

移动群智感知多任务参与者优选方法研究

刘 琰 郭 斌 吴文乐 於志文 张大庆

(西北工业大学计算机学院 西安 710072)

摘 要 该文主要研究面向移动群智感知的任务分发方法,提出了一种新的参与者选择:面向多任务并发的参与者优选,不同于其他参与者选择,文中选择出的参与者不再局限于只能完成一个任务,每个参与者可以在规定时间内尽可能的完成多个任务,由此降低群智平台的成本,并提出 MultiTasker 方法,其目标是选择出最佳的参与者集合,使参与者完成任务所移动的总距离最短以降低成本,并且完成任务的参与者人数最少以优化用户资源.为了实现这个目标,文中设计了3种算法:T-Random、T-Most 和 PT-Most. T-Random 和 T-Most 是以任务为中心进行参与者选择,而 PT-Most 是以用户为中心进行参与者选择.通过一个大规模的真实数据集对设计的3种算法进行实验评估,同时研究了参与者选择情况与各种因素之间的关系,如任务分布和任务执行时间等.

关键词 移动群智感知;任务分发;多任务并发;参与者选择;优化;物联网;信息物理融合系统

中图法分类号 TP393 **DOI号** 10.11897/SP.J.1016.2017.01872

Multitask-Oriented Participant Selection in Mobile Crowd Sensing

LIU Yan GUO Bin WU Wen-Le YU Zhi-Wen ZHANG Da-Qing

(School of Computer Science, Northwestern Polytechnical University, Xi'an 710072)

Abstract Participant selection or task allocation is a key issue in Mobile Crowd Sensing (MCS). While previous participant selection approaches mainly focus on selecting a proper subset of users for a single MCS task, multitask-oriented participant selection is essential and useful for the efficiency of large-scale MCS platforms. This paper studies the task allocation issue in MCS, and we propose a new multitask-oriented participant selection problem. The difference between our work and other studies in participant selection is that each participant can complete as many tasks as possible within the given time period, rather than merely one task. This participant selection problem has some advantages: First, each participant can complete multiple tasks, and this will bring benefits to the running of the MCS platform. It is particularly useful in cases that user resources are not sufficient regarding to the published tasks. Second, there exist geographical proximity among the published tasks, and we can optimize the MCS platform performance by assigning the tasks as a whole. Third, each participant can perform as many tasks as possible, and this may improve personal income and participation enthusiasm. To address this participant selection issue, we propose MultiTasker, the objective of which is to minimize the total distance the selected participants move while minimizing the number of participants, subjecting to that all tasks are allocated with the needed number of participants and the tasks are completed by each selected participant within the given time period. In order to achieve the objective, we propose three algorithms: T-Random, T-Most and PT-Most. T-Random and T-Most select participants

收稿日期:2015-12-25;在线出版日期:2016-05-04. 本课题得到国家“九七三”重点基础研究发展计划项目基金(2015CB352400),国家自然科学基金(61332005,61373119,61222209)资助. 刘 琰,女,1993年生,硕士,主要研究方向为移动群智感知. E-mail: ly_liuyan_ly@163.com. 郭 斌,男,1980年生,博士,教授,中国计算机学会(CCF)高级会员,主要研究领域为普适计算、移动群智感知. 吴文乐,男,1990年生,硕士研究生,主要研究方向为移动群智感知. 於志文,男,1977年生,博士,教授,中国计算机学会(CCF)高级会员,主要研究领域为普适计算、社会感知计算. 张大庆,男,1964年生,博士,教授,主要研究领域为普适计算.

in a task-centric way, while PT-Most selects participants in a people-centric manner. We evaluated the three algorithms using a large-scale real-world dataset, and studied the relationship between participants and other factors, such as the distribution of tasks and task completion time.

Keywords mobile crowd sensing; task allocation; multitask-oriented; participant selection; optimization; Internet of Things; Cyber-Physical System

1 引言

近年来,移动群智感知(mobile crowd sensing)成为普适计算领域的前沿研究问题^[1],它依赖大量普通用户的移动设备及其具备的多元感知能力来完成大规模感知任务,例如,检测空气质量^[2]、收集交通信息^[3]等.在进行移动群智感知时,需要执行一系列的工作,比如,发布感知任务、选择用户完成任务、收集感知数据等.因此,发展一个通用的移动群智感知平台对完成多种不同的感知任务来说是必不可少的.目前,已经有一些移动群智感知平台,如 Campaignr^①和 Medusa^[4].但是,大部分平台都是主要用于任务发布和数据收集,这些平台中的用户自己决定完成某些任务,而不是由平台根据不同的需求或者优化目标(如移动距离最短、成本最小等)统一分配任务.

参与者选择是移动群智感知中的关键挑战之一,如何进行任务分发对数据采集的全面性、任务完成率和数据采集质量等都具有重要影响^[5].目前,已经有很多工作研究移动群智感知中的参与者选择问题^[6-9]. Cardone 等人^[10]研究移动群智感知平台中的参与者选择机制,在参与者个数一定的情况下,最大限度的提高感知任务的空间覆盖范围. Li 等人^[11]研究团队形成问题,即寻找一个有特定技能的专家小组,每个人完成一个给定的任务,同时最小化团队成员之间的交流成本.但这些工作主要面向单个群智感知任务进行任务分配.在大型群智感知平台上,如 Amazon Mechanical Turk^②,往往会有大量并发出现的任务.出于以下一些原因,需要对同时间段出现的任务集在现有候选用户集基础上进行统一分配.第一,通过单个用户完成多个任务,可以实现群智平台对用户资源的有效利用,特别是在用户资源可能不足的情况下(如任务大量涌现或目标区域可选用户较少);第二,任务之间具有时空邻近性,通过统一分配可以优化系统的性能;第三,单个用户在给定时间内完成多个任务也可以提高个人激励收益和参与积极性.因此,本文针对移动群智感知多任务并发环

境下的参与者选择问题进行了研究.

在参与者选择问题中,最终选择出的参与者人数、参与者完成的任务个数和完成任务所移动的总距离等因素都是主要的优化目标.但是现有的研究工作主要是基于单目标优化来选择参与者,如最大化完成任务数或者最小化移动距离.很多情况下,在参与者选择研究中,考虑多个优化目标也是很有必要的,比如同时需要最小化选择的参与者人数和完成任务所移动的总距离.考虑到这个问题,本文研究了基于多目标优化的多任务参与者选择问题,图 1 阐述了研究该问题的实际价值.

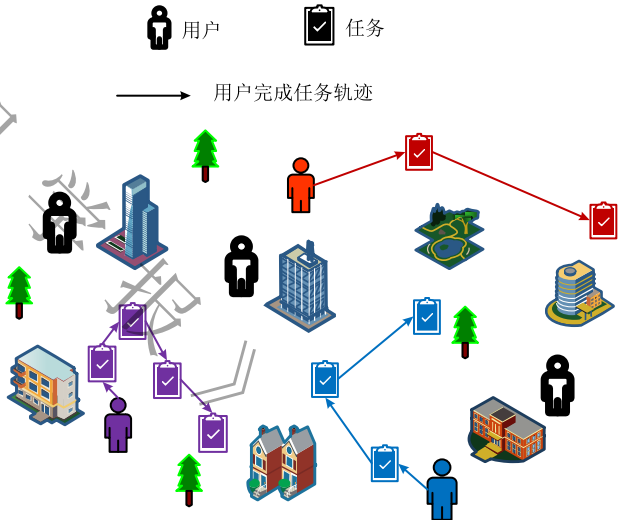


图 1 感知任务完成情景

假设在移动群智感知平台上有很多已经注册的用户,由于一些激励机制,这些用户可以随时完成多个感知任务.当任务平台将不同的任务分配给不同的用户,被选择的用户,即参与者需要利用随身携带的设备来收集感知收据,如利用智能手机拍照收集信息.不同请求者可以在感知平台上发布多个任务,比如某条道路发生交通事故,需要收集交通事故的情况;某个时刻城市突然下暴雨,需要了解某些道路的积水情况;某个地区举办活动,需要收集当前活动

① <http://research.cens.ucla.edu/urban/>
② <https://www.mturk.com/mturk/>

的信息等. 对于这些紧急任务(需要尽快完成的任务)来说, 平台需要根据任务和用户的信息选择合适的参与者来完成任务, 比如为了按时完成任务, 收集有用的信息, 平台要求所有参与者在一定时间范围内完成发布的任务(如两个小时). 对于感知平台来说, 为了降低系统成本, 需要最小化参与者的激励成本. 对于参与者来说, 希望可以最大化每个参与者的收益, 即在任务个数一定的情况下, 选择最少的参与者完成任务.

本文提出 MultiTasker 方法以实现多任务并发环境下的参与者优选. 其优化目标是最小化参与者的激励成本(激励来源包括两部分: 一部分是完成一个任务的固定激励; 另一部分是动态激励, 其大小与参与者完成任务所移动的总距离成正比), 同时最小化完成任务的参与者人数(每个参与者在规定时间内尽可能完成多个任务, 以增加个人收益, 并优化用户资源). 针对该参与者选择问题, 本文提出 3 种算法: T-Random、T-Most 和 PT-Most. T-Random 和 T-Most 是以任务为中心进行参与者选择, 而 PT-Most 是以用户为中心进行参与者选择. 在此基础上, 通过一个大规模的真实数据集对设计的 3 种算法进行实验评估, 通过对选择出的参与者人数、平均每个人完成的任务数、完成任务所移动的总距离等实验结果的对比, 根据不同的情况选择出性能最好的算法. 同时, 还研究了参与者的选择情况与各种因素之间的关系, 如任务分布和任务执行时间等.

2 相关工作

任务分配从本质上来说是一个群组生成问题, 即如何选择一组最合适的参与者来完成任务. 目前的研究工作主要分为在线和离线两个方面.

2.1 面向在线社交网络的群组生成

在线社交网络的群组生成问题主要考虑参与者的社会关系、专业技能等因素. 如 Lappas 等人^[12]研究在考虑社会关系的情况下组建一个团队, 即形成一个专家团队来完成给定的任务, 同时尽可能的减少团队成员之间的沟通成本和通信成本. 一个项目完成的成功与否不仅与团队成员的专业技能有关, 还与成员之间的合作、交流有关, 所以选择参与者时还需要考虑成员之间的社会关系. 比如在一个公司里, 社会关系包括上下级关系, 同一部门之间的同事关系, 不同部门之间的同事关系等. 如果两个团队在技术水平上都有能力完成一个项目, 一个团队是随

机选择互不认识的成员, 而另一团队是多年一起工作的成员, 那么第二个团队有更好的交流和协作能力去完成工作. Sozio 等人^[13]研究在已知社会关系图和给定查询集合的情况下, 寻找包含给定查询集合的密集连通子图. 其提出的方法除了考虑社会关系外, 还考虑空间位置等物理因素. Yang 等人^[14]研究在考虑空间位置和社会关系的情况下, 寻找一个社会团体, 使这个团体到给定位置的距离最小, 并且有一定的社会联系. 他们提出基于位置社交网络(LBSN)的社会空间组查询, 目的是在同时考虑社会关系和空间位置的基础上, 为即将举办的活动寻找一组合适的参与者.

2.2 面向移动群智感知的工作者选择

不同于在线社交网络的群组生成, 面向移动群智感知的工作者选择是以物理空间位置为基础进行选择. Papadias 等人^[15]研究在已知给定集合点的情况下, 寻找其他的点使其到给定集合点的距离最小. Zhang 等人^[16]提出一种新的参与者选择框架 CrowdRecruiter, 在满足给定空间覆盖率的情况下, 通过选择最少的参与者来最大限度地减少激励支付. CrowdRecruiter 首先根据用户的历史移动轨迹, 预测出用户当前的轨迹. 由于总激励成本与参与者的总人数呈正比, 所以希望选择最少的参与者以减小激励成本. 接下来计算由多个用户组成的参与者集合的联合覆盖率, 在满足要求覆盖率的基础上选择最少的参与者完成任务. Reddy 等人^[17]主要研究在考虑空间位置、时间要求以及参与者行为习惯的情况下, 选择合适的参与者完成任务, 并提出基于参与式感知的参与者选择架构. 该架构包含三个部分, 首先根据要求对参与者进行初步排除, 然后计算剩余参与者的效用和成本, 在成本满足约束条件的情况下, 选择效用较大的参与者完成任务, 最后检查、核对参与者的移动性是否满足既定轨迹和时间要求. 比如选择校园里的学生记录学校垃圾箱的情况, 那么首先选择的是经常路过垃圾箱的学生, 然后在这些学生里尽量选择较少的人, 同时保证能记录到校园中所有垃圾箱的情况. 在这些学生中, 有的人是步行经过垃圾箱, 而有的人是骑车经过, 两者的移动方式不同, 任务的完成时间也就不同, 相对于步行来说, 骑车的效率会更高一些.

无论是面向在线社交网络的群组生成, 还是面向移动群智感知的工作者选择, 选择出的参与者都只在某一时间段内完成一个任务. 对群智平台来说, 由于对参与者有只能完成一个任务的限制, 可能会

浪费参与者的资源（特别是在用户资源不足情况下），且增加任务的完成时间. 对参与者而言，通过完成多个任务可以获得更多报酬，提高参与积极性. 所以，对于参与者和任务发布者来说，单个用户在给定时间内完成多个任务是一个双赢的情况. 同时，为了降低激励成本，希望参与者移动的总距离最短（动态激励成本与移动距离成正比）. 因此，本文提出了面向多任务并发的参与者选择问题，并针对于该问题设计了 MultiTasker 方法，其目标是选择出最佳的

参与者集合，使参与者完成任务所移动的总距离最短以降低成本，并且最小化参与者人数以优化用户资源.

3 MultiTasker 架构

本节主要介绍 MultiTasker 的总体架构，如图 2 所示. 该架构主要包含三个部分：任务发布、任务分配和任务执行.

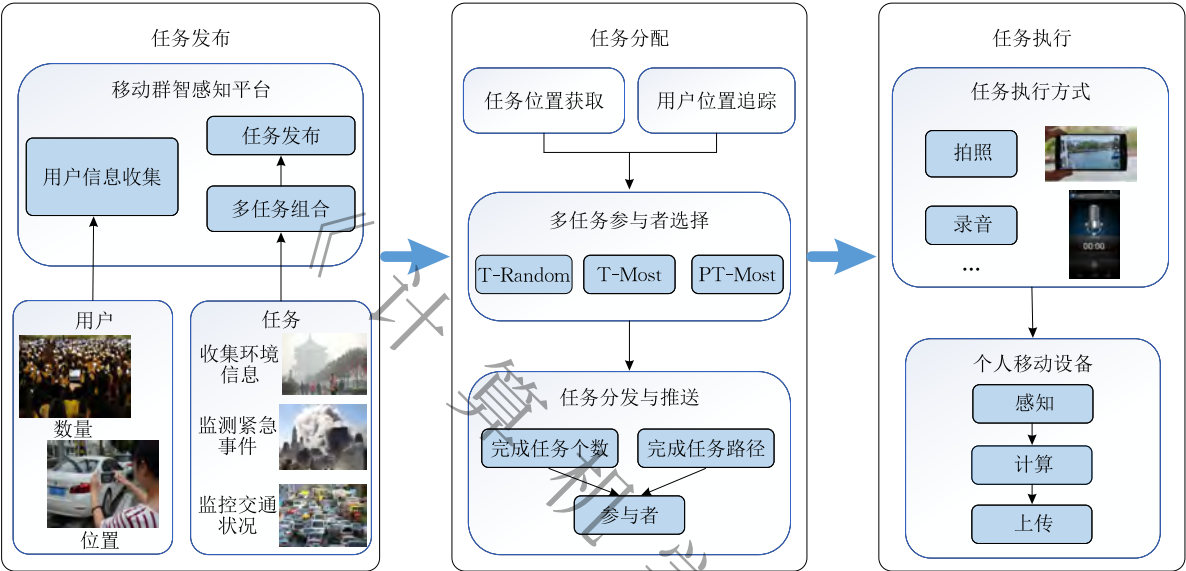


图 2 MultiTasker 的总体架构

3.1 任务发布

在群智感知平台上，会有很多的请求者发布任务. 这些任务主要分为两类，一类任务对时间的要求不高，比如给正在花期的植物拍照，可能只需要参与者在一周内完成，并且不需要参与者专门到达任务的位置立即完成，只需要选择一些顺便路过的用户完成任务，以降低任务成本和减少用户负担. 这类任务对系统带来的负担较小，可在系统负载较小时选人完成. 另一类任务则对时间有较高的要求，需要及时完成，如某个地方发生暴雨灾害、火灾、大型集会、交通迟滞等. 这类任务由于面向紧急或热点事件，可能会大量地涌现，并且用户需要在一定时间内完成以收集有用的信息. 因此可能会出现用户资源不足的情况. 针对这种情况，系统的组织者需要根据任务的紧急性、平台中候选者人数等因素对任务进行适当的组合，在某一时间内可能同时发布多个任务. 由于群智感知平台处于多任务并发的环境下，考虑到系统性能，为优化用户资源和提高用户参与性，需要对多任务并发环境下的参与者选择问题进行研究.

3.2 任务分配

以一定时间（如 10 分钟）为限，对该时间段内出现的任务视为并发任务进行统一分配. 移动群智感知平台首先基于一定的激励机制确定可以完成任务的用户，同时这些用户允许群智平台获取自己的某些信息，如当前位置等. 如果平台给其分配多个任务，参与者也可以随时完成. 在任务分配中，又包含三个阶段：首先获取平台中用户和任务的具体位置，然后根据要求选择合适的参与者，最后将任务分配给每个参与者完成.

(1) 任务位置获取和用户位置追踪. 由于该移动群智感知平台中的感知任务都是基于位置，因此，对于任务请求者来说，需要在发布任务时提供任务的位置信息以使用户完成. 对于平台中的用户来说，被选择的用户需要移动到任务的位置完成任务. 因此，平台需要确定用户当前的具体位置，以位置为基础选择参与者. 目前，在移动群智感知中，最为常见的确定位置的两种方法是：一是通过用户个人智能手机的 GPS 定位；二是通过用户打电话时所在基站

的位置确定用户的位置,两者各有优缺点.通过GPS定位获得的用户位置比较精确,但是如果用户长时间使用GPS,会耗费手机大量的能量,增加任务完成的成本.由于用户在日常生活中都会自发的打多个电话,所以通过打电话所在的基站位置确定用户的位置,不会消耗额外的能量.但是由于基站覆盖一定的范围,所以通过基站获取用户的位置不精确.总的来说,平台可以根据任务的性质、成本等因素利用不同的方式来确定用户的位置.

(2)多任务参与者选择.组织者可能在同一时间发布多个任务,在这种情况下,可能会出现用户资源不足情况,为优化用户资源和提高用户参与性,希望最小化参与者的激励成本以减少系统成本和最小化参与者人数以提高每个参与者的收益.同时,针对紧急任务,希望所有任务都能够在一定的时间范围内完成.参与者的激励来源包括两部分:一部分是完成一个任务的固定激励;另一部分是动态激励,其成本与参与者完成任务所移动的总距离成正比.因此,该问题的目标是选择出最佳的参与者集合,使参与者完成任务所移动的总距离最短,并且参与者人数尽可能的少.针对于该参与者选择问题,本文提出3种解决方法:T-Random、T-Most和PT-Most. T-Random和T-Most是以任务为中心进行参与者选择,而PT-Most是以用户为中心进行参与者选择.

(3)任务分发与推送.在该群智平台中,用户是被动完成任务,即平台给用户分配特定的任务,而不是用户自己选择完成的任务.对于选择的参与者,平台可能给每个人分配多个任务,要求参与者专门完成这些任务.因此,平台将需要完成的任务以及完成任务的路线信息发送给参与者.

3.3 任务执行

参与者接收到来自群智平台的任务信息后,需要自己的移动设备上完成感知、计算、上传等工作.由于针对时间限定任务及资源不足情况,所以要求用户有意识去访问各个任务地点执行分配的任务,而不必遵循原有的路径.首先,参与者根据平台提供的路线信息,到达任务地点完成感知任务.由于任务的完成形式可能不同,参与者需要根据任务的性质选择不同的感知形式,比如拍照、录视频等.同时为了保证收集到数据的质量和全面性,希望参与者在同一地点收集多个数据反映情况,如完成收集城市公共设施的损害情况任务时,参与者可以从不同的角度拍照,以此来完整的反映公共设施的损害情况.参与者完成感知任务后,将收集的数据信息先

存储在手机中,最后统一由个人移动终端对数据进行简单处理和计算,在适当的时候将数据上传至服务器.

3.4 场景实例

下面通过一个应用实例来说明MultiTasker的功能:某城市在某个时间下起了暴雨,由于下暴雨,某些道路可能出现积水的情况;同时,道路的能见度比较低,所以交通事故的发生率也相应增加.针对这两个紧急情况,由于采集数据的时间一致,所以任务发布者将这两个任务同时发布,希望可以选择一些参与者通过手机拍照或者录视频的方式记录当下某个地方的情况.由于一些低凹的路段容易形成积水,影响人们出行,所以需要记录容易发生积水路段的情况.而交通事故一般发生在一些主干道、车流量大的地方,所以需要记录这些地方的交通事故情况.由于是紧急情况,所以希望选择一些参与者可以在1小时内专门完成这些任务,并且每个人尽可能完成多个任务.为了保证收集到的数据的质量和多样性,希望每个任务由5个用户完成.假设完成每个任务的代价相同,完成每一个任务需要5分钟(由于完成任务时,需要通过拍照或者录音的方式采集数据,所以认为花费5分钟完成一个任务是合理的).因为完成每个任务的代价相同,所以,激励成本与移动距离成正比,平台为了降低成本,希望选出的参与者完成这些任务所移动的距离最短.总的来说,我们需要选择最佳的参与者集合,使参与者在完成所有任务的基础上移动的总距离最短,同时最小化完成任务的参与者人数,并且满足两个约束条件:每个任务由5个用户完成,每个参与者在1小时内完成多个任务.

4 问题定义

4.1 问题描述

MultiTasker的主要优化目标是在最小化激励成本和参与者人数的情况下优化用户资源选择,在规定时间内完成并发任务集.对于群智平台来说,希望完成任务的成本较小.成本主要包含设备消耗能量和参与者激励,本文主要研究参与者选择问题,不考虑参与者设备的消耗能量,所以希望选择的参与者激励成本最小以减少系统成本.由于参与者在一段时间内完成多个任务,所以需要改变原有的移动轨迹专门完成任务.对于参与者来说,一方面要完成任务,另一方面还要为完成任务移动一定的距离.因此,本文假设参与者激励成本的来源包括两部分:一

部分是完成任务获得的激励,在任务个数确定情况下,任务激励总成本不变;另一部分为动态激励成本,假设其与参与者移动距离成正比.参与者完成任务移动的距离越短,动态激励成本越小,系统的成本越小.另一方面,为优化用户资源,系统希望选择尽可能少的参与者完成任务,同时也增加了每个参与者的收益.

已知有 n 个感知任务 $T = \{t_1, t_2, \dots, t_i, \dots, t_n\}$, 为了保证收集到数据的质量和全面性,任务请求者规定每个任务 t_i 由 s_i 个人来完成;由于要通过拍照、录音等方式收集数据,同时为了便于分析,假设每个任务需要花费 q 分钟完成.假设当前平台中有 m 个用户 $U = \{u_1, u_2, \dots, u_j, \dots, u_m\}$, 由于是紧急任务,要求参与者在 h 小时内完成所有任务,假设每个用户移动的速度为 v m/min. 用 $UT_i = \{u_{i1}, u_{i2}, u_{i3}, \dots\}$ 表示完成任务 t_i 的参与者集合,则完成任务集合 T 中 n 个任务的参与者集合表示为 $P = \{p_1, p_2, p_3, \dots\}$ ($P = UT_1 \cup UT_2 \cup \dots \cup UT_n$). 每个参与者 u_j 完成的任务集合是 $TU_j = \{t_{j1}, t_{j2}, t_{j3}, \dots\}$, 完成这些任务所移动的总距离为 $D(TU_j)$.

MultiTasker 有两个优化目标,一个优化目标是最小化参与者完成任务所移动的总距离(式(1)),移动总距离越短,参与者的激励成本越小,则系统的成本越小;另一个优化目标是最小化选择出的参与者的人数(式(2)),在平台用户资源不足的情况下,使每个参与者尽可能完成多个任务,以充分利用用户资源,同时最大化每个参与者的收益.另外,还需要满足两个约束条件:为了保证参与者收集到的数据能反映每个任务的真实情况,避免出现不准确的数据,要求每个任务 t_i 由 s_i 个人来完成(式(3));同时,由于发布的感知任务是紧急任务,对时间要求较高,所以平台要求每个参与者在 h 小时内完成任务(式(4)),完成任务的时间包括两部分,一部分是完成每个任务所花费的时间,另一部分是移动一定的距离到达任务位置所花费的时间.形式化定义如下:

$$\min \sum_{j=1}^m D(TU_j) \quad (1)$$

$$\min |P| \quad (2)$$

约束条件:

$$|UT_i| = s_i, 1 \leq i \leq n \quad (3)$$

$$|TU_j| \times q + \frac{D(TU_j)}{v} \leq h \times 60, 1 \leq j \leq m \quad (4)$$

解决该参与者选择问题主要有两个挑战:一个是每个参与者需要在规定时间内完成多个任务;另

一个是该问题有两个优化目标.针对于第一个挑战,相对于每个参与者完成一个任务来说,该问题是一个动态的任务完成过程,参与者需要不断的改变位置完成感知任务,同时完成前一个任务也会对接下来的任务完成产生影响.针对第二个挑战,需要同时优化完成任务所移动的总距离和参与者人数,即该问题是一个多目标优化问题^[18].但是,对于多目标优化问题来说,通常最优解不止一个,而是一个最优解集^[19] (Pareto-optimal solutions).因为有多目标时,由于多个目标之间无法比较和冲突现象,不一定有在所有目标上都是最优解.一个解可能在某个目标上是最好的,但在其他目标上是最差的.多目标优化问题往往是通过转化为单目标问题求解.本文的参与者选择问题有两个优化目标,一个是最小化移动距离,另一个是最小化参与者人数.因此,我们可以将一个目标函数作为主要的优化目标,在优化主要目标的基础上,考虑另一个优化目标.

4.2 问题分析

由于该参与者选择问题是 NP 难问题^[20],因此,本文采用贪心启发算法解决问题,分别考虑问题中的两个优化目标.在以任务为中心的参与者选择中,主要优化目标是最小化参与者人数;在以用户为中心的参与者选择中,主要优化目标是最小化移动距离.

4.2.1 以任务为中心

在参与者选择中,选择出的参与者可能需要在某一时间段内完成多个任务,即参与者完成任务时位置不断变化.由于参与者的位置变化,所以如果不能确定参与者位置变化的轨迹,很难选择下一个需要完成的任务.但是,从另一个角度分析,虽然参与者的位置不断变化,但是任务的位置始终不变,所以可以通过任务组合的方式,先根据任务和用户之间的距离确定合适的参与者,然后针对该参与者选择一个可以在规定时间内完成的任务集合.同时考虑到需要参与者完成任务所移动的总距离最小,所以希望选择出离任务最近的候选者完成任务.因此,在任务组合中,可以先选择一个任务作为初始任务,然后以距离为标准,不断选择最近的任务加入任务集合.但是新加入的任务不能和任务集合中已有的任务相同,直到选择出离初始任务最近的参与者能在规定时间内完成的任务的集合.这样进行任务组合,避免了参与者位置不断变化,将参与者与任务位置联系起来,即参与者的位置根据任务集合中任务的位置变化.

4.2.2 以用户为中心

在以任务为中心的参与者选择中,主要是先确定不同的任务集合,然后选择合适的参与者完成,不需要确定每个用户完成的任务集合.不用选择每个用户完成的任务集合,会在一定程度上减少计算量,但是由于没有考虑到所有用户,所以可能选择出来的参与者性能不佳.因此,考虑参与者的整体性能,可以以用户为中心进行选择.由于参与者需要完成任务,所以每次选择完成的任务后,参与者的位置都会发生变化,完成任务的位置就是参与者的当前位置.如果参与者每次都选择离自己最近的任务完成,这样就可以保证移动的距离最短,但是,还需要同时考虑参与者在规定时间内完成多个任务.因此,以每个用户为中心,选择离其最近的任务,如果完成该任务后,用户还有时间去完成其他任务,那么继续选择离用户当前位置最近的任务去完成.这样,每个用户都有了需要完成任务的集合,并且保证移动距离较短.同时,考虑到一个参与者应该尽可能完成多个任务,选择任务个数最多的用户作为参与者完成任务.

5 算法设计

针对于该参与者优选问题,本文提出 3 种解决方法: T-Random、T-Most 和 PT-Most. T-Random 和 T-Most 是以任务为中心进行选择, PT-Most 是以用户为中心进行选择.

5.1 算法 1: T-Random

T-Random 是以任务为中心进行选择,将最小化参与者人数作为主要优化目标,即要求每个参与者在规定时间内尽可能完成多个任务.同时考虑到最小化移动距离,因此以距离最小为基准选择每个参与者完成的任务集合.先选择一个任务作为初始任务,然后选择离初始任务最近的用户完成该任务,接下来针对该参与者,选择在规定时间内完成的任务集合,即每次选择离该参与者当前位置最近的任务完成.

具体的参与者选择过程如下.随机选择一个初始任务,然后选择离该初始任务最近的参与者,由于参与者不能在同一个地方多次完成同一个任务,所以接下来选择的任务不能和前面选择的任务相同.接着选择离初始任务最近的任务,作为该参与者完成的第二个任务,选择离第二个任务最近,并且与前几个任务不同的任务作为该参与者完成的第三个任务, ..., 以此类推,按照该方法,直到选择出该参与者

在 h 个小时内完成的任务集合.这样,选择出第一个参与者完成的任务集合,然后剔除该参与者和减少其完成任务需要的人数.按照上面的方法,依次选择出参与者完成的任务集合,直到所有任务都完成.

算法 1. T-Random.

- 输入: 任务集合 T , 用户集合 U
输出: 参与者集合 P 及其完成的任务集合 TU
1. 随机选择初始任务 $t_{ij} (j=1)$
 2. 选择离初始任务 t_{ij} 最近的参与者 $p_i (i \geq 1)$
 3. 选择离任务 t_{ij} 最近的任务 $t_{i(j+1)} (j \geq 1)$
 4. 循环执行步 3、步 4, 直至参与者 p_i 完成这些任务的时间大于 $h \times 60$ 分钟
 5. 输出 $TU_i = \{t_{i1}, t_{i2}, t_{i3}, \dots\}$
 6. 循环执行步 1~6, 直至所有任务都完成
 7. 输出参与者集合 $P = \{p_1, p_2, p_3, \dots\}$ 及其完成的任务集合 $TU = TU_1 \cup TU_2 \dots \cup TU_i$
 8. 结束

5.2 算法 2: T-Most

在以任务为中心的算法中,如何选择初始任务对结果影响很大,因为初始任务决定任务如何进行组合、任务之间的距离是否最小等.考虑这些因素,本文设计了 T-Most. 与 T-Random 不同, T-Most 每次选择剩余任务中需要人数最多的任务作为初始任务.因为每次选择该任务,可以保证每个任务在整体迭代中都能被选择为初始任务,避免了某些任务在进行多次迭代后都没有被选择的情况,也就是从整体上来对任务进行组合.

选择剩余任务中需要完成人数最多的任务作为初始任务,然后选择离该初始任务最近的参与者.接着选择离初始任务最近的任务,作为该参与者完成的第二个任务,然后选择离第二个任务最近,并且与前几个任务不同的任务作为该参与者完成的第三个任务, ..., 以此类推,按照该方法,直到选择出该参与者在 h 小时内完成的任务集合.选择第一个参与者完成的任务集合 TU_1 , 然后剔除该参与者和减少任务集中任务需要的人数.按照上面的方法,依次选出参与者和其完成的任务集合 $TU_2, TU_3, \dots, TU_i$, 直到所有任务都完成.

算法 2. T-Most.

- 输入: 任务集合 T , 用户集合 U
输出: 参与者集合 P 及其完成的任务集合 TU
1. 选择剩余任务中需要人数最多的任务作为初始任务 t_{ij}
 2. 选择离初始任务 t_{ij} 最近的参与者 $p_i (i \geq 1)$
 3. 选择离任务 t_{ij} 最近的任务 $t_{i(j+1)} (j \geq 1)$
 4. 循环执行步 3、步 4, 直至参与者 p_i 完成这些任务的

时间大于 $h \times 60$ 分钟

5. 输出 $TU_i = \{t_{i1}, t_{i2}, t_{i3}, \dots\}$
6. 循环执行步 1~6, 直至所有任务都完成
7. 输出参与者集合 $P = \{p_1, p_2, p_3, \dots\}$ 及其完成的任务集合 $TU = TU_1 \cup TU_2 \cup \dots \cup TU_i$
8. 结束

5.3 算法 3: PT-Most

T-Most 和 T-Random 都是以任务为中心进行参与者选择, 考虑参与者在规定时间内完成多个任务, 主要优化目标是 minimized 参与者人数; 而 PT-Most 主要是以用户为中心进行参与者选择, 主要在最大化每个参与者完成任务的基础上, 最小化完成任务所移动的总距离。

PT-Most 先在候选者集合中随机选择一个候选者, 然后选择离该候选者最近的任务, 作为完成的第一个任务, 再选择离第一个任务最近的任务作为第二个任务, 直到选出该候选者在 h 个小时内完成的任务集合。每个候选者都有一个任务集合, 然后选择任务集合中任务个数最多的候选者作为第一个参与者, 如果任务数最多的候选者有多个, 再根据移动距离的远近作为选择标准。选择出第一个参与者完成的任务集合, 然后剔除该参与者和减少其完成任务需要的人数。按照上面的方法, 依次选出参与者完成的任务集合, 直到所有任务都完成。

算法 3. PT-Most.

输入: 任务集合 T , 用户集合 U

输出: 参与者集合 P 及其完成的任务集合 TU

1. 选择用户集合 U 中的一个用户作为候选者 p_{ji}
2. 选择离候选者 p_{ji} 最近的任务 $t_{ik} (k=1)$
3. 选择离任务 t_{ik} 最近的任务 $t_{i(k+1)} (k \geq 1)$
4. 循环执行步 3、步 4, 直至参与者 p_{ji} 完成这些任务的时间大于 $h \times 60$ 分钟
5. 输出 $TU_{ji} = \{t_{i1}, t_{i2}, t_{i3}, \dots\}$
6. 循环执行步 1~6, 选择出每个候选者 p_{ji} 在 $h \times 60$ 分钟完成的任务集合 TU_{ji}
7. 选择最大的 $|TU_{ji}|$ 作为参与者 p_j 完成的任务集合 TU_j
8. 循环执行步 1~8, 确定每个参与者完成的任务集合
9. 输出参与者集合 $P = \{p_1, p_2, p_3, \dots\}$ 及其完成的任务集合 $TU = TU_1 \cup TU_2 \dots \cup TU_j$
10. 结束

6 实验评估

6.1 数据集和实验设置

6.1.1 数据集

本文采用 D4D 数据集^[21]进行实验评估。D4D

数据集由 the Orange Group 提供, 主要包含两类数据, 一类是基站的位置信息, 即科特迪瓦(非洲西部国家)的基站分布位置, 其中包含基站的 ID, 以及基站的经度和纬度; 另一类是用户详细的通话记录, 即科特迪瓦 50 000 个匿名用户在 2011.12.01~2012.04.28 期间的通话记录(科特迪瓦的总人口为两千万), 每条记录包含用户的 ID, 通话的日期和时间以及所在基站的 ID。由于数据集中的基站数目较多, 为方便实验研究, 本文选取科特迪瓦的首都阿比让进行实验, 其基站的分布位置如图 3 所示。



图 3 基站的分布位置

6.1.2 实验设置

由于本文采取的数据集是 D4D 数据集, 其中用户的位置是通过用户的电话记录先获得基站 ID, 进而根据基站的位置来确定的。因此, 在以下实验中, 我们认为用户的位置就是当前打电话所在的基站位置。同时, 为了与用户位置对应, 我们将任务位置与基站位置关联起来, 即任务的位置就是某一个基站的位置。则用户与任务之间的距离就是两个基站之间的距离, 需要注意的是用户在一个基站范围内移动的距离为 0。

在以下的实验中, 如果没有特别的说明, 假设每个任务需要 5 个用户完成 ($s_i = 5$), 由于完成任务时, 需要通过拍照或者录音的方式采集数据, 所以我们认为花费 5 分钟完成一个任务是合理的 ($q = 5$)。本文提出的参与者选择问题主要针对紧急任务, 对时间要求高, 所以假设每个参与者需要在 1 个小时内完成多个任务 ($h = 1$)。由于以下实验主要研究任务不同对参与者选择的影响, 所以每个参与者完成任务的移动方式相同, 即本文认为参与者都是通过步行的方式完成任务, 每个参与者移动的速度为 70 m/min ($v = 70$)。

如果需要在某个时间段完成任务, 如 10:00~11:00, 应该将 10:00 候选者所在的位置作为他们的当前位置, 选择合适的参与者完成任务。但是, 由

于在 10:00 这一刻不一定有足够多的人打电话,所以选择的参与者可能不是最优的. 考虑到人的移动速度不是很快,所以可以将 10:00 之前的一段时间的位置作为候选者的当前位置,如 9:50~10:00,并且假设候选者在这一段时间内不移动. 考虑到实验的准确性,以下的实验数据都是通过多次实验平均而来.

6.2 实验评价指标

本文针对于参与者选择问题提出 3 种解决方法: T-Random、T-Most 和 PT-Most. T-Random 和 T-Most 是以任务为中心进行参与者选择,而 PT-Most 是以用户为中心进行参与者选择. 3 种方法的解决思路不同,由此选择的参与者不同. 因此,需要通过实验来验证 3 种方法的性能. 在参与者选择中,要求最小化参与者人数,并且最小化完成任务所移动的总距离. 因此,可以对比 3 种方法所选择出的参与者人数,平均每个人完成的任务数,以及完成任务所移动的总距离. 在任务个数一定的情况下,参与者总人数和平均每个人完成的任务数成反比,选择的参与者人数越少,平均每个人完成的任务数越多. 我们希望参与者在规定时间内尽可能完成多个任务,即希望选择的参与者人数越少,平均每个人完成的任务数越多. 另一个需要对比的数值是完成任务所移动的总距离. 由于在任务个数和任务位置确定的情况下,选择的参与者越少,意味着参与者需要移动较长的距离完成任务,总距离就会增大. 本文希望完成任务所移动的总距离也最短,但是,参与者人数和移动的总距离是反比关系,如果希望两者同时保持一个较好的水平,可以观察参与者人数和移动总距离的乘积,两者乘积越小越好. 所以,下面实验可以通过对比,选择出的参与者人数、平均每个人完成的任务数、完成任务所移动的总距离以及参与者人数和移动总距离的乘积,然后根据不同的情况选择出性能最好的算法.

群智平台中的任务是多种多样的,如任务位置不同、任务个数不同、任务分布不同等,由此选择出的参与者也就不同. 因此,在下面的实验中研究了任务个数、任务分布、任务完成时间等因素对参与者选择的影响. 实验研究的目的是: 如果提前了解各种因素对参与者选择的影响,在任务发布时,就可以考虑到这些因素,选择出最佳的参与者完成任务. 比如在完成个数相同,但分布不同的两个任务时,集中式和分散式任务(实验结论是,分散式任务相对于集中式

任务选择出的参与者人数更多,移动的距离更长),由于分散式任务移动的距离较长,可能有些用户不愿意移动较长的距离去完成任务,这样平台就可以提前排除这些候选者,选择出较为合适的人来完成分散式任务. 如果有些任务对时间要求不高,白天或者晚上完成都可以(实验结论是,同一个任务,白天相对于晚上选择出的参与者人数更少,移动的距离更短),由于白天完成任务对系统带来的负担较少,所以,可以选择让参与者白天完成任务.

总的来说,下面的实验主要是对设计的 3 种算法进行评估,通过对选择出的参与者人数、平均每个人完成的任务数、完成任务所移动的总距离、参与者人数和移动总距离的乘积等实验结果进行对比,根据不同的情况选择出性能最好的算法. 同时,还研究了参与者的选择情况与各种因素之间的关系,如任务分布和任务执行时间等.

6.3 实验结果

6.3.1 任务个数变化

本实验随机选择阿比让地区其中 5 天的时间进行研究,保持其他变量不变,控制任务个数变化. 以市中心的基站为中心,以距离为基准,依次选择多个任务. 将完成任务的时间设为 15:00~16:00,以 15:00 出现在该区域的用户为候选者,一共有 300 候选者.

从图 4(a)可看出,随着任务个数的增加,选择出的参与者人数不断增加,并且增加的速度越来越快. 从图 4(b)可看出,平均每个人完成的任务个数

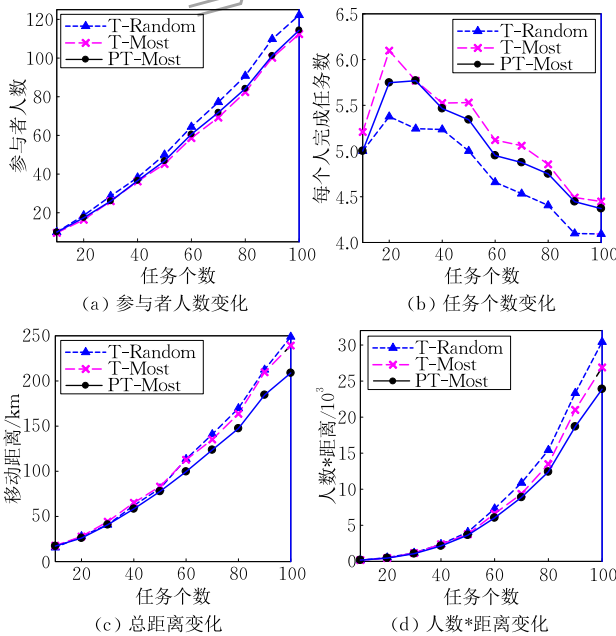


图 4 任务个数增大对 3 种算法性能的影响

先增加再不断减小. 由于任务个数增加, 任务范围相应增加, 参与者完成任务移动的距离就会增加, 因此在规定时间内完成的任务个数减小. 我们希望每个参与者尽可能的完成多个任务, 并且移动的总距离最短. 从图 4(a) 和 (c) 中可看出, T-Most 选出的参与者人数最少, 但是 PT-Most 选出的参与者移动的总距离最小. 参与者人数和移动总距离在本质上成反比, 因为任务的个数和位置一定, 如果选择较少的参与者, 那么需要移动更多的距离. 如果希望两者都保持在一个较为适合的数值, 从图 4(d) 中可以看出, PT-Most 性能会更好, 因为 PT-Most 在选择时同时考虑了这两个因素. 所以, 针对于不同的情况, 可以选择不同的算法. 如果比较注重参与者的人数, 可以选择 T-Most, 反之, PT-Most 的性能会更好.

6.3.2 运行时间对比

评价一个算法性能的好坏还需要考虑该算法的运行时间, 也就是算法的时间复杂度. 本实验研究算法的运算时间与任务个数之间的关系.

从图 5 可看出随着任务个数增加, 算法的运行时间也在增加. PT-Most 的运算时间远大于 T-Random 和 T-Most 的运算时间. 并且, PT-Most 的运行时间随着任务个数的增加增幅更大. 由于 PT-Most 是以用户为中心进行参与者选择, 如果候选者人数较多, 那么迭代一次就需要较长的时间. 虽然 PT-Most 的性能在整体上比较好, 但是, 如果对时间要求严格, T-Most 的性能会更好一些.

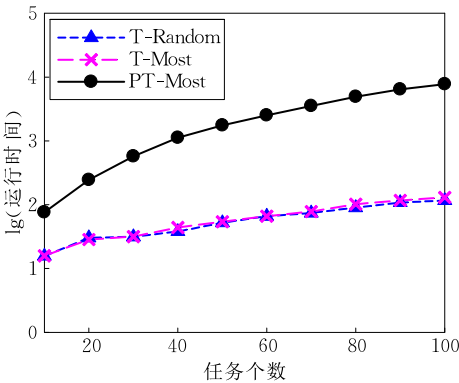


图 5 算法运行时间与任务个数的关系

6.3.3 候选者人数不同

在参与者选择中, 有两个因素对选择出的结果影响较大, 一个是任务个数, 另一个是候选者人数. 任务个数越多, 选择出的参与者就越多, 完成任务所移动的总距离就越大. 但是候选者越多, 选择的机会越多, 选择最佳参与者的机会也就越大. 本次实验主

要研究 3 种算法对参与者选择的影响. 一共 50 个任务, 保持其他因素不变, 选择多个参与者在 15:00~16:00 这个时间段完成任务.

由于候选者人数增加, 就可能选择更近的参与者完成任务, 移动总距离也会降低, 如图 6(c) 所示. 从图 6(a) 和 (b) 可看出, 候选者人数增加对选择的参与者人数没有很大影响. 因为候选者人数变化本质上是位置的变化, 但是由于候选者位置只对初始任务有影响, 而对接下来的任务完成情况没有影响, 所以对整体的任务选择影响较小, 参与者人数基本不变. 如果同时考虑候选者人数和移动总距离两个因素, 从图 6(d) 可看出, 随着候选者人数增加, 总人数与总移动距离的乘积越来越小.

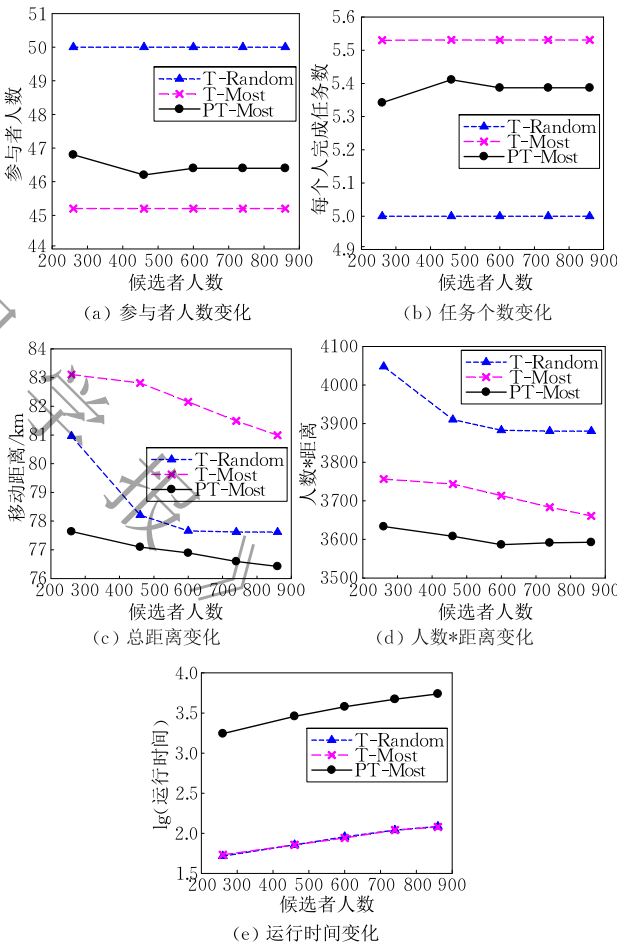


图 6 候选者人数增加对 3 种算法性能的影响

PT-Most 是以用户为中心进行参与者选择, 所以, 候选者人数对 PT-Most 的运算性能有较大的影响. 由图 6(e) 可看出 PT-Most 的运算时间随着候选者人数的增加而增加, 增加的速度比 T-Random 和 T-Most 的增加速度大. 由于 PT-Most 每次选择参与者都需要对所有候选者进行遍历, 选择出每个候

选者的任务集合,会花费较多时间. 所以,PT-Most 的运算时间主要由候选者人数决定.

6.3.4 任务分布不同

实际情况中,任务性质不同,分布也就不同. 总的来说,主要分为集中式和分散式,如火灾、地震等事件属于集中式任务,测量噪音、记录空气质量等属于分散式任务. 如图 7 所示,以市中心的基站为中心,按照距离远近,由近及远依次选择 30 个任务作为集中式任务;同样,以市中心的基站作为中心,按照其余基站到市中心基站的距离,由近及远进行排序,然后每隔 5 个基站选择一个作为分散式任务,一共选择 30 个任务.

为,集中式分布相对于分散式分布,任务之间的距离较短,所以,1 个小时内每个参与者完成的任务数比较多,选择出的参与者人数较少.

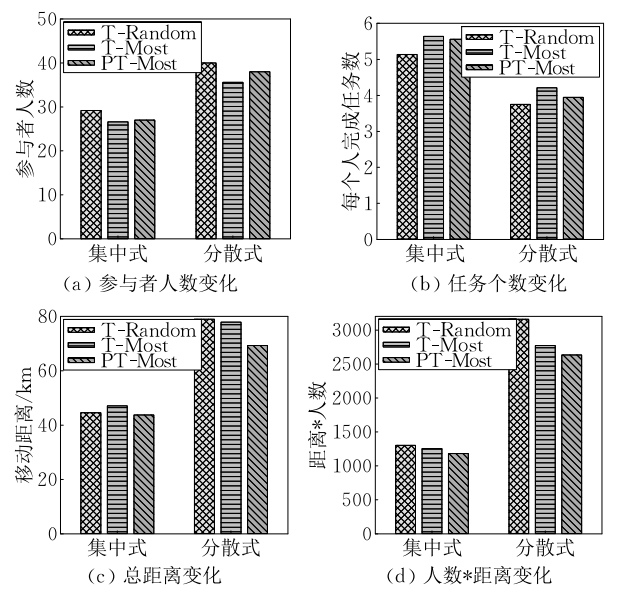


图 8 集中式和分散式任务对比

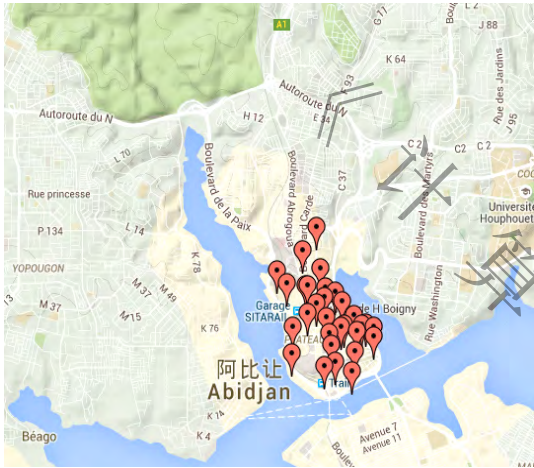
6.3.5 任务时间不同

不同的任务有不同的完成时间,完成紧急任务的时间是提前不可知的,如火灾、地震等事件. 而对于一些对时间要求不高的任务来说,如测量空气的质量、城市噪音等,只需要参与者在空闲时完成. 所以,本次实验研究任务的完成时间对参与者选择的影响,通过周内和周末的对比,白天和晚上的对比,分析实验结果的不同.

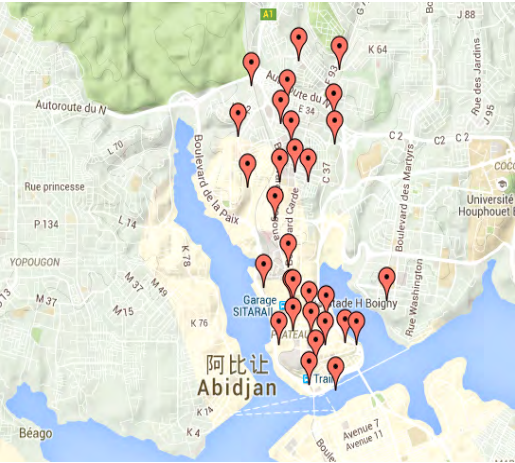
(1) 周内和周末

对于大部分用户来说,周内和周末打电话次数不同,移动轨迹不同,相应用户的分布就不同,所以同一个任务,周内和周末选择的参与者可能不同. 选取 30 个任务,分别在周内两天和周末两天进行研究,完成任务的时间段为 15:00~16:00.

从图 9 可以看出周内和周末完成任务有很大的差别. 在任务个数相同、任务分布相同的情况下,相对于周末,周内选择出完成任务的参与者较少,完成任务所移动的距离也较短,如图 9(a)和(c). 由于周末选择出的参与者人数较多,所以每个参与者完成的任务数相应的较少,如图 9(b). 如果同时考虑参与者人数和完成任务所移动的总距离,周内选择出的参与者性能更优,如图 9(d)所示. 因为,周内的候选者比较活跃,打电话的次数多,移动的范围较大,所以在任务个数相同的情况下,周内的候选者人数较多,选择出的参与者性能较好.



(a) 集中式任务



(b) 分散式任务

图 7 任务的分布位置

从图 8(a)和(c)可看出,分散式相对于集中式分布选择的参与者更多,移动距离更大. 同时由于分散式分布选择出的参与者较多,在任务个数不变的情况下,相应的每个用户完成的任务数少,如图 8(b)所示. 如果同时考虑参与者人数和移动总距离两个因素,集中式分布选出的参与者性能更优. 因

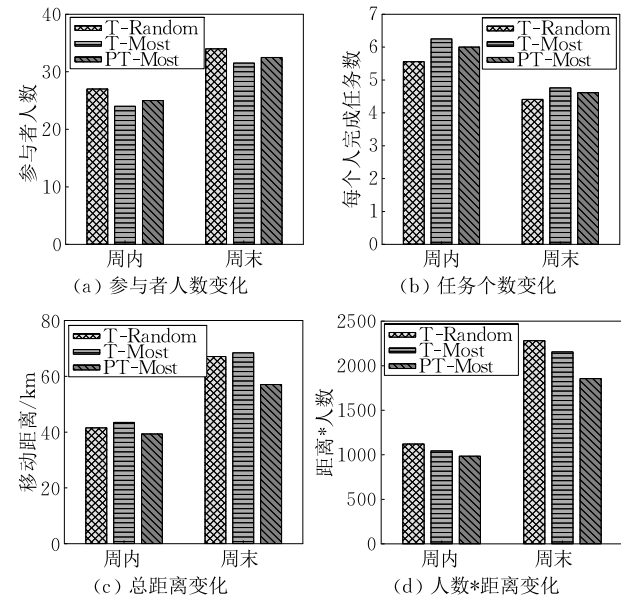


图 9 周内和周末对比

(2) 白天和晚上

上面的实验研究了周内和周末完成任务对参与者选择的影响,实验结果表明周内选择出的参与者性能较好.虽然任务完成的时间不同,分别在周内和周末,但还是在同一个时间段.而本实验主要研究在同一天,不同时间段内完成任务对参与者选择的影响.本实验总共需要完成 30 个任务,分别选取白天 15:00~16:00 完成任务,晚上 21:00~22:00 完成任务.

从图 10 中可看出,在任务个数和任务分布相同的情况下,白天选择出的参与者较少,平均每个用户完成较多的任务,如图 10(a)和(b),并且完成任务

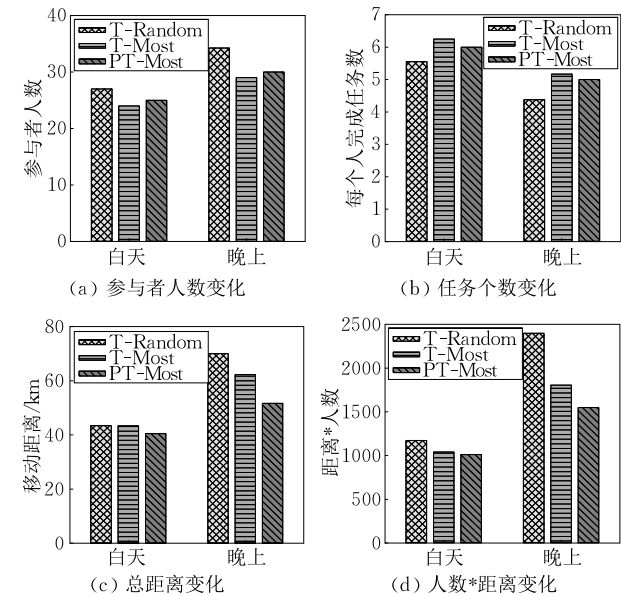


图 10 白天和晚上对比

所移动的总距离也较短,如图 10(c). 相对于中午而言,大部分用户在晚上打电话的次数比较少,移动的范围较小,因此不能获得大多数候选者的位置.与中午相比,晚上候选者的人数比少,分布范围较小,由此导致晚上选择出的参与者人数较多,移动的总距离较长.

6.3.6 时间限制不同

本文要求参与者在 h 小时内尽可能的完成多个任务,所以 h 的大小对参与者完成的任务数有很大的影响.为了研究参与者性能与 h 大小的关系,在实验中选择 3 个值,即 $h=1$ 小时, $h=1.5$ 小时, $h=2$ 小时.

如图 11 所示,当 h 增加时,平均每个人完成的任务数也增加,因此参与者人数就相应的减少.因为, h 增加,参与者移动速度不变,因此参与者有更多的时间完成多个任务.从图 11(c)可看出,参与者移动的总距离随着 h 的增加只有小幅度的增加.因为在本次实验中,任务的个数和位置都没有变化,候选者的人数也没有发生变化,只是完成任务的时间增加,但并没有影响到参与者到任务的距离,和任务到任务之间的距离,所以,完成这些任务所移动的距离变化不大.

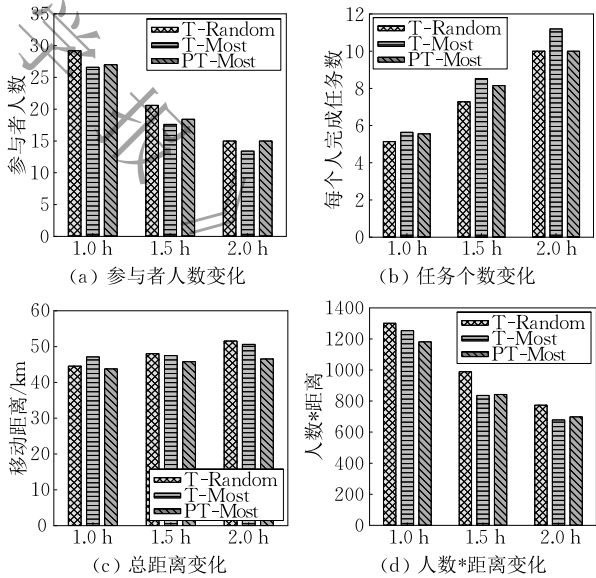


图 11 时间限制不同

6.3.7 每个任务需要的人数不同

前面的实验都是假设每个任务由 5 个参与者完成,任务之间是同构的.但在实际情况中,任务之间是有差别的,有些任务可能需要较多的用户完成以保证数据的准确性,有些任务考虑到成本就需要较少的用户完成,任务之间是异构的.考虑到这种情

况,本次实验研究同构任务和异构任务参与者选择的不同.

(1) 20 个任务

本次实验共 20 个任务需在 15:00~16:00 内完成,同构任务中每个任务需要 19 个用户完成. 在异构任务中分别选择 5、15、35 这 3 个值,占总任务的百分比分别是 30%、35%、35%,也就是说,有 6 个任务分别需要 5 个用户完成,7 个任务分别需要 15 个用户完成,7 个任务分别需要 35 个用户完成.

从图 12 可以看出在同构任务中,T-Most 选择的参与者相对较少,而在异构任务中,PT-Most 选择的参与者相对较少. 无论对于同构任务,还是异构任务而言,PT-Most 选择出的参与者完成任务所移动的距离最短. 如果同时考虑参与者总人数和完成任务所移动的总距离,PT-Most 的性能最优. 通过同构和异构任务相比较可以发现,异构任务选择的参与者人数大于同构任务所选择的参与者人数,平均每个人完成的任务数较少,如图 12(a)和(b),移动的总距离也较大,如图 12(c).

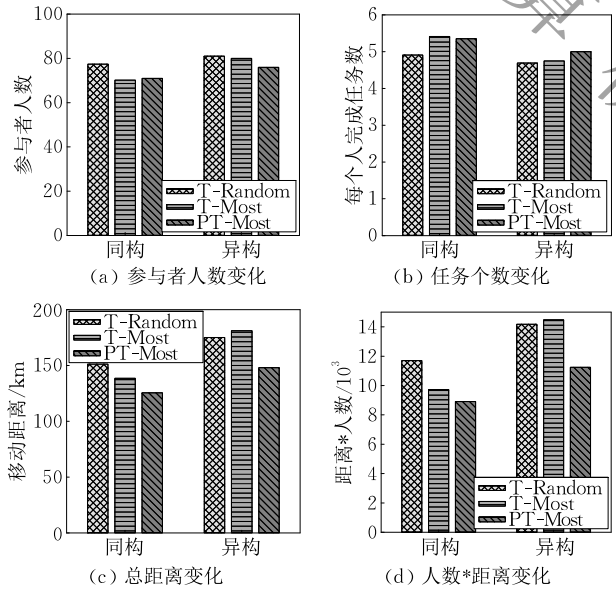


图 12 同构任务和异构任务对比

(2) 40 个任务

前面实验研究了任务个数为 20 个,同构任务和异构任务的差别. 由于任务个数的不同也会对同构任务和异构产生不同的影响,所以本次实验,一共 40 个任务需在 15:00~16:00 内完成,同构任务中每个任务需要 12 个人完成. 在异构任务中分别选择 5、15、35 这 3 个值,占总任务的百分比分别是 30%、35%、35%. 也就是说,有 12 个任务分别需要 5 个人来完成,14 个任务分别需要 15 个人来完成,14 个任

务分别需要 35 个人来完成.

通过同构和异构任务相比较可以发现,异构任务选择的参与者人数大于同构任务所选择的参与者人数,相应的平均每个人完成的任务数较少,如图 13(a)和(b)所示. 与任务个数为 20 相比较,任务个数为 40 时,同构和异构选择出的参与者认识相差较大. 在同构任务中,每个任务之间是平衡关系,而在异构任务中,每个任务需要的参与者人数相差较大,所以最后可能导致剩余需要参与者人数最多的任务,由此需要选择出更多的参与者,移动的总距离增加.

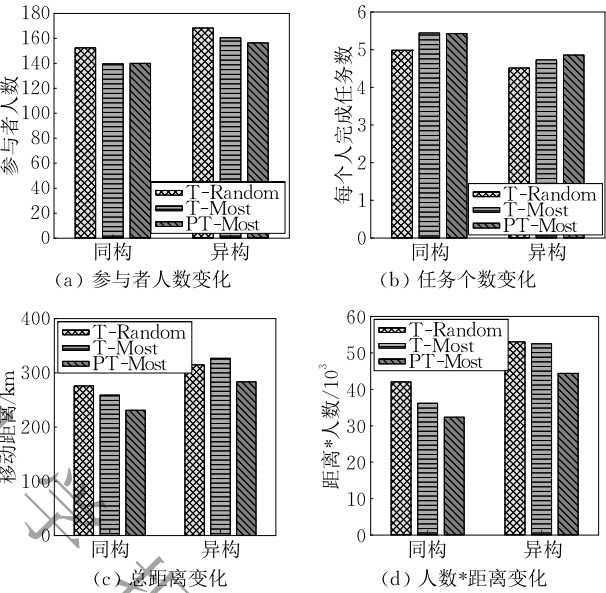


图 13 同构任务和异构任务对比

6.3.8 应用实例

以上实验研究了参与者的选择情况与各种因素之间的关系,如任务分布和任务执行时间等,接下来通过一个应用实例说明在多种因素的影响下,参与者的选择情况.

一个城市中的商业区和住宅区有明显的差异,如商业区人口分布密集,人员流动性强、移动范围大,而住宅区人员较少,人员分布较广等. 由于多种因素的影响,在商业区完成任务和住宅区完成任务是有差异的,比如需要测量城市的空气质量,就需要分别在商业区和住宅区选择参与者完成任务,由于人员分布的不同,所以选择的参与者人数也就不同. 本次实验主要研究商业区和住宅区对参与者选择的影响.

在阿比让地区,商业区和住宅区的基站数目和分布都不同. 为了保证实验的一致性,在商业区和住宅区都分别选择 30 个任务,任务分布如图 14 所示,任务完成的时间为 15:00~16:00.

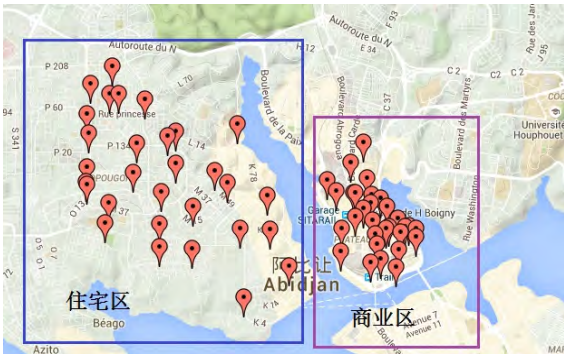


图 14 商业区和住宅区任务的分布位置

商业区和住宅区基站分布不同,主要体现在商业区的基站分布比较密集,而住宅区分布比较分散.也就是说,商业区的任务分布形式是集中式,住宅区的任务分布形式是分散式.除了基站分布情况不同,商业区和住宅区还有一个最大的不同点是,商业区中人们打电话的次数比较多,流动性强,这样获取候选者的位置就会很容易,也就是说相对住宅区而言,商业区的候选者人数较多.在这两个因素的影响下,商业区和住宅区的实验结果相差很大,如图 15 所示.在任务个数相同的情况下,商业区选择的参与者人数小于住宅区选出的参与者人数,并且商业区的移动总距离小于住宅区移动的总距离,如图 15(a)和(c).从整体来看,同时考虑参与者总人数和移动距离,在商业区和住宅区中完成任务,PT-Most 的性能相对于其他两个算法更好一些.

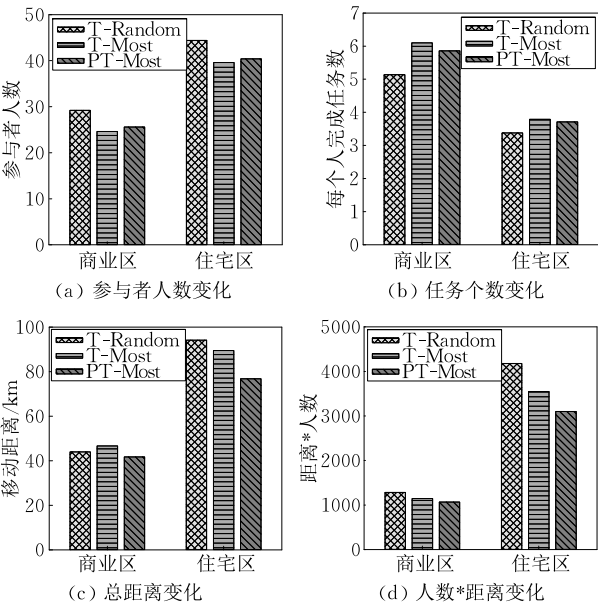


图 15 商业区和住宅区的对比

6.4 讨论

本文提出 MultiTasker 方法以实现多任务并发环境下的参与者优选.其优化目标是最小化完成任

务移动的总距离和参与者人数,同时满足参与者在规定时间内完成所有任务.但是,由于处于研究的初级阶段,所以本文还存在一些不完善的地方,可以作为未来研究的重点:

(1) 参与者移动性.城市中每个人的移动方式不同,如开车、坐公交车、骑车、步行等.在本文中,认为参与者的移动方式相同,即每个参与者都相同,没有考虑参与者之间的差异性.在以后的工作中,群智平台可以将每个参与者作为一个单独的个体,考虑到每个人的移动方式可能不同.在任务分配时,任务分发系统就可以根据参与者的移动性分配合适的任务,比如紧急任务就可以优先分配给移动速度较快的参与者,他们可以开车去完成.

(2) 时间约束不同.本文解决的问题是参与者在规定时间内尽可能完成多个任务,并且完成任务所移动的总距离最短.由于需要参与者在 h 小时内完成所有任务,所以要求参与者有意识的去访问各任务地点执行任务,而不必遵循原有规划路径.本文主要考虑在 h 较小的情况下选择参与者,但 h 较大时,文中提出的算法就不适用,因为在这种情况下,参与者有足够的时间完成任务.如果参与者能够不改变原有的轨迹去完成任务,就可以减少参与者的负担.

(3) 任务执行时间不同.本文假设完成每个任务的代价相同,并且完成每一个任务都需要 5 分钟,没有考虑在实际情况中,每个任务具有差异性,某些任务比较简单,需要的时间短,而复杂的任务需要的时间长.所以,在以后的工作中,可以结合实际设置任务的完成时间.

7 总结与展望

本文主要研究面向移动群智感知的任务分发方法,提出并解决了面向多任务并发的参与者选择问题.针对该问题,设计了 3 种解决方法: T-Random、T-Most 和 PT-Most. T-Random 和 T-Most 是以任务为中心进行参与者选择,不同之处是 T-Random 随机选择一个任务作为初始任务进行任务组合,而 T-Most 选择剩余任务中需要人数最多的任务作为初始任务进行任务组合. PT-Most 是以用户为中心进行参与者选择.最后,通过大规模的数据集对 3 种算法进行评估,实验结果显示,如果希望每个参与者尽可能完成多个任务,同时移动的总距离最短,那么 PT-Most 的性能更好一些,但是算法的运行时间却

是最长的.同时,还研究了参与者的选择情况与各种因素之间的关系,如任务分布和任务执行时间等.

在选择参与者时,本文只考虑参与者的位置关系,没有考虑其他因素,即认为每个参与者都是相同的个体,没有很好的利用参与者本身具有的专业技能.因此,在以后的研究中,可以在参与者选择时将每个人具有的专业技能作为一个选择标准,以此提高任务完成的质量和效率.

致 谢 感谢我的导师郭斌老师的指导与督促,不厌其烦的帮我修改论文,感谢吴文乐师兄的热心帮助,帮我解决实验中出现的问题!

参 考 文 献

- [1] Guo Bin, Yu Zhiwen, Zhang Daqing, Zhou Xingshe. From participatory sensing to mobile crowd sensing//Proceedings of the IEEE Pervasive Computing and Communication Workshops (PerCom'14). Budapest, Hungary, 2014: 593-598
- [2] Zheng Yu, Liu Furui, Hsieh Hsun-Ping. U-Air: When urban air quality inference meets big data//Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'13). New York, USA, 2013: 1436-1444
- [3] Coric V, Gruteser M. Crowdsensing maps of on-street parking spaces//Proceedings of the 2013 IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS'13). Massachusetts, USA, 2013: 115-122
- [4] Ra Moo-Ryong, Liu Bin, La Porta T F, Govindan R. Medusa: A programming framework for crowd-sensing applications//Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services (MobiSys'12). New York, USA, 2012: 337-350
- [5] Reddy S, Burke J, Estrin D, et al. A framework for data quality and feedback in participatory sensing//Proceedings of the 5th International Conference on Embedded Networked Sensor Systems. New York, USA, 2007: 417-418
- [6] Cheung M H, Southwell R, Hou Fen, Huang Jianwei. Distributed time-sensitive task selection in mobile crowdsensing //Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'15). New York, USA, 2015: 157-166
- [7] He Shibo, Shin Dong-Hoon, Zhang Junshan, Chen Jiming. Toward optimal allocation of location dependent tasks in crowdsensing//Proceedings of the 2015 IEEE Conference on Computer Communications (INFOCOM'14). Hong Kong, China, 2014: 745-753
- [8] Wang Leye, Zhang Daqing, Pathak A, et al. CCS-TA: Quality-guaranteed online task allocation in compressive crowdsensing//Proceedings of the 2015 ACM International

- Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'15). New York, USA, 2015: 683-694
- [9] Yan Zhixian, Banerjee N, Chakraborty D, et al. PUCAA: 1st international workshop on pervasive urban crowdsensing architecture and applications//Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication (UbiComp'13 Adjunct). Zurich, Switzerland, 2013: 1055-1062
- [10] Cardone G, Foschini L, Bellavista P, et al. Fostering participation in smart cities: A geo-social crowdsensing platform. IEEE Communications Magazine, 2013, 51(6): 112-119
- [11] Li Cheng-Te, Shan Man-Kwan. Team formation for generalized tasks in expertise social networks//Proceedings of the 2010 IEEE Second International Conference on Social Computing. Washington, USA, 2010: 9-16
- [12] Lappas T, Liu Kun, Terzi E. Finding a team of experts in social networks//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09). Paris, France, 2009: 467-476
- [13] Sozio M, Gionis A. The community-search problem and how to plan a successful cocktail party//Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'10). Washington, USA, 2010: 939-948
- [14] Yang De-Nian, Shen Chih-Ya, Lee Wang-Chien, Chen Ming-Syan. On socio-spatial group query for location-based social networks//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'12). Beijing, China, 2012: 949-957
- [15] Papadias D, Shen Qiongmao, Tao Yufei, Mouratidis K. Group nearest neighbor queries//Proceedings of the 20th International Conference on Data Engineering. Washington, USA, 2004: 301-312
- [16] Zhang Daqing, Xiong Haoyi, Wang Leye, Chen Guanling. CrowdRecruiter: Selecting participants for piggyback crowdsensing under probabilistic coverage constraint//Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'14). Seattle, USA, 2014: 703-714
- [17] Reddy S, Estrin D, Srivastava M. Recruitment framework for participatory sensing data collections//Proceedings of the Pervasive Computing. Helsinki, Finland, 2010: 138-155
- [18] Sugihara H, Komoto J, Tsuji K. A multi-objective optimization model for urban energy systems in a specific area//Proceedings of the 2014 IEEE International Conference on Systems, Man and Cybernetics. California, USA, 2002: 6
- [19] Wang Jing, Zhou Yongsheng, Yang Haoxiong, Zhang Hao. A trade-off pareto solution algorithm for multi-objective optimization//Proceedings of the 5th International Joint Conference on Computational Sciences and Optimization (CSO'12). Harbin, China, 2012: 123-126

[20] Cardei M, Thai M T, Li Y, et al. Energy-efficient target coverage in wireless sensor networks//Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies(INFOCOM'14). Hong Kong, China, 2005; 1976-1984

[21] Blondel V D, Esch M, Chan C, et al. Data for development: The D4D challenge on mobile phone data. arXiv preprint arXiv: 1210. 0137, 2012



LIU Yan, born in 1993, M. S. Her research interest is mobile crowd sensing.

GUO Bin, born in 1980, Ph.D. , professor. His research interests include pervasive computing and mobile

crowd sensing.
WU Wen-Le, born in 1990, M.S. candidate. His research interest is mobile crowd sensing.
YU Zhi-Wen, born in 1977, Ph. D. , professor. His research interests include pervasive computing and social sensing computing.
ZHANG Da-Qing, born in 1964, Ph. D. , professor. His research interest is ubiquitous computing.

Background

This research is partly supported by the National Basic Research Program of China (973 Program) under Grant No.2015CB352400, the National Natural Science Foundation of China under Grant Nos.61332005, 61373119, 61222209.

With the rapid proliferation of sensor-equipped mobile devices, mobile crowd sensing has become a popular research field. The proliferation of sensor-equipped mobile devices and the mobility of people are making mobile crowd sensing an effective way to sense and collect information at a low deployment cost. This paper studies the participant selection issue in mobile crowd sensing.

In recent works, several researches on selecting participants, such as, team formation, community search, and online task assignment, have been studied. Reddy et al. first study participant selection problem, they propose a coverage-based recruitment strategy to select a predefined number of participants in order to maximize the spatial coverage. Chien-Ju Ho and Jennifer Wortman Vaughan study the problem of assigning heterogeneous tasks to workers with

different, unknown skill sets in crowdsourcing markets such as Amazon Mechanical Turk. De-Nian Yang et al. propose Socio-Spatial Group Query to select a group of nearby attendees with tight social relation.
We propose a new multitask-oriented participant selection problem. The difference between our work and other studies in participant selection is that each participant can complete as many tasks as possible within the given time period, rather than merely one task. To address this issue, we propose MultiTasker, the objective of which is to minimize the total distance the selected participants move, subjecting to that all tasks are allocated with the needed number of participants and the number of tasks completed by each selected participant is maximized within the given time period. In order to achieve the objective, We propose three algorithms: T-Random, T-Most and PT-Most. We evaluated the three algorithms using a large-scale real-world dataset, and studied the relationship between participants and other factors, such as the distribution of tasks.