最优的完成任务路径,每个任务由三个参数(位置、截止时间和周围可检验工人数量)来描述。

为了综合考虑任务的多个因素,从而为 MSA 确定最优的任务完成顺序,平台借助区域熵的概念为每个任务权值赋值。区域熵是用来衡量某个区域内工人分布情况的指标,即工人总数和工人未来访问该区域的相对比例。区域熵的公式可以具体表示为:

$$\text{Entropy}(l) = - \sum_{w \in W_l} P_l(w) \times \log P_l(w) \quad (4)$$

其中, $|O_l|$ 为访问区域 l 的总数, $O_{w,l}$ 为访问区域 l 的工人 w 总数。从 O_l 中属于 $O_{w,l}$ 的一个随机的概率为

$$P_l(w) = \frac{|O_{w,l}|}{|O_l|}.$$

利用区域熵提出了任务权值 k 的具体计算公式如下:

$$k = -\alpha \cdot \sum_{w \in g_s} (P_l(s) \times \log P_l(s)) + \beta \cdot (P_l(s) \times \log P_l(s)) + \gamma \cdot (P_d(s) \times \log P_d(s)) \quad (5)$$

其中, α, β, γ 代表归一化系数,用于平衡任务的周围可检验工人的数量、旅行成本和截止时间。公式 1 中的变量具体计算公式如下:

$$P_s(w) = \frac{|O_{w,s}|}{|O_s|} \quad (6)$$

其中, $O_{w,s}$ 和 O_s 分别代表任务区域 g_s 和可信工人区域 g_m 内的工人集合的数量, $P_s(w)$ 代表同时属于 O_s 和

2 CDRP 算法与 MSA-RC 算法

基于可信工人路径规划的声望检验算法框架如图 1 所示,该框架包括三种角色:任务、工人和 MSA。首先平台初始化工人的声望水平,当任务动态出现时信息被上传到平台;接下来 MSA 与工人一起完成任务,为了减少成本,对 MSA 的路径进行规划,并将 MSA 与工人的结果对比确认工人的可靠性;最后根据检验结果对工人的声望值进行更新。

$O_{w,s}$ 的工人概率。

$$P_l(s) = \frac{l}{\|l_s - l_a\|_2} \quad (7)$$

其中, l 代表在可信工人当前位置和范围内所有任务距离之和, $\|l_s - l_a\|_2$ 代表任务 s 与可信工人当前位置 l_a 之间的欧氏距离。

$$P_d(s) = \frac{t_{\max}}{d_s - a_s} \quad (8)$$

其中, t_{\max} 代表要完成的全部任务的截止时间和当前时间的最大差值, d_s 为任务 s 的截止时间, a_s 代表当前时间。

路径规划的目的是为 MSA 规划出同时满足截止时间、距离成本以及周围可检验工人数量最优的完成任务路线,路径规划模块具体分为两个阶段:

(a) 聚类匹配:为了降低路径规划的时间复杂度,在聚类匹配算法中,采用了 k 均值聚类算法,根据任务与 MSA 之间的欧氏距离是否小于 MSA 的有效时空区域距离限制,在区域内的任务被聚成簇用来代表的是每个 MSA 可完成的任务集,即 $S_m = \{s_1, s_2, \dots, s_n\}$ 。

(b) 动态插入:对于动态到达的任务,将聚类匹配后的 MSA 与任务序列的每个插入位置分别根据公式 5 计算权值,并求出最大权值差,找到动态任务的最佳插入位置同时一边插入一边检查时间的可行性,可行即插入,反之则舍弃,最后更新任务序列。

插入可行性:动态插入新的任务时,并不是所有的任务都能插入到现有任务序列中。具体原因是,当插入当前任务序列时,后续任务可能无法在截止时间内

完成,导致插入任务后匹配的任务数量不会增加。具体来说,在部分可行路线上的 s_{j-1} 和 s_j 之间插入一个新任务 s 的必要条件 $R_m = \{s_1, s_2, \dots, s_n\}$ 是(假设每个任务完成的时间为 0):

$$t_d + a(s_i) \leq d_{s_i} \quad (9)$$

其中, $j \leq i \leq m$, t_d 表示插入任务 s 所造成的旅行延迟,即 $t_d = c(s, s_{j-1}) + c(s_j, s) - c(s_{j-1}, s_j)$ 。

算法 1:CDRP 算法。

输入: Mobile security agent m , clusters p and tasks set $S = \langle s_1, s_2, \dots, s_n \rangle$ and worker reputation ρ_i^j .

输出: The updated route R_m and Reputation Level ρ_i^j .

```

1: With  $p$  mobile security agents as the initial mean pointers
 $\langle m_1, m_2, \dots, m_p \rangle$ 
2: For ( $j < n$ ) {
3:   Calculate the task  $s_j$  and the Euclidean distance of each
mean variable  $m_i: c_{ij} = |s_j - m_i|_2$ 
4:   If ( $c_{ij} < g_m$ )
5:     Tasks are assigned to the appropriate cluster  $S_{m_i} =$ 
 $S_i \cup s_j$ 
6:   }
7: Repeat
8:    $\delta_{\max} \leftarrow 0$ 
9:   For (all unscheduled tasks  $s \in S$ ) {
10:    For (all insertion positions  $\text{pos} \in [0, |R_m|]$ ) {
11:     If (position  $\text{pos}$  can be inserted) {
12:       $\delta \leftarrow k(R_m \cup s, m) - k(R_m, m) / * k(\cdot)$  函数代表将
变量带入到公式 5 中计算权值 */
13:      If ( $\delta > \delta_{\max}$ ) {
14:         $\delta_{\max} \leftarrow \delta$ 
15:        the best insertion position and task  $\text{pos}^* \leftarrow$ 
 $\text{pos}, t^* \leftarrow s$ 
16:      }
17:    }
18:  }
19:  }
20:  Insert  $t^*$  into  $R_m$  at the  $\text{pos}^*$  place
21:  remove  $t^*$  from  $S$ 
22: Until no more task could be inserted into  $R_m$ 
23: Return the updated route  $R_m$ 

```

算法 1 给出了 MSA-RC 算法的伪代码描述,首先输入可信工人 m 、 p 个聚类中心、 n 个任务组成的任务集 S 和工人原有的声望值。第 2~6 行循环比较每个 MSA 与任务之间的欧氏距离 $c(m_i, s_j)$, 如果距离小于该 MSA 的区域限制 g_m , 则将该任务 s_j 聚类到 m_i 中, 否则继续循环, 直到任务全部比较结束, 并返回可信工人 m_i 所对应的任务集 $S_{m_i} = \{s_1, s_2, \dots, s_n\}$ 。第 8 行初始化最大权值差 δ_{\max} , 第 9, 10 行循环检验每个未被调度的任务 $s \in S_m$ 和可以被插入的 $|R_m| + 1$ 位置。第 11, 12 行首先利用插入可行性进行判断, 找到可以插

入的位置并调用公式 5 计算新插入任务后的 k 值与原 k 值的增量。第 13~19 行如果增量大于当前最大权值差, 则将该增量赋给 δ_{\max} , 当前位置为最佳插入位置。第 20~22 行将任务 t^* 插入到任务序列 R_m 的 pos^* 位置, 并将其从原任务序列 S 中移除, 当没有任务可以被插入时循环结束。最后返回更新后的任务序列 R_m 。

2.2 声望检验模块

声望检验模块是声望检验算法的另一个核心模块, 该模块主要是用于为工人的声望水平进行建模, 每个工人的声望水平由可信、不可信和不确定因素的期望值来描述。由于成本限制, MSA 相对于普通工人的数量存在较大的差距, 而抽样检验的方式又无法完全覆盖所有普通工人, 因此存在一部分工人在一段时间内无法及时接受检验, 从而产生一定程度的不确定性。为了解决工人声望水平的不确定性难以度量的问题, 采用了 JφSang's 信任模型进行建模和描述。在这个模型中, 使用包含信任、不信任和不确定因素的向量来描述实体的声望水平。

为了将公式 1 中的声望投影到一维概率空间并产生概率分布值, 将声望水平的 ω 期望值 $E\{\omega\}$ 定义为:

$$E\{\omega\} = \frac{b + u}{b + d + 2u} \quad (10)$$

平台对工人的可靠性进行初始化后, 根据 MSA 提交检验工人的可靠性结果进一步更新该工人的声望水平, 具体的更新公式如下。

结果一致: 在这种情况下, MSA 检验出工人在本次检测中诚实地完成任务, 并向平台发送了正确的信息。则提高工人的可信值并降低不可信与不确定因素, 工人声望更新公式如下:

$$\rho_i^j = [(b_i^j)^{\alpha}, (d_i^j)^{2-\alpha}, (u_i^j)^{2-\alpha}] \quad (11)$$

其中, α 是一个恒定的学习率因子。详细地说, α_b 代表可信程度通过提高而被“强化”, 而其他两项 $2 - \alpha_b$ 代表降低不可信与不确定而被“弱化”。

结果不一致: 与前面的情况相反, 工人被检测出没有诚实地完成任务, 并向平台发送了错误的信息, 则更新公式如下:

$$\rho_i^j = [(b_i^j)^{2-\alpha}, (d_i^j)^{\alpha}, (u_i^j)^{2-\alpha}] \quad (12)$$

如果没有可用的 MSA, 即在本次检测回合未检测出该工人的声望等级, 则提高了工人的不确定因素同时降低了可信与不可信, 具体公式如下:

$$\rho_i^j = [(b_i^j)^{2-\alpha}, (d_i^j)^{2-\alpha}, (u_i^j)^{\alpha}] \quad (13)$$

最后, 计算工人的声望最终意见的期望值, 如下式表示。

$$r_i^j = E\{\rho_i^j\} \quad (14)$$

算法 2:MSA-RC 算法。

输入: The authenticity of Scheduling results a and worker

reputation ρ_i^j 。

输出: The Updated Reputation Level ρ_i^j 。

1: If($a = 1$) {

2: Consistent results ρ_i^j

3: Calculate the final expected value by using Equation 11

4: }

5: If($a = 0$) {

6: Inconsistent results ρ_i^j

7: Calculate the final expected value by using Equation 12

8: else {

9: Not identified ρ_i^j

10: Calculate the final expected value by using

Equation 13

11: }

12: }

13: Return Get the reward value based on the final expected

value by using Equation 14

算法 2 给出了 MSA-RC 算法的伪代码描述, 首先输入工人原有的声望值并输出更新后的声望值。第 1~12 行对工人 w_i 进行循环判断, 第 1~4 行如果 $a=1$ 即工人 w_i 被 MSA 检验是诚实的, 则将信任增加, 不确定与不信任减小; 第 5~7 行如果 $a=0$ 即工人 w_i 被 MSA 检验是不诚实的, 则将不信任增加, 不确定与信任减小; 否则, 第 8~12 行如果工人 w_i 没有被检验, 则该工人的不确定增加, 同时信任与不信任减小, 直到所有工人检验完成则结束循环, 第 13 行将更新后的参数代入公式 14 计算出更新后工人的声望水平。

2.3 时间复杂度分析

在整个优化过程中, 所提出的可信工人路径规划的声望检验算法的时间复杂度主要由两部分组成: 路径规划和声望检验。其对应的复杂度分别为 $O(np \cdot |S| \cdot |R_m|^2)$ 以及 $O(1)$, 其中 n 代表任务的数量, p 代表任务簇集的数量, S 代表未调度的任务, R_m 代表任务序列。因此, 可信工人路径规划的声望检验算法的时间复杂度可以表示为 $O(np \cdot |S| \cdot |R_m|^2)$ 。

3 实验

3.1 实验数据

该文使用了真实数据集 gMission 进行实验评估。gMission 是一个通用的空间众包平台。在 gMission 数据集中, 每个任务都包括任务描述、位置、截止时间(分钟), 而每个工人都与位置、截止时间(分钟)、最大活动范围(公里)以及其历史任务完成成功率相关联。需要注意的是, 该数据集中并未提供有关 MSA 的相关信息, 因此根据每个工人的声望水平, 选择了声望水平排名前 10% 的工人作为 MSA。表 1 提供了真实数据集的统计信息, 其中 $|S|$, $|W|$ 和 $|M|$ 分别代表任

务数、工人数和可信工人数。除了真实数据集外, 该文还使用了模拟数据集来评估不同算法的性能。实验参数设置如表 2 所示, 其中默认参数已用粗体标出, 包括模拟改变参数 α , β , γ , α_b 和 α_u 对权值 k 与声望水平的影响。

表 1 真实数据集

平台	S	W	M
gMission	17 113	5 312	1 000

表 2 参数设置

参数	设定
S	1k, 2k, 3k, 4k, 5k
M	200, 400, 600 , 800, 1k
α	0.2, 0.4, 0.5, 0.6
β	0.1, 0.2 , 0.3, 0.6
γ	0.1, 0.2 , 0.3, 0.6
α_b	0.1, 0.3 , 0.6, 0.8
α_u	0.9 , 0.7, 0.4, 0.2

3.2 实验方法

为了保证实验公平, 所有对比算法均采用默认的参数值, 例如, 使用 1k, 10k 等不同数量的任务和其他默认参数进行评估。该文将每个任务与 MSA 的位置映射到道路网络中, 根据 MSA 的区域限制对任务进行聚类分区, 并对任务进行路径规划, 通过对比 MSA 与工人完成任务的结果得到工人的可信度, 当 MSA 规模越大, 工人的可信度判断越准确。该文采用总旅行成本(简称成本)指标来评估 CDRP 算法的性能, 同时, 采用任务完成质量(简称质量)指标来评估 MSA-RC 算法的效果。所有算法均在 Visual Studio 2022 上使用 C++ 实现, 并在 12th Gen Intel(R) Core(TM) i5-12500H 2.50 GHz 的服务器和 16 GB 主存上进行实验, 在相同实验设置下, 用 C++ 生成 20 次不同位置的任务初始数据, 并报告同一设置下的平均值。

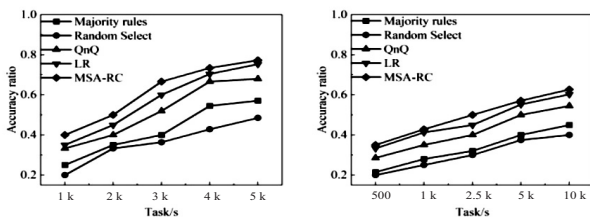
3.3 实验结果与分析

通过和 Greedy、DeDp^[17]、BES^[14] 和 PR-share^[18] 等方法进行比较来评估 CDRP 算法的性能。Greedy 是在路径规划中每次贪婪地选择局部最优的任务优先完成。DeDp 是一种动态规划算法, 与 Greedy 不同的是该算法考虑每次选择全局最优的任务。BES 利用秃鹰算法模拟了秃鹰捕食的三个阶段, 选择区域、搜索猎物、俯冲捕猎。随着迭代次数的增加, 秃鹰会改变其搜索方式。PR-share 是利用概率路由调度算法先对任务进行聚类后对路线进行分区。此外, 该文还通过与 Majority rules、Random Select、QnQ^[19] 和 LR^[20] 等方法进行比较来验证 MSA-RC 算法的性能。Majority rules 是多个工人完成同一任务利用多数原则的选取

最终的结果。Random Select 多个工人完成同一任务的最终结果通过随机选择的方法确定最终结果。QnQ 是一个事件信任和用户声誉模型,以隔离不同的用户类别,用户声誉分数是基于他们贡献的“质量”和“数量”。LR 是基于机器学习的声誉系统,通过提取一组对消费者可靠性有很大影响的因素来构建一个新的消费者档案数据集,这些因素作为机器学习算法的输入。

3.3.1 任务数量变化对任务完成质量的影响

为了验证 MSA-RC 算法的有效性,图 2 描绘了在真实数据集上改变任务数量对任务完成质量的影响,图中横轴 Task(s) 代表任务的量,纵轴表示任务完成质量 quality,任务完成质量用实验中被认定为诚实的工人占全部工人的比例来衡量。由图 2(a) 与图 2(b) 可以看出,随着任务数量的增加,任务完成质量呈现递增的趋势。这是因为随着任务数量的增加,覆盖工人数量也相应增加,这意味着更多工人可以接受检验,从而提高了任务完成质量。由于 Majority rules 与 Random Select 没有建立声望模型,所以相对于其他三种建立声望系统的方法完成任务的质量较低,同时 MSA-RC 的任务完成质量优于 QnQ 和 LR。例如:在真实数据集任务数量为 1k 的情况下, Majority rules、Random Select、QnQ、LR 和 MSA-RC 的任务完成质量分别为 0.28, 0.25, 0.35, 0.412 3 和 0.438 6。与上述 4 种算法相比, MSA-RC 算法的任务完成质量分别提高了 15.66%, 17.54%, 12.53% 和 10.64%。这是由于 MSA-RC 使用 MSA 对工人的可信程度进行检验,同时考虑了可信、不确定与不可信对实验的影响,可以有效地提高任务完成质量。



(a) 模拟数据的质量 (b) 真实数据的质量

图 2 模拟数据与真实数据的任务完成质量变化

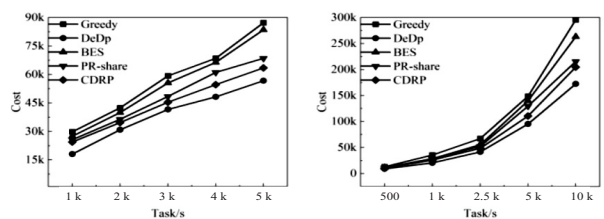
具体的实验结果见表 3。

表 3 真实数据的质量变化

Task(s)	Majority rules	Radom Select	QnQ	LR	MSA-RC
500	0.215	0.2	0.286	0.333 3	0.35
1k	0.28	0.25	0.35	0.412 3	0.438 6
2.5k	0.32	0.3	0.4	0.44	0.5
5k	0.4	0.375	0.5	0.542 5	0.571 4
10k	0.45	0.4	0.545 4	0.582 3	0.626 7

3.3.2 任务数量变化对成本的影响

图 3 中描绘了在模拟数据集和真实数据集上改变任务数量对旅行成本的影响。图中横轴 Task(s) 代表任务的量,纵轴 Cost 代表旅行成本。由图 3(a) 与图 3(b) 可以看出,随着任务数量的增加 MSA 需要完成更多的任务,所以成本呈现快速递增的趋势。Greedy 虽然每次选择最优的任务完成,但无法保证全局的成本导致成本较高。由于 DeDp 采取全局最优的算法每次选择的任务是基于全局成本综合考虑,从而相对于其他四个算法时间花费最少成本。CDRP 与 BES 和 PR-share 相比具有较低的距离成本,例如:在真实数据集任务数量为 1k 的情况下, Greedy、DeDp、BES、PR-share 和 CDRP 的成本分别为 35 506.2, 20 379.7, 28 797.4, 27 896 和 26 068.9。与 BES 和 PR-share 算法相比, CDRP 分别降低了 11.04% 和 10.70%, 这是由于 CDRP 利用权值 k 对簇内的 MSA 进行规划具有较低的成本,而其他算法完成任务时预先无法获得最优路径产生绕道成本,具体的实验结果见表 4。



(a) 模拟数据的成本 (b) 真实数据的成本

图 3 模拟数据与真实数据的成本变化

表 4 真实数据的成本变化

Task(s)	Greedy	DeDp	BES	PR-share	CDRP
500	12 937.354	9 078	11 384.6	11 163.5	10 206.9
1k	35 506.2	20 379.7	28 797.4	27 896	26 068.9
2.5k	67 258.4	41 650.98	55 593.6	52 496.3	48 458.3
5k	148 776.35	95 232.1	139 802.6	129 680.6	110 689.2
10k	295 790	172 517.5	263 357	215 481.2	204 583.6

4 结束语

为了解决工人自选择任务模式下无法确定路径规划算法工人可信度的问题,提出了一种基于可信工人路径规划的声望检验算法。(1)利用 k 均值聚类的方法将任务进行聚类,生成基于 MSA 的可完成任务集。(2)利用权值 k 对可信工人路径优化,确定最优任务完成路径。(3)利用抽查检验的方式将 MSA 与工人提交的数据结果进行对比,获取最终声望值。通过在模拟数据集和真实数据集上进行比较,验证了该算法不仅能有效地降低 MSA 完成任务的时间复杂度,还能进一步提高工人完成任务的质量。但是实验中的 MSA 选择存在一定随机性,未来将侧重对于 MSA 选择的算法优化。

参考文献:

- [1] KAZEMI L, SHAHABI C. A privacy-aware framework for participatory sensing[J]. ACM SIGKDD Explorations Newsletter, 2011, 13(1): 43-51.
- [2] BULUT M F, YILMAZ Y S, DEMIRBAS M. Crowdsourcing location-based queries[C]//2011 IEEE international conference on pervasive computing and communications workshops (PERCOM workshops). Beijing: IEEE, 2011: 513-518.
- [3] BHATTACHARJEE S, GHOSH N, SHAN V K, et al. QnQ: a reputation model to secure mobile crowdsourcing applications from incentive losses[C]//2017 IEEE conference on communications and network security (CNS). Tokyo: IEEE, 2017: 1-9.
- [4] HUANG Y, CHEN M. Improve reputation evaluation of crowdsourcing participants using multidimensional index and machine learning techniques[J]. IEEE Access, 2019, 7: 118055-118067.
- [5] ZHANG Q, DING Q, ZHU J, et al. Blockchain empowered reliable federated learning by worker selection: a trustworthy reputation evaluation method[C]//2021 IEEE wireless communications and networking conference workshops (WCNCW). Austin: IEEE, 2021: 1-6.
- [6] SUN L, YANG Q, CHEN X, et al. RC-chain: reputation-based crowdsourcing blockchain for vehicular networks[J]. Journal of Network and Computer Applications, 2021, 176: 102956.
- [7] RESTUCCIA F, DAS S K. FIDES: a trust-based framework for secure user incentivization in participatory sensing[C]//Proceeding of IEEE international symposium on a world of wireless, mobile and multimedia networks 2014. Beijing: IEEE, 2014: 1-10.
- [8] YU Y, LIU S, GUO L, et al. CrowdR-FBC: a distributed fog-blockchains for mobile crowdsourcing reputation management[J]. IEEE Internet of Things Journal, 2020, 7(9): 8722-8735.
- [9] TONG Y, ZHOU Z, ZENG Y, et al. Spatial crowdsourcing: a survey[J]. The VLDB Journal, 2020, 29: 217-250.
- [10] ASGHARI M, SHAHABI C. On on-line task assignment in spatial crowdsourcing[C]//2017 IEEE international conference on big data (big data). Osaka: IEEE, 2017: 395-404.
- [11] DENG D, SHAHABI C, DEMIRYUREK U, et al. Task selection in spatial crowdsourcing from worker's perspective[J]. GeoInformatica, 2016, 20: 529-568.
- [12] COSTA C F, NASCIMENTO M A. Online in-route task selection in spatial crowdsourcing[C]//Proceedings of the 28th international conference on advances in geographic information systems. Seattle: ACM, 2020: 239-250.
- [13] ZHENG L, CHEN L, YE J. Order dispatch in price-aware ridesharing[J]. Proceedings of the VLDB Endowment, 2018, 11(8): 853-865.
- [14] NAGAPPAN K, RAJENDRAN S, ALOTAIBI Y. Trust aware multi-objective metaheuristic optimization based secure route planning technique for cluster based IIoT environment[J]. IEEE Access, 2022, 10: 112686-112694.
- [15] BOUATOCHE M, BELKADI K. GRASP-tabu search algorithms for the route planning problem in spatial crowdsourcing[J]. International Journal of Applied Metaheuristic Computing, 2022, 13(1): 1-18.
- [16] LIU B, LONG J, DENG M, et al. An adaptive route planning method of connected vehicles for improving the transport efficiency[J]. ISPRS International Journal of Geo-Information, 2022, 11(1): 39.
- [17] SHE J, TONG Y, CHEN L. Utility-aware social event-participant planning[C]//Proceedings of the 2015 ACM SIGMOD international conference on management of data. Melbourne: ACM, 2015: 1629-1643.
- [18] 蔡文广, 刘佳旭, 张小欣. 基于概率路由的出租车共乘调度算法[J]. 计算机应用研究, 2022, 41(2): 1-6.
- [19] BHATTACHARJEE S, GHOSH N, SHAN V K, et al. QnQ: quality and quantity based unified approach for secure and trustworthy mobile crowdsensing[J]. IEEE Trans Mob Comput, 2020, 19: 200-216.
- [20] ALQWADRI A, AZZEH M, ALMASALHA F. Application of machine learning for online reputation systems[J]. International Journal of Automation and Computing, 2021, 18(3): 492-502.