



研究与开发

联合区域热度和社交属性感知的 移动群智感知参与者选择机制

向罗勇, 陈文, 张陆洋

(重庆信息通信研究院, 重庆 401336)

摘要: 针对群智感知中的平台在用户稀疏区域获取的任务数据可靠性低且任务难以按时完成的问题, 提出一种联合区域热度和社交属性感知的参与者选择机制。首先, 考虑不同区域热度对任务完成程度的影响, 根据区域活跃用户数、用户平均停留时间及区域历史感知任务完成情况评估区域热度。其次, 为了分析用户社交属性对任务完成程度的影响, 结合用户的状态信息和用户历史感知任务记录计算用户意愿度、信誉度以及活跃度; 综合考虑上述因素, 分别为高热度和低热度区域设计了以最大化任务完成质量和最大化任务完成数量两种不同的社交属性感知的参与者选择机制。结果表明, 所提机制能够显著提升总体数据质量, 在低热度区域也可以及时可靠地完成感知任务。相比于 SUR 和 GGA-I 两种算法, 失败率分别降低了 66.7%和 50.6%。

关键词: 移动群智感知; 参与者选择; 区域热度; 用户特征

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2020050

Regional heat and social attribute aware participant selection mechanis in mobile crowd sensing

XIANG Luoyong, CHEN Wen, ZHANG Luyang

Chongqing Academy of Information and Communications Technology, Chongqing 401336, China

Abstract: Aiming at the problem of tasks that are low reliability acquired by platform and difficult to accomplish on time in user sparse area. A participant selection mechanism that combines regional heat and social attribute aware was proposed. Firstly, considering the influence of different regional heat on task completion, the regional heat was evaluated according to the number of active users, the average residence time of users and the completion of historical task. Secondly, in order to analyze the impact of user social attributes on task completion, the user willingness, reputation and activity were calculated by combining the status information of users and the historical task record of users. Finally, by taking the above factors into account, two different mechanisms of participant selection for social attribute perception were designed for high and low heat areas to maximize quality and number of task completion respectively. The results show that the proposed mechanism can significantly improve the overall data quality, and can also perform sensing tasks in sparse areas on time. Meanwhile, compared with SUR and GGA-I, the failure rate is reduced by 66.7% and 50.6% respectively.

Key words: mobile crowd sensing, participant selection, regional heat, social attribute

收稿日期: 2019-10-18; 修回日期: 2020-02-10

1 引言

随着无线通信的快速发展以及无线移动终端设备的爆炸式普及,移动感知和众包思想的结合产生了移动群智感知(mobile crowd sensing, MCS)范例^[1]。MCS将用户携带的移动终端设备作为基本感知单元,依托众多用户组成大规模感知网络,从而完成一些仅依靠个体很难实现的复杂感知任务^[2]。MCS中的服务平台将感知任务分配给选择的参与者,参与者利用携带的手机或者其他终端设备完成相应的感知任务。进而,通过移动蜂窝网络或短距离无线通信方式(如蓝牙、Wi-Fi等)将所感知的数据上传到服务平台。服务平台对所有获取的数据进行处理和分析,并以此实现智慧交通、环境监测和数字地图等MCS应用^[3]。

传统的无线传感器网络中,数据产生过程不依赖于参与者,而参与者仅作为最终的消费者应用感知数据进行决策^[4]。但在MCS系统下,参与者将会参与感知过程的各个环节,包括感知数据的收集、上传和应用决策等过程。其中,MCS系统的任务感知是以大量的参与者为基础的,在感知过程中,由于参与者对不同任务的积极性等因素的影响,导致每个参与者提供数据的可信度不相同。因此,为了获取可信度更高的数据,需要系统匹配最合适参与者来执行任务。若在任务分配中无法将任务与参与者进行准确的匹配,将直接降低任务完成的数量和质量,导致参与者提供的数据可信度下降,同时还会消耗平台更多的计算资源。因此,参与者选择是MCS应用中十分重要的问题^[5-6]。

服务平台在选择参与者时总有预算和时间限制^[7-8]。同时,服务平台需要有足够的用户参与感知任务。相关研究表明城市地区的用户分布服从独立的泊松散射,或者通过排斥散射^[9]或吸引^[10],从而导致在某一时刻大量用户聚集或离散在某一个区域,即高热度区域和低热度区域。在高热度

区域,多个用户在同一时间以一定的概率竞争相同的任务。然而,在低热度区域,往往不能确定用户到达间隔,从而任务完成时间也是不确定的。此外,对于以用户为中心的MCS,用户具有明显的社会特征。用户的参与意愿程度将影响其是否愿意参与,对于服务平台而言,将影响是否有足够的参与者完成其发布的感知任务;用户履行合约的能力代表其完成任务的概率,直接影响平台的任务完成率;用户在现实社区和物理社区的活跃状态对服务平台的任务执行效果也有较大的影响。在为用户分配感知任务之后,由于环境条件的变化可能会发生任务失败,并且难以保证自私移动用户提交的数据质量。用户信誉与传感数据的可靠性密切相关,并通过合理设计的模型进行评估^[11]。接收传感任务之后,高信誉的用户可以委派任务给其他用户,仍然可以得到报酬,从而导致不可靠的遥感数据^[12]。用户活动对服务平台的任务执行性能也有很大影响。具有较大区域活动的用户更有可能访问更多位置,从而能够完成更多的感知任务^[13]。此外,用户与来自不同社区的其他人交互以形成覆盖现实世界社区的传感网络。具有高水平社交活动的用户具有广泛的覆盖范围,更稳定可靠,并且可以完成更多的感知任务^[14-15]。因此,MCS参与者选择过程应充分考虑区域热点的差异和用户的社交属性。

针对上述问题,提出一种联合区域热度和社交属性感知的参与者选择机制。首先对城市环境中不同区域的热度进行了评估,分析了用户的意愿度、信誉度以及活跃度对服务平台获取数据质量的影响。同时,结合用户的意愿度和活跃度设计了一种新颖的任务排队策略,在此基础上,结合用户的社会网络关系设计了社区用户协助策略。最后,为高热度和低热度区域分别设计了两种不同的参与者选择机制,旨在预算和时间限制下,保证平台分配的任务都能被参与者及时可靠地完成,提升平台的服务质量。



2 系统模型

MCS 参与者选择过程的基本原理如图 1 所示, MCS 网络主要包括三类实体: 服务请求者、服务平台以及移动用户。服务请求者将关于某些区域的感知请求发送到服务平台; 服务平台接收感知请求之后通过发布感知任务收集感知区域的信息, 每个任务对需要的数据类型和截止时间具有明确的限制; 移动用户在接收任务消息后通过特定的应用软件向服务平台提交配置文件, 被服务平台选择的用户作为参与者完成感知任务, 之后将收集的感知数据传输至服务平台。服务平台还负责用户身份、任务偏好类型等信息的注册, 用户位置、投标等实时状态信息的获取以及对感知数据的收集、评估和分析。因此为了便于服务

平台处理各类繁琐的工作, 将服务平台根据职责分为 3 个部分: 信息管理模块 (information management module, IMM)、参与者选择模块 (participant selection module, PSM) 和数据服务模块 (data service module, DSM)。其中, IMM 负责用户注册、实时状态信息获取和用户信誉度的管理, PSM 负责感知任务的发布和参与者选择, DSM 负责服务请求的接收、任务数据报告的验证、整合和评估。

系统具体运行过程如下: 服务平台将感知请求进行整理、分类, 然后发布感知任务集合 $\mathcal{S}=\{\theta_1, \theta_2, \dots, \theta_j\}$, 集合中每个感知任务的预算为 $\mathcal{B}_j$, 当参与者完成该任务时获得的报酬为 $\gamma_{n_j}^{\theta_j}$, 用户获得的报酬不能超过相应的预算, 从而服务平台能够获得的收益 v_{θ_j} 为预算与报酬之差。任务

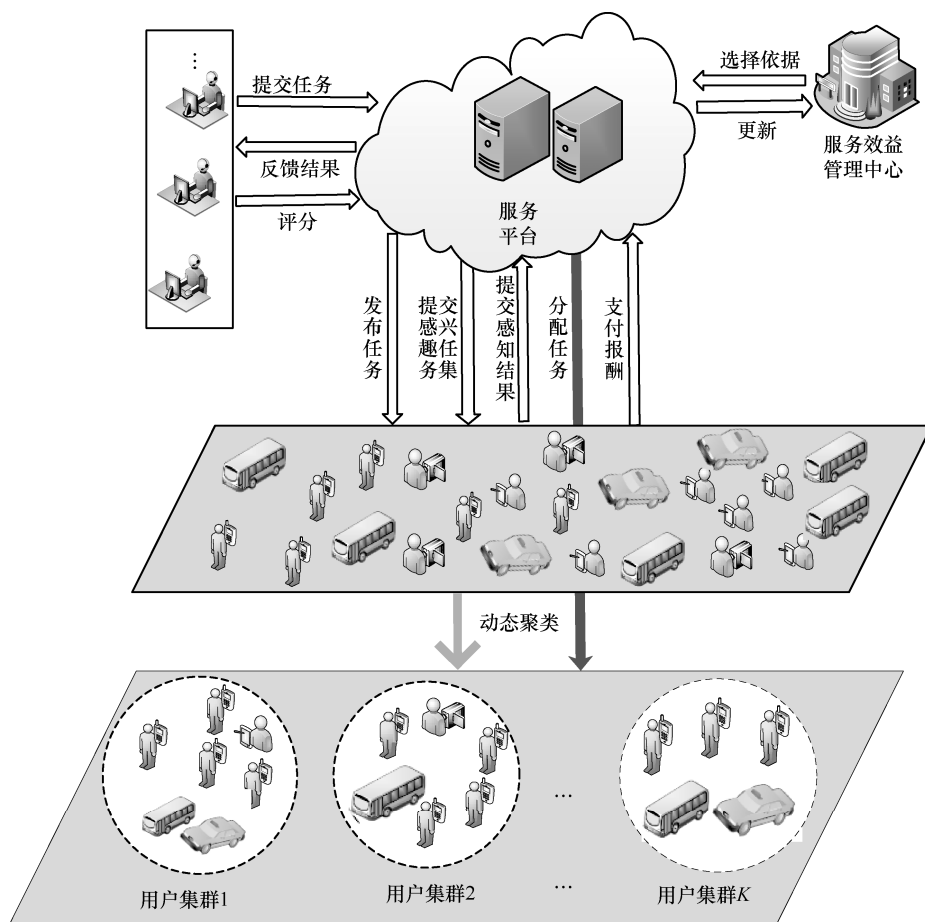


图 1 区域热度和社交属性感知的 MCS 系统模型

θ_j 的相关信息用多元组 $\langle l_j^{\theta_j}; \theta_j^{\text{typ}}; \theta_j^{\text{num}}; D_{\text{cover}}^{\theta_j}; t_{\text{max}}^{\theta_j} \rangle$ 表示, 多元组参数对应感知任务的位置、类型、需要的参与者数量、覆盖半径以及数据的截止接收时间。区域内接收任务消息后通过特定的应用软件向服务平台提交配置文件的用户集合表示为 $N = \{n_1, n_2, \dots, n_z\}$, 提交的配置文件表示为 $\langle l_{n_i}; b_{n_i}^{\theta_j}; p_{\text{data}}; v_{n_i} \rangle$, 分别代表其位置、投标、单位流量花费以及移动速度。IMM 根据用户实时状态信息评估的用户意愿度, 在此基础上, PSM 结合存储的信誉度以及投标信息在高热度区域选择取得最大综合数据质量的参与者, 并将选择的参与者集合表示为 $U = \{u_1, u_2, \dots, u_z\}$; 而在低热度区域选择及时完成感知任务并获得可靠数据的参与者, 选择的参与者集合表示为 $C = \{c_1, c_2, \dots, c_y\}$ 。

3 区域热度评估

将城市空间分成不同热度的感知区域, 并结合区域的活跃用户数、用户平均停留时间以及区域历史感知任务完成情况综合评估每个感知区域的热度。首先, 采用观察 T 时刻内该区域每个时刻的平均用户数量来表示活跃用户数 $\overline{P}_T$, 活跃用户数 $\overline{P}_T$ 如式 (1) 所示:

$$\overline{P}_T = \frac{\sum_{i=1}^T \text{total}_{t_i}^k}{Y_{\text{region}} \times T} \quad (1)$$

其中, $\text{total}_{t_i}^k$ 表示 t_i 时刻在区域 Y_{region} 的总用户数, T 表示观察的时刻总数, Y_{region} 表示所观察的区域面积。

所有用户 N 在 Y_{region} 的平均停留时间 $\overline{U}_{\text{dur}}(Y_{\text{region}}, L_k)$ 如式 (2) 所示:

$$\overline{U}_{\text{dur}}(N, Y_{\text{region}}) = \frac{\sum_{n_i \in N} [\mathcal{L}(n_i, Y_{\text{region}}) - \mathfrak{Z}(n_i, Y_{\text{region}})]}{N} \quad (2)$$

其中, $\mathfrak{Z}(n_i, Y_{\text{region}})$ 表示用户 n_i 的到达时间, $\mathcal{L}(n_i, Y_{\text{region}})$ 表示用户 n_i 的离开时间。

感知任务完成数量和完成比例都与区域热度

紧密相关。其中, 感知任务完成比例对于平台服务质量的影响更为显著, 从而区域热度的评估应注重感知任务的完成比例。同时, 不同感知区域内活跃用户数、停留时间以及完成任务数量都存在巨大差异; 此外, 由于边际效用, 随着 3 个参数值的增大, 对区域热度的影响将逐渐降低。针对此种情况采用对数方法能更好地体现区域热度 H_{region} , 假设 T 个时刻内该区域内完成的感知任务总数为 $\theta_{Y_{\text{region}}}^{\text{complete}}$, 服务平台发布在该区域的任务总数为 $\mathcal{G}_{Y_{\text{region}}}^{\text{total}}$, 那么 H_{region} 如式 (3) 所示:

$$H_{\text{region}} = \frac{\lg \frac{\overline{P}_T \times \overline{U}_{\text{dur}}(N, Y_{\text{region}}) \times \theta_{Y_{\text{region}}}^{\text{complete}}}{\lg (\overline{P}_T \times \overline{U}_{\text{dur}}(N, Y) \times \theta_Y^{\text{complete}})_{\text{max}}}}{\lg (\overline{P}_T \times \overline{U}_{\text{dur}}(N, Y) \times \theta_Y^{\text{complete}})_{\text{max}}} \times \frac{\theta_{Y_{\text{region}}}^{\text{complete}}}{\mathcal{G}_{Y_{\text{region}}}^{\text{total}}} \quad (3)$$

其中, $\lg (\overline{P}_T \times \overline{U}_{\text{dur}}(N, Y) \times \theta_Y^{\text{complete}})_{\text{max}}$ 表示在所有区域中 $\lg \overline{P}_T \times \overline{U}_{\text{dur}}(N, Y) \times \theta_Y^{\text{complete}}$ 能够取得的最大值。

对于服务平台而言, 感知任务的完成比例以及完成任务的及时率都将显著影响其服务质量, 只有当该区域任务及时完成比例不少于平台期望最少完成的任务数量时, 平台的服务质量才能得到保障。从而将不少于该期望值时, 对应的最小热度值设置为热度高低的动态平衡点 H_{dep} , 并将其表示为式 (4) 所示:

$$H_{\text{dep}} = \frac{\lg \frac{\overline{P}_T \times \overline{U}_{\text{dur}}(N, Y) \times \theta_Y^{\text{complete}}}{\lg (\overline{P}_T \times \overline{U}_{\text{dur}}(N, Y) \times \theta_Y^{\text{complete}})_{\text{max}}} (\theta_{Y_{\text{region}}}^{\text{complete-timely}} \geq \mathcal{G}_Y^{\text{expectation-total}})}{\lg (\overline{P}_T \times \overline{U}_{\text{dur}}(N, Y) \times \theta_Y^{\text{complete}})_{\text{max}}} \quad (4)$$

式 (4) 中 $\lg \frac{\overline{P}_T \times \overline{U}_{\text{dur}}(N, Y) \times \theta_Y^{\text{complete}}}{\lg (\overline{P}_T \times \overline{U}_{\text{dur}}(N, Y) \times \theta_Y^{\text{complete}})_{\text{max}}} (\theta_{Y_{\text{region}}}^{\text{complete-timely}} \geq \mathcal{G}_Y^{\text{expectation-total}})$ 表示在所有区域中及时完成的任务数 $\theta_{Y_{\text{region}}}^{\text{complete-timely}}$ 不少于平台最低期望完成的任务数 $\mathcal{G}_Y^{\text{expectation-total}}$ 时, $\lg \frac{\overline{P}_T \times \overline{U}_{\text{dur}}(N, Y) \times \theta_Y^{\text{complete}}}{\lg (\overline{P}_T \times \overline{U}_{\text{dur}}(N, Y) \times \theta_Y^{\text{complete}})_{\text{max}}}$ 能够取得的最小值。当 H_{region} 小于 H_{dep} 时, 表明该区域为低热度区域, 而当 H_{region} 大于 H_{dep} 时, 表示该区域为高热度区域。

4 用户社交属性评估

4.1 用户意愿度感知

服务平台发布的任务具有时间和具体位置的



限制,用户需要在给定的时间内,到达指定位置完成感知任务。当任务位置与用户所处的位置相距太远时,用户不仅需要改变自己的行程安排,还要浪费时间到达任务位置,导致没有足够时间完成感知任务。用户参与感知任务移动的距离越少时,参与意愿越高;而随着移动距离的增长,参与意愿将降低,直至超过用户的接受范围,导致用户不再愿意参与感知任务。从而将用户的移动距离与其参与意愿的关系 $w_{dis}^{n_i, \theta_j}$ 表示为式(5):

$$w_{dis}^{n_i, \theta_j} = \begin{cases} 1, D_{n_i}^{\theta_j} \leq D_{cover}^{\theta_j} \\ 1 - \max(\log_{D_m}^{(D_{n_i}^{\theta_j} - D_{cover}^{\theta_j})}, 0), D_{cover}^{\theta_j} < D_{n_i}^{\theta_j} \leq D_m \\ 0, D_{n_i}^{\theta_j} > D_m \end{cases} \quad (5)$$

其中, $D_{n_i}^{\theta_j}$ 表示用户位置 l_{n_i} 和任务位置 l_{θ_j} 之间的欧几里得距离, $D_{cover}^{\theta_j}$ 表示任务覆盖半径,覆盖范围内的用户都可以参与完成感知任务, D_m 表示用户接受的最远移动距离。

此外,服务平台发布可能包含图像和视频以提供丰富信息的感知任务,同时用户也可能上传图像或视频数据至服务平台,从而引起较高的流量开销。如果用户在接受感知任务或者上传数据的过程中产生较高的流量费用,而用户的收益由服务平台支付的报酬减去其流量等开销,那么用户的收益将会降低,甚至为负。在此情况下,用户的参与意愿将会显著地降低。意味着当用户获得的收益越高时,其参与感知任务的意愿更强。此处将用户集合 N 分为流量免费的用户集合 ϕ_1 和流量付费的用户集合 ϕ_2 , 流量充足用户可以直接使用流量上传数据;而流量付费用户使用流量时需支付高额的费用。用户 n_i 完成任务 θ_j 的总开销 $C_{cost}^{n_i, \theta_j}$ 包含流量开销 C_{data} 、终端设备电量损耗以及内存占用等开销 C_{else} 。由用户 n_i 提交的配置文件可知单位流量成本为 p_{data} , 用户消耗的流量包含接受任务产生的流量 o_{θ_j} 以及上传数据产生的流量 $\int_0^{\tau_{trans}} o_{trans} dt$, o_{θ_j} 可以由平台直接统计,

$\int_0^{\tau_{trans}} o_{trans} dt$ 与用户提交数据的精度紧密相关, o_{trans} 表示数据发送速率,从而用户的总开销定义如式(6)所示:

$$C_{cost}^{n_i, \theta_j} = \begin{cases} c_{else}, n_i \in \phi_1 \\ p_{data}(o_{\theta_j} + \int_0^{\tau_{trans}} o_{trans} dt) + c_{else}, n_i \in \phi_2 \end{cases} \quad (6)$$

用户完成 θ_j 获得的报酬为 $\gamma_{n_i}^{\theta_j}$, 从而用户参与意愿与总开销 $C_{cost}^{n_i, \theta_j}$ 的关系表示如式(7)所示:

$$w_{cost} = \max \left[0, \frac{\gamma_{n_i}^{\theta_j} - C_{cost}^{n_i, \theta_j}}{\gamma_{n_i}^{\theta_j}} \right] \quad (7)$$

进一步地,用户在参与感知活动时倾向完成偏好任务类型,从而造成当平台发布的感知任务与用户偏好任务类型的相似度越高时,用户越有意愿参与。每个任务类型包含各自的类标识,属于聚类,从而采用互信息的思想度量两个聚类之间的相似性。将 σ_1 作为用户的偏好类型任务, σ_2 是服务平台发布的感知任务, $v_i, i \in \{1, \dots, \delta\}$ 和 $w_h, h \in \{1, \dots, \phi\}$ 分别为 σ_1 和 σ_2 的类标识,则 σ_1 和 σ_2 之间的相似度表示为式(8):

$$I(\sigma_1, \sigma_2) = \sum_{v_i \in \sigma_1} \sum_{w_h \in \sigma_2} p(v_i, w_h) \log \frac{p(v_i, w_h)}{p(v_i)p(w_h)} \quad (8)$$

其中, $p(v_i) = |v_i|/n$, $p(w_h) = |w_h|/n$, $p(v_i, w_h) = |v_i \cap w_h|/n$, n 表示样本的个数。

互信息对多值特征有偏倚,而标准化后的互信息能够修正其对多值特征的偏倚,并且使其值域为 $[0, 1]$, 即两者相似度越高则越趋近于 1, 反之,则趋近于 0。结合 σ_1 和 σ_2 各自对应的信息熵以及两者的相似度完成互信息的标准化,信息熵是某类感知任务出现概率的体现,从而得到互信息的标准化形式 $\Omega(\sigma_1, \sigma_2)$ 如式(9)所示:

$$\Omega(\sigma_1, \sigma_2) = \frac{I(\sigma_1, \sigma_2)}{\sqrt{H(\sigma_1) * H(\sigma_2)}} \quad (9)$$

其中, $H(\sigma_1)$ 、 $H(\sigma_2)$ 分别表示 σ_1 、 σ_2 的信息熵。

通过 σ_1 与 σ_2 的相似度衡量用户的参与意愿,用户 n_i 的参与意愿就可以表示为 $w_{preference} = \Omega(\sigma_1, \sigma_2)$ 。

综上所述, 用户的移动距离、流量开销以及任务类型都将影响用户的意愿度, 从而将用户的意愿度表示为式 (10):

$$\mathcal{W}_{n_i} = w_{dis}^{n_i, \theta_j} \times \frac{1 + w_{preference}}{2} \times (e^{w_{cost}} - 1) \quad (10)$$

4.2 用户信誉度感知

完成用户的意愿度评估之后, 还需要进一步分析用户的信誉度, 因为在 MCS 场景中, 服务平台的感知数据完全依靠用户提供, 同时, 服务平台并不知道真实的感知结果。因此, 服务平台需要保证选择的用户能够提交真实的感知数据, 而信誉度作为衡量感知数据可靠性的重要参数, 信誉度越高意味着用户感知数据可靠性越强。

将用户 n_i 参与感知任务的时间跨度记为 T_{day} (以天为单位), 即最开始参与感知任务的日期记为 T_{day} 天前。在更新用户的累积得分时, 越近的记录应占据更大的权重, 而越久远的记录则相反。同时, 随着用户参与时间的延长, 数据记录越多的用户, 其经历的评分阶段越多, 累计得分的评估越准确。应用以自然常数为底的对数能够有效的将用户参与记录划分为 T_{stage}^{max} 个时间段, T_{stage}^{max} 为 $\lceil \ln T_{day} + \ell \rceil$ 的最大值, 能够较快地更新用户最近记录对其信誉度的影响, 而减缓更新陈旧记录对信誉度的影响。由指数函数性质可知, 引入衰减系数之后, 5 个时间阶段以外的任务记录已不具有多少权重, 符合数据价值衰减特性, 从而将用户的累计得分定义为式 (11):

$$\delta_{n_i} = \sum_{T_{stage}=1}^{\min(T_{stage}^{max}, 5)} \sum_{\chi=1}^Q \left(\frac{1}{2} \right)^{T_{stage}} \overline{\delta_{n_i, score}^{T_{stage}, \chi}} \quad (11)$$

其中, Q 表示用户 n_i 在 T_{stage} 个阶段内完成的任务次数, $\overline{\delta_{n_i, score}^{T_{stage}, \chi}}$ 表示 n_i 在第 T_{stage} 阶段的平均得分。再将其累计得分与服务平台的所有注册用户的累计得分均值相比, 所得比值越大表明其历史提交的任务数据真实性越高, 此时比值作为增益系数

增大其信誉度; 相反如果用户完成了所有平台分配的任务, 但是比值低, 表明其提交虚假数据多, 那么比值可以作为惩罚系数降低其信誉度。从而将比值 α 作为信誉度叠加系数更能准确地衡量参与者的信誉度, 如式 (12) 所示:

$$\alpha = \frac{\delta_{n_i}}{\sum_{i=1}^Z \delta_{n_i} / Z} \quad (12)$$

为了防止用户短时间内连续提交真实任务数据提高信誉度而在之后上传虚假数据的现象, 用户信誉度应随着连续提交真实任务数据的次数缓慢地增加。当用户提交虚假任务数据时, 信誉度应该显著下降, 从而将虚假任务数和用户完成的任务数相比作为衰减系数, 并以指数的形式使其信誉度快速下降。对于用户没能完成感知任务的情况, 服务平台则减缓信誉度下降的幅度。结合用户 n_i 参与感知任务的记录以及信誉度叠加系数将其信誉度评估为如式 (13) 所示:

$$\mathfrak{R}_{n_i}^{reputation} = \min \left\{ 1, \alpha \frac{g_{n_i}^{True}}{g_{n_i}^{total} \times e^{\frac{g_{n_i}^{Fake}}{g_{n_i}^{Complete}}}} \right\} \quad (13)$$

其中, $g_{n_i}^{total}$ 是用户接受的任务总数, $g_{n_i}^{True}$ 表示完成的感知任务中数据满足服务平台预估结果的任务数量, $g_{n_i}^{Complete}$ 表示用户完成的任务总数, $g_{n_i}^{Fake}$ 表示用户完成的任务中提交的是虚假数据的任务数量。

4.3 用户活跃度感知

用户活跃度体现其感知能力的大小。通过用户的历史访问地点记录推算其区域活跃度。在现实世界中, 通过观察可发现用户的轨迹和其行为活动之间存在相关性。用户的移动是受到事件驱动的, 用户访问区域 m_k 的次数以及平均停留的时间都可以反映其在区域内的活跃状态, 访问次数越多, 在后续的时间到达区域的概率越大, 而平均停留越长, 用户完成感知任务的可能性就进一



步提高。从而根据用户访问区域 m_k 的次数以及平均停留时间分析其区域活跃度。

用户访问区域 m_k 的次数 $N_{\text{num}}(n_i, m_k)$ 与其访问所有区域的总次数 $\sum_{m_j \in M} N_{\text{num}}(n_i, m_j)$ 相比, 可以得出用户访问区域 m_k 的概率 $P(n_i, m_k)$, 如式 (14) 所示:

$$P(n_i, m_k) = \frac{N_{\text{num}}(n_i, m_k)}{\sum_{m_j \in M} N_{\text{num}}(n_i, m_j)} \quad (14)$$

用户 n_i 在 m_k 的每一次停留时间由离开时间 $e(n_i, m_k)$ 与到达时间 $s(n_i, m_k)$ 之差表示, 从而在 T 次访问中停留在 m_k 的平均时间 $\mathcal{H}_{\text{dur}}(n_i, m_k)$ 由式 (15) 表示:

$$\mathcal{H}_{\text{dur}}(n_i, m_k) = \frac{\sum_{m_k \in M} [e(n_i, m_k) - s(n_i, m_k)]}{T} \quad (15)$$

用户的访问概率与平均停留时间都与用户的区域活跃度直接正相关, 因此, 将用户 n_i 在 m_k 的访问概率与平均停留时间之积作为区域活跃度因子 $\mathcal{R}_{\text{region}}^{n_i}$ 。

$$\mathcal{R}_{\text{region}}^{n_i} = P(n_i, m_k) \times \mathcal{H}_{\text{dur}}(n_i, m_k)$$

为了更加直观地体现用户区域活跃度的差异, 同时使用户的区域活跃度能够限定在某个区间, 将 $\mathcal{R}_{\text{region}}^{n_i}$ 通过 Max-Min 标准化映射为在区间 $[0, 1]$ 中的值, 并将其作为用户区域活跃度 $\mathcal{R}_{n_i}$, 如式 (16) 所示:

$$\mathcal{R}_{n_i} = \frac{\mathcal{R}_{\text{region}}^{n_i} - \mathcal{R}_N^{\min}}{\mathcal{R}_N^{\max} - \mathcal{R}_N^{\min}} \quad (16)$$

其中, $\mathcal{R}_N^{\max}$ 和 $\mathcal{R}_N^{\min}$ 分别是 N 个用户中区域活跃度因子最大和最小的用户。

5 参与者选择策略

5.1 高热度区域参与者选择策略

通过第 3 节可知, 城市空间被划分为具有不同热度的感知区域, 发布在高热度区域的感知任

务同时存在多个用户都有空闲时间参与的情况, 因此, 将用户 n_i 能够取得数据质量 $\text{QoC}_{n_i}^{\theta_j}$ 定义如式 (17) 所示:

$$\text{QoC}_{n_i}^{\theta_j} = \frac{\mathfrak{R}_{n_i}^{\text{reputation}}}{\gamma_{n_i}^{\theta_j} \times t_{\text{total}}} \quad (17)$$

其中, $\gamma_{n_i}^{\theta_j}$ 表示用户 n_i 完成任务 θ_j 获得的报酬, 此处服务平台选择用户所要支付的报酬是用户的投标 $b_{n_i}^{\theta_j}$ 。 t_{total} 表示用户 n_i 从最开始与平台交互到完成数据上传的总时间, 将其定义为如式 (18) 所示:

$$t_{\text{total}} = D_{n_i}^{\theta_j} / v_{n_i} + \overline{t_{\text{interactive}}} \times \mathcal{A}_{n_i} + t_{\text{collect}} \quad (18)$$

其中, $\overline{t_{\text{interactive}}}$ 表示用户与平台的平均交互时间, t_{collect} 表示用户收集任务 θ_j 数据的时间。

根据第 4.1 节的分析, 用户的意愿度体现其参与完成任务的概率, 从而服务平台期望用户 n_i 取得的综合数据质量表示为式 (19):

$$\overline{\text{QoC}_{n_i}^{\theta_j}} = w_{n_i}^{\theta_j} \times \frac{\mathfrak{R}_{n_i}^{\text{reputation}}}{\gamma_{n_i}^{\theta_j} \times t_{\text{total}}} \quad (19)$$

在高热度区域, 用户 n_i 每次参与拍卖时只能对一个感知任务提交配置文件, 从而保证单个任务能够取得较高的数据质量。如果当前发布的感知任务 θ_j 分配给了用户 n_i , 用户 n_i 继续选择参与感知任务 θ_k 时, 其可用时间发生了变化, 完成任务 θ_k 的时间变为完成两个任务的时间之和。此时, 用户 n_i 完成任务 θ_k 能够实现的综合数据质量如式 (20) 所示:

$$\overline{\text{QoC}_{n_i}^{\theta_k}} = w_{n_i}^{\theta_k} \times \frac{\mathfrak{R}_{n_i}^{\text{reputation}}}{\gamma_{n_i}^{\theta_k} \times \sum_{j=1}^{Q_{n_i}^{\text{fact}}} t_{\text{total}}} \quad (20)$$

从而当用户 n_i 被分配不少于两个感知任务时, 依照排队策略完成感知任务。服务平台评估用户在某个感知任务能够实现的综合数据质量, 再根据数据质量大小选择参与者。此时, 参与者选择转换为找到一个集合 N 的子集 Θ , 在满足限制条件下使集合 $\mathcal{G}$ 中的感知任务被集合 Θ 覆盖时

取得最大的总体数据质量。转换之后的问题被证明是经典的 NP-hard 集合覆盖问题。对于给定集合 N 的子集 Θ ，用效用函数 $\mathcal{F}_{\Theta}(\mathcal{G})$ 表示选择的参与者集合 Θ 完成集合 $\mathcal{G}$ 中所有的感知任务能够取得的总体数据质量。 $\mathcal{F}_{\Theta}(\mathcal{G})$ 是非负、非减少和子模块函数，其能够取得的最大值与预算及数据截止接收时间紧密相关。根据子模块函数的性质，对于任意 $\Theta \subseteq N$ ， $\mathcal{F}_{\Theta}(\mathcal{G}) > 0$ ，使用贪婪算法能够找出接近最优解的近似解，从而服务平台可以采用贪婪算法迭代搜索选择参与者，在每一轮迭代中，选择提高效用函数 $\mathcal{F}_{\Theta}(\mathcal{G})$ 最多的用户，即第 i 次迭代时选择的用户 n_i 满足式 (21) 的条件：

$$\mathcal{F}_{\Theta_i}(\mathcal{G}) = \mathcal{F}_{\Theta_{i-1}}(\mathcal{G}) \cup \left\{ \arg \max \Delta \mathcal{F}_{(n_i | \Theta_{i-1})}(\mathcal{G}) \right\} \quad (21)$$

其中， $\Delta \mathcal{F}_{(n_i | \Theta_{i-1})}(\mathcal{G}) = \mathcal{F}_{(\Theta_{i-1} \cup \{n_i\})}(\mathcal{G}) - \mathcal{F}_{\Theta_{i-1}}(\mathcal{G})$ 。当服务平台完成所有感知任务分配或者预算为零时，参与者选择过程终止。

5.2 低热度区域参与者选择策略

服务平台在低热度区域收集数据时，潜在的参与者较少，但是区域内大部分用户停留时间较长、访问的地点较多，从而可以选择完成多个感知任务的参与者。在低热度感知区域，服务平台面临用户以随机顺序到达感知区域、必须在截止时间前完成一定数量的感知任务以及需要获得可信的感知数据共 3 个挑战。多阶段抽样—接收方法可以动态地扩大用户集合的数量并可以根据用户的特征和平台的预算得出每一阶段的选择阈值。完成每一阶段的参与者选择之后，动态更新选择阈值，从而在每一个选择阶段能够很好地解决这 3 个挑战。具体来说，将参与者选择过程的分为 $1, 2, \dots, \lfloor \text{lb}^T \rfloor, \lfloor \text{lb}^T \rfloor + 1$ 个阶段，每一阶段 k 在 $T' = \left\lfloor 2^{k-1} T / 2^{\lfloor \text{lb}^T \rfloor} \right\rfloor$ 结束，阶段 k 的预算为 $\mathcal{B}' = \left\lfloor 2^{k-1} \mathcal{B} / 2^{\lfloor \text{lb}^T \rfloor} \right\rfloor$ 。在初始阶段观察提交配置文件的用户能够实现的收益，并根据收益均值设置阈值，在后续的选择阶段，选择超过设置阈值的用户。

根据第 4.3 节的描述，活跃度高的用户使用任务队列策略和社区用户协助策略能够提高完成任务的数量和概率，并保证得到可靠的任务数据。因此，当用户选择完成的任务数量不超过该阶段剩余任务数量时，用户实际分配的任务数量为 $Q_{n_i}^{\text{fact}}$ 。完成感知任务后，会给平台带来一定的收益，而真实且完整的数据给平台带来的收益越高，从而将用户能实现的可靠收益总量表示为式 (22)：

$$\zeta_{\text{total}}^{n_i} = \mathfrak{R}_{n_i}^{\text{reputation}} \times \sum_{j=1}^{Q_{n_i}^{\text{fact}}} (\theta_j \times v_{\theta_j}) \quad (22)$$

根据观察阶段 κ 个用户提交的配置文件得出选择阈值，并将阈值定义为式 (23)：

$$\rho = \frac{\sum_{i=1}^{\kappa} (\zeta_{\text{total}}^{n_i})}{\kappa} \quad (23)$$

动态的增加用户数量和预算，当用户 n_i 在时间 t 提交配置文件时，服务平台对用户能够取得收益的总量与阈值进行比较，检验用户 n_i 是否值得选择。完成每一阶段的参与者选择之后，根据该阶段全部选择的参与者能够实现的收益总量均值更新选择阈值，保证服务平台在低热度区域能够收集足够且可靠的感知数据，并且减少需要选择的参与者总数，从而降低平台的开销。

6 数值结果及性能分析

下面使用真实的数据集验证提出机制的高效性和有效性。首先给出数据集的来源、环境设置和基准方法，然后，将提出的机制与其他基准机制在任务分配率、综合数据质量以及完成时间 3 个方面进行了比较；最后，对不同方法的仿真结果进行了详细的分析。

6.1 验证环境参数

借助 MATLAB 平台完成仿真，使用来源于 Brightkite 的真实数据集。Brightkite 是一个基于位置的社交网络服务提供商。用户通过签入来分享他们的位置，通过公共 API 收集用户签入的信息，构成



基于无向图的节点关系网络,网络由 58 228 个用户和 214 078 条边组成。用户接入的信息包括用户编号、接入时间、经纬度以及签到地点 ID。在此基础上,提取两个子集:100 个地点和 1 224 个用户来表示模型中的任务位置和参与者,利用这两个子集构建真实的 MCS 场景,以验证所提出的算法优于基准算法。主要仿真参数设备见表 1。

表 1 主要仿真参数设置

参数	参数值
区域面积/km ²	1~2
区域热度	0~1
任务的截止时间/min	5~45
单个任务需要的参与者数量/个	1~5
每个参与者接受的任务数/个	0~8
服务平台任务数量/个	1~100
任务覆盖半径/m	200~1 600
参与者的信誉度	0~1
参与者的活跃度	0~1
参与者的移动速度/(km·h ⁻¹)	10~50
感知任务的种类/个	5

为了便于比较,在安全参与者选择算法 (secure user recruitment)^[8]和有意改变路径的贪婪增强型遗传算法 (greedy-enhanced genetic algorithm for intentional movement)^[4]的基础上添加约束条件,然后将两种算法作为基准算法。SUR 算法选择实现数据质量最大的用户, GGA-I 算法优先选择完成多个感知任务所移动距离最短的用户。

6.2 性能分析

图 2 反映了参与者信誉度随着参与时间以及服务平台发布任务次数的变化情况。将参与者分为总是积极感知、总是消极感知和间断性消极感知 3 种,分别对应于第 25 号参与者、第 58 号参与者以及第 96 号参与者,参与者初始信誉度设置为 0.5。25 号参与者在第 0~16 次的感知过程中总是进行积极感知时,信誉度从 0.5 缓慢上升至

0.96,由于其没有参与第 16~21 次感知任务,信誉度衰减到 0.95,之后随着其参与次数的增加,信誉度缓慢增加。96 号参与者连续 6 次消极感知,其信誉度从 0.5 快速衰落到 0.35,此后,服务平台再也没选择该参与者。58 号参与者进行间断性的消极感知,信誉度随着其感知行为上下波动,但随着参与者提交虚假数据次数的增多,服务平台对其惩罚力度也在逐渐加大,其信誉值下降幅度相应增大,信誉度的恢复周期也逐渐增长,原因在于信誉度更新机制对参与者的消极感知实施了惩罚措施,消极感知次数越多,信誉度下降速度越快。此外,引入了衰减因子,参与者参与任务的历史记录参考价值随着时间会产生一定的衰减。

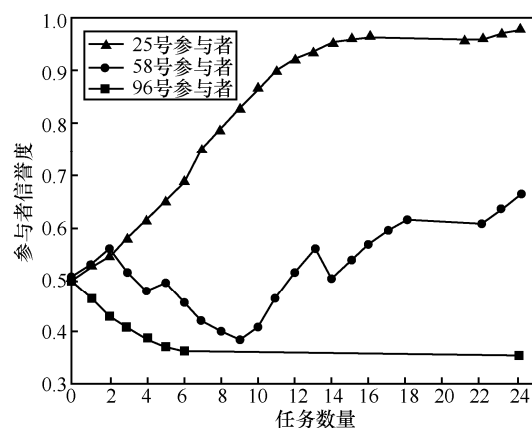


图 2 用户信誉度的变化

图 3 描述了在低热度区域随着任务总数的增加,不同算法中任务失败率的变化情况。任务失败率由用户没有完成的任务数和服务平台没有分配的任务数之和与总的任务数量相比而得。在任务数量少于 80 个时, GGA-I 的任务失败率始终高于 SUR 和 TUM-P,而在任务总数达到 90 个时, SUR 的性能急剧恶化,任务失败率超过 16.7%。其原因在于 GGA-I 算法中一个参与者需要完成多个感知任务,任务数量较少且分布的区域较广,难以找到合适的用户,用户移动距离偏长,导致任务失败率高。而服务平台发布的任务超过 80 个时, SUR 由于在前期选择质量较高的用户消耗了更多的预

算, 导致部分任务不能分配, 失败率快速增加。相反, TUM-P 算法通过用户的活跃度确定其完成任务的能力, 结合用户的信誉度选择参与者, 从而保证任务的完成。当服务平台分配的任务总数达到 100 个时, 相比于 SUR 和 GGA-I 两种算法, 失败率分别降低了 66.7%和 50.6%, 在有限的预算中取得较多的数据结果。

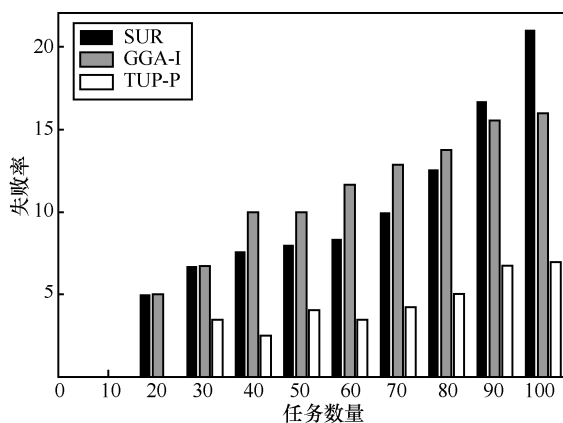


图3 任务数量与失败率的关系

在图3的基础上, 只限制平台的预算。此外, 在100个感知任务加入10个视频任务, 进一步分析用户意愿度和信誉度对任务完成比例的影响。由图4可知, 用户的意愿度和信誉度都对任务完成率有积极的促进作用, 但是单个特征处于较高水平时, 任务完成率远低于用户两个特征都处于较高水平时的任务完成率。当用户的意愿度和信誉度都为1时, 任务完成率达到了96%。

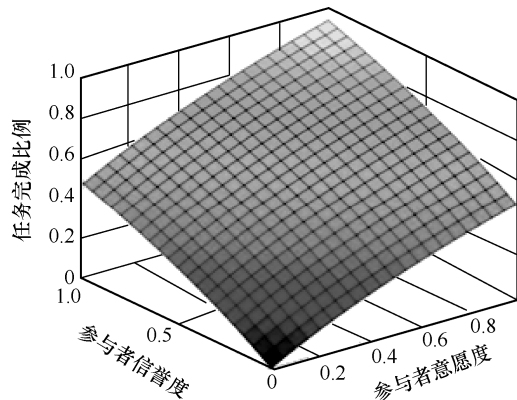


图4 用户意愿度和信誉度对分配任务完成比例的影响

7 结束语

针对 MCS 中服务平台在用户稀疏区域获取的任务数据可靠性低以及任务难以按时完成的问题, 提出一种联合区域热度和社交属性感知的参与者选择机制。根据区域活跃用户数、用户平均停留时间及区域历史感知任务完成情况评估区域热度, 结合用户的状态信息和用户历史感知任务记录计算用户意愿度、信誉度以及活跃度; 综合考虑上述4个因素合理选择任务参与者。在高热度区域选择取得综合数据质量最高的参与者, 从而使服务平台的总体质量最大化; 而在低热度区域建立整数线性规划模型, 将时间分成不同的阶段, 在每一阶段选择能实现效用最大的参与者, 从而高效可靠地完成所有的感知任务。在未来的工作中, 将继续研究 MCS 多任务场景中其他影响参与者选择的因素, 深入分析用户投标与其提供数据的质量关系, 进一步增强参与者选择算法的效用。

参考文献:

- [1] GANTI R K, YE F, LEI H. Mobile crowdsensing: current state and future challenges[J]. IEEE Communications Magazine, 2011, 49(11): 32-39.
- [2] WU D, SI S, WU S, et al. Dynamic trust relationships aware data privacy protection in mobile crowd-sensing[J]. IEEE Internet of Things Journal, 2018, 5(4): 2958 - 2970.
- [3] SUN X S, HU S, SU L, et al. Participatory sensing meets opportunistic sharing: automatic phone-to-phone communication in vehicles[J]. IEEE Transactions on Mobile Computing, 2016, 15(10): 2550-2563.
- [4] GUO B, LIU Y, WU W, et al. Active crowd: a framework for optimized multitask allocation in mobile crowdsensing systems[J]. IEEE Transactions on Human-Machine Systems, 2017, 47(3): 392-403.
- [5] KOUTSOPOULOS I. Optimal incentive-driven design of participatory sensing systems[C]//International Conference on Computer Communications, Turin, April 14-19, 2013, Turin, Italy. Piscataway: IEEE Press, 2013: 1402-1410.
- [6] ZHANG M, YANG P, TIAN C, et al. Quality-aware sensing coverage in budget-constrained mobile crowdsensing network[J]. IEEE Transactions on Vehicular Technology, 2016, 65(9): 7698-7707.



- [7] ZHANG X, YANG Z, LIU Y, et al. Towards efficient mechanisms for mobile crowdsensing[J]. IEEE Transactions on Vehicular Technology, 2017, 66(2): 1760-1771.
- [8] XIAO M, WU J, ZHANG S, et al. Secret-sharing-based secure user recruitment protocol for mobile crowdsensing[C]// International Conference on Computer Communications, May 1-4, 2017, Atlanta, GA, USA. Piscataway: IEEE Press, doi: 10.1109/INFOCOM.2017.8057032.
- [9] ESTRADA R, MIZOUNI R, OTROK H, et al. A crowd-sensing framework for allocation of time-constrained and location-based tasks[J]. IEEE Transactions on Services Computing, doi:10.1109/TSC.2017.2725835.
- [10] FIANDRINO C, ANJOMSHOA F, KANTARCI B, et al. Sociability-driven framework for data acquisition in mobile crowdsensing over fog computing platforms for smart cities[J]. IEEE Transactions on Sustainable Computing, 2017, 2(4): 345-358.
- [11] YANG B, LEI Y, LIU J, et al. Social collaborative filtering by trust[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(8): 1633-1647.
- [12] SUN J, PEI Y, HOU F, et al. Reputation-aware incentive mechanism for participatory sensing[J]. IET Communications, 2017, 11(13): 1985-1991.
- [13] GUO B, LIU Y, WU W, et al. ActiveCrowd: a framework for optimized multitask allocation in mobile crowdsensing systems[J]. IEEE Transactions on Human-Machine Systems, 2017, 47(3): 392-403.
- [14] 龙浩, 张书奎, 张洋, 等. 移动群智感知中基于社区的任务分发算法[J]. 通信学报, 2019, 40(10): 42-54.
- LONG H, SHANG S K, ZHANG Y, et al. Task distribution algorithm based on community in mobile crowd sensing[J]. Journal on Communications, 2019, 40(10): 42-54.
- [14] 冯友宏, 杨志, 丁绪星, 无线协作网络中的最优中继选择方案及中断概率分析[J]. 电信科学, 2018, 34(11): 87-95.
- FENG Y H, YANG Z, DING X X. Optimal relay selection scheme and outage probability analysis in cooperative wireless networks[J]. Telecommunications Science, 2018, 34(11): 87-95.

[作者简介]



向罗勇(1986-), 重庆信息通信研究院(中国信息通信研究院西部分院)中级工程师, 主要研究方向为无线通信检测、无线电频谱管理、移动通信天线设计与检测。



陈文(1990-), 重庆信息通信研究院(中国信息通信研究院西部分院)中级工程师, 主要研究方向为无线电传输技术以及无线电产品测试认证。



张陆洋(1988-), 重庆信息通信研究院(中国信息通信研究院西部分院)中级工程师, 主要研究方向为无线通信电磁兼容、工业互联网节点标识解析。