

# 基于数据价值的无人机数据收集方法\*

徐丹, 李伟, 王安文, 范浩楠, 龚晓庆, 陈晓江, 房鼎益



(西北大学 信息科学与技术学院, 陕西 西安 710100)

通讯作者: 陈晓江, 房鼎益, E-mail: xjchen@nwu.edu.cn, dyf@nwu.edu.cn

**摘要:** 数据收集是无线监测网络的关键环节. 利用无人机进行数据收集的本质是通过无人机的移动代替网络中的转发节点, 减少数据从源节点到基站的转发次数, 有效节约监测网络能量, 从而成为未来发展的趋势. 现有的研究关注如何利用无人机有限的能量获得更多的数据, 缺乏对获取数据的价值评估, 从而导致无人机数据收集能效比不高. 如何利用无人机最少的能量付出在监测区域获取最大的数据价值, 其难点在于数据价值是针对不同应用的主观评价, 而不同节点获取的数据价值如何比较, 目前还缺乏统一的标准. 可以发现, 数据相似节点的数据价值存在相似性. 在此基础上, 提出了一种数据收集方法 OnValueGet, 利用关键性代表节点的数据, 最大程度地近似代表整个监测区域的数据, 从而在能量约束下获得最大数据价值. 核心思想在于: 从分析感知数据的时空相似性入手, 确定数据价值较高的感知节点, 即数据关键节点. 在应用的误差范围内, 它们采集的数据可以近似表示全部网络感知节点采集的数据. 无人机以数据关键节点为数据采集的核心目标, 在能量有限的情况下, 根据遇到的障碍物和节点感知到数据的异常与否, 动态地规划数据收集路线, 从而使收集到的数据具有最大价值, 显著提升数据收集的能效比.

**关键词:** 无人机; 数据收集; 数据相似; 数据价值; 数据关键节点

**中图法分类号:** TP311

中文引用格式: 徐丹, 李伟, 王安文, 范浩楠, 龚晓庆, 陈晓江, 房鼎益. 基于数据价值的无人机数据收集方法. 软件学报, 2017, 28(12): 3206–3222. <http://www.jos.org.cn/1000-9825/5248.htm>

英文引用格式: Xu D, Li W, Wang AW, Fan HN, Gong XQ, Chen XJ, Fang DY. UAV data collection method based on data value. Ruan Jian Xue Bao/Journal of Software, 2017, 28(12): 3206–3222 (in Chinese). <http://www.jos.org.cn/1000-9825/5248.htm>

## UAV Data Collection Method Based on Data Value

XU Dan, LI Wei, WANG An-Wen, FAN Hao-Nan, GONG Xiao-Qing, CHEN Xiao-Jiang, FANG Ding-Yi

(School of Information Science and Technology, Northwest University, Xi'an 710100, China)

**Abstract:** Data collection is the most crucial problem of wireless monitoring networks. The UAV based data collection methods have become the trend, as they can reduce the relaying times of data from the source to the sink, and improve the efficiency of network energy by replacing traditional self-organized transmission nodes with UAV. Current UAV based data collection schemes, however, focus on how to maximize the quantity of data using limited energy without consideration of data value, and hence perform poorly in energy efficiency of UAV. The challenge for achieving maximum data value with minimum UAV energy consumption is to measure the value of the data, as the value of data is subjective evaluation of applications, and there is no uniform measurement to compare the value of data that collected by different nodes. This paper introduces the first data value based data collection method OnValueGet that collect the most valuable data under the energy constraint. The intuition underlying the design is that nodes with similar data experience similar data value. The paper

\* 基金项目: 国家自然科学基金(61272461, 61672428, 61572402, 61170218, 61672427); 国家科技支撑计划(2013BAK01B02); 西北大学科学研究基金(15NW32)

Foundation item: National Natural Science Foundation of China (61272461, 61672428, 61572402, 61170218, 61672427); Project National Key Technology R&D Program (2013BAK01B02); Scientific Research Foundation of North West University (15NW32)

收稿时间: 2016-07-27; 修改时间: 2016-09-14, 2016-11-04; 采用时间: 2016-11-26; jos 在线出版时间: 2017-03-24

CNKI 网络优先出版: 2017-03-24 17:09:28, <http://kns.cnki.net/kcms/detail/11.2560.TP.20170324.1709.009.html>

defines and selects the most valuable nodes (called data-critical nodes) by analyzing and comparing the temporal and spatial similarity of data. The data sensed by data-critical nodes can approximately represent all nodes' sensing data within a certain error. Aiming to collect the data of these data-critical nodes, the paper then adapts greedy algorithm to programing the route of UAV with the limited energy, and significantly improves the energy efficiency.

**Key words:** UAV; data collection; data similarity; data value; data-critical node

无线监测网络部署在灾害监测与预警、生态监测、建筑结构健康监测、文化遗产保护等野外环境大规模监测场景中,持续获取各种应用所需要的数据,成为社会生活不可或缺的重要部分.野外环境复杂的通信条件和一定规模的监测覆盖要求,使得保证长期且有效的数据收集成为一个监测网络成败的关键.

利用无线传感网络,部署大量数据转发节点,进行自组织传输网络构建,形成数据传输通道,成为较为常见的数据收集方法<sup>[1]</sup>,其核心是:数据感知节点将采集到的数据经过一跳或者多跳中转节点传递给 Sink 节点,Sink 节点通过 3G 或者卫星接入,从而将数据返回监测中心.其关键挑战在于:当节点在野外部署时间过长时,就会出现能量空洞问题<sup>[2]</sup>,Sink 节点周围转发节点会过早地进入死亡状态,导致数据采集失败.更为严重的是:为了保证可靠传输,需要在大规模区域保证一定的覆盖,从而导致需要大量的数据转发节点.传感网节点有效通信距离在 50m 左右,成本 150 元,如果覆盖 20km<sup>2</sup> 范围的野外遗址保护区,且要保证覆盖率在 90%以上,仅通信节点就需要约 5 000 个,费用在 76 万元左右<sup>[3]</sup>,如此巨大的部署成本和代价,难以在大规模监测场景中应用.

对此,利用移动 Sink 节点进行数据收集,成为了一个相对有效的解决方案.常见的思路是:将 Sink 节点放置在可移动小车上,通过车的移动收集数据感知节点所采集的数据.

这种方法可以很好地解决能量空洞问题<sup>[4]</sup>,但在诸多真实场景下难以实用.车的移动依赖完善的道路,大部分野外监测在缺乏道路的无人区,移动小车缺乏移动的条件,我们称之为 Sink 移动范围受限.

最近兴起的无人机(UAV)技术成为 Sink 节点移动数据收集的有效解决方案<sup>[5-7]</sup>.无人机搭载 Sink 作为移动数据收集节点,根据路径规划进行移动数据收集,我们将这种方法称为基于无人机的数据收集方法.目前,已经开始在农作物监测、灾后救援等方面展开初步应用<sup>[8-10]</sup>.其本质是用无人机的飞行代替了中间数据转发节点,避免了过大的部署代价和成本以及能量空洞问题.同时,这种方法将数据收集和数据采集分离在不同的空间维度展开,也避免了“Sink 移动受限问题”.因此,利用基于无人机的数据收集方法成为大规模区域监测数据收集的有效手段.由于无人机主要收集的是感知节点已经采集并存储了一段时间的数据,这些数据的实时性无法得到保证.因此,无人机数据收集系统无法应用在实时性要求较高的应用中,如火灾监控、防盗监控等.

现有的无人机由于能量有限导致飞行时间和里程受限,如何在有限的能量下最大程度地收集数据,成为目前研究的热点.一个解决思路是在监测区域内部署多个无人机进行协同工作<sup>[6,11]</sup>,但是成本高,协同工作更带来了额外的复杂度.另一个思路是建立路径优化模型,使无人机沿最短路径访问每一个需要收集数据的采集点,在有限的能量情况下,尽可能多地收集数据.此类研究主要聚焦于如何在有限时间让数据收集数量最大,而忽略了采集到的数据质量优化问题.采集到数据冗余性和相似度高使得数据价值低,难以对整个部署区域数据进行有效表征,数据收集失去了既定的意义.

如何利用无人机最少的能量付出在监测区域获取最大的数据价值,其难点在于数据价值是针对不同应用的主观评价,而位于不同地理位置的节点获取的数据价值如何比较,目前缺乏统一的标准.我们发现,数据相似节点的数据价值存在的相似性.在此基础上,我们提出了一种数据收集方法——OnValueGet,利用关键性代表节点的数据,最大程度地近似代表整个监测区域的数据,从而在能量约束下获得最大数据价值.其核心思想在于:从分析感知数据的时空相似性入手,确定数据价值较高的感知节点,本文称为数据关键节点,在应用的误差范围内,它们采集的数据可以近似表示全部网络感知节点采集的数据.无人机以数据关键节点为数据采集的核心目标,在能量有限的情况下,根据遇到的障碍物和节点感知到数据的异常与否,动态地规划数据收集路线,从而使收集到的数据具有最大价值,显著提升数据收集的能效比.

## 1 相关工作

以文献[12-18]为代表的静态数据收集方法中,主要使用数据中转节点建立自组织网络,把数据传输到基站.核心在于通过最小化网络通信开销以及良好的能量负载平衡方法,最大化网络生命周期,保证数据可靠收集.

动态数据收集方法可以有效地避免能量空洞问题,但主要面临数据延迟的挑战.当感知节点移动<sup>[19]</sup>时,研究通过选择高概率与 Sink 通信的下一跳节点,保证低功耗和延迟.当 Sink 节点移动<sup>[4,20]</sup>时,研究者侧重于如何离线规划好移动 Sink 节点数据收集路径保证低延迟,如 Shah<sup>[21]</sup>,或者在线如何动态地规划 Sink 节点的数据收集路径保证低延迟.

基于无人机的数据收集方法<sup>[5-7]</sup>本质上可以视作基于移动 Sink 节点的数据收集方法的特定实例,因此,其可以避免能量空洞问题.空中飞行不用依赖道路,因此也避免了 Sink 移动受限问题.无人机数据收集有两种典型的工作方式:一种是在无人机上搭载数据感知节点,让无人机在监测区域内航行并采集数据;另一种是在无人机上搭载 Sink 节点,飞行收集地面感知节点获取的数据.两种不同的工作方式对应两种不同的数据收集网络:FANET(flying ad-hoc network)<sup>[7]</sup>和 FGNET(flying and ground network)<sup>[5,6]</sup>.前者不依赖于地面节点的部署,具有较强的灵活性,但是系统实现难度较大,因此难以实用;而后者更具有实用性,成为目前数据收集方法的主流,研究的焦点在于能耗、延迟和传输质量.

由于无人机的能量受限,现有的无人机大都由电池供电,航行时间有限,难以有效地在大规模区域进行数据收集.如何在线地规划无人机数据收集的路线,使得无人机在能量约束下获得最大数据量,成为研究的热点.

为了解决无人机能量受限,研究者们常常引入多无人机进行相互协同的完成任务<sup>[22]</sup>.例如,Ahmed 等人<sup>[5,6]</sup>提出将簇头节点与无人机通信时间的分配问题抽象成一个势博弈(potential game)问题,通过簇头节点之间的博弈,既保证了节点之间的公平性和网络能量的高效,又保证了整个网络在数据收集的过程中能效最大化.文献[23]提出了一种多无人机相互协同采集图像信息进行目标定位的方法.文献[24]利用多无人机相互协同工作完成了对未知环境探测的任务.此类方法可以解决由于无人机能量受限而引起的数据缺失问题,但成本开销巨大,而且大都假设周围环境对无人机的影响很小甚至没有<sup>[5,6,11,23,24]</sup>,真实场景不存在.

针对基于无人机的数据收集方法所带来的延迟问题,在文献[11]中,由 Gu 等人提出了一种集数据采集和收集为一体的数据收集方法.他们分别研究了在 Karam<sup>[25]</sup>模型中数据收集延迟和数据信息延迟问题,最终提出了一种基于层次的多飞行器相互协作的数据收集方法.文献[26]提出一种 DC-Collection 算法,不仅能够均衡各节点的能量消耗,延长网络生命周期,而且能够缩短移动数据收集器收集数据行走的路径长度,从而缩短数据收集延迟.在文献[27]中,作者提出了一种基于相对距离感知的动态数据传输方法 RDAD(relative distance-aware data delivery scheme),有效地降低了数据传输延迟.文献[28]提出了一种旅行动态决策算法 SOFCOM,使用详尽的搜索技术来计算全局最优决策序列,从而达到减少延迟的目的.

目前,由于广泛存在的各种无线信号,使得利用卫星或者 4G 网络进行通信的无人机信号极易受到干扰,导致传输质量下降.针对这个问题,国内外的研究者从视频编码、空间频谱检测和信道状态的估计等多方面展开了深入的研究.文献[29]通过选择适合当前网络带宽的视频编码比特率,保证实时连续的视频传输.文献[30]提出了一种分布式的移动算法,使无人机群在空间和频率未知的情况下收集信道感知的指标,如感知相关距离等,以最大限度地提高检测精度.文献[31]在给定的最小信噪比内,控制器控制无人机在可能的最大传输距离内盘旋,通过观察 SNR 的变化情况,让无人机飞到最适合的 SNR 地方进行数据的收集,从而保证数据传输质量.

以上研究极大地提升了无人机数据收集的效率和性能,但是针对收集到的数据的价值特性考虑不足.在一个既定的监测区域中,现有研究只考虑到数据沿何种收集线路获取较好的数量、质量或者延迟,但是没有考虑到收集的数据价值特性.换言之,收集到的数据冗余度过大,或者相似度过高,对于应用而言,并不能真正解决问题.例如,在一个草原监测中,一部分区域温湿度变化很大,一部分区域温湿度变化不大,如果根据延迟方法和能量方法,取回的数据可能恰恰是温湿度变化不大的部分,那么对于应用而言,则丧失了有效的数据价值.

## 2 基于数据关键节点的数据收集方法 OnValueGet

本节具体介绍基于数据关键节点的数据收集方法 OnValueGet 的基本思想以及具体实现过程。

### 2.1 OnValueGet 框架

OnValueGet 的核心在于进行数据关键节点选择,并在此基础上动态规划数据收集路径.OnValueGet 整体框架如图 1 所示。

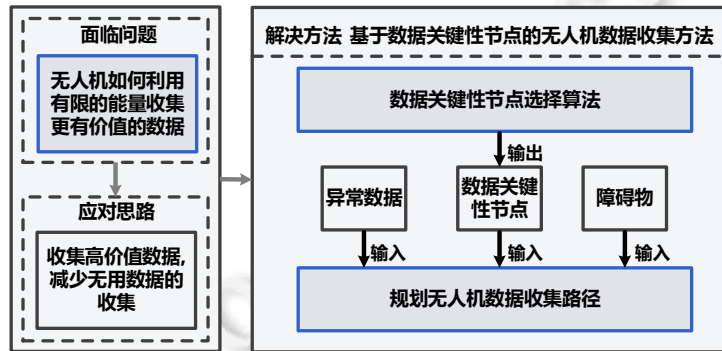


Fig.1 Framework of OnValueGet

图 1 OnValueGet 框架

针对无人机如何利用有限的能量收集更有价值的信息问题,OnValueGet 从分析感知数据的时空相关性入手,提出数据关键节点的概念,并设计了数据关键节点选择算法(data-critical nodes selection algorithm,简称 DcNS).OnValueGet 将选出的数据关键节点作为 UAV 数据收集的主要目标,并根据遇到的障碍物和节点感知到数据的异常与否情况,利用 Greedy 算法动态地规划 UAV 的航行路线,使得 UAV 在网络区域内收集少量可以近似代表全网监测数据的数据关键节点数据.在保证采集数据价值的情况下,减少了 UAV 数据收集过程中的任务量,实现在能量约束的情况下最大限度地提升所收集数据的价值。

### 2.2 数据关键节点相关概念

**定义 1(感知数据序列).** 对于任意数据感知节点  $n_i$ , 将其在  $[t, t+\alpha]$  时间段内采集的带有时序意义的数字序列  $\{d_{i,t}, d_{i,t+1}, \dots, d_{i,t+\alpha}\}$  称为  $n_i$  在  $[t, t+\alpha]$  上的感知数据序列. 并将其记为  $I_i^{t,t+\alpha}$ , 其中,  $d_{i,t+j}$  表示节点  $n_i$  在时刻  $t+j, 0 \leq j \leq \alpha$  时采集的数据。

**定义 2(数据相似节点).** 在时间段  $[t, t+\alpha]$  上, 数据感知节点  $n_i$  和  $n_j$  之间是否相似, 是根据在时间段  $[t, t+\alpha]$  上, 节点  $n_i$  和  $n_j$  所采集的数据序列  $I_i^{t,t+\alpha}$  和  $I_j^{t,t+\alpha}$  之间的相似度来判断. 若数据序列  $I_i^{t,t+\alpha}$  和  $I_j^{t,t+\alpha}$  之间的相似度在给定的阈值范围内, 则称节点  $n_i$  和  $n_j$  相似, 记为  $Sim(n_i, n_j)$ 。

**定义 3(数据关键节点).** 在监测网络的感知节点  $N$  中选择节点集合  $N_x$ , 若  $N_x$  满足以下 3 个条件, 则称  $N_x$  中的元素为数据关键节点. 第一, 对于网络中节点  $\forall n_a \in N - N_x, \exists n_b \in N_x$  与节点  $n_a$  相似; 第二, 当  $N_x$  中的所有节点在处于一条直线时, 保证相距最远的两个节点之间的距离(关键长度)最小; 当  $N_x$  中的所有节点没有处于一条直线时, 当保证由  $N_x$  中所有节点所对应的最小外接矩形的面积(关键面积)最小; 第三, 对于任何满足条件 1 和条件 2 的节点集合  $N_x = \{N_{x1}, N_{x2}, \dots, N_{xm}\}$ , 从中选择出元素数目最少的集合  $\min(|N_{x1}|, |N_{x2}|, \dots, |N_{xm}|)$  作为数据关键节点集合  $N_x$ . 关键长度和关键面积的物理意义在于使得采集数据相似的节点在地理位置上尽可能地聚集在一起, 从而有效地减少 UAV 的飞行距离。

**定义 4(节点覆盖范围).** 假设网络中数据关键节点的集合是  $N_x, N_x = \{cn_1, cn_2, \dots, cn_m\}$ , 对于每个数据关键节点

$cn_i$ , 设与其相似的节点所组成的集合为  $S_i$  (节点自身与自身之间也存在相似关系), 设  $S_i^{sub}$  为  $S_i$  的子集, 即  $S_i^{sub} \subseteq S_i$ . 如果  $\forall cn_j \in N_x, \forall cn_k \in N_x, S_j^{sub} \cap S_k^{sub} = \emptyset$ , 并且  $S_1^{sub} \cup S_2^{sub} \cup \dots \cup S_m^{sub} = N$ , 其中,  $N$  表示所有数据感知节点的集合. 此时, 将  $S_i^{sub}$  称为数据关键节点  $cn_i$  的节点覆盖范围.

**定义 5 (数据价值).** 数据价值是每个数据感知节点所具有的属性, 具体表示由其采集的数据在具体应用中的使用价值决定. 本文中假设: 在正常状况下, 所有感知节点所采集的数据具有相同的使用价值, 为了简便起见将其设为 1. 因此, 对于数据关键节点  $cn_k$ , 如果所对应的节点覆盖范围为  $\Gamma$ , 则其数据价值为  $|\Gamma|$ .

**定义 6 (应用误差  $\epsilon$ ).**  $\epsilon$  指应用可以接受的数据误差. 假设应用误差  $\epsilon = a$ , 当最终收集到的数据关键节点  $cn_i$  在时间段  $[t, t+a]$  内采集的数据序列  $I_i^{t,t+a}$  与其覆盖范围  $S_i^{sub}$  内所有感知节点采集到的数据序列  $I = \{I_j^{t,t+a}, I_k^{t,t+a}, \dots, I_n^{t,t+a}\}$  相差在  $[-a, a]$  之间时, 应用都是可以接受的.  $\epsilon$  是根据应用要求设定.

### 2.3 数据相似节点判断方法

感知节点采集数据的价值是 OnValueGet 方法实现的重要依据. 我们利用数据相似度作为数据价值的评估依据, 即数据相似、数据价值相似. 由数据价值的定义可以得出, 数据关键节点的价值等于其覆盖范围内所有节点的价值之和.

因此, OnValueGet 核心在于找到可以近似表示整个监测网络中感知节点采集数据的数据关键节点. 根据数据关键节点的定义可知: 要判断某一节点是否为数据关键节点, 必须求出该节点与其他节点之间的相似度. 计算节点之间的相似度, 在本质上可以归结为计算两个节点采集的数据相似度. 数据相似度的计算可以进一步分解为两个数据序列之间距离的求解 (两个数据序列之间的距离越大, 相似性越小) 和相似度阈值设定问题.

本文先对数据序列之间的距离问题进行讨论, 公式 (2.1) 是求解两个节点  $n_i$  和  $n_j$  的数据序列  $I_{i,t+\lambda \times \theta}^i$  和  $I_{j,t+\lambda \times \theta}^j$  之间距离最直接的方法:

$$dis_{i,t+\lambda \times \theta}^{i,j} = |d_{i,t} - d_{j,t}| + |d_{i,t+\theta} - d_{j,t+\theta}| + \dots + |d_{i,t+\lambda \times \theta} - d_{j,t+\lambda \times \theta}| \quad (2.1)$$

但是公式 (2.1) 方法的缺陷在于要求节点  $n_i$  和  $n_j$  必须保证数据采集严格同步, 或者同时采集, 或是同时不采集. 这种情况在真实网络环境中很难保证. 原因在于无线信号受到影响的不确定性和节点真实采集周期的不固定性<sup>[32,33]</sup>. 图 2 显示了具有相同采集周期的不同节点, 每天采集数据的数量统计情况. 从图 2 中可以明显地看出: 不同节点在同一段时间内采集到的数据数量是不同的; 而即使是同一个节点, 在不同的时间段内采集数据数量也会有所不同.

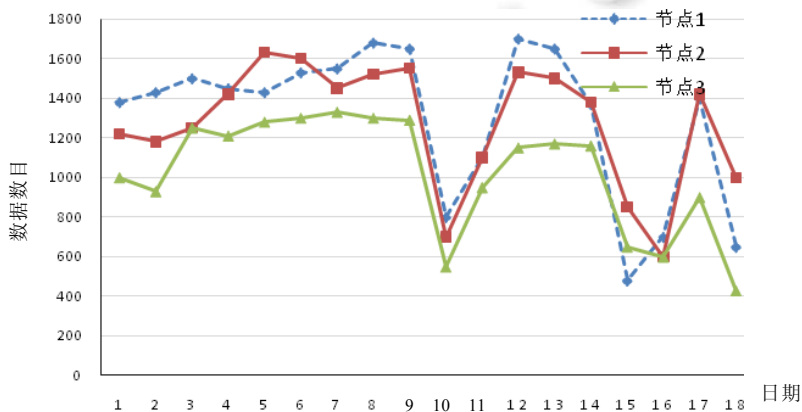


Fig.2 Number of sensing data from different nodes in a same period

图 2 不同节点同一时间段内感知数据数量统计图

针对如何有效地计算数据序列  $I_{i,t+\lambda \times \theta}^i$  和  $I_{j,t+\lambda \times \theta}^j$  之间距离的问题, 通过分析发现, 类似于数据序列长度不同

而导致难以有效计算两者之间距离的问题也出现在语音识别领域.对于同一个单词,不同的人发音的长短以及音调的高低往往是不同的,即使是同一个人,对同一个单词的两次发音,其音调高低以及发音长短也往往是不同的.认识到语音识别问题和数据序列距离计算问题的相同之处,我们借鉴语音识别领域应用非常广泛的 DTW (dynamic time warping)算法<sup>[34]</sup>来解决两个长度不一样数据序列之间的距离计算问题.

利用 DTW 设计的数据序列之间距离计算方法描述如下.

假设有两个数据序列  $A=\{a(1),a(2),\dots,a(k)\}$ ,  $B=\{b(1),b(2),\dots,b(m)\}$ ,序列  $A$  和序列  $B$  的维度分别是  $k$  和  $m$ ,为了不失一般性,令  $k \neq m$ ,对于它们每一维的元素可以有两种形态:一种是  $a(i)$  和  $b(i)$  是简单的数字,也可以将其看成是维度为 1 的序列;另一种就是  $a(i)$  和  $b(i)$  都是维度不为 1 的序列,并且  $|a(i)|=|b(i)|$ ,即,必须保证  $a(i)$  和  $b(i)$  的维度是相同的.运用 DTW 算法计算这两个不同维度序列之间的距离可以分两步来进行.

第 1 步:计算序列  $A$  每一个维度与序列  $B$  每一个维度之间的距离,最终可以形成一个  $k \times m$  维的矩阵,记为  $C_{A,B}$ ,  $C_{A,B}$  中第  $i$  行第  $j$  列的元素表示序列  $A$  中第  $i$  维元素与  $B$  中第  $j$  维元素之间的欧氏距离,即:  $C_{A,B}(i,j)=|a(i)b(j)|$ ;

第 2 步:设置一个与  $C_{A,B}$  同维度的空矩阵  $R_{A,B}$ ,然后,按照公式(2.2)扩展  $R_{A,B}$  矩阵,最终,矩阵  $R_{A,B}$  右下角的值即为序列  $A,B$  之间的距离,即:  $dis^{A,B}=R_{A,B}(k,m)$ :

$$R_{A,B}(i,j) = \min \begin{cases} R_{A,B}(i-1,j) + C_{A,B}(i,j) \\ R_{A,B}(i-1,j-1) + C_{A,B}(i,j) \\ R_{A,B}(i,j-1) + C_{A,B}(i,j) \end{cases} \quad (2.2)$$

从上文的描述可知,OnValueGet 利用 DTW 算法解决了数据序列之间的距离计算问题.接下来讨论相似度阈值的设定问题,即:当数据序列之间的距离在多少范围内,OnValueGet 判断两个节点是相似的.相似度阈值的设定直接影响着节点相似与否的判断,从而影响数据关键节点的选择以及 UAV 节点最终采集到的数据价值.在本文中,我们通过以下的证明分析给出适当的相似度阈值.

假设在时间段  $[t,t+\alpha]$  内,数据感知节点总共有  $H$  个采集周期(每个采集周期内采集一个数据).设  $\beta$  表示在时间段  $[t,t+\alpha]$  内的每个采集周期内,节点  $n_k$  和节点  $n_j$  采集到的数据之差的绝对值所构成的集合:

$$\beta = \{\beta_1, \beta_2, \dots, \beta_H\} \quad (2.3)$$

$$\beta_i = |d_{j,\Delta t_i} - d_{k,\Delta t_i}| \quad (2.4)$$

$$dis_{t,t+\alpha}^{k,j} = \sum_{\beta_i \in \beta} \beta_i \leq \varepsilon \times \max(|l_{t,t+\alpha}^k|, |l_{t,t+\alpha}^j|) = \varepsilon \times H \quad (2.5)$$

假设  $\beta$  中小于  $\varepsilon$  的元素集合为  $\beta_z$ , 大于等于  $\varepsilon$  的元素集合为  $\beta_Q$ ,  $|\beta_z|/|\beta_Q| = \mu/\nu$ .

**定理 1.** 当  $\mu/\nu$  存在,且  $\mu/\nu \rightarrow \max(|l_{t,t+\alpha}^k|, |l_{t,t+\alpha}^j|)$ , 或  $\mu/\nu$  不存在时,若数据序列  $l_{t,t+\alpha}^k$  和  $l_{t,t+\alpha}^j$  之间的距离小于等于  $\varepsilon \times \max(|l_{t,t+\alpha}^k|, |l_{t,t+\alpha}^j|)$ , 即  $dis_{t,t+\alpha}^{k,j} \leq \varepsilon \times \max(|l_{t,t+\alpha}^k|, |l_{t,t+\alpha}^j|)$ , 则节点  $n_k$  和节点  $n_j$  在  $[t,t+\alpha]$  上是相似的;当  $\mu/\nu$  存在,且  $\mu/\nu \ll \max(|l_{t,t+\alpha}^k|, |l_{t,t+\alpha}^j|)$  时,若  $dis_{t,t+\alpha}^{k,j} \leq \nu\varepsilon \times \max(|l_{t,t+\alpha}^k|, |l_{t,t+\alpha}^j|)/(\mu+\nu)$ , 则称节点  $n_k$  和节点  $n_j$  相似,记为  $Sim(n_k, n_j)$ .

证明:当  $|\beta_Q|=0$  时,则  $\beta$  中的所有元素都小于  $\varepsilon$ .在物理意义上表示在时间段  $[t,t+\alpha]$  内的任意采集周期  $\Delta t_i$  内,节点  $n_k$  和节点  $n_j$  采集的数据  $d_{k,\Delta t_i}$  和  $d_{j,\Delta t_i}$ , 它们之间的相差都在应用误差范围内,即  $|d_{j,\Delta t_i} - d_{k,\Delta t_i}| < \varepsilon$ , 因此在任意时刻,节点  $n_k$  采集的数据都与节点  $n_j$  采集的数据相似,所以  $Sim(n_k, n_j)$ .

当  $|\beta_z| \gg |\beta_Q|$  时,则  $\beta$  中的几乎所有元素都小于  $\varepsilon$ .在物理意义上表示在时间段  $[t,t+\alpha]$  中存在  $s$  个采集周期  $\{\Delta t_{r_1}, \Delta t_{r_2}, \dots, \Delta t_{r_s}\}$ , 节点  $n_k$  和节点  $n_j$  在这些采集周期内采集的数据  $\{d_{k,\Delta t_{r_1}}, d_{k,\Delta t_{r_2}}, \dots, d_{k,\Delta t_{r_s}}\}$  和  $\{d_{j,\Delta t_{r_1}}, d_{j,\Delta t_{r_2}}, \dots, d_{j,\Delta t_{r_s}}\}$ , 它们之间相差都大于给定的应用误差,即  $|d_{k,\Delta t_{r_x}} - d_{j,\Delta t_{r_x}}| > \varepsilon, 1 \leq x \leq s$ , 但由于  $s \rightarrow 0$ , 所以在这种情况下,我们也认为节点  $n_k$  采集的数据与节点  $n_j$  采集的数据相似,所以  $Sim(n_k, n_j)$ .

当  $|\beta_z| \ll |\beta_Q|$  时,则  $\beta$  中的几乎所有元素都大于等于  $\varepsilon$ .假设  $\beta$  中所有小于  $\varepsilon$  的元素之和为  $|\beta_{small}|$ , 则有:

$$\lim_{|\beta_Q| \rightarrow H} \frac{\varepsilon \times H - |\beta_{small}|}{|\beta_Q|} - \varepsilon = 0 \quad (2.6)$$

公式(2.6)的物理意义在于,此时虽然  $\beta$  中的几乎所有元素都大于等于  $\varepsilon$ , 但是  $\beta$  中元素与  $\varepsilon$  相差趋近于 0, 即  $\beta - \varepsilon \rightarrow 0$ ,

$\beta_i \in \beta$ . 在这种情况下,我们也认为节点  $n_k$  采集的数据与节点  $n_j$  采集的数据相似,所以  $Sim(n_k, n_j)$ .

当  $|\beta_Q|$  与  $|\beta_Z|$  相当时,可分两种情况分别讨论.

- 若  $\beta_Z$  与  $\beta_Q$  中的元素与  $\varepsilon$  之差都趋近于 0,即两个数据序列之间的相似性稳定,在这种情况下,我们认为  $n_k$  采集的数据与  $n_j$  采集的数据相似,所以  $Sim(n_k, n_j)$ ;
- 若  $\beta_Z$  与  $\beta_Q$  中的元素与  $\varepsilon$  相差极大,即  $\sum_{\beta_i \in \beta_Z} \beta_i \approx 0$ , 那么  $\sum_{\beta_j \in \beta_Q} \beta_j \approx H\varepsilon$ ,  $E(\beta_j) \approx \varepsilon(\mu + \nu)/\nu$ . 此时,虽然两个数据序列的整体误差均值仍为  $\varepsilon$ ,但是  $\beta_Q$  集合中几乎  $\nu H/\mu + \nu$  个数据的误差均值为远大于应用误差  $\varepsilon$ . 我们认

为两个数据序列之间的相似性不稳定,此时应将相似度阈值设置为  $\nu\varepsilon \times \max(|l_{t,t+\alpha}^k|, |l_{t,t+\alpha}^j|)/(\mu + \nu)$ , 即:

当  $dis_{t,t+\alpha}^{k,j} \leq \nu\varepsilon \times \max(|l_{t,t+\alpha}^k|, |l_{t,t+\alpha}^j|)/(\mu + \nu)$  时,我们认为  $Sim(n_k, n_j)$ .

因此,如果在时间段  $[t, t+\alpha]$  内,节点  $n_k$  和节点  $n_j$  的采集周期是完全同步的,而且两者都没有丢失数据,并且数据序列  $l_{t,t+\alpha}^k$  和  $l_{t,t+\alpha}^j$  之间的距离小于  $\varepsilon \times H$ , 则节点  $n_k$  和节点  $n_j$  是相似的,即  $Sim(n_k, n_j)$ .

如果在时间段  $[t, t+\alpha]$  内,节点  $n_k$  和节点  $n_j$  的采集周期没有完全同步,或者出现了丢失数据的情况,则利用 DTW 算法将维度较短的向量进行了拉伸,然后与维度较长的向量进行距离求解,得出  $l_{t,t+\alpha}^k$  和  $l_{t,t+\alpha}^j$  之间的距离. 因此,同理可以得到:如果数据序列  $l_{t,t+\alpha}^k$  和  $l_{t,t+\alpha}^j$  之间的距离小于  $\varepsilon \times \max(|l_{t,t+\alpha}^k|, |l_{t,t+\alpha}^j|)$ , 则节点  $n_k$  和节点  $n_j$  是相似的,记作  $Sim(n_k, n_j)$ . 证毕.  $\square$

## 2.4 数据关键节点选择方法 DeNS

第 2.3 节讨论了判断节点相似的方法与阈值的设定,接下来我们要在此基础上找出的数据采集的核心目标点:数据关键节点.其难点在于如何从给定的网络节点中,快速准确地找出符合数据关键节点要求的节点集合.我们通过对问题的分析,提出了基于贪心策略的数据关键节点选择方法.

根据数据关键节点的定义,数据关键节点的选择问题可描述为:假设整个网络是一个无向图  $G=(V, E)$ , 网络中的感知节点是图  $G$  中的顶点集合  $V$ . 如果通过对感知数据分析得知,在应用的误差要求下,节点  $u$  和  $v$  相似,则两个顶点之间存在边  $e=(u, v), e \in E$ ; 反之则不存在边.我们将图  $G$  称为全网节点的相似关系图.如何从网络中所有的数据感知节点中选择出数据关键节点,可以理解为在  $V$  中找到一个节点集合  $N_x$ , 使得网络中剩余的任意一个顶点  $u \in V - N_x$  至少有一个邻居顶点  $v \in N_x$ , 同时保证  $N_x$  中的顶点数目最少,并且  $N_x$  中的顶点所对应的关键长度或者关键面积最小.

**定理 2.** 网络中数据关键节点的选择问题是 NP-hard.

证明:由于图论中最小支配集问题是典型的 NP-hard<sup>[35]</sup>问题,因此只要能将最小支配集问题等价转换成数据关键节点选择问题的一个特例,我们就可以证明数据关键节点选择问题是 NP-hard 问题.下面我们通过一个多项式时间的规约来证明.

首先,给定无向图  $G=(V, E)$ , 将图  $G$  作为数据关键节点选择问题的一个实例.其中,  $V$  表示网络中所有的感知节点,  $E$  表示任意两个节点之间的边.若两节点之间有边,则表示两节点相似.然后,构造一个拓扑结构与  $G$  相同的无向图  $G'=(V', E')$ , 并将其作为最小支配集问题的一个实例.

我们从  $V$  中找出元素个数最少的节点集合  $N_x$ , 使得对于  $V - N_x$  中任意一个点  $u$ , 集合  $N_x$  中存在一点  $v$ , 使得  $Sim(n_k, n_j)$ . 也就是说,对于  $V$  中的任意一个点  $u$ , 要么在集合  $N_x$  中, 要么与集合  $N_x$  中的点  $v$  相似.当不要求  $N_x$  中的节点对应的关键长度或关键面积最小时,从  $V$  中要找的节点集合  $N_x$  与从  $V'$  中要找出所有支配集中顶点个数最少的支配集  $N'_x$  是相同的.因此,数据关键节点选择问题是 NP-hard 问题.证毕.  $\square$

OnValueGet 对于数据关键节点的选择,可以分为 3 个步骤.

第 1 步.根据节点之间的相似性构造全网节点的相似性关系图;

第 2 步.找出网络中所有的最小支配集;

第 3 步.在这些最小支配集中找出关键面积最小的作为数据关键节点.

从上文中我们可以看出:数据关键节点求解问题最核心的部分是从一个无向图中找出其所有的最小支配集,即 LMDS 问题.而到目前为止,对于 LMDS 问题最优的精确求解算法,其时间复杂度为  $O(2^{0.78n})$ <sup>[36]</sup>.为了 UAV 能够高效地完成数据关键节点的数据收集工作,必须寻求其他方法来选择数据关键节点.针对上述问题,本文在文献[37]所给出的解法基础上提出了数据关键节点的选择算法 DcNS,见算法 1.

**算法 1.** 数据关键节点选择算法 DcNS.

Input:

- $D_{pre}$ :网络中感知节点采集历史数据;
- $P_{n^*}$ :网络中所有节点的位置;
- $\epsilon$ :应用误差;

Output:数据关键节点集合.

```

1:  $dtw(D_{pre}, \epsilon) \rightarrow Sim_{graph}$ 
2:  $cnodeset = \{\}$ 
3: While  $Sim_{graph}.|V| > 0$ 
4:   for all  $v \in Sim_{graph}.V$  do
5:      $Sim_{graph}[v].worth = \frac{Sim_{graph}[v].|neighbor|}{dis(Sim_{graph}[v], cnodeset)}$ 
6:   end for
7:    $cn = findmaxworthnode()$ 
8:    $cs.cnode = cn$ 
9:    $cs.coverage = Sim_{graph}[cn].neighbor$ 
10:   $cnodeset.push\_back(cs)$ 
11:   $Sim_{graph}.V -= Sim_{graph}[cn]$ 
12:   $Sim_{graph}.V -= Sim_{graph}[cn].neighbor$ 
13: end while
14: Return  $cnodeset$ 

```

DcNS 算法的主要输入有:网络中所有感知节点所采集的历史数据集  $D_{pre}$ 、网络中所有感知节点的地理位置  $P_{n^*}$  以及应用的误差要求  $\epsilon$ .算法的输出是网络中数据关键节点的集合以及每个数据关键节点所对应的节点覆盖范围.

在 DcNS 算法的第 1 行中,以历史数据  $D_{pre}$  为输入,利用 DTW 算法计算节点采集数据序列之间的距离,然后根据应用的误差要求判断节点之间是否相似,进而构建全网节点的相似关系图  $Sim_{graph}$ .DcNS 算法的第 3 行~第 14 行是求解数据关键节点的过程.在算法的第 4 行~第 6 行,DcNS 根据公式(2.7)计算  $Sim_{graph}$  中每个节点相对应的价值(worth)属性.

$$Sim_{graph}[v].worth = \frac{Sim_{graph}[v].|neighbor|}{dis(Sim_{graph}[v], cnodeset)} \quad (2.7)$$

公式(2.7)的分子表示  $Sim_{graph}$  中节点  $Sim_{graph}[v]$  的邻居节点数目,分母表示  $Sim_{graph}[v]$  与现有的数据关键节点集合  $cnodeset$  中节点距离的平均值.因此,节点  $Sim_{graph}[v]$  的价值属性  $worth$  与其邻居节点的数目  $Sim_{graph}[v].|neighbor|$  成正比,与现有数据关键节点的距离  $dis(Sim_{graph}[v], cnodeset)$  成反比.算法第 7 行找出  $Sim_{graph}$  中  $worth$  属性最大的节点  $cn$ ,其最终目的是找出节点  $cs$ ,使得节点  $cs$  的数据价值最大,并且与现有的数据关键节点距离最近;然后,将此节点加入到现有的数据关键节点集合  $cnodeset$  中,并设置该数据关键节点  $cs$  的  $coverage$  属性为它所有的邻居节点(数据关键节点的  $coverage$  属性为其覆盖范围);最后,将节点  $cs$  及其相对应的邻居节点从顶点集合  $Sim_{graph}.V$  中删除.

下面根据 DcNS 算法执行的两个阶段详细分析 DcNS 算法的时间复杂度.DcNS 算法选择数据关键节点的



第 1 阶段是利用 DTW 算法,构建全网节点的相似关系图.由于在构建相似关系图的过程中,要判断每个节点与网络中其他节点之间是否相似,因此其时间复杂度为  $O(n^2)$ .假设节点对应的数据序列的平均长度为  $a$ ,则第 1 阶段整体的时间复杂度为  $O(a^2 \times n^2)$ .DcNS 算法的第 2 阶段是根据相似关系图选择数据关键节点,其核心部分是对所有节点进行两次循环判断过程.因此,第 2 阶段的时间复杂度为  $O(n^2)$ .因此,DcNS 算法的时间复杂度为  $O(a^2 \times n^2)$ .

假设全网节点相似性关系图  $Sim_{graph}$  中有  $N$  个顶点和  $M$  条边,且最终由 DcNS 算法计算得出的数据关键节点的数目为  $\kappa$ ,由文献[37]中的结论可以得知,  $\kappa \leq N + 1 - \sqrt{2M + 1}$ .因此,如果将数据关键节点作为 UAV 数据收集的目标节点,则 UAV 数据收集的任务量最少可以减少  $\sqrt{2M + 1} - 1$ .

DcNS 算法计算出的数据关键节点在网络运行过程中并不是固定不变的.由于数据关键节点是 UAV 进行数据收集的主要目标节点,所以其会与 UAV 频繁地通信.因此,如果数据关键节点一直固定不变,则其能量会快速耗尽,影响网络的生存周期以及 UAV 收集数据的可靠性.针对此问题,本文提出的解决方法是:为数据关键节点设置一个能量阈值,当其能量小于此阈值时,则在其节点覆盖范围内找到与现有的数据关键节点平均距离最小的那个节点,将其设定为新的数据关键节点.由于篇幅的限制,本文不详细讨论该能量阈值的具体设置情况.

## 2.5 基于数据关键节点的无人机数据收集

选出  $m$  个可以近似表示整个监测网络中感知节点采集数据的数据关键节点  $N_x = \{cn_1, cn_2, \dots, cn_m\}$  之后, OnValueGet 将  $N_x = \{cn_1, cn_2, \dots, cn_m\}$  中的节点作为数据收集的主要目标,并根据数据关键节点的具体位置以及数据收集过程中可能出现的异常情况和障碍物情况,动态地规划 UAV 的飞行路线.

为了便于描述 UAV 数据收集过程,本文假设每个 UAV 都有一个数据收集任务队列 DCTQ(data collection task queue),  $DCTQ = \{T_1, T_2, \dots, T_n\}$ ,用来存储其当前状况下所有的数据收集任务.DCTQ 中的每一个数据收集任务  $T_i \in DCTQ$  都由该任务对应的节点号  $Node\_id$  和该节点对应的数据价值  $Node\_worth$  两项组成,即  $T_i = \{Node\_id, Node\_worth\}$ .在 UAV 进行数据收集的初始阶段,利用集合  $N_x = \{cn_1, cn_2, \dots, cn_m\}$  中节点的节点号和其数据价值对 DCTQ 进行初始化.如果 UAV 完成了 DCTQ 中的某项任务,则将此任务从 DCTQ 中删除.

针对监测网络中可能出现的不在数据关键节点监测范围内异常情况,OnValueGet 根据数据平凡节点发送的异常感知数据请求 ADCR(abnormal data collection request)来进行异常数据的收集.数据平凡节点是网络中除去数据关键节点外的数据感知节点,它并不是 UAV 进行数据收集的目标节点,只有当其采集到异常数据的时候才有可能成为 UAV 的数据收集目标.数据平凡节点将采集到的数据暂存在自己的存储模块,如果存储模块已满,则使用 FIFO 方法进行替换.如果数据平凡节点采集的数据不在预先设定的正常范围内,则标记这条数据为异常数据,然后计算此时节点相应的数据价值,接下来调大发射功率,发送一个短小的控制报文,即 ADCR 给 UAV,控制报文的内容包括节点号和节点此时对应的数据价值.当 UAV 收集到 ADCR 时,将其存储在自己的 DCTQ 中.

本文将 UAV 从起始点出发到再次回到起始点记为一个工作周期  $w$ , UAV 在一个工作周期内理论上最大的可用能量为  $E_w$ .设在一个工作周期内,向 UAV 传输数据的所有数据感知节点的集合为  $l = \{n_j, n_k, \dots, n_m\}$ ,所以在一个工作周期内 UAV 收集的所有数据,其数据价值总和见公式(2.8):

$$V = \sum_{n_i \in l} v(n_i) \quad (2.8)$$

其中,  $v(n_i)$  表示节点  $n_i$  的数据价值.

综上所述,UAV 数据收集的路径规划问题可表述为:在数据收集的过程中,如何动态地选择 DCTQ 中数据收集任务执行的先后顺序,保证其在一个工作周期内收集到的数据具有最大的数据价值,而且在数据收集的过程中满足以下限制条件:网络中节点位置信息已知;UAV 能量有限;数据关键节点数据价值已知;数据平凡节点数据价值未知.

从以上的描述中可以分析得出,UAV 数据收集的路径规划问题是一个 Online Knapsack Problem,这是一个 NPC 问题,不存在有效的近似算法<sup>[38]</sup>.

为了解决此问题,本文引入收集效益(collection gain,简称 CG).CG 的物理意义为:UAV 在当前状况下,收集一个感知节点的数据所产生的直接投入产出比和最大间接投入产出比的数学期望之和.它反映了 UAV 的能量使用效率.因此,如果一条数据收集任务的 CG 值越高,则 UAV 执行该任务的优先级越高.

对于 UAV 的 DCTQ 中的第  $i$  条任务  $T_i$  而言,其对应的收集效益可以根据公式(2.9)计算得到:

$$CG_i = (1 - \delta) \frac{C_i}{\|P_{UAV} - P_{n_i}\|} + (1 - \delta)^2 f_{max}^i \tag{2.9}$$

在公式(2.9)中, $\delta$ 表示在数据收集过程中数据平凡节点采集到异常数据的概率, $C_i$ 表示  $T_i$  所对应的数据价值, $P_{UAV}$ 表示 UAV 的实时位置, $P_{n_i}$ 表示  $T_i$ 对应的数据感知节点  $n_i$ 的位置.因此, $(1 - \delta)C_i/\|P_{UAV} - P_{n_i}\|$ 表示 UAV 执行任务  $T_i$ 的直接投入产出比的数学期望.函数  $f_{max}^i$ 表示 UAV 执行完任务  $T_i$ 后,UAV 任务队列中所有任务投入产出比的最大值, $(1 - \delta)^2 f_{max}^i$ 表示 UAV 执行  $T_i$ 的间接最大投入产出比的数学期望.因此,公式(2.9)表示 UAV 执行某个数据收集任务时(收集某个节点的数据)所对应的收集效益.

根据收集效益,本文采用贪心方法对 UAV 收集数据的路径进行规划,使得 UAV 最终收集到尽可能大的数据价值.

### 3 仿真实验结果与分析

#### 3.1 仿真环境与真实数据来源分析

本文所提出的 OnValueGet 方法主要关注于如何选择数据关键节点.因此,我们假设 MAC 层数据碰撞问题可以得到较好的解决.我们利用 MATLAB 进行仿真实验数据分析.仿真实验的硬件环境为:处理器,Intel i5 2.8GHz;内存,8GB;硬盘,1TB;软件环境为,Windows7 64bit,mysql 5.5.

仿真实验使用的数据为真实采集的室外数据和室内数据.其中室外数据集是采用了 2015 年 12 月 1 日~2015 年 12 月 20 日之间,部署在榆林镇北台长城遗址内的 41 个数据感知节点所采集的长城遗址土壤内部 15cm 深度的温度值.数据采集的范围是-20°C~40°C 之间,部署图如图 3 所示.

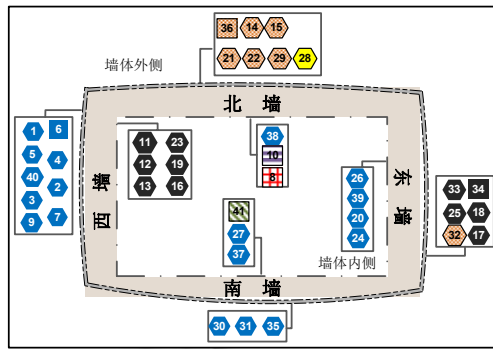


Fig.3 Outdoor deployment of nodes

图 3 室外节点部署图

室内数据集是采用了 2004 年 2 月 28 日~2004 年 4 月 5 日之间,部署在 Intel Berkeley Research 实验室的 54 个数据感知节点所采集的周围环境中的温度、湿度、光照和自身的电压数据,部署图如图 4 所示.对于室内实验采集到的 4 种数据,本文只利用节点采集的温度数据.由于在本文所选择的时间段内,5 号节点没有数据,所以在进行实验时实际使用了其他 53 个节点的温度数据.



Fig.4 Indoor deployment of nodes

图 4 室内节点部署图

### 3.2 DcNS算法验证与分析

下面分别利用室外和室内数据集,对本文提出的数据关键节点的选择算法 DcNS 的正确性进行验证.

图 3 的部署图中标出了将应用误差设置为 0.01 时,室外数据集所对应的数据关键节点及其相应的节点覆盖范围.图中正方形表示数据关键节点,六边形表示数据平凡节点.相同的颜色背景节点表示在同一颜色背景的数据关键节点的覆盖范围内.从图 3 中可以看到,整个网络中有 6 个数据关键节点,因此在无异常感知数据的情况下,整个数据收集过程中的任务量将会减少 85.37%.这 6 个数据关键节点中,6 号、34 号、36 号数据关键节点所对应的节点覆盖范围比较大,分别是 19,12,7,占到了整个网络节点数目的 92.67%.

对于 6 号数据关键节点所对应的节点覆盖范围,它们在位置上正好位于太阳每天照射的地方,因此这些地方的土壤温度都相对较高,所以这些节点所采集的数据具有很大的相似性;对于 34 号数据关键节点所对应的节点覆盖范围,由于镇北台西墙地势较高,东墙地势较低,西墙的外侧和东墙的内侧基本都挡住了太阳,而且墙体很厚,所以在西墙的内侧和东墙的外侧采集的土壤温度值都比较低,所以这些节点采集的数据具有巨大的相似性;对于 36 号数据关键节点所对应的节点覆盖范围,由于北墙的外侧是条路,所以北墙外侧土壤的温度受外界影响较大与其他墙体都有一定的差别,所以这些节点采集的数据具有巨大的相似性.

图 4 的部署图中标出了将应用误差设为 0.0001 时,室内数据集所对应的数据关键节点及其相应的节点覆盖范围.从图 4 可以看出:共有 7 个数据关键节点,45 号数据关键节点的节点覆盖范围为 42,占到了整个网络节点总数的 70.77%;其他数据关键节点的覆盖范围则零散地分布在监测区域的左下角和右上角.整体看来,数据关键节点以及其对应的节点覆盖范围,在地理分布上没有任何显著的特征.这是因为在室内环境下,环境非常稳定,并且室内的空气也是相互流通的,因此几乎在各个地方空气温度值都是相同的,数据之间基本都是相似的,所以才会出现上述结果.

图 5 是将应用误差  $\epsilon$  分别设置为 0.001,0.002,0.005,0.01,0.02,0.05 时,室外数据集对应的数据关键节点的数目分布图.图 6 是将应用误差  $\epsilon$  分别设置为 0.000 1,0.000 2,0.000 5,0.001,0.002,0.005 时,室内数据集对应的数据关键节点的数目分布图.我们利用文献[11]中提出的 Greedy 算法和第 2.4 节提出的数据关键节点选择算法 DcNS 分别对节点数据集进行处理,在不同的应用误差下选择数据关键节点,并对关键节点的数量进行比较.

从图 5 和图 6 中可以看出:无论是使用室外数据集还是室内数据集,数据关键节点的数目都会随着应用误差的增加而减少.这是因为当应用误差增加的时候,会促使更多的节点在同一时刻采集的数据在误差的范围内,所以更多的节点将会在一个数据关键节点的节点覆盖范围内.

从实验结果可知,通过 Greedy 算法处理节点数据得到的关键节点数量明显大于通过 DcNS 处理节点数据得到的关键节点数量.这是因为 Greedy 算法只是利用节点相似关系图找出与最多节点相似的节点作为数据关键节点,而 DcNS 在此基础上将已存在的数据关键节点的位置作为新的限制条件,使得新选择的数据关键节点的位置尽可能地靠近已经选择的数据关键节点,同时也减少了数据关键节点的数量.又因为关键节点数量越

多,UAV 节点就需要采集更多任务节点的数据,UAV 的数据采集任务就越大,Greedy 算法对应的 UAV 任务量明显比 DcNS 对应的 UAV 任务量大.综上所述,若采用 DcNS 来处理节点数据,能够很好地减少关键节点数量,从而减少 UAV 数据采集任务.

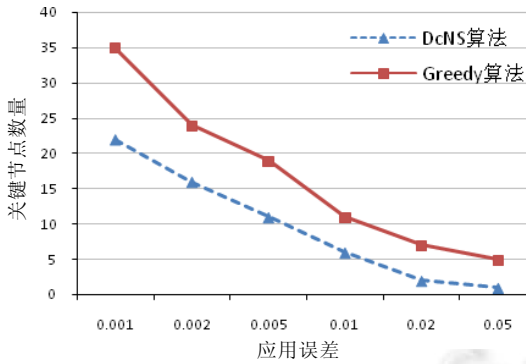


Fig.5 Number of data-criticalnodes of DcNS and Greedy algorithm with different application error from outdoor dataset

图 5 室外数据集不同应用误差下,DcNS 和 Greedy 算法选择数据关键节点数量比较

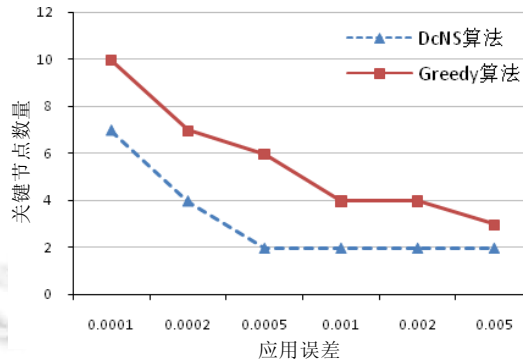


Fig.6 Number of data-criticalnodes of DcNS and Greedy algorithm with different application error from indoor dataset

图 6 室内数据集不同应用误差下,DcNS 和 Greedy 算法选择数据关键节点数量比较

图 7、图 8 分别是在使用室外数据集和室内数据集的环境中,不同应用误差下,UAV 收集所有节点数据花费时间和只收集关键节点数据花费时间的对比.根据现实环境中的情况,本文假设 UAV 节点每飞行一个单位长度耗费 0.02 个单位能量,在航行的过程中,总是以单位时间内(这里设定为 1s)飞行 1.6 个单位长度的速度匀速运行.且在 UAV 节点数据收集之前,它已经获取了数据关键节点的所有信息以及网络中所有数据感知节点的地理位置.

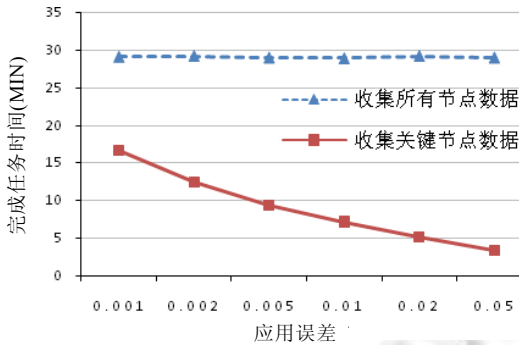


Fig.7 Latency of data collection with different application error from outdoor dataset

图 7 室外数据集下,不同应用误差对应的数据收集时间

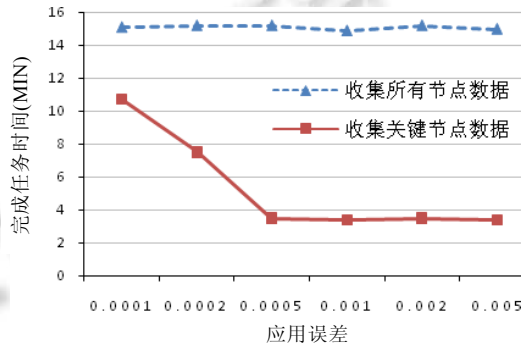


Fig.8 Latency of data collection with different application error from indoor dataset

图 8 室内数据集下,不同应用误差对应的数据收集时间

因为应用误差只对关键节点个数有影响,所以随着应用误差的增大,关键节点个数变小,UAV 完成数据收集任务航行的路径变短,完成数据收集任务的时间也随之变少.而 UAV 收集所有节点数据的时间和应用误差是没有直接关系的,所以不管应用误差如何变化,UAV 收集所有节点数据花费的时间保持不变.

由图 7、图 8 可以看出:室内数据集和室外数据集的仿真环境中,无论应用误差如何变化,UAV 收集所有节

点数据花费时间总是远远大于只收集关键节点数据花费时间.在仿真实验中,因为 UAV 以匀速飞行,单位飞行长度和单位飞行时间内 UAV 耗费能量不变,UAV 耗费能量和 UAV 飞行时间是呈正比关系.所以,UAV 收集所有节点数据的能耗大于只收集关键节点数据的能耗.

图 9、图 10 所示为在室外数据集和室内数据集环境中,关键节点数量从 1 递增至 10 时,DcNS 算法选择的数据关键节点采集的数据与数据关键节点覆盖范围内数据平凡节点采集的数据之间的平均误差变化(在此称为数据误差).可以看出:随着关键节点数量的增加,数据误差会迅速变小,最终保持在一个非常小的数量级上(室内数据误差在小数点后 3 位,室外数据误差在小数点后两位).

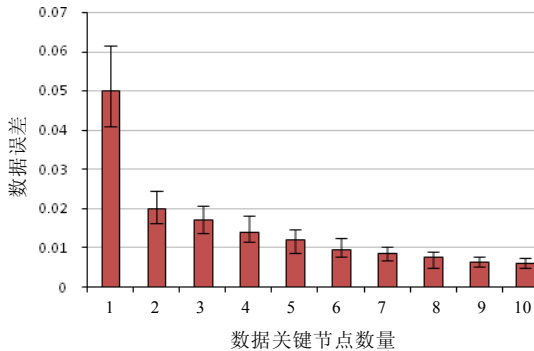


Fig.9 Data error with different number of data-criticalnodes from outdoor dataset

图 9 室外数据集下,不同数量数据关键节点对应的数据误差

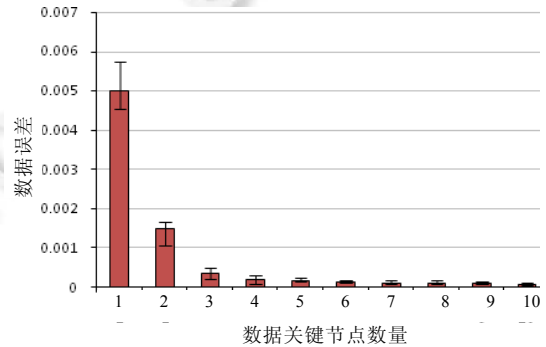


Fig.10 Data error with different number of data-criticalnodes from indoor dataset

图 10 室内数据集下不同数量数据关键节点对应的数据误差

因为数据关键节点是在给定的应用误差下对比网络内节点采集数据的相似性进行判断,所以从结果可以看出,数据关键节点采集的数据和其覆盖范围内的数据平凡节点采集的数据的平均误差在给定的应用误差范围内.虽然有个别数据的误差大于给定的应用误差,从整体来看,数据误差的分布较为稳定,没有出现太大的波动.除此之外,虽然相对于网络中所有的节点来说数据关键节点不能采集全部的信息,但是收集数据关键节点采集的数据和所有平凡节点采集的数据之间的误差并不是很大,这样就使得系统可以在给定的应用误差范围内有效的减少相似冗余的数据.

表 1 是将应用误差分别设置为 0.001,0.002,0.005,0.01,0.02,0.05 时,数据关键节点选择算法的运行时间的变化情况.从表 1 中可以看出:无论应用误差设置多少,数据关键节点选择算法的运行时间不会有相关性的变化.在第 2.4 节中我们已经对算法的时间复杂度进行了分析,结果表示:算法时间复杂度与待分析的数据量及节点数目有关,与应用误差范围没有关系.

Table 1 Running time of data-critical nodes selection algorithm with different application error

表 1 不同应用误差下数据关键节点选择算法时间

应用误差	运行时间(s)	应用误差	运行时间(s)
0.01	2.047	0.001	2.033
0.02	1.982	0.002	1.994
0.05	2.029	0.005	1.942

从本节的实验结果可知:应用误差要求与数据关键节点的数目以及关键面积大小有着密切的联系,从而影响到 UAV 进行数据收集时的任务量和航行路径长度.因此,选择合适的应用误差,对于求解数据关键节点至关重要.通过对比可以得出:UVA 节点收集数据关键节点采集的数据可以有效地减少采集时间,进而减少采集能耗.同时,数据关键节点采集的信息和其覆盖范围内所有节点采集的数据相比,误差在允许的应用误差范围内.通过实验分析可以验证,OnValueGet 数据收集方法可以有效地提高 UVA 节点收集数据的效能.

## 4 总 结

面对目前无线监测网络中主要的数据收集方法在实际应用中严重受限的问题,本文提出基于 UAV 的数据收集方法 OnValueGet,设计了基于数据关键节点的发现方法.实验结果表明:OnValueGet 在应用规定的误差范围内有效地减少数据收集量,从而减少数据收集能耗和时间.在 UAV 节点能量限制的情况下,提高收集数据的总价值.但是,数据关键节点的选择以感知节点采集的数据内容为主要依据,忽略了节点位置信息所占的价值比重.因此,研究数据关键节点为网络代表的方法在部分物联网应用(如位置服务应用)中存在很大的限制,因此主要适用于位置信息已知的、以感知节点采集数据内容为主要目标的大规模物联网监测应用.

除此之外,在真实场景应用中,本文的许多工作还有进一步的提升空间.

- 首先,本文利用感知数据在物理意义上的时空相关性以及周期变化性,对传感网的历史数据进行了分析,然后得出了网络中的数据关键节点.然而如何选取历史数据,对于分析得到数据关键节点是非常重要的部分.对此类数据集进行挑选的方法,是下一步要进行的工作;
- 其次,在对 UAV 节点进行数据收集的仿真过程中,本文尽量去模拟 UAV 节点在真实环境中遇到的一些状况,但是与真实情况相比还是有一些噪声需要进一步模拟,所以最终的结果与真实环境可能还会有一定差距;
- 最后,对于数据关键节点能量阈值的设置,本文没有进行详细的讨论.因此,关于数据关键节点能量阈值的设置以及更新方法,需要在下一步工作中继续进行深入研究.

## References:

- [1] Ren FY, Huang HN, Lin C. Wireless sensor networks. Ruan Jian Xue Bao/Journal of Software, 2003,14(7):1282–1291 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/14/1282.htm> [doi: 10.13328/j.cnki.jos.2003.07.013]
- [2] Li J, Mohapatra P. Analytical modeling and mitigation techniques for the energy hole problem in sensor networks. Pervasive and Mobile Computing, 2007,3(3):233–254. [doi: 10.1016/j.pmcj.2006.11.001]
- [3] Cao F, Liu LP, Wang Z. A new energy-efficient WSN deployment algorithm. Information and Control, 2006,35(02):147–153 (in Chinese with English abstract). [doi: 10.3969/j.issn.1002-0411.2006.02.004]
- [4] Di Francesco M, Das SK, Anastasi G. Data collection in wireless sensor networks with mobile elements: A survey. ACM Trans. on Sensor Networks, 2011,8(1):7:1–7:31. [doi: 10.1145/1993042.1993049]
- [5] Abdulla AEAA, Md Fadlullah Z, Nishiyama H, Kato N. An optimal data collection technique for improved utility in UAS-aided networks. In: Proc. of the 2014 IEEE INFOCOM. IEEE, 2014. 736–744. [doi: 10.1109/INFOCOM.2014.6848000]
- [6] Abdulla AEAA, Md Fadlullah Z, Nishiyama H, Kato N. Toward fair maximization of energy efficiency in multiple UAS-aided networks: A game-theoretic methodology. IEEE Trans. on Wireless Communications, 2015,14(1):305–316. [doi: 10.1109/TWC.2014.2343219]
- [7] Bekmezci I, Sahingoz OK, Temel S. Flying ad-hoc networks (FANETs). Ad Hoc Networks, 2013,11(3):1254–1270. [doi: 10.1016/j.adhoc.2012.12.004]
- [8] Adams SM, Friedland CJ. A survey of unmanned aerial vehicle (UAV) usage for imagery collection in disaster research and management. In: Proc. of the 9th Int'l Workshop on Remote Sensing for Disaster Response. Stanford University, 2011. 14–16.
- [9] Casbeer DW, Beard RW, McLain TW, Li SM. Forest fire monitoring with multiple small UAVs. In: Proc. of the 2005 IEEE American Control Conf. IEEE, 2005. 3530–3535. [doi: 10.1109/ACC.2005.1470520]
- [10] Honkavaara E, Saari H, Kaivosoja J, Pölonen I, Hakala T, Litkey P, Mäkynen J, Pesonen L. Processing and assessment of spectrometric, stereoscopic imagery collected using a lightweight UAV spectral camera for precision agriculture. Remote Sensing, 2013,5(10):5006–5039. [doi: 10.3390/rs5105006]
- [11] Gu Z, Hua QS, Wang Y, Lau F. Reducing information gathering latency through mobile aerial sensor network. In: Proc. of the 2013 IEEE INFOCOM. IEEE, 2013. 656–664. [doi: 10.1109/INFOCOM.2013.6566851]

- [12] Liu M, Cao JN, Chen GH, Chen LJ, Wang XM, Gong HG. EADEEG: An energy-aware data gathering protocol for wireless sensor networks. *Ruan Jian Xue Bao/Journal of Software*, 2007,18(5):1092–1109 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/18/1092.htm> [doi: 10.1360/jos181092]
- [13] Shi GT, Liao MH. Movement-Assisted data gathering scheme with load-balancing for sensor networks. *Ruan Jian Xue Bao/Journal of Software*, 2007,18(9): 2235–2244 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/18/2235.htm> [doi: 10.1360/jos182235]
- [14] Zhang Q, Xie ZP, Ling B, Sun WW, Shi BL. A maximum lifetime data gathering algorithm for wireless sensor networks. *Ruan Jian Xue Bao/Journal of Software*, 2005,16(11):1946–1957 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/16/1946.htm> [doi: 10.1360/jos161946]
- [15] Liang JB, Wang XJ, Li TS, Chen JE. Maximum lifetime algorithm for precise data gathering based on tree in wireless sensor networks. *Ruan Jian Xue Bao/Journal of Software*, 2010,21(9):2289–2303 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/3684.htm> [doi: 10.3724/SP.J.1001.2010.03684]
- [16] Zhang CQ, Li ML, Wu MY. An approach for constructing load-balancing networks for data gathering wireless sensor networks. *Ruan Jian Xue Bao/Journal of Software*, 2007, 18(5):1110–1121 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/18/1110.htm> [doi: 10.1360/jos181110]
- [17] Liu M, Gong HG, Mao YC, Chen LJ, Xie L. A distributed energy-efficient data gathering and aggregation protocol for wireless sensor networks. *Ruan Jian Xue Bao/Journal of Software*, 2005,16(12):2106–2116 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/16/2106.htm> [doi: 10.1360/jos162106]
- [18] Zheng GQ, Li JD, Zhou ZL. Energy-Efficient data gathering protocol for multihop wireless sensor networks. *Ruan Jian Xue Bao/Journal of Software*, 2010,21(9): 2320–2337 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/3702.htm> [doi: 10.3724/SP.J.1001.2010.03702]
- [19] Zhu JQ, Liu M, Gong HG, Chen GH, Xu FL, Song C. Selective replication-based data delivery for delay tolerant mobile sensor networks. *Ruan Jian Xue Bao/Journal of Software*, 2009,20(8):2227–2240 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/3323.htm>
- [20] Luo J, Hubaux JP. Joint sink mobility and routing to maximize the lifetime of wireless sensor networks: The case of constrained mobility. *IEEE/ACM Trans. on Networking*, 2013,18(3):871–884. [doi: 10.1109/TNET.2009.2033472]
- [21] Shah R, Roy S, Jain S, Brunette W. Data mules: Modeling a three-tier architecture for sparse sensor networks. In: *Proc. of the 2003 1st IEEE Int'l Workshop on Sensor Network Protocols and Applications*. IEEE, 2003. 30–41. [doi: 10.1109/SNPA.2003.1203354]
- [22] Watts AC, Ambrosia VG, Hinkley EA. Unmanned aircraft systems in remote sensing and scientific research: Classification and considerations of use. *Remote Sensing*, 2012,4(6):1671–1692. [doi: 10.3390/rs4061671]
- [23] Tisdale J, Ryan A, Zu K, Tomqvist D. A multiple UAV system for vision-based search and localization. In: *Proc. of the American Control Conf. IEEE*, 2008. 1985–1990. [doi: 10.1109/ACC.2008.4586784]
- [24] Pongpunwattana A, Rysdyk R, Berg M. Real-Time planning for multiple autonomous vehicles in dynamic uncertain environments. *Journal of Aerospace Computing Information & Communication*, 2012,1(12):580–604. [doi: 10.2514/1.12919]
- [25] Dantu K, Kate B, Waterman J, Bailis P, Welsh M. Programming micro-aerial vehicle swarms with karma. In: *Proc. of the 9th ACM Conf. on Embedded Networked Sensor Systems (SenSys 2011)*. New York: ACM Press, 2011. 121–134. [doi: 10.1145/2070942.2070956]
- [26] Liang JB, Zou SJ, Chen NJ, Li T. Delay-Constrained data collection without aggregation in wireless sensor network with mobile collector. *Ruan Jian Xue Bao/Journal of Software*, 2016,27(7):1822–1840 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4926.htm> [doi: 10.13328/j.cnki.jos.004926]
- [27] Xu FL, Liu M, Gong HG, Chen GH, Li JP, Zhu JQ. Relative distance-aware data delivery scheme for delay tolerant mobile sensor networks. *Ruan Jian Xue Bao/Journal of Software*, 2010,21(3):490–504 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/3459.htm> [doi: 10.3724/SP.J.1001.2010.03459]
- [28] Simon T, Mitschele-Thiel A. Next-Hop decision-making in mobility-controlled message ferrying networks. In: *Proc. of the 1st Workshop on Micro Aerial Vehicle Networks, Systems, and Applications for Civilian Use*. ACM Press, 2016. 9–14.

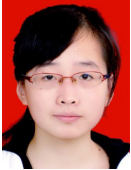
- [29] Yoon J, Liu P, Banerjee S, Kim S. PI in the sky. In: Proc. of the 2nd Workshop on Micro Aerial Vehicle Networks, Systems, and Applications for Civilian Use. ACM Press, 2016. 53–54. [doi: 10.1145/2935620.2935626]
- [30] Trotta A, Bedogni L, Di Felice M, Bononi L, Natalizio E. Enhancing TV white-spaces database with unmanned aerial scanning vehicles (UASVs). In: Proc. of the 2nd Workshop on Micro Aerial Vehicle Networks, Systems, and Applications for Civilian Use. ACM Press, 2016. 23–28. [doi: 10.1145/2935620.2935621]
- [31] Hamza A, Keppitiyagama C, De Zoysa K, Iyer V, Hewage K, Voigt T. A quadcopter controller to maintain radio link quality. In: Proc. of the 1st Workshop on Micro Aerial Vehicle Networks, Systems, and Applications for Civilian Use. ACM Press, 2015. 21–26. [doi: 10.1145/2750675.2750678]
- [32] Yang Z, Cai L, Liu Y, Pan J. Environment-Aware clock skew estimation and synchronization for wireless sensor networks. In: Proc. of the 2012 IEEE INFOCOM. IEEE, 2012. 1017–1025. [doi: 10.1109/INFOCOM.2012.6195457]
- [33] Jin M, Chen XJ, Fang DY, Tang ZY, Liu C, Xu D, Wang W. Temperature-Adaptive time synchronization for wireless sensor networks. Ruan Jian Xue Bao/ Journal of Software, 2015,26(10):2667–2683 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4792.htm> [doi:10.13328/j.cnki.jos.004792]
- [34] Salvador S, Chan P. Toward accurate dynamic time warping in linear time and space. Intelligent Data Analysis, 2007,11(5): 561–580.
- [35] Johnson DS. The NP-completeness column: An ongoing guide. Journal of Algorithms, 1985,6(3):434–451. [doi: 10.1016/0196-6774(82)90032-3]
- [36] Lu G, Zhou MT, Tang Y, Wu ZQ, Qiu GY, Yuan L. A survey on exact algorithms for dominating set related problems in arbitrary graphs. Chinese Journal of Computers, 2010,33(6):1073–1087 (in Chinese with English abstract). [doi: 10.3724/SP.J.1016.2010.01073]
- [37] Parekh AK. Analysis of a greedy heuristic for finding small dominating sets in graphs. Information Processing Letters, 1991,39(5): 237–240. [doi: 10.1016/0020-0190(91)90021-9]
- [38] Marchetti-Spaccamela A, Vercellis C. Stochastic on-line knapsack problems. Mathematical Programming, 1995,68(1-3):73–104. [doi: 10.1007/BF01585758]

#### 附中文参考文献:

- [1] 任丰原,黄海宁,林闯.无线传感器网络.软件学报,2003,14(7):1282–1291. <http://www.jos.org.cn/1000-9825/14/1282.htm>
- [3] 曹峰,刘丽萍,王智.能量有效的无线传感器网络部署.信息与控制,2006,35(2):147–153. [doi: 10.3969/j.issn.1002-0411.2006.02.004]
- [12] 刘明,曹建农,陈贵海,陈力军,王晓敏,龚海刚.EADEEG:能量感知的无线传感器网络数据收集协议.软件学报,2007,18(5): 1092–1109. <http://www.jos.org.cn/1000-9825/18/1092.htm> [doi: 10.1360/jos181092]
- [13] 石高涛,廖明宏.传感器网络中具有负载平衡的移动协助数据收集模式.软件学报,2007,18(9):2235–2244. <http://www.jos.org.cn/1000-9825/18/2235.htm> [doi: 10.1360/jos182235]
- [14] 张卿,谢志鹏,凌波,孙未未,施伯乐.一种传感器网络最大化生命周期数据收集算法.软件学报,2005,16(11):1946–1957. <http://www.jos.org.cn/1000-9825/16/1946.htm> [doi: 10.1360/jos161946]
- [15] 梁俊斌,王建新,李陶深,陈建二.传感器网络中基于树的最大生命精确数据收集.软件学报,2010,21(9):2289–2303. <http://www.jos.org.cn/1000-9825/3684.htm> [doi: 10.3724/SP.J.1001.2010.03684]
- [16] 张重庆,李明禄,伍民友.数据收集传感器网络的负载平衡网络构建方法.软件学报,2007,18(5):1110–1121. <http://www.jos.org.cn/1000-9825/18/1110.htm> [doi: 10.1360/jos181110]
- [17] 刘明,龚海刚,毛莺池,陈力军,谢立.高效节能的传感器网络数据收集和聚合协议.软件学报,2005,16(12):2106–2116. <http://www.jos.org.cn/1000-9825/16/2106.htm> [doi: 10.1360/jos162106]
- [18] 郑国强,李建东,周志立.多跳无线传感器网络的高能效数据收集协议.软件学报,2010,21(9):2320–2337. <http://www.jos.org.cn/1000-9825/3702.htm> [doi: 10.3724/SP.J.1001.2010.03702]
- [19] 朱金奇,刘明,龚海刚,陈贵海,许富龙,宋超.延迟容忍移动传感器网络中基于选择复制的数据传输.软件学报,2009,20(8): 2227–2240. <http://www.jos.org.cn/1000-9825/3323.htm> [doi: 10.3724/SP.J.1001.2009.03323]



- [26] 梁俊斌,邹绍军,陈宁江,李韬.传感网中延迟限定的非汇聚数据移动式收集.软件学报,2016,27(7):1822-1840. <http://www.jos.org.cn/1000-9825/4926.htm> [doi: 10.13328/j.cnki.jos.004926]
- [27] 许富龙,刘明,龚海刚,陈贵海,李建平,朱金奇.延迟容忍传感器网络基于相对距离的数据传输.软件学报,2010,21(3):490-504. <http://www.jos.org.cn/1000-9825/3459.htm> [doi: 10.3724/SP.J.1001.2010.03459]
- [33] 金梦,陈晓江,房鼎益,汤战勇,刘晨,徐丹,王薇.一种温度自适应无线传感网络时间同步方法.软件学报,2015,26(10):2667-2683. <http://www.jos.org.cn/1000-9825/4792.htm> [doi:10.13328/j.cnki.jos.004792]
- [36] 路纲,周明天,唐勇,吴振强,裘国永,袁柳.任意图支配集精确算法回顾.计算机学报,2010,33(6):1073-1087. [doi: 10.3724/SP.J.1016.2010.01073]



徐丹(1988-),女,陕西西安人,博士生,工程师,主要研究领域为无线网络数据传输方法.



龚晓庆(1974-),女,博士,副教授,主要研究领域为软件工程,软件安全,工作流.



李伟(1991-),男,硕士,主要研究领域为无线网络数据传输算法研究.



陈晓江(1973-),男,博士,教授,博士生导师,CCF 专业会员,主要研究领域为无线网络,物联网,定位与无源感知.



王安文(1979-),男,博士生,讲师,主要研究领域为计算机网络,智能控制.



房鼎益(1959-),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为网络与信息安全,数字内容与计算机软件安全保护,无线传感器网络及其应用,移动计算,分布计算系统.



范浩楠(1991-),男,硕士,主要研究领域为无人机数据传输算法研究.