

基于 K-means 的三维无线传感器网络分簇算法研究

苗丽娟¹ 陈晓节¹ 程超凡¹ 李奕宽¹ 马钦程¹

MIAO Lijuan CHEN Xiaojie CHENG Chaofan LI Yikuan MA Qincheng

摘要

针对三维无线传感器网络节点分布复杂, 能耗问题突出, 而传统分簇算法在分簇时未充分结合能量与距离因素等局限性, 导致三维场景下网络能耗较大的问题, 文章提出了一种基于改进 K-means 聚类的三维分簇算法 E-K-means 算法, 通过引入节点能量与三维欧氏距离的加权因子, 优化网络节点的入簇机制, 增强簇内节点与簇头节点之间的能耗均衡性, 从而降低节点的能耗, 延长网络的生存时间。仿真结果表明, 与传统的 K-means 分簇算法相比, 改进的分簇算法 E-K-means 的分簇效果更好, 网络生存时间更长。

关键词

三维无线传感器网络; 分簇算法; E-K-means 算法; 网络生存时间

doi: 10.3969/j.issn.1672-9528.2025.08.011

0 引言

随着科技的飞速发展, 无线传感器网络 (wireless sensor network, WSN) 作为一种集成了传感器、嵌入式计算、网络和无线通信四大技术的新型信息获取与处理技术, 已在众多领域展现出广泛的应用前景^[1]。无线传感器网络由大量的传感器节点组成, 这些节点具备感知、处理和通信能力, 能够实时协作感知、采集各种环境或监测对象的信息, 并进行相

应处理后传送给有需求的用户, 而三维无线传感器网络 (3D wireless sensor network, 3D-WSN) 是一种在三维空间中部署的无线传感器网络, 与传统的二维无线传感器网络相比, 3D-WSN 能够提供更立体和全面的监测覆盖, 尤其适用于需要在高度或深度方向进行监测的应用场景。三维传感器网络的应用领域更广泛, 包括但不限于环境监测、灾害预警与监测、智能建筑与城市管理、工业自动化、农业领域等^[2]。

与传统二维网络相比, 三维传感器网络在空间维度上具有更广泛的覆盖范围和更丰富的数据采集能力, 但三维空间中的传感器节点之间的通信距离和路径更为复杂。在实际应用中, 无线传感器网络节点通常依靠电池供电, 且往往部署在不易更换电池的位置, 节点能量有限^[3], 三维场景下受到空间结构、障碍物、信号衰减等因素的影响, 能量消耗更大, 因此, 如何高效利用有限的能量, 延长网络的生命周期, 提

1. 太原学院 山西太原 030000

[基金项目] 2024 年山西省高等学校大学生创新创业训练计划项目 (20241445); 太原学院 2024 年大学生创新创业训练计划项目 (TYX2024063); 2023 年山西省高等学校一般性教学改革创新立项项目 (J20231428、J20231426); 2024 年山西省高等学校一般性教学改革创新立项项目 (J20241478)

[19] DEVLIN J, CHANG M W, LEE K, et al.BERT: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Brussels: ACL, 2019: 4171-4186.

[20] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural computation, 1997,9(15):1735 - 1780.

[21] LI Y R, SU H, SHEN X Y, et al.DailyDialog: a manually labelled multi - turn dialogue dataset[C]//Proceedings of the Eighth International Joint Conference on Natural Language Processing. Brussels: ACL, 2017: 986 - 995.

[22] 巩固. 基于 Transformer 的生成式对话系统 [D]. 成都: 电

子科技大学, 2021.

[23] 时代. 基于生成式对抗网络的开放领域多轮对话 [D]. 北京: 北京邮电大学, 2020.

[24] BENGIO Y, DUCHARME R, VINCENT P, et al.A neural probabilistic language model[J].Journal of machine learning research, 2003,3:1137 - 1155.

[25] 王春喻. 面向多轮对话的情感强度回复生成方法研究 [D]. 内蒙古: 内蒙古工业大学, 2022.

【作者简介】

李凯伟 (1996—), 男, 山西晋城人, 硕士, 研究方向: 自然语言处理对话生成, email: likaiwei@sxist.edu.cn.

(收稿日期: 2025-03-19 修回日期: 2025-07-31)

高网络能量效率,成为三维无线传感器网络研究的关键问题之一。

分簇算法^[4]在解决上述问题方面具有重要应用,通过将无线传感器网络划分成若干个簇,每个簇选出一个簇头,簇头负责收集簇内节点传来的数据并进行融合处理后再传送给基站,这种方式可以有效减少网络能量消耗、降低节点能耗、延长网络寿命,同时还能在信息广播、数据查询等方面发挥积极作用,有助于提升网络的整体性能。

1 典型的分簇算法

分簇算法可以从不同的角度进行分类,按照在分簇过程中簇头的选取是否随机,可以将分簇算法归结为两大类^[5],LEACH (low - energy adaptive clustering hierarchy)^[6]算法、TEEN (threshold sensitive energy efficient sensor Network protocol)^[7]算法等是随机选取簇头节点的典型分簇算法,HEED (hybrid energy efficient distributed clustering)^[8]、EEUC (energy efficient unequal clustering)^[9]算法等是根据节点信息选取簇头的典型分簇算法。

LEACH 是一种经典的随机选择簇头的分簇算法。它采用随机循环的方式来确定簇头。在每一轮开始时,每个传感器节点都会生成一个介于 0 到 1 之间的随机数。这个随机数跟预先设定的阈值比较,若随机数小于设定的阈值,则该节点就成为本轮的簇头。TEEN 主要应用于需要实时监测环境数据变化的场景。与 LEACH 随机选择簇头的方式类似,不同之处在于节点的数据传输阶段,TEEN 算法通过感知数据的硬阈值和软阈值来触发数据传输,只有当监测数据超过硬阈值并且变化幅度超过软阈值时,节点才会向簇头发送数据。

HEED 是一种分布式的分簇算法,主要基于节点的剩余能量和节点间的通信代价来选择簇头。在初始阶段,每个节点都会计算自己成为簇头的概率,这个概率与节点的剩余能量成正比,剩余能量高的节点有更高的概率被初步选为簇头。EEUC 是为无线传感器网络设计的分簇路由协议,在簇头选举过程中,考虑节点的剩余能量和节点的竞争半径。节点首先计算自己的竞争半径,该半径与节点的剩余能量和到基站的距离有关。剩余能量高的节点和距离基站远的节点具有更大的竞争半径。

目前,大多数无线传感器网络分簇算法多基于二维环境下,但实际应用中存在大量三维应用场景。相较于二维情况,三维无线传感器网络的分簇面临着更多挑战,由于三维空间中求解问题的复杂性,二维的研究成果往往不能直接应用于三维网络。K-means 聚类算法作为一种有效的数据聚类方法,能够依据节点的多种属性特征对其进行合理聚类分组,进而优化网络的拓扑结构,将聚类算法应用于网络分簇具有实际意义。本文在 K-means 聚类算法的基础上,综合考虑网络节

点能量和距离因素改进网络节点的分簇方法,提出 E-K-means 聚类算法。通过该算法对三维无线传感器网络进行分簇,实现更精准的簇划分、更均衡的能量消耗以及更高效的数据传输,延长网络寿命、从而更好地满足现实应用中复杂多样的三维监测场景需求。

2 E-K-means 分簇算法

2.1 基本思路

E-K-means 分簇算法是一种综合考虑节点能量和节点间距离的分簇算法,旨在将三维无线传感器网络中的节点划分为不同的簇,每个簇由一个簇头节点领导。这种分簇方式有助于减少节点间的通信距离和能量消耗,提高网络的能效和生存时间。

E-K-means 分簇算法保留了 K-means 聚类算法中的簇头选举机制,对节点加入簇的过程进行了改进,首先,初始簇头节点的选择:随机选择 K 个节点作为初始簇头节点。然后,计算能量距离并进行分簇:计算每个节点与 K 个簇头节点的能量距离,将网络中的每个节点分配到距离其最近的簇头节点所在的簇中。簇头节点更新:对于每个簇,重新计算该簇内所有节点的簇内平方误差和,并将该均值作为新的簇头节点。最后不断迭代:直到簇头节点的位置不再发生显著变化,或者达到预定的迭代次数。在迭代过程中,节点的分配和簇头节点的位置会不断更新,直到算法收敛。

2.2 无线通信能耗模型

三维无线传感器网络的能耗模型是分析节点能量消耗、优化网络寿命的重要工具,而通信能耗是网络最大的能耗来源,本文使用如下无线通信能耗模型^[10],节点发送 k 比特数据到距离 d 的接收节点,其所消耗的能量 $E_{TX}(k, d)$ 为:

$$E_{TX}(k, d) = \begin{cases} kE_{elec} + k\epsilon_{fs}d^2 & d \leq d_0 \\ kE_{elec} + k\epsilon_{mp}d^4 & d > d_0 \end{cases} \quad (1)$$

$$d_0 = \sqrt{\epsilon_{fs} / \epsilon_{mp}} \quad (2)$$

式中: d_0 为传输距离阈值,当通信距离大于传输距离阈值时,使用多径衰减模型,否则采用自由空间模型; ϵ_{fs} 、 ϵ_{mp} 为信道模型传输增益; E_{elec} 为发送或接收每比特数据消耗的能量。

节点接收 k 比特数据消耗的能量 $E_{RX}(k)$ 为:

$$E_{RX}(k) = kE_{elec} \quad (3)$$

如果节点为簇头节点,发送 k 比特数据到距离 d 的节点需要消耗的能量 $E_{TX}(k, d)$ 为:

$$E_{TX}(k, d) = \begin{cases} kE_{elec} + kE_{DA} + k\epsilon_{fs}d^2 & d \leq d_0 \\ kE_{elec} + kE_{DA} + k\epsilon_{mp}d^4 & d > d_0 \end{cases} \quad (4)$$

式中: E_{DA} 为簇头节点进行数据融合的能量。

2.3 算法实现

算法分为以下几个阶段实现:

(1) 节点分布, 在三维区域 $M \times M \times M$ 的空间范围内, 随机分布 N 个网络节点, 同时随机设置网络节点的初始能量, 每个节点保存自身的位置坐标信息和能量信息 (x_i, y_i, z_i, E_i) , 在 $(M/2, M/2, M/2)$ 的位置放置 sink 节点。

(2) 选择 K 值, E-K-means 分簇算法的目标是将三维无线传感器网络节点划分为 K 个簇, 使得每个节点属于能量距离最近的簇头节点。

K 值的确定通常采用肘部法则或轮廓系数的方法, 本文采用轮廓系数的方法。

轮廓系数: 分别计算不同 K 值下所有节点的平均轮廓系数, 轮廓系数公式为:

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}} \quad (5)$$

式中: a_i 表示节点 i 的内聚度, 即节点 i 到同簇其他节点的平均距离; b_i 表示节点 i 的分离度, 即节点 i 到最近簇中所有节点的平均距离, 计算公式分别为:

$$a_i = \frac{1}{n-1} \sum_{j \in C_i, j \neq i} \text{distance}(i, j) \quad (6)$$

式中: j 代表与 i 在同一个簇内的其他节点; C_i 代表节点 i 所在的簇; n 代表簇内节点个数; distance 代表 i 与 j 之间的距离, a_i 的值越小, 说明该簇越紧密。

$$b_i = \min_{C \neq C_i} \left(\frac{1}{|C|} \sum_{j \in C} \text{distance}(i, j) \right) \quad (7)$$

式中: j 是与 i 在不同簇的节点; C 是其他簇; $|C|$ 是 C 簇内的节点数。通过遍历其他簇得到多个值并从中选择最小的值作为最终的结果。

将式 (6) 和 (7) 代入式 (5) 计算 S_i 的值, S_i 取值范围在 -1 到 1 之间, 值越接近 1, 聚类效果越好, 即节点与其所属簇头节点的距离越近, 而与其他簇头节点的距离越远。

初始簇头节点的选择, 随机选择 K 个节点作为初始簇头节点。

分配剩余网络节点, 对非簇头节点计算其到各个簇头节点的能量距离, 并将其归入到能量距离最小的簇头节点所在的簇。用公式表示为:

$$d(x, \mu) = \sqrt{\sum_{j=1}^n (x_j - \mu_j)^2} \cdot e^{-E_j} \quad (8)$$

式中: n 表示维度; x_j 表示节点在 j 维度上的坐标; μ_j 表示簇头在 j 维度上的坐标; e^{-E_j} 表示节点在 j 维度上的能量。

更新簇头节点, 当所有的节点都聚类到所属簇内后, 根据各簇节点重新计算各个簇的簇头节点所在位置, 用公式表示为:

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x \quad (9)$$

式中: μ_i 表示第 i 个簇的簇头节点; $|C_i|$ 表示簇 C_i 中的节点

数量; x 表示 C_i 中的节点。

迭代优化, 重复执行式 (3) 与 (4), 直到网络节点的划分情况不再发生变化, 即当簇中心的变化小于某个阈值, 此阈值根据式 (10) 计算综合考虑能量因素的簇内误差平方和得到, 或达到最大的迭代次数, 标志着算法收敛。此时, E-K-means 算法完成, 最终的簇划分为算法的输出结果。

$$\sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \cdot e^{-E_j} \quad (10)$$

3 仿真结果与分析

三维无线传感器网络的应用领域广泛, 为了验证本文所提出的算法的有效性, 利用 MATLAB 仿真软件模拟实际应用场景, 得出在特定网络场景下, 网络节点的分簇情况, 以合理的分簇达到减少网络能耗, 提升网络的生存时间和数据传输效率, 提升网络的性能的目的。

3.1 仿真环境

本文设置在 $100 \text{ m} \times 100 \text{ m} \times 100 \text{ m}$ 的三维网络场景下, 均匀随机布置 100 个传感器节点, 各节点初始能量为随机值 $E_0(1+\text{rand})$, E_0 为网络节点初始能量的最小值, rand 为 (0,1) 之间的随机数, 应用场景部分参数设置如表 1 所示。

表 1 仿真参数设置

参数	取值
场景范围 (长 × 宽 × 高)	$100 \text{ m} \times 100 \text{ m} \times 100 \text{ m}$
汇聚节点位置	(50, 50, 50)
节点个数 N	100
信道带宽	10 MHz
信号频率	2.4 GHz
E_0	0.5 J
E_{elec}	$50 \text{ nJ} \cdot \text{bit}^{-1}$
E_{DA}	$5 \text{ nJ} \cdot \text{bit}^{-1}$
ϵ_{fs}	$10 \text{ pJ} \cdot \text{bit}^{-1} \cdot \text{m}^2$
ϵ_{mp}	$0.0013 \text{ pJ} \cdot \text{bit}^{-1} \cdot \text{m}^4$
k	4 000 bits

3.2 仿真结果

网络仿真场景的空间范围为 $100 \text{ m} \times 100 \text{ m} \times 100 \text{ m}$, 网络节点的通信半径为 50 m, 本次仿真采用均匀随机撒点的方式。

图 1 为 E-K-means 分簇效果图, 利用改进的 E-K-means 分簇算法将三维无线传感器网络节点进行簇划分, 100 个网络节点被划分在 7 个网络簇中, 每个簇由一个簇头领导, 簇头节点负责收集簇内节点的数据并传输到 sink 节点。图 2 为关于 K-means 算法及其改进分簇算法 E-K-means 的分簇效果对比图, 横坐标为网络节点划分簇的个数 K , 纵坐标为轮廓系数, 轮廓系数是评估聚类性能的重要指标, 其取值范围为

$[-1, 1]$, 值越接近 1 表示聚类效果越好。

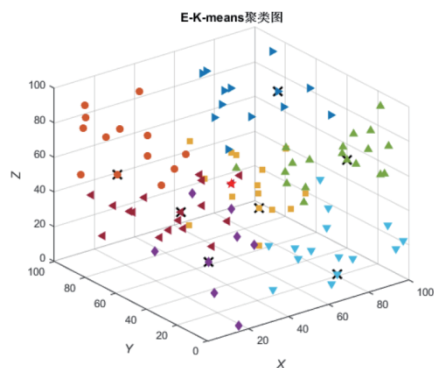


图 1 E-K-means 分簇效果图

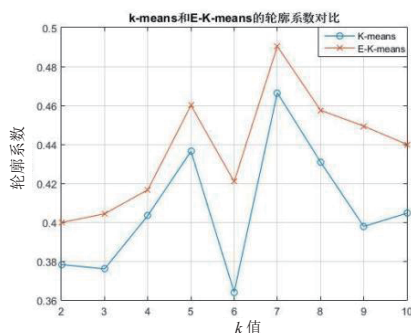


图 2 分簇效果对比图

从图 2 可以看出, 在不同 K 值下, E-K-means 算法的轮廓系数普遍高于传统的 K-means 算法, 表明 E-K-means 算法聚类效果较好。E-K-means 算法考虑节点之间的距离以及节点的能量, 提升了聚类的准确性和稳定性, 直观地证明了 E-K-means 算法在优化聚类效果方面的有效性。

图 3 为关于 K-means 算法及其改进分簇算法 E-K-means 算法的簇内网络生存时间对比图, 横坐标为网络划分的簇号, 纵坐标为簇内网络生存时间, 此网络生存时间统计簇内 20% 的网络节点能量耗尽时所持续的时间。从图中可以看出, 对于大部分簇而言, E-K-means 算法的簇内生存时间普遍高于传统的 K-means 算法, 表明 E-K-means 算法网络生存时间较长, 说明 E-K-means 分簇算法能够有效地组织节点以减少能量消耗并延长网络寿命。

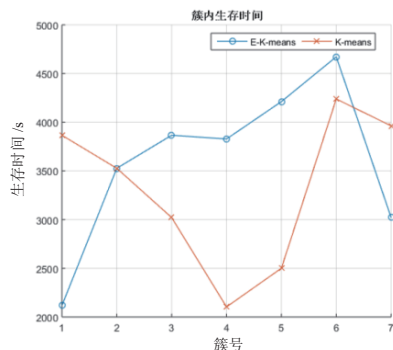


图 3 簇内网络生存时间对比图

4 结论

本文在传统 K-means 算法的基础上, 提出 E-K-means 分簇算法, 综合考虑能量与距离因素, 改进了网络节点的入簇机制, 将能量大且距离近节点加入簇头节点所在的簇中, 仿真结果表明, E-K-means 分簇算法的分簇效果更好, 网络生存时间更长。

参考文献:

- [1] 程海晟. 基于改进型 K 聚类算法的无线传感器网络拓扑控制研究 [D]. 上海: 东华大学, 2020.
- [2] CAI H, ZHANG Y, YAN H H, et al. A delay-aware wireless sensor network routing protocol for industrial applications[J]. Mobile networks and applications. 2016, 21:879-889.
- [3] 蒋鹏, 刘孝卿. 基于组合加权分簇的三维无线传感器网络覆盖控制方法 [J]. 计算机应用研究, 2011, 28(5):1842-1846.
- [4] 徐晶晶, 张欣慧, 许必宵, 等. 无线传感器网络分簇算法综述 [J]. 计算机科学, 2017, 44(2):31-37.
- [5] 何泉宇, 姚彦鑫. 无线传感器网络中分簇算法综述 [J]. 电脑知识与技术, 2019, 15(09):28-30.
- [6] 维明世, 何潘金. 无线传感器网络 LEACH 算法的改进与仿真分析 [J]. 无线通信技术, 2023, 32(4):38-42.
- [7] RAO P C S, JANA P K, BANKA H. A particle swarm optimization based energy efficient cluster head selection algorithm for wireless sensor networks[J]. Wireless networks, 2017, 23(7):2005-2020.
- [8] 杨彩霞. 基于无线传感器网络的集中式分簇算法研究 [D]. 兰州: 兰州交通大学, 2016.
- [9] KRISHNAMOORTHY S. Enhanced adaptive clustering mechanism for effective cluster formation in WSN[EB/OL]. (2017-10-25)[2025-05-12]. <https://www.semanticscholar.org/paper/Enhanced-Adaptive-Clustering-Mechanism-for-Cluster-Krishnamoorthy/7390ec07c69f78ca8f74f2cc7140a7b9f15cbb92>.
- [10] ARBEL E, HEL-OR H. Shadow removal using intensity surfaces and texture anchor points[J]. IEEE transactions on pattern analysis machine intelligence(T-PAMI), 2011, 33(6): 1202-1216.

【作者简介】

苗丽娟 (1993—), 女, 山西怀仁人, 硕士研究生, 讲师、实验室主任, 研究方向: 无线传感器网络功率控制。

(收稿日期: 2025-03-18 修回日期: 2025-07-31)