

群智感知中基于公交系统的任务分发机制研究

安 健^{1,2)} 彭振龙^{1,4)} 桂小林^{1,2),3)} 向乐乐¹⁾ 梁丹薇^{1,3)}

¹⁾(西安交通大学电子与信息工程学院 西安 710049)

²⁾(西安交通大学深圳研究院 广东 深圳 518057)

³⁾(西安交通大学陕西省计算机网络重点实验室 西安 710049)

⁴⁾(云计算物联网电子商务智能福建省高校工程研究中心 福建 泉州 362000)

摘 要 任务分发作为实现群智感知的重要环节,为了激励更多用户参与数据的采集和共享,已有研究通常利用机会网络进行任务扩散并降低参与者的执行成本,但在节点选择过程中并未充分考虑感知任务的多样性将对节点接触时间、参与数量、感知区域类型等带来的挑战.基于此,该文利用城市中公交载体的轨迹可预测、活动覆盖范围大、乘客节点自主聚集且交互时间有保证等优势,提出了一种基于公交系统的任务差异化分发方法.首先,利用泰森多边形的划分思想,实现感知任务与目标区域的合理覆盖.其次,对感知区域内一定预算约束下的任务差异化分发问题进行分析,并分别提出了两种分发算法:基于覆盖差异的分发算法(COV-DA)和基于扩散差异的分发算法(SCR-DA).最后,利用真实数据集,通过仿真实验从分发准确性和平均移动距离对两种算法的性能进行比较分析.实验结果表明,SCR-DA算法具有更好的分发准确性,而使用 COV-DA 算法时,移动节点在完成任务时则需要更短的移动距离.

关键词 群智感知;任务分发;公交系统;预算约束

中图法分类号 TP393 DOI号 10.11897/SP.J.1016.2019.00295

Research on Task Distribution Mechanism Based on Public Transit System in Crowd Sensing

AN Jian^{1,2)} PENG Zhen-Long^{1,4)} GUI Xiao-Lin^{1,2),3)} XIANG Le-Le¹⁾ LIANG Dan-Wei^{1,3)}

¹⁾(School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049)

²⁾(Xi'an Jiaotong University Shenzhen Research School, Shenzhen, Guangdong 518057)

³⁾(Shaanxi Province Key Laboratory of Computer Network, Xi'an Jiaotong University, Xi'an 710049)

⁴⁾(E-Commerce Intelligence of Engineering Research Center of University of Fujian Province Based on Cloud Computing and Internet of Things, Quanzhou, Fujian 362000)

Abstract As mobile smart devices become increasingly popular and are equipped with increasingly powerful sensors, they have been pervasively applied in crowdsensing as effective tools to solve large-scale sensing tasks in urban areas. Task requesters can allocate sensing tasks to mobile nodes through a crowdsensing platform, eliminating the cost of deploying and maintaining large numbers of fixed sensors. However, some crowdsourcing tasks (e. g. audio and video transmission) that generate a large amount of sensed data may bring higher network traffic costs to participants using a 3G/4G network, which may affect their enthusiasm and cause enormous transmission pressure on the network. To reduce communication costs, if sensed data does not require real-time uploading, researchers have applied opportunistic network and “store-carry-forwarding” to

收稿日期:2017-07-04;在线出版日期:2018-03-20. 本课题得到国家自然科学基金(61502380)、深圳市科技计划项目(JCYJ20170816100939373)、福建省科技计划重点项目(2016H0029)资助. 安 健,男,1983年生,博士,高级工程师,中国计算机学会(CCF)会员,主要研究方向为群智感知、物联网. E-mail: anjian@mail.xjtu.edu.cn. 彭振龙,男,1977年生,博士研究生,主要研究方向为群智感知. 桂小林(通信作者),男,1966年生,博士,教授,主要研究领域为物联网、云计算. E-mail: xlgui@mail.xjtu.edu.cn. 向乐乐,男,1993年生,硕士研究生,主要研究方向为群智感知. 梁丹薇,女,1995年生,硕士研究生,主要研究方向为群智感知.

transmit data. But they have ignored a significant problem that it is impossible for a large amount of data to be transferred in such a short momentary encounter. With the increasing sensing capabilities of smart devices, the types and sizes of sensed data have changed greatly. The data transmission paradigm in an opportunistic network is based on the assumption that data exchange can be completed as long as the mobile nodes can connect with each other. It implies that the data size can be ignored or that the data can be exchanged in a transient encounter between mobile nodes. In this paper, we built a data diffusing and transmission paradigm in crowdsensing based on a City Public Traffic System (PTS), and emphatically thoroughly discuss a paradigm for Multi-Task diffusion and transmission within budget constraints. This paradigm makes full use of the advantages of a bus to realize the rapid transmission of large-scale sensed data; predictable trajectory, wide coverage area, fast moving speed and long contact duration among passengers. First, we use the division of voronoi diagram to divide the target area and achieve the reasonable coverage of sensing tasks and target area. Secondly, we analyze the problem of task differential distribution under certain budget constraints, and propose two distribution algorithms: distribution algorithm based on coverage difference (COV-DA) and distribution algorithm based on diffusion difference (SPR-DA). Both transmission schemes are under budget constraint that aims to maximize the overall transmission utility of all data brought into the PTS. By diffusing and scheduling all data in the PTS, the data can be uploaded to the crowdsensing platform by proper mobile passengers. Finally, to effectively evaluate the performance of our algorithm, the COV-DA and SPR-DA algorithms are compared with the classical Greedy and Epidemic algorithms in detail from the accuracy of task distribution, the average moving distance and the time of task distribution. The experimental results show that the SPR-DA algorithm has better distribution accuracy, while using the COV-DA algorithm, the mobile node needs a shorter moving distance when the task is completed. Compared with Greedy and Epidemic algorithm, COV-DA and SPR-DA have better performance in distributing accuracy and average mobile distance, and can achieve a tradeoff between overall transmission utility and transmission redundancy, and save more network traffic costs and other resources for mobile nodes.

Keywords crowd sensing; task assignment; public transit system; budget constraint

1 引 言

任务分发作为实现群智感知^[1-2]的重要环节,它是指组织或个人利用不同的激励机制和约束手段,以竞争和合作等多种自主协同方式来共同完成众包任务.其中,感知任务的合理分配将是影响用户参与和任务完成质量的一个关键因素.为了减少对通信基础设施的依赖,以及降低参与者的任务开销,很多研究将机会网络^[3-6]引入到了数据传输过程中,即依靠移动节点间的机会接触,在弱连接的网络环境下进行数据的传输.

考虑到群智感知服务的多样性,任务信息的类型、大小都不断发生变化,已有基于节点间短暂交互的数据机会式传输手段很难保证服务质量.一方面,

大部分研究建立在数据大小可以忽略不计这一假设条件之上^[7-9],没有考虑节点间的交互时间,但在现有短距离传输技术条件下,感知任务中涉及的图片、视频这一类较大数据的传输需要保证节点间存在比较稳定且持久的接触时间.此外,实际情况下,节点间的交互时间是较难预估的,已有的部分研究仅将其作为一个必要因素加入路由节点的选择标准中,但对交互时间的长短却没有提出合理的预估方法,这无疑降低了研究的实用性.另一方面,忽略了任务特征与感知区域内部的相关性差异,如感知任务主要聚集在几个具有明显属性差异的区域内,如果不对区域的特征因素进行分析就直接进行任务分发,则会导致任务分发缺乏合理性,最终使得节点执行任务时的代价过大,影响节点参与感知服务的积极性.

通过观察发现,在城市公交系统中,公交设施

(公交站点、公交线路)遍布整个城市区域,是一种天然的城市覆盖资源,公交又是城市居民的一种重要出行方式,巨大的客流量足以满足群智感知中对移动节点的数目要求. Zhou 等人^[10]以城市公交系统为背景,利用手机中收集的各种上下文环境信息和节点间的协作与共享,实现了对公交到站时间的有效预测. Liu 等人^[11]基于公交乘客“参与式”的数据感知和轨迹映射,进一步地实现了对城市交通的监测. 同时公交乘客间存在稳定持续的接触时间,且由于在公交固定移动线路的限制下,乘客间的接触时长也可以转化为计算公交载体在两个停靠点间的运行时间,这些研究成果对于本文解决群智感知中多媒体数据的机会传输有着重要的指导意义.

基于此,本文将城市公交系统引入到任务分发过程中,利用公交载体在城市活动过程中的轨迹可预测、覆盖范围大、移动速度快以及乘客节点交互时间有保证等优势. 首先,构建基于公交系统的任务差异化分发模型,利用泰森多边形的划分思想,实现感知任务与目标区域的合理覆盖;其次,对感知区域内一定预算约束条件下的任务分发问题进行分析,并针对性提出两种算法:基于覆盖差异的分发算法(Coverage Difference Based Distribution Algorithm, COV-DA)和基于扩散差异的分发算法(Spread Difference Based Distribution Algorithm, SPR-DA),以解决节点交互时间、任务与区域相关性等因素导致的分发效率低下等问题. 最后利用真实数据集,从分发效率、准确性和平均移动距离等多个指标进行对比分析.

2 相关工作分析

群智感知实现了对移动用户自身和服务交互过程中产生的多源、异构信息的快速收集. Guo 等人^[12]系统性的对群智感知相关概念、关键技术、应用类别、计算模式等进行了说明,考虑到任务执行过程中可能带来的对用户正常活动的干扰以及额外的资源消耗,Guo 等人指出了服务节点的招募和选择是进行合理、有效任务分发的关键因素,这也是近年来该研究方向中的一个热点问题. 通过分析近年来领域代表性论文^[3,5-6,13-14]可以看出,针对群智感知环境下的任务分发与节点选择可以归结为两类:(1)基于复制策略的选择算法,这种策略主要通过提高数据副本的冗余度来增加端到端传输的成功概率,降低传输延迟,这一类算法中最具代表性的是传

染病路由(Epidemic routing)算法、洪泛算法等;(2)基于启发策略的选择算法,感知任务候选节点主要根据网络拓扑动态变化、上下文环境信息等动态制定节点策略,并按照一定的优化目标进行任务扩散. 代表性的如遗传算法(Genetic routing)、贪婪算法(Greedy routing)等.

Han 等人^[6]提出利用机会通信来促进数据的扩散,重点研究了目标节点集合的选择问题,并提出了 Greedy、Heuristic 和 Random 三种算法来进行合理选择. Tuncay 等人^[13]针对在以人为中心的移动感知网络,提出了一种利用机会感知的分布式数据扩散办法,实现对感知任务的快速分发. Gao 等人^[14]则针对节点的移动性和连接间断性的特点,提出了一种基于社会交互模式的数据传播方式,实现感知数据在非持续连接网络中的可靠传递. Tian 等人^[15]则将 Epidemic 扩散路由算法应用于车联网中,研究了不同路网环境下的数据扩散效率. 以上方法虽然可以直接减少任务分发过程中的资源消耗,然而,目标节点集的选择仅仅考虑了节点的接触频率,在任务从目标集合中节点扩散出去的过程中,会产生巨大数量的信息副本. 同时,对于任务大小与节点间接触时间的内在关联性也未过多涉及.

Guo 等人^[16]重点研究了面向多任务的群智感知服务节点选择问题. 针对感知任务的时间敏感性,基于改进的遗传算法,分别设计了基于任务驱动的“参与式”任务分发机制(GGA-I)和基于轨迹预测的“机会式”任务分发机制(GGA-U). Luo 等人^[17]提出了一种利益最大化的众包任务竞拍激励机制,针对机会感知任务的应用特点设计了贡献依赖激励函数,并对其进行优化,使得主体利益最大化,但是研究关注的重点都是节点自身,任务类型、区域特征等属性往往处于被动地位. 南文倩等人^[18]为了提高任务众包用户的参与积极性,提出了动态激励模型 CSII,利用跨空间数据分析并通过系统、任务发布者和参与者三者之间的多元交互来激励用户. Karaliopoulos 等人^[19]则利用贪婪算法思想,在最小传输时延和最大化任务收益之间进行目标优化,以实现群智感知环境下众包节点的最优化选择.

在任务属性聚类方面,Wang 等人^[20]提出了 CCS-TA 的任务分发框架,该框架利用空间和时间相关性,动态的选择最小数目的感知子区域用于收集数据,同时在一定概率的准确性要求下推导未分配感知节点的子区域的缺失数据. Dang 等人^[21]提出了一种 Zoom 任务分发框架,该框架利用地图来

对任务信息进行编码,包括感知的对象、内容以及地点.该方法将感知区域划分为多个大小相等的正方形区域,但是这种栅格化的划分方法并不能体现区域内部的差异性,感知平台也就无法根据实际情况对任务进行差异性的分发. An 等人^[9]利用层次分析法,基于服务质量因子、服务链路因子和感知区域热度因子的三层服务选择模型,实现对众包节点的合理分配,但对服务节点与众包任务之间的内在关联关系也并未过多涉及.

综上所述,现有研究主要通过一定的约束手段和激励机制,基于节点的活动轨迹和机会式接触来实现感知任务扩散.对于任务内涵、区域属性、节点接触时间等因素往往考虑不足.本文在引入城市公交系统的基础上,构建了众包任务的差异化分发模型,并分别提出两种不同算法,以期进一步提高任务的分发效率和准确性.

3 基于公交系统的任务差异化分发模型

如图 1 所示,基于公交系统的任务差异化分发模型包含三个组件:感知平台、公交系统以及分发控制模块.公交系统主要包含乘客节点和公交载体,乘客节点会进入到公交载体内,构成一个信息分发扩散的封闭单元,任务信息只能在内部进行扩散流动.由于在公交载体的移动过程中,每一个站点处都可能节点的加入和离开,因此该封闭单元从时间角度来看又是动态变化的.感知平台会通过分发控制模块来指导任务的扩散分发,分发控制模块内部则主要包含三个部分:区域划分、区域覆盖以及任务的

扩散调度.

首先,当感知平台得到任务请求者提出的服务需求后,会根据所有感知目标的位置分布,在分发控制模块中将整个感知区域划分为多个子区域,并计算每一个子区域与任务的相关性以及所占比重,作为差异化分发的依据;其次,平台会根据可用的分发预算、公交线路的移动轨迹和时间表选择出一定数目的公交线路来完成对整个感知区域的覆盖,并进行任务信息的投放,使得每一个被选择的目标线路上都存在可扩散分发的任务信息.这里将所选择出来的线路集称为目标线路集合,对应的公交载体内直接接收到任务信息的乘客节点构成的集合称为目标节点集合.最后,分发控制模块会根据节点间的交互时间以及子区域的任务相关性来设计合理的扩散调度方案,指导任务信息在对应公交载体内部的间接扩散.从而达到任务信息差异化分发的效果.本文设定乘客节点间通过所携带移动设备上的蓝牙传感器完成信息的扩散传输.

乘客的下车信息则是通过节点的历史乘车记录来进行推断,相关研究^[22-23]指出,公交乘客在日常出行中主要频繁来往于几个“偏好站点”,且具有一定的规律性,通过分析乘客节点在一段时间内的公交出行行为,就大致可以推断出乘客出行的下车站点,具体分析乘客在一段时间(如一个月)内的主要出行起讫对、出行时段(如上午、下午、晚上)和所乘车次,当处于某时段的一个起讫对的出行次数达到规定的阈值(如一月五次),则将该出行起讫对定义为该乘客节点的一个出行模式,当乘客节点有新的出行行为时,通过 NFC 和 GPS 可以获取所乘车次、上车站点以及时段因素,因此可以选择出最符合条件的出行模式,进而推断出乘客的出行目的地,再通过公交线路的营运时刻表,就可以获得节点的下车时刻点以及在公交载体内的停留时间.

分发控制模块是该模型中感知平台和公交系统进行交互的纽带,在分发模块的三个步骤中,区域划分是任务分发的前提步骤,用于获取差异化分发的依据,而后两部分则是与分发算法直接关联的部分,其中区域覆盖是在预算约束下选择直接分发的对象,即目标线路集合的选择,任务的扩散调度则是在区域覆盖基础上,生成每一个公交载体内任务信息的间接扩散方案.

3.1 区域划分

在群智感知中,常见的区域划分方法是根据经纬度,对目标区域进行栅格化处理,即将被感知区域

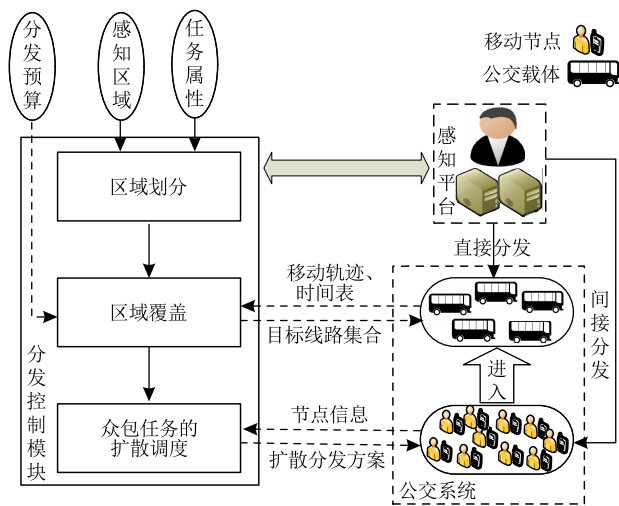


图 1 基于公交系统的任务信息分发模型

划分为若干个大小相同的正方形,若一个用户出现在某个正方形区域内,则称这个正方形区域被该用户覆盖.这种对整个区域无差别的划分方法简单易操作,能够覆盖整个感知区域,但并不能体现不同区域内部之间差异性.感知任务的属性类别、粒度、时效性经常会有所不同,这些都与区域内部特征相关联,而传统无差别的栅格划分方法很难做到这一点.泰森多边形也被称为 Voronoi 图,最早由荷兰气象学家提出,用于根据离散分布的气象站的观测数据,计算平均降雨量,现多用于统计学中,用来完成定向分析、统计分析等^[24].泰森多边形可用于对含有若干离散点的平面进行几何剖分,能够根据区域分布特征进行差异化划分.

基于此,本文提出了一种基于感知目标类别的区域划分方法,该方法从聚类角度出发,根据任务类别的位置分布,来获取区域内多个聚类中心点,随后根据具体目标类别的特征属性,利用泰森多边形的思想对子区域进行细粒度的划分,并计算每一个子区域的任务相关性以及所占比重.这里,感知目标类别是根据任务请求者给出的具体感知任务来确定的.感知需求在发布时首先需要对任务类别进行分析,来确定对应的目标客体,而感知任务所需要获取的信息则可看作是对目标客体相关属性的采集和分析.例如,当感知任务为收集某一城市区域内的餐饮信息时,则感知目标类别为餐厅;若要收集某一区域的景点分布状况时,则感知目标类别对应为旅游景点.

将整个需要进行感知的区域记为 A , A 的范围由任务提出者确定.假设目标类别 γ_i 对应的位置坐标记为 (x_i, y_i) , 则位置坐标集合可记为 $L = \{(x_i, y_i)\}_{i=1}^n$. 选用 DBSCAN 算法对整个感知区域内的相同目标类别进行聚类,并得到包含多个坐标点的小集合,记为 $C = \{c_1, c_2, \dots, c_m\}$. 随后求出每一个聚集区的中心坐标,这里直接通过集合内节点的坐标平均值来表示,对于聚集区 c_i 而言,中心点 $o_i = (x_i^*, y_i^*)$ 的计算如式(1)所示,记 C 对应的中心点集合为 $O = \{o_i\}_{i=1}^m$.

$$x_i^* = \frac{\sum_{(x_i, y_i) \in c_i} x_i}{|c_i|}, y_i^* = \frac{\sum_{(x_i, y_i) \in c_i} y_i}{|c_i|} \quad (1)$$

DBSCAN 算法仅是将任务目标聚类成为多个集合,并未进行明确的区域划分.在得到每一个聚集区的中心坐标点后,则利用泰森多边形法对整个感知区域进行几何意义上的划分.记每一个划分出来的泰森多边形称为子区域,并将所有的子区域集合

记为 $A = \{a_i\}_{i=1}^m$, 对应的每一个子区域的中心点为 $O = \{o_i\}_{i=1}^m$. 由于对目标区域进行了不规则划分,导致划分后不同子区域覆盖范围不一致.因此,将区域内相关事物的数量作为衡量子区域差异性的方法并不合理.基于此,本文将任务特征类型和相关区域内事物密度两种因子结合起来,对子区域的任务相关性比重进行计算.

记感知区域内不同任务的特征所需的感知权重集合为 $F = \{f_i\}_{i=1}^z$. 则各特征类型任务在不同子区域的相关性比重为 $P = \{\{p_{i,j}\}_{i=1}^z\}_{j=1}^m$, 其中,

$$p_{i,j} = f_i \rho_{ij} \quad (2)$$

其中, ρ_{ij} ($1 \leq i \leq z, 1 \leq j \leq m$) 是指第 j 个子区域内特征类别 i 所对应感知事物的分布密度,且满足 $\sum_{j=1}^m \rho_{ij} = 1$.

例如:以获取感知区域内餐厅信息为例,如任务请求者给出的餐厅的特征类型为高、中档餐厅,且给出的对应感知比例分别为 0.6 和 0.4. 针对特定的任务感知需求,需首先统计每一子区域内不同特征类别事物在整个目标区域的密度.如目标区域包含子区域为 3 个,且不同子区域内高级餐厅的密度分别为 0.5、0.3、0.2,中档餐厅分布密度为 0.6、0.2、0.2,则对应的 3 个子区域的高级餐厅任务相关性比重为 0.3、0.18、0.12,中档餐厅任务相关性比重为 0.24、0.08、0.08.

3.2 感知区域覆盖

由任务分发模型可知,感知平台在初始分发过程中首先会选择少量的公交线路构成目标线路集合,并直接通过 3G/4G 将任务信息投放到该集合中每一条线路上,而任务信息的间接分发过程都是位于公交载体之内,即在直接获得任务信息后,目标节点只能在公交载体内进行任务信息的间接分发,因此目标节点集的选择问题就转化成为公交线路的选择问题.

假设所需要分发的任务信息 q 的大小为 $w(q)$, 且直接通过 3G/4G 网络分发所耗费的移动数据费用记为 b ($b = \Phi \times w(q)$), 其中, Φ 为任务信息大小和节点携带费用之间的转换参数.由于任务信息不能无节制的通过 3G/4G 网络分发,需要在一定分发预算范围内.本文将分发预算记为 B , 由任务请求者给出,在此条件下,可直接分发的任务信息数量为 N ($N = B/b$), 即最终选择的进行数据分发扩散的公交线路为 N 条.设定感知区域 A 内的公交线路集合为 R_A , 将每一条线路 r ($r \in R_A$) 位于区域 A 内的公交站点集合记为 $E(r)$, 同样将每一子区域 a_i 内存在的

公交站点集合记为 V_i , 在此基础上, 定义区域覆盖的概念如下:

定义 1(区域覆盖). 若公交线路在某一区域内的公交站点停靠, 则认为公交线路覆盖这一区域. 对于某一公交线路 $r(r \in R_A)$ 和子区域 a_i 而言, 覆盖需要满足如式(3)所示.

$$E(r) \cap V_i \neq \emptyset \quad (3)$$

根据如上定义, 使用 $G(r, a_i)$ 来表示线路是否覆盖子区域 a_i , 如式(4)所示.

$$G(r, a_i) = \begin{cases} 1, & E(r) \cap V_i \neq \emptyset \\ 0, & \text{反之} \end{cases} \quad (4)$$

将最终选择的覆盖感知区域 A 的公交线路集合记为 R'_A , 由于可能会将某一线路选择多次, 而车次不同, 移动载体内乘客节点也是不同的, 因此它们之间是存在差异的, 为了进行区别. 这里将这些相同的线路分别按照顺序记为 $\{r_1^1, r_1^2, \dots, r_1^{\kappa}\}$, 其中 r_i 为 R'_A 内的第 i 条公交线路, 而 κ 为 R'_A 内 r_i 线路被选择的次数.

在完成公交线路集合 R'_A 选定后, 需要在公交线路 $r(r \in R'_A)$ 的移动过程中选择明确的乘客节点作为任务信息的直接分发者, 本文直接选择 r 线路上出行区间最长的乘客节点作为直接分发的目标节点, 记为 $u^{long}(r)$, 并从该节点开始将任务信息在移动载体内扩散分发出去.

3.3 任务的扩散调度

为了简化问题, 将公交线路 r 上的所有公交站点按照顺序分别编号为 $\{1, 2, \dots, H(r)\}$, 其中 $H(r)$ 为 r 线路上的所有站点的总数目. 定义乘客节点 u 的上下车站编号分别为 e_u^1 和 e_u^2 . 首先定义公交线路 r 上在编号为 $h(1 \leq h \leq H(r))$ 的站点处上车的乘客集合为 $Q_h(r)$, 该值只能在公交载体实际到达 h 站点处才能得到. 当位于 h 站点时, 定义在子区域 a_i 内下车的乘客节点集为 $U_h^i(r)$, $U_h^i(r)$ 的公式表示如式(5)所示.

$$U_h^i(r) = \{u \mid u \in \bigcup_{j=1}^{h-1} Q_j(r), e_u^2 \in V_i\} \quad (5)$$

可以看出, $U_h^i(r)$ 由下车站点 h 之前所有已在公交载体内但目的地属于子区域 a_i 内公交站点的乘客节点组成. 由于每一个公交站点都可能存在乘客节点的加入和离开, 而这些加入的节点中就可能存在感知区域内站点处下车的乘客, 这些都是任务信息的潜在扩散对象, 因此在每一个站点处, 都需要对在目的地位于感知区域内的乘客节点集进行更新.

假设某一公交线路 r 上的到站时间是确定可知的, 记为 $T(r) = \{t_1, t_2, \dots, t_{H(r)}\}$, 其中 t_i 为公交载体在第 i 个公交站点的到达时间, 若某一乘客节点 u 进入到该公交载体, 则该节点在移动载体内的停留时间为 $t_{e_u^2} - t_{e_u^1}$. 设定蓝牙方式下的数据传输速度是一致不变的, 记为 $speed$. 由于乘客间不同的出行计划, 乘客节点间在移动载体内的接触时间是不同的, 任务信息的转发所需的近距离接触时间也是不能忽略的, 因此需要对任务信息在节点间的扩散复制时间进行定量分析. 对于任务信息 q 来说, 在蓝牙传输方式下转发所需的时间为 $t_{trans}(q)$, 如式(6)所示.

$$t_{trans}(q) = \omega(q)/speed \quad (6)$$

在数据的扩散分发过程中, 对任务信息的携带者来说, 需要寻找交互时间满足数据转发时间的节点对象进行扩散. 为了对节点间的交互时间进行描述, 本文定义空闲时间对和可调度序列这两个有关的概念.

定义 2(空闲时间对). 描述节点间关于某一时段的空闲关系. 若节点 a 和节点 b 均在某一时段区间 $[t, t']$ 是空闲的, 则称两者存在关于时间段 $[t, t']$ 的空闲时间对 FTP(Free Time Pair). 其中空闲是指用户的蓝牙设备无数据传输任务.

定义 3(可调度序列). 描述两个节点间关于某一数据所存在的空闲时间对, 记为 TP(Transmission Pair). 首先将发送数据的发送方节点称为出节点, 而接收数据的称为入节点. 针对从出节点 a 发送的数据 q , 若出节点 a 和入节点 b 在时刻 t_{min} 之后存在关于区段 $[t_s, t_e]$ 的空闲时间对, 其中 t_s 是数据传输的起始时刻且 $t_s \geq t_{min}$, 而 t_e 等于 t_s 和 $t_{trans}(q)$ 之和, 则称两者存在时刻 t_{min} 之后关于数据 q 的可调度序列, 记为 $TP = \langle q, t_{min}, a, b, [t_s, t_e] \rangle$.

如图 2 所示, u_1 和 u_2 在移动载体内的停留时间分别为 $[t_0, t_5]$ 和 $[t_0, t_6]$, 其中深色段为节点的忙碌时段, 即携带的蓝牙设备存在数据传输任务. 时段 $[t_1, t_2]$ 和 $[t_3, t_4]$ 是这两个节点间的空闲时间对, 若存在某一由 u_1 携带的数据需要向 u_2 复制扩散, 且所需的扩散时间段 $[t_s, t_e] \subseteq [t_1, t_2]$ 或者 $[t_s, t_e] \subseteq$

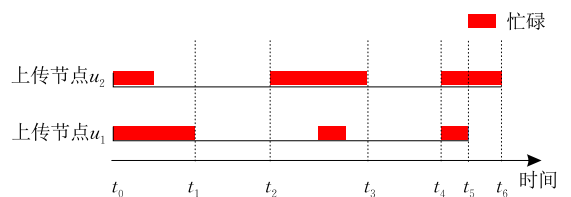


图 2 可调度序列示意图

$[t_3, t_4]$, 则称 u_1 和 u_2 之间存在关于数据 q 的可调度序列. 通过上面的例子可以看出, 可调度序列不止一个, 将其中 t_i 值最小的称为最早可调度序列, 记为 ETP (Earliest Transmission Pair).

公交载体在特定线路的整个移动过程中, 乘客节点是流动的, 即每一个公交站点都可能会有节点的加入和离开, 而节点间的接触时间作为珍贵的竞争资源, 需要充分利用, 因此每一个移动载体在到达一个停靠点 h , 感知平台需要根据新增节点的目的地所属的子区域, 执行动态的扩散调度算法, 生成扩散调度方案, 指导任务信息在新增的可用节点集中进行差异化的分发扩散. 在公交线路的起始点处, 就有可能存在目的站点位于感知区域内的节点对象, 而随着公交载体在固定线路上的移动, 当其离开在感知区域 A 内的最后一个站点时, 则不再存在可用的扩散分发对象, 考虑到以上情况, 动态分发调度算法的执行只需限定在移动线路上的 $[1, h_A^{\max}(r)]$ 站点间即可, 其中 $h_A^{\max}(r)$ 为 $E(r)$ 的移动线路上最后一个位于感知区域 A 内的公交站点.

定义 4 (可调度序列集). 描述两个节点集之间在某一时刻存在的可调度序列的集合. 若节点集 Y 内均为携带数据 q 的节点集, 而 Z 内均为未携带数据 q 的节点集合, 则 Y 和 Z 内节点间存在的所有 ETP 的集合称为可调度序列集, 记为 $SCH(Y, Z)$. 特别地, 当 Y 和 Z 任意一个为空时, $SCH(Y, Z)$ 也为空.

与定义下车节点集一样, 本文定义 $r(r \in R'_A)$ 线路上在 h 站点运行任务信息分发调度算法后, 携带数据 q 且目的地站点属于子区域 a_i 的节点集合为 $C_h^i(r)$, 由于 $h_A^{\max}(r)$ 为 $E(r)$ 中最后一个站点, 则 $C_{h_A^{\max}(r)}^i(r)$ 表示最终 r 线路上被分发到任务信息且目的站点位于子区域 a_i 内的乘客节点, 并记 $C^i(r)$ 为线路 r 在子区域 a_i 内接收到任务信息的节点集合.

定义 5 (扩散调度方案). 描述 R'_A 集合中每一个移动载体内任务的扩散对象以及时机, 感知平台在每一个移动线路 $r(r \in R'_A)$ 内都会建立一个任务信息 q 的扩散调度子方案 $\Omega(r)$. 该子方案由有多个可调度序列组成. 在 R'_A 内移动载体的移动过程中, 总的扩散调度方案则为这些子方案的并集, 记为 $\Gamma = \bigcup_{r \in R'_A} \Omega(r)$.

4 任务差异化分发算法

这一部分首先对感知区域内一定预算约束下任

务的差异化分发问题进行定义, 随后针对该问题提出了两种不同的分发算法: 基于扩散差异的分发算法 (Spread Difference Based Distribution Algorithm, SPR-DA) 和基于覆盖差异的分发算法 (Coverage Difference Based Distribution Algorithm, COV-DA), 并对两种算法进行了详细的介绍.

4.1 任务差异化分发问题定义

本文的主要目的是进行任务信息的差异化分发, 而差异化依据则是 3.1 节中得出的每一个子区域的任务相关性比重, 比重值越大的子区域则应该分发到的任务信息的数目就越多, 从而实现感知效率的最大化.

定义 6 (预算约束下任务的差异化分发问题). 记移动线路 r 最终在子区域 a_i 内分发扩散的任务信息数量为 $\#C_{h_{\max}(r)}^i(r)$, 则目标线路集合 R'_A 在子区域 a_i 内分发的任务信息数量为 $\#\bigcup_{r \in R'_A} C_{h_{\max}(r)}^i(r)$, 则定义该子区域占感知区域内分发的任务信息总数量的比重为 $p'_i (p'_i = \#\bigcup_{r \in R'_A} C_{h_{\max}(r)}^i(r) / \sum_{i=1}^m \#\bigcup_{r \in R'_A} C_{h_{\max}(r)}^i(r))$. 在此基础上, 定义任务信息的差异化分发问题如下:

$$\text{Minimize } \sum_{i=1}^m |p'_i - p_i| \quad (7)$$

$$\text{Subject to } |R'_A| = N, N = B/b \quad (8)$$

$$SCH(\bigcup_{i=1}^m C_h^i(r), \bigcup_{i=1}^m U_h^i(r) - \bigcup_{i=1}^m C_h^i(r)) = \emptyset \quad (9)$$

其中, 式(7)用来衡量分发算法最终达到的差异化分发效果, 需要首先计算最终分发到任务信息的乘客节点位于每一子区域内的数量以及占总数量的比重, 而后计算所有子区域的任务分发比重与任务相关性的差异程度来表示最终的差异化分发效果.

需要注意的是, 一方面任务信息的直接分发必须满足分发的预算条件, 即直接分发的移动流量花费不能超过分发预算 B , 对应的选择出来的公交线路覆盖集合的数量必须是与可直接分发的任务信息数量是相等的, 也就是须满足式(8). 另一方面, 由于任务信息分发的基本目标在于将任务扩散给更多的节点从而保证能够收集到足够的数据, 因此移动载体内的任务信息的分发过程中, 需要保证载体内携带任务信息的节点集和未携带的节点集的可调度序列集为空, 也就是需要满足式(9)中的条件, 其中 $\bigcup_{i=1}^m C_h^i(r)$ 表示在 r 线路上的 h 站点处进行分发算法后, 感知区域内所扩散到任务信息的节点集合, 而相应的 $\bigcup_{i=1}^m U_h^i(r) -$

$\bigcup_{i=1}^m C_h^i(r)$ 则表示感知区域内未携带任务信息的节点集合.

4.2 算法设计

由 3.1 节可知,区域划分仅是根据特定的任务要求以及相关事务的物理位置对整个感知区域所做的平面划分,并计算了每一个子区域的任务相关性比重,为任务信息的差异化分发提供依据,但该部分没有实际对任务信息的分发扩散提出相应的策略.区域覆盖和任务信息分发调度则是决定差异化分发效果的重要部分,一方面,区域覆盖用于选定直接分发的线路集合,而该集合又决定了子区域内目标节点集合的差异;另一方面任务的扩散调度用于生成在单个封闭的公交载体内部任务信息在节点间的扩散调度方案,包括扩散的对象以及时机,该部分决定了单个分发源的分发效果,因此在本文中,任务信息的差异化分发算法由目标线路集合的选择和任务的扩散调度这两部分构成.

一般情况下,如果选择的公交线路集合对某一个子区域的覆盖程度高,则在该子区域内部最终分发到的任务信息数目就越多,因此在分发预算有限的情况下,需要设定覆盖策略来选择公交线路的目标集合.同样地,单个公交载体内部,在节点间接触时间有限的限制条件下,需要对任务信息进行扩散,使得最终的分发性能最优.本文提出了两种将覆盖策略和分发调度策略结合起来的分发算法,两种算法都需要完成一定预算下目标线路集合的选择来确定感知区域内的分发源,以及生成单个公交载体内任务信息的分发调度方案,只是在选择和调度的标准上存在差异性,因此在介绍两种算法之前,首先定义覆盖需求向量 $\mathbf{K} = \{k_i\}_{i=1}^m$ 用来设定每一个子区域所需要的被覆盖比重,同时定义扩散调度需求向量 $\mathbf{W} = \{w_i\}_{i=1}^m$ 为单个公交载体内设定的每一个子区域所需要的扩散比重.

(1) 基于覆盖差异的分发算法

COV-DA 算法结合了区域的平均覆盖策略和公交载体内任务信息的差异化扩散调度策略,即首先选择合适的公交线路集合,保证每一子区域被覆盖的次数是均等的,这样也就保证了每一个子区域内任务信息的分发扩散源数目是相当的,随后针对每一个封闭的公交载体,再依照子区域的任务相关性比重,控制分发调度的优先次序,在节点间交互时间的限制条件下,有差别地进行选择性扩散.

在区域覆盖部分,将覆盖需求向量 \mathbf{K} 中的每一

个值设定为 $1/m$,保证每一个子区域的被覆盖次数是均等的,目标线路集合的选择算法则以定义 1 为基础,确定每一个子区域与公交线路的关系,以覆盖需求向量 \mathbf{K} 为目标,进行覆盖线路选择.在 COV-DA 算法中,该问题等同于一个负载均衡问题,是一个 NP 难问题,本文使用常见的贪婪式算法来进行目标线路集合的选择,下面给出区域内目标线路集合的选择算法的简要实现过程,如算法 1 所示.

算法 1. 目标线路集合选择算法.

1. 输入子区域集合 A , A 内公交线路集合 R_A , 可选的公交线路数目 N , 覆盖需求向量 \mathbf{K} ;
2. 计算每一条线路 $r(r \in R_A)$ 加入到目标线路集合 R'_A 与覆盖需求向量之间的差异值 $\sum_{i=1}^m \left(\frac{\sum_{r' \in R'_A \cup \{r\}} G(r', a_i)}{\sum_{i=1}^m \sum_{r' \in R'_A \cup \{r\}} G(r', a_i)} - k_i \right)^2$;
3. 选择与覆盖需求向量间差异值最小的线路 r_{\min} 加入到目标线路集合 R'_A 中;
4. 重复步骤 2 和步骤 3, 直到目标线路集合中元素数目为 N 个;
5. 输出目标线路集合 R'_A ;
6. 结束.

在任务的扩散调度部分,最优子区域是指在该子区域分发一个任务信息可以最小化与扩散调度需求向量之间的差异.由于任务信息分发的另一个重要目的就是使更多的节点接收到任务信息,因此在最优子区域内无法找到合适的分发对象,则应继续在次优内寻找,直到节点间不存在可调度序列.具体算法如算法 2 所示.

算法 2. 任务信息的扩散调度算法.

1. 输入任务信息 q , 目标线路集合 R'_A , 扩散调度需求向量 \mathbf{W} , 乘车节点集 $\{\{U_h^i\}_{h=1}^{k_A^{\max}}\}_{i=1}^m$, 初始分发节点 $u^{long}(r)$;
2. 针对每一条线路 r , 在停靠点 h 处从 R'_A 中选择最优扩散子区域 ms ;
3. 计算子区域 ms 内的可调度序列集 $TSCH$;
4. 从 $TSCH$ 中选择距离子区域 ms 中心点最近的乘客节点加入到该子区域的扩散节点集合中,并将对应的可调度序列 $mETP$ 加入到扩散调度子方案;
5. 重复步骤 2, 步骤 3 和步骤 4;
6. 输出扩散调度方案 Γ ;
7. 结束.

(2) 基于扩散差异的分发算法

SPR-DA 算法结合了区域的差异化覆盖策略和公交载体内任务的平均扩散调度策略,即首先选择合适的公交线路集合,保证每一子区域被覆盖的次数比重与子区域的任务相关性比重的差异度最小,

这样保证每一个子区域内任务信息的分发扩散源数目达到一定的差异化, 随后针对每一个封闭的公交载体, 则是在节点间接触时间的限制条件下, 在每一个子区域内进行均等的扩散调度。

在区域覆盖部分, 将覆盖需求向量 \mathbf{K} 设定为等同于子区域的任务相关性比重, 具体的选择过程则也是依照算法 1 中的感知区域覆盖算法来选择目标线路集合。在任务信息分发调度部分, 则设定扩散调度需求向量 \mathbf{W} 内的每一个元素均为 $1/m$, 使得在每一个公交载体移动过程完成后, 最终每一个子区域内分发到的任务信息数目是均等的。具体的扩散调度子方案的生成也依照算法 2 中任务信息的扩散调度算法得到。

5 实验结果与分析

5.1 实验数据

本文随机选取北京市内 100 条公交的实际运行线路数据, 包括线路沿途的站点位置信息以及该线路的班次信息, 并通过 TaxiQuery 生成不同密度级别的乘客出行记录模拟乘客乘车过程。TaxiQuery 是根据北京市内 33000 辆出租车在 2011 年 3 月到 5 月的载客记录而建立的一个实验平台, 用于模拟城市内用户的出行行为。该平台可以根据需要产生不同数目, 不同时段的用户打车记录, 每一条记录包含打车的始发点, 目的点以及打车时间这几个关键字。为了体现公交乘客在不同站点的出行分布, 本实验首先通过该平台产生若干条打车记录, 再利用百度地图的 API 将每一条打车出行记录转化为公交出行记录, 并为 100 路公交线路筛选出相关的乘客出行记录, 这样就模拟得到了每一条公交线路上的乘客出行记录。

实验将感知任务设为“获取北京市三环内所有餐馆的信息, 包括餐馆外观、菜单类型以及推荐菜品等图片信息”。因此, 餐馆位置就是感知任务的目标区域。餐厅数据通过爬虫获取大众点评网的 4778 家餐厅数据来进行实验验证, 包括餐厅的类别(西餐、中餐等)、餐厅的位置坐标等。

5.2 评价指标

一是分发准确性, 由 3.3 节可知, 基于公交系统的任务分发的目标是在一定预算约束下每一个子区域内所分发到的任务信息数目所占的比重接近于对应的任务相关性比重。任务分发的准确性由最终的分发差异度来确定, 差异度越小, 则分发的准确性越

高, 反之则越低, 因此定义分发准确性 DA 如式(10)所示。

$$DA = \frac{\sum_{i=1}^m \left(1 - \frac{|p'_i - p_i|}{p_i}\right)}{m} \times 100\% \quad (10)$$

另一个则是平均的感知移动距离, 即任务信息的接受者到达任务地点(本实验指餐厅)的移动距离, 由于无法确定接收到任务信息的节点中哪些愿意去完成任务以及前往哪个任务点进行任务感知, 因此本文假设所有接收到任务信息的节点都作为感知节点存在, 并将其到达子区域中心点距离的平均值作为节点执行任务的平均移动距离。设每一个子区域 $a_i (1 \leq i \leq m)$ 内分发到的节点集合为 CU_i , 则定义平均感知移动距离 AMD 如式(11)所示。

$$AMD = \frac{\sum_{i=1}^m \sum_{u \in CU_i} dis(e_u^2, o_i)}{\sum_{i=1}^m |CU_i|} \quad (11)$$

选择分发准确性作为评价指标是因为分发的准确性越高, 则说明该分发计划越接近根据任务要求所计算出来的任务相关性比重, 即越重要的区域内分发到的任务信息越多, 反之则越少, 从而能够在整个感知区域内获得更高的感知效率; 选择平均的感知移动距离作为评价指标是因为通常感知节点在完成感知任务时会对自身的移动轨迹造成影响, 因此在任务分发后节点前往执行任务时的移动距离越小, 则越能调动节点的积极性。

5.3 对比方法

为了验证所提方法的有效性, 针对现有群智感知节点选择策略, 分别与 Epidemic 算法^[15]和 Greedy 启发式算法^[19]进行比较。其中, Epidemic 算法主要是通过随机转发实现感知任务的快速扩散。Greedy 的思想则是在最小传输时延和最大化任务收益之间进行目标优化。考虑到本文研究场景与已有方法的不同, 因此无法在相同的环境下与现有这两个算法进行直接比较, 我们主要将已有文献中的节点选择策略引入到公交场景下, 在不改变原算法核心思想的前提下, 引入感知任务扩散过程中的交互时间、可用时间序列等因素, 进一步保证实验的公平性和有效性。本文主要比较了任务分发准确性、平均移动距离、任务分发时间等指标。

5.4 实验结果

5.4.1 区域划分结果

首先通过 DBSCAN 聚类算法对北京市三环内的所有餐厅进行聚类。设定聚类过程中的聚类半径为 1000 m, 子区域内最少节点数为 50, 根据上述两

个参数能够划分出 57 个子区域,如图 3(a)所示,每一个红点表示一个餐厅位置,通过 DBSCAN 聚类得到 57 个聚类中心点,在此聚类基础上,根据泰森多边形法则对该区域进行划分,最终得到 57 个子区域,划分结果如图 3(b)所示.为了方便计算,在对

感知区域 A 完成划分后,本实验将每一子区域内的餐厅数目的比例作为感知任务区域任务相关性比重.随后通过 Taxiquery 模拟北京市三环内不同密度的公交出行客流量,来对算法的分发性能进行验证.

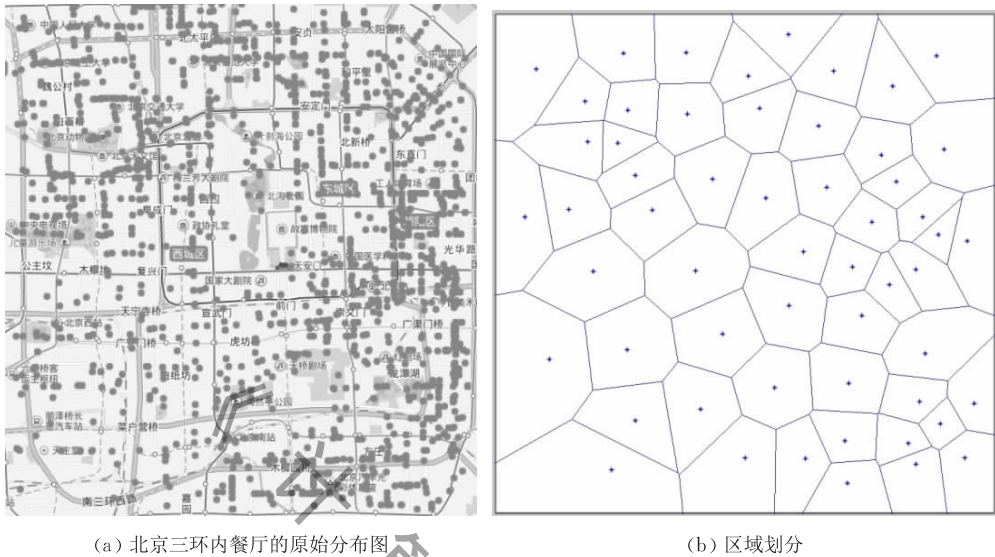


图 3 基于泰森多边形的感知区域划分结果

5.4.2 任务信息分发准确性测试

设定任务的信息包大小为 9 MB 且固定不变时,不同乘客流量密度下的分发准确性如图 4 所示.随着乘客流量密度的增加,4 种算法下的分发准确性都有缓慢的增长,并且最终都趋于平衡状态,这是因为更多节点的存在可以满足分发扩散算法的选择,使得不同子区域内的分发差异性更加明显,从而分发的准确性更高.后期逐渐趋于平衡则是因为任务总体预算、可利用的分发扩散时间是有限的,对于每一个节点来说,无论是作为出节点还是入节点,都需要消耗可用的接触时间,因此即使载体内存在足

够多的乘客节点,也无法进一步提升准确性.针对本文两种算法的性能比较而言,SPR-DA 算法的平均准确性要高于 COV-DA 约 8.4%,高于 Greedy 算法约 13.2%,这是因为各子区域内活动节点的密度是不同的,节点密度大的区域,则公交客流量也大,因此在公交载体内部更利于差异化的分发,从而分发的准确性也相对于 COV-DA 要高. Greedy 算法并未考虑区域差异性对在节点选择带来的影响,在节点选择过程中遵循一定的启发式规则,即在满足最小传输延时的情况下,同时保证任务收益的最大化.虽然能够找到覆盖目标区域候选节点集,但从分发准确性上与本文算法还是有一定差距. Epidemic 算法在任务转发中只是单纯的选择可用的下一跳节点,对于区域属性、节点轨迹等都未涉及,故准确性最低.

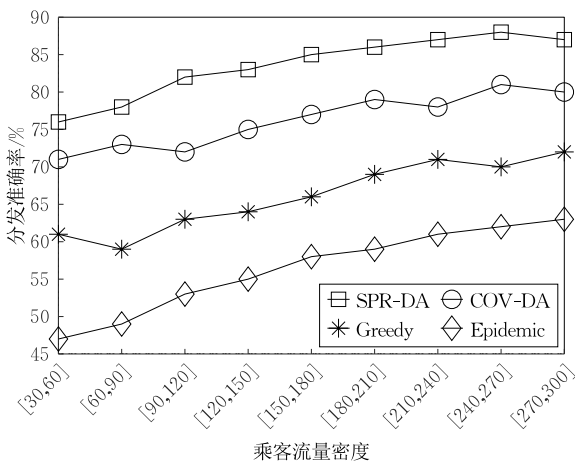


图 4 不同乘客流量密度下的分发准确性比较

设定乘客流量密度为 $[240,270]$ 区间,而在任务的信息包大小变化的情况下,4 种算法下的分发准确性如图 5 所示.随着任务信息大小的增加,4 种算法的分发准确性都呈现一定的降低趋势,这是由于任务信息越大,任务转发成本越高,在进行节点选择时,可选的调度序列就越少,从而无法满足差异化分发需求.其中 SPR-DA 算法平均可达到 84.5%,COV-DA 算法平均为 78.4%,Greedy 算法平均 65.3%.对于本文两种算法而言,依旧是 SPR-DA

算法的分发准确性要大于 COV-DA 算法,这是因为各子区域内活动节点的密度是不同的,节点密度大的区域,则公交客流量也大,因此移动载体内部更倾向于差异化的分发,从而分发的准确性也相对于 COV-DA 算法要高.考虑到 Epidemic 算法在节点选择过程中的随机性,其准确性最低.

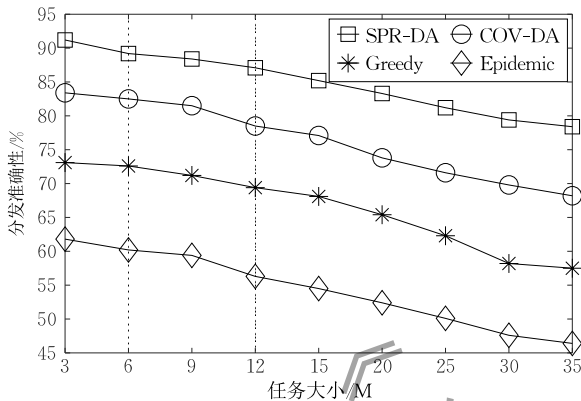


图 5 不同任务信息大小下节点平均移动距离比较

5.4.3 任务感知的平均移动距离测试

实验分别在任务信息大小固定和乘客流量密度固定两种条件下对节点感知任务时的平均移动距离进行了测试.

当任务的信息包大小固定为 9 MB 时,节点的平均移动距离随乘客流量密度大小的变化趋势如图 6 所示,随着乘客流量密度的增加,4 种算法下的平均移动距离都呈现出不同程度的减少并逐渐趋于平稳,这是因为当乘客流量较小时,可选择的合适候选节点数量有限,几乎每一个可用节点都可被分发到任务转发信息,因此平均移动距离较大.随着乘客流量密度的增加,距离目标距离相对较近的可选节点变多,故 4 种算法的平均移动距离都有减少.相较于

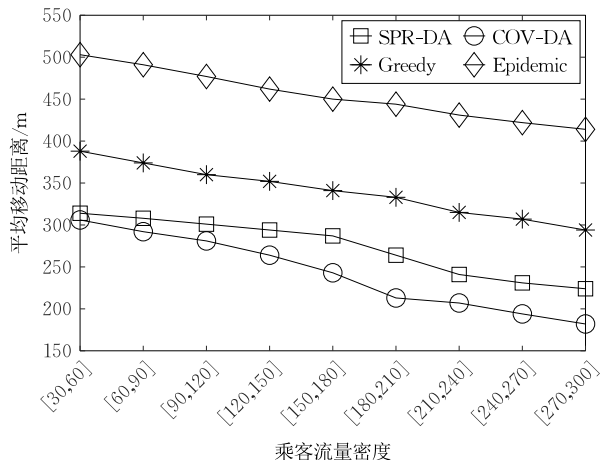


图 6 不同乘客流量密度下的节点平均移动距离比较

本文所提方法,COV-DA 的指标表现最好,节点平均移动距离为 242.5 m,SPR-DA 为 286.4 m,Greedy 为 340.5 m,Epidemic 最差,平均达到了 454.8 m.

值得注意的是,当乘客流量较小时,COV-DA、SPR-DA 两种算法的平均移动距离大致相当,但随着乘客流量的增加,COV-DA 算法的性能要好于 SPR-DA,这是因为 COV-DA 算法设定移动载体内平均地为每一个子区域选择分发对象,因此可以充分的选择距离每一个子区域中心点较近的乘客节点,从而最终的移动距离就比较近,随着乘客流量的进一步增加,节点间接触时间有限,因此最终两种算法的平均移动距离性能又重新趋于平稳.

当乘客流量密度固定为 [240,270] 区间时,节点平均移动距离随任务信息大小的变化趋势如图 7 所示.随着任务信息不断变大,4 种算法下节点的感知移动距离都不断增加,这是因为较小质量的任务信息需要较小的接触时间,因此,公交载体内存在足够多的分发对象可供选择,故平均移动距离比较小.当任务信息质量逐步变大时,受到节点可用时间的限制,合适的候选节点集也不断减少,因此一些较远的节点被选中,则平均移动距离也逐渐增大.随着任务质量的进一步增大,可选的分发对象严重不足,既存在距离子区域中心点较近节点,又存在较远节点,此时的平均移动距离则趋向不变.针对本文两种算法来说,COV-DA 算法下的节点平均移动距离要小于 SPR-DA 算法,具体原因与图 6 中分析相同.

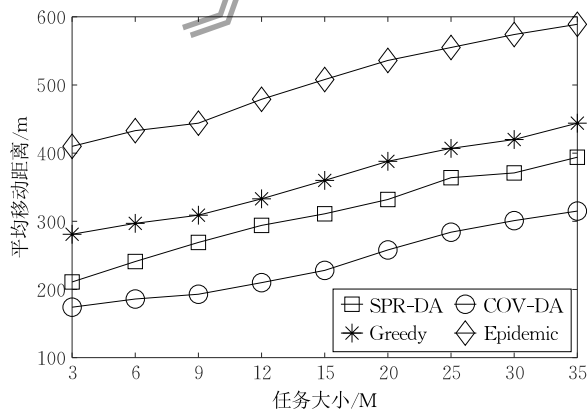


图 7 不同任务信息大小下的分发准确性比较

由于节点在完成感知任务时会对自身的移动轨迹造成影响,因此大多数群智感知应用中都以节点前往任务位置进行感知时的移动距离作为分配感知激励的重要因素.因此,当任务的感知激励预算比较充足时,则优先使用 SPR-DA 算法,获得更高的分

发准确性;若激励预算较少时,则使用 COV-DA 算法以求收集到更多的数据。

5.4.4 不同任务分发算法时间性能测试

实验在任务信息大小为 15 MB、乘客流量密度不断变化的条件下测试不同算法的任务平均分发时间。如图 8 所示,在任务分发方面,Epidemic 算法时间最短,Greedy 其次,这是由于本文所提两种算法对区域属性、节点位置、分发预算等因素都纳入了考虑,在任务扩散时对转发节点的选择要求将更高,故分发时间有一定延长。但结合任务分发准确性、平均移动距离等指标来看,本文所提方法都明显由于前者,能够进一步提高感知服务质量和任务执行效率。

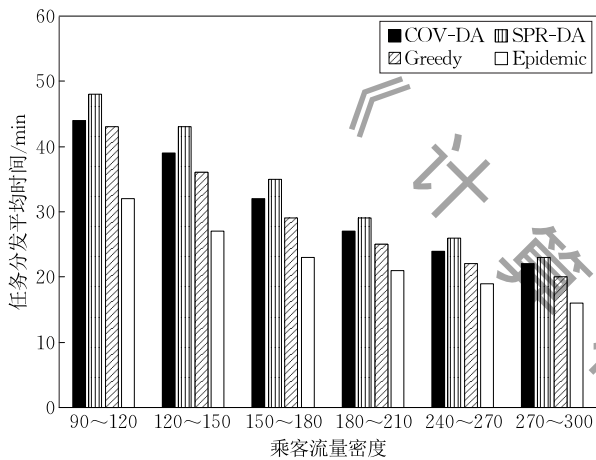


图 8 不同乘客流量密度下任务分发时间比较

6 结 论

本文提出了一种基于公交系统的差异化分发方法,在该方法中构建了基于公交系统的任务分发模型,并分析了预算约束下任务的差异化分发问题,并给出了两种差异化分发算法:基于扩散差异的分发算法(SPR-DA)和基于覆盖差异的扩散算法(COV-DA),SPR-DA 算法是在每一个封闭的公交载体内进行差异化的分发调度,而 COV-DA 则是在区域覆盖部分就完成差异化。SPR-DA 算法具有较高的分发准确性,使用 COV-DA 算法时则节点则需要较短的移动距离。当任务的感知激励预算比较充足时,则优先使用 SPR-DA 算法,获得更高的分发准确性;若激励预算较少时,则使用 COV-DA 算法以求收集到更多的数据。

文章重点是在引入公交系统的基础上,完成任务在感知区域内的差异化分发,并未考虑到实际接

受任务的感知节点数目是否满足要求,下一步工作中将结合基于节点任务选择的反馈情况来对任务分发模块进行完善,进一步提高服务质量。

参 考 文 献

- [1] Liu Yun-Hao. Crowdsourcing computing. Communications of the China Computer Federation, 2012, 8(10): 38-41(in Chinese)
(刘云浩. 群智感知计算. 中国计算机学会通讯, 2012, 8(10): 38-41)
- [2] Yu Zhi-Wen, Yu Zhi-Yong, Zhou Xing-She. Socially aware computing. Chinese Journal of Computers, 2012, 35(1): 16-26(in Chinese)
(於志文, 於志勇, 周兴社. 社会感知计算: 概念、问题及其研究进展. 计算机学报, 2012, 35(1): 16-26)
- [3] Ma H, Zhao D, Yuan P. Opportunities in mobile crowd sensing. IEEE Communications Magazine, 2014, 52(8): 29-35
- [4] Li Y, Qian M, Jin D, et al. Multiple mobile data offloading through disruption tolerant networks. IEEE Transactions on Mobile Computing, 2014, 13(7): 1579-1596
- [5] Ma Hua-Dong, Yuan Pei-Yan, Zhao Dong. Research progress on routing problem in mobile opportunistic networks. Journal of Software, 2016, 26(3): 600-616(in Chinese)
(马华东, 袁培燕, 赵东. 移动机会网络路由问题研究进展. 软件学报, 2015, 26(3): 600-616)
- [6] Han B, Hui P, Kumar V S A, et al. Mobile data offloading through opportunistic communications and social participation. IEEE Transactions on Mobile Computing, 2012, 11(5): 821-834
- [7] Zhang M, Yang P, Tian C, et al. Toward optimum crowdsensing coverage with guaranteed performance. IEEE Sensors Journal, 2016, 16(5): 1471-1480
- [8] Wang L, Zhang D, Yan Z, et al. effSense: A novel mobile crowd-sensing framework for energy-efficient and cost-effective data uploading. IEEE Transactions on Systems, Man, and Cybernetics; Systems, 2015, 45(12): 1549-1563
- [9] An J, Gui X, Wang Z, et al. A crowdsourcing assignment model based on mobile crowd sensing in the Internet of Things. IEEE Internet of Things Journal, 2015, 2(5): 358-369
- [10] Zhou P, Zheng Y, Li M. How long to wait?: Predicting bus arrival time with mobile phone based participatory sensing// Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services. New York, USA, 2012: 379-392
- [11] Liu Z, Jiang S, Zhou P, et al. A participatory urban traffic monitoring system: The power of bus riders. IEEE Transactions on Intelligent Transportation Systems, 2017, 99: 1-14

- [12] Guo B, Wang Z, Yu Z, et al. Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm. *ACM Computing Surveys*, 2015, 48(1): Article 7
- [13] Tuncay G S, Benincasa G, Helmy A. Autonomous and distributed recruitment and data collection framework for opportunistic sensing//*Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*. Istanbul, Turkey, 2012: 407-409
- [14] Gao W, Cao G, La P T, et al. On exploiting transient social contact patterns for data forwarding in delay-tolerant networks. *IEEE Transactions on Mobile Computing*, 2013, 12(1): 151-165
- [15] Tian D, Yang Y, Xia H, et al. An evaluation model for epidemic routing in VANETs//*Proceedings of the 12th IEEE International Conference on Ubiquitous Intelligence and Computing*. Beijing, China, 2015: 1661-1664
- [16] Guo B, Liu Y, Wu W, et al. ActiveCrowd: A framework for optimized multitask allocation in mobile crowdsensing systems. *IEEE Transactions on Human-Machine Systems*, 2016, 47(3): 392-403
- [17] Luo T, Tan H P, Xia L. Profit-maximizing incentive for participatory sensing//*Proceedings of the 33rd IEEE International Conference on Computer Communications*. Toronto, Canada, 2014: 127-135
- [18] Nan Wen-Qian, Guo Bin, Cheng Hui-Hui, et al. A cross-space, multi-interaction-based dynamic incentive mechanism for mobile crowd sensing. *Chinese Journal of Computers*, 2015, 38(12): 2412-2425(in Chinese)
(南文倩, 郭斌, 陈荟慧等. 基于跨空间多元交互的群智感知动态激励模型. *计算机学报*, 2015, 38(12): 2412-2425)
- [19] Karaliopoulos M, Telelis O, Koutsopoulos I. User recruitment for mobile crowdsensing over opportunistic networks//*Proceedings of the 34th IEEE International Conference on Computer Communications*. Hong Kong, China, 2015: 2254-2262
- [20] Wang L, Zhang D, Pathak A, et al. CCS-TA: Quality-guaranteed online task allocation in compressive crowdsensing//*Proceedings of the International Conference on Pervasive and Ubiquitous Computing*. Osaka, Japan, 2015: 683-694
- [21] Dang T, Feng W C, Bulusu N. Zoom: A multi-resolution tasking framework for crowdsourced geo-spatial sensing//*Proceedings of the 30th IEEE International Conference on Computer Communications*. Shanghai, China, 2011: 501-505
- [22] Mhirech A, Zahraouy H, Ismaili A A. Effect of damaged vehicle evacuation on traffic flow with open boundaries. *Physical Review E*, 2010, 81(1): 011132
- [23] Caliendo C, Guglielmo M L, Guida M. A crash-prediction model for road tunnels. *Accident Analysis & Prevention*, 2013, 55: 107-115
- [24] Deritei D, Lazar Z I, Papp I, et al. Community detection by graph voronoi diagrams. *New Journal of Physics*, 2014, 16(6): 063007



AN Jian, born in 1983, Ph. D., senior engineer. His research interests include crowd computing, Internet of Things.

PENG Zhen-Long, born in 1993, Ph. D. candidate. His research interests focus on crowd computing.

GUI Xiao-Lin, born in 1966, Ph. D., professor. His research interests include Internet of Things, cloud computing.

XIANG Le-Le, born in 1993, M. S. candidate. His research interests focus on crowd computing.

LIANG Dan-Wei, born in 1995, M. S. candidate. Her research interests focus on crowd computing.

Background

With the increasing number of intelligent mobile devices and the embedding of more sensors, Crowd sensing is becoming an effective means to solve the large-scale sensing tasks in urban areas. The task requesters can distribute multitudinous tasks through the corresponding crowd sensing platform to the mobile terminal nodes in the city to complete, without the need for a large number of fixed sensors for data collection. In order to reduce the high mobile data costs caused by the transmission of information directly through the 3G/4G network and lower the huge load pressure on the 3G/4G

network, many studies have proposed to use the opportunity network to indirectly complete the data transmission, and reduce the communication costs, so as to attract more mobile nodes to participate. However, with the growing of mobile terminal's sensing ability, the size and type of task information and sensed data are constantly changing, and the data transmission mode based on the short interaction time between nodes is difficult to guarantee the quality of data service. But most of the studies didn't analyze the issue in depth or propose corresponding solutions.

In view of the problem of crowdsourcing service quality in crowd sensing, in this paper, we focus on the two aspects of crowdsourcing task distribution and sensed data backhaul, and propose crowdsourcing service mechanism based on urban public transport system, which takes advantage of the urban public transportation (e. g. the trajectory can be predicted, the coverage area is extensive, the moving speed is fast, and the interaction time is guaranteed).

Aiming at the problem of crowdsourcing task distribution, firstly, we propose the differentiated distribution model of crowdsourcing task based on the public transportation system. Secondly, we take account into the problem of differentiated distribution under certain budget constraint and propose two differentiated distribution algorithms: the distribution algorithm based on coverage difference (COV-DA) and the spread

algorithm based on diffusion difference (SPR-DA). Finally, using the real data set, the distribution performance of the two algorithms is verified by the distribution accuracy and the average moving distance. The experiment shows that the SPR-DA algorithm has higher distribution accuracy, but in the COV-DA algorithm, the mobile node needs a smaller moving distance. Compared with classical algorithm (e. g. Greedy and Epidemic), COV-DA and SPR-DA have better performance in distributing accuracy and average mobile distance.

This work was partly supported by the NSFC (No. 61502380), the Science and Technology Program of Shenzhen (No. JCYJ20170816100939373), the Key Projects of Science and Technology Plan in Fujian Province (No. 2016H0029).

计算机学报