

考虑时空相关性的无线传感器网络数据分簇聚类方法

杨志杰¹
YANG Zhijie

摘要

由于无线传感器网络数据之间具有时空关联性,而现有方法往往仅考虑单方面,无法更全面地捕捉数据的特征,导致聚类的效果难以得到保障。为此,提出考虑时空相关性的无线传感器网络数据分簇聚类方法研究。利用势函数描述节点间的数据流动趋势和能量分布状态,确定 Sink 节点(数据收集点)和 Source 节点(数据产生点)的势函数值。基于此,根据与 Sink 和 Source 节点的相对位置及距离动态计算其他普通节点的势函数值,实现无线传感器网络数据场的构建。根据空间相关和时间相关构建数据挖掘机制,更全面地捕捉数据特征,并通过配置信息分发机制启动分簇过程,比较出发感兴趣事件节点中具有时空相关性节点的势函数值,完成聚类。在测试结果中,DB 指数始终稳定在 0.165 以下,DB 指数均值分别低于粒子群聚类方法 0.006 2,低于改进 PSO 聚类方法 0.009 2。

关键词

时空相关性;无线传感器网络;数据分簇聚类;势函数;数据场模型;配置信息分发机制

doi: 10.3969/j.issn.1672-9528.2024.11.021

0 引言

无线传感器网络(wireless sensor networks, WSN)是由大量分布在监测区域内的微型传感器节点组成的网络系统,这些节点通过无线通信协作感知、收集、处理和传输监测区域内的信息^[1]。WSN 在环境监测、健康护理、智能家居等领域有着广泛的应用。数据分簇聚类是 WSN 中一种重要的数据管理和处理技术,其核心目标是通过聚类,将具有相似特征的数据点归为一类,从而过滤掉噪声和异常值,提高数据的整体准确性^[2]。同时,考虑到在数据量庞大的 WSN 中,直接处理所有数据可能会导致计算资源不足,聚类也可以将数据分组,减少后续处理的数据量,从而降低计算复杂度,在更好地理解数据分布规律和变化趋势基础上,为网络优化提供依据,以提高网络的能量效率和可扩展性。文献[3]中根据蜂窝移动网络的结构、数据存储结构以及数据特征,对数据属性聚类。然而该方法未能充分获取更全面的数据特征,导致最终聚类结果不佳。文献[4]中以粒子群算法为基础进行无线传感网络大数据聚类优化。将传感器节点作为粒子,根据不同节点对应数据与整体数据特征之间的适应度,迭代优化具有相似特征的数据,并能够较快地收敛到较优的聚类结果,提高聚类效率。尽管 PSO 算法具有全局搜索能力,但在某些复杂情况下仍可能陷入局部最优解。文献[5]中以改进 PSO 为基础进行无线传感器网络数据自适应聚类,引入了

自适应机制,根据聚类过程中的实际情况动态调整参数,提升方法适应性。但该方法在实现过程中仅考虑数据单方面关联,导致最终聚类效果不佳。文献[6]中提出基于改进引力搜索机制 GSA 的数据聚类算法。设计一种基于汉明距离的引力搜索粒子距离度量方法,并引入加速因子到粒子速度更新策略中,利用最优粒子位置代表的聚类解加速局部开发过程,完成聚类。但该方法在实现过程中未能充分考虑数据的时空相关性,数据特征捕捉不全面,影响了最终的聚类效果。

在 WSN 中,每个传感器节点都位于特定的空间位置,这些节点所采集的数据往往与它们所处的环境特征紧密相关。通过对节点间空间位置关系的分析,可以揭示数据在空间上的分布规律,这对于理解环境特性、优化网络布局和提高数据准确性具有重要意义。其次,WSN 节点采集的数据通常包含时间戳,这使得数据具有时间序列特征。在时间上,同一节点或不同节点在不同时间点的数据往往呈现出一定的关联性。通过对时间相关性的分析,可以预测数据的变化趋势,为数据预测、异常检测和节能优化等提供有力支持。因此,基于该背景,为提高聚类效果,本文提出考虑时空相关性的无线传感器网络数据分簇聚类方法研究,并通过对比测试的方式,分析验证了设计方法的聚类效果。

1 无线传感器网络数据分簇聚类方法设计

对于无线传感器网络而言,受各节点之间交互关系的影响,对应的数据之间也存在一定的相互作用。为此,本文构建了无线传感器网络数据场^[7],以描述节点间的数据流动趋

1. 金肯职业技术学院实训与信息中心 江苏南京 211156

势和能量分布状态,应用势函数对这种相互作用关系进行细化表示^[8],为后续的分簇聚类提供基础。

1.1 无线传感器网络数据场构建

在具体的无线传感器网络数据场构建过程中,本文主要利用势函数定义网络中每个节点的数据吸引或排斥能力。具体而言,设定 Sink 节点(数据收集点)的势函数值为 0,表示最低势能点;Source 节点(数据产生点)的势函数值设为 1,表示最高势能点。其他普通节点的势函数值则根据它们与 Source 和 Sink 节点的相对位置及距离动态计算确定,这样就在整个网络中形成了一个数据势场^[9]。其中,普通节点的势函数值计算方式可以表示为:

$$\varphi^p(q) = \mu_p \cdot \exp\left(-\frac{d(p,q)^2}{2\delta^2}\right) \quad (1)$$

式中: $\varphi^p(q)$ 表示在 Sink 节点为 q 、Source 节点为 p 的无线传感器网络中,普通节点的势函数值; μ_p 表示 Source 节点的势函数值,即为“1”; $d(p,q)$ 表示普通节点位置与 Source 和 Sink 节点的相对距离; δ 表示普通节点与 Source 和 Sink 节点的作用强度,即数据交互程度。结合式(1)所示的计算方式可以看出,在本文设计的无线传感器网络数据场中,每个节点根据其势函数值决定其数据转发策略^[10]。节点会倾向于将数据转发给势函数值更低的邻居节点,即向势能更低的方向移动数据,直至最终到达 Sink 节点^[11]。

综上所述,结合势函数构建无线传感器网络数据场,描述节点间的数据流动趋势,为后续的分簇聚类提供基础。

1.2 基于时空相关性的无线传感网络数据特征挖掘机制构建

鉴于无线传感网络数据表现出的时空相关性,以及这种相关性蕴含的深层关联关系和上述数据场中势函数值所呈现的数据流动趋势,本文充分考虑了无线传感网络数据的空间相关和时间相关,以此实现对无线传感网络数据特征的充分挖掘,减少分簇聚类阶段冗余数据的处理。

针对无线传感网络数据的空间相关性,当无线传感网络中某区域内所有节点的感知数据均表现出一定的相似属性时,则定义该区域为相关区域^[12-13]。其中,具体相关区域的大小仅受感知数据相似范围内用户规模及事件的具体特征影响。

针对无线传感网络数据的时间相关性,本文定义在每个源节点中,均持有上一次感知并发送的数据信息。判定是否发送当前数据的方式可以表示为:

$$\left(\frac{|\varphi_n - \varphi_o|}{\varphi_o}\right) \times 100 > t_{ct} \quad (2)$$

式中: φ_n 表示当前节点势能, φ_o 表示上一次感知并发送的数据信息对应节点势能, t_{ct} 表示无线传感器网络的时间相关容

忍度。其中,节点势能均可通过式(1)计算获取。

按照式(2)所示的判定方式,当源节点的当前读数与上一次读数之间满足式(2)所示的关系时,则发送源节点的当前数据信息;当源节点的当前读数与上一次读数之间不满足式(2)所示的关系时,则抑制源节点的当前数据信息,不对当前感知的数据进行发送处理。

按照上述所示的方式,实现对无线传感网络数据时空特征的充分挖掘,为后续的分簇聚类提供基础。

1.3 无线传感器网络数据分簇聚类

在对无线传感器网络(WSN)中的数据进行分簇聚类处理时,本文从高效、节能的角度出发,提出了一种针对性的分簇策略。其中,策略的核心在于动态地组织网络节点,以应对特定事件监测任务,同时优化资源利用,特别是能量消耗。由此,基于上述数据特征挖掘结果,本文通过配置信息分发机制启动分簇过程。其中, Sink 节点作为中心,向整个无线传感器网络广播包含事件类型、自身位置及上一跳节点信息的配置包。同时,这些配置包被无线传感器网络中的每个节点接收,并存储于本地,为后续的分簇决策奠定基础。具体如图 1 所示。

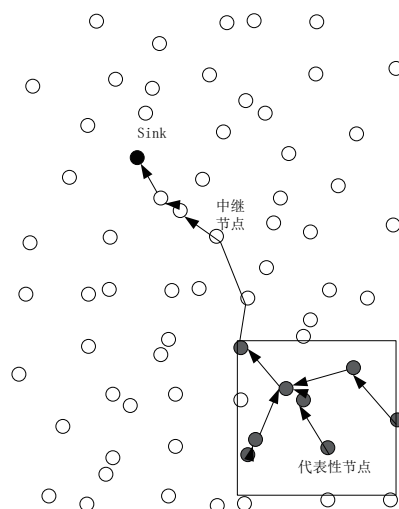


图 1 无线传感器网络数据分簇聚类

按照图 1 所示的方式,当无线传感器网络中的传感器节点探测到预设的感兴趣事件时,即会触发事件声明流程。在此过程中,节点不仅声明事件的发生,同时根据上述公式(1)评估自身的剩余势函数值,并基于挖掘到的数据特征,与接收到的声明消息中具有时空相关性发送节点势函数值进行比较。若某节点发现其剩余势函数值低于声明消息的发起者,则自动定位为成员节点,即从属角色,并继续转发事件声明,以此实现对相同事件感知节点的聚集。其中,结合以节点之间的空间相对位置的单个对象数据势值的表示方式,在对任意两个节点的感知属性相似度进行分析时,其可以表示为:

$$\nabla\varphi = \frac{|\varphi^p(q)_i - \varphi^p(q)_j|}{\eta_{ij}} \quad (3)$$

式中： $\nabla\varphi$ 表示任意两个节点的感知属性相似度， η_{ij} 为数据时空相关性特征。

无论是从代表性节点至簇头的数据传递，还是簇头向 Sink 节点的数据上报，均遵循最短路径原则。随着事件感知节点的汇聚，一个基于势函数值优化的分簇结构逐渐形成。具体如图 2 所示。

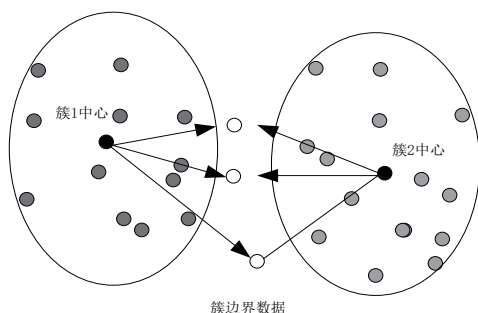


图 2 分簇结构

在此结构中，剩余势函数值最充裕的节点被选举为领导节点（或称簇头），负责整个簇的协调与管理。簇内其余节点则依据上述公式（3）的感知属性相似度结果划分为代表性节点和非代表性节点，前者直接参与数据传输，后者则通过代表性节点的数据间接共享信息，实现数据的有效压缩与估计，进一步节省能量。

按照上述所示的方式，实现对无线传感器网络数据的分簇聚类处理。

2 测试与分析

2.1 测试准备

在进行测试时，进行了无线传感器网络（WSN）环境的仿真处理，通过设定一系列关键参数以模拟实际网络环境。在具体的仿真阶段，对网络范围大小、Sink 节点坐标、节点初始能量、兴趣区域中心位置信息及半径进行具体设置后，以节点数量作为自变量的设置。

设定网络覆盖范围为 500.0 m×500.0 m 的正方形区域。Sink 节点作为数据收集的中心点，其坐标为 (250, 250)，即网络覆盖区域的中心点。利用 Sink 节点接收来自各个传感器节点的数据。在此基础上，对于每个传感器节点的初始能量，设置其为 0.5 J。以此确保网络在初始阶段能够稳定运行一段时间，同时也便于后续观察数据的聚类效果。对于兴趣区域的中心位置，设置其具体坐标为 (300, 350)，作为一个在网络范围内的特定点，利用其监测特定事件及网络环境参数。兴趣区域的半径设为 50.0 m，即围绕中心位置 (300, 350) 的圆形区域。该半径参数决定了哪些传感器节点将直接参与对兴趣区域的监测。具体如图 3 所示。

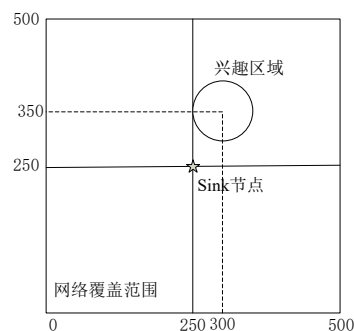


图 3 测试无线传感器网络环境

为了更加全面地分析分簇聚类方法的应用效果，设置节点数量作为自变量参数，从 100 个节点开始，逐步增加到 500 个节点，增加梯度为 100 个节点。具体的阶段按照随机分布的方式布局在 500.0 m×500.0 m 的正方形无线传输区域内。通过改变节点数量，观察不同密度下的聚类效果。

按照对无线传感器网络的设置，对不同条件下的聚类性能进行全面评估。

2.2 测试方案

具体的测试阶段，分别采用以粒子群算法为基础的无线传感网络大数据聚类优化方法，以及以改进 PSO 为基础的无线传感器网络数据自适应聚类算法作为测试的对照组。在评估不同方法的聚类性能时，本文考虑到理想的聚类结果通常展现出簇内紧密度高而簇间分离度大的特性，即同一簇内的数据点彼此高度相似，而不同簇间的数据点则呈现低相似性。为此，本文将 DB 指数（davies-bouldin index）作为衡量聚类效果的关键指标。具体通过计算簇内平均距离与簇间最近距离之比的总和的平均值量化聚类特性。其计算方式可以表示为：

$$DB = \frac{1}{K} \sum \max \left[\frac{S(c_k) + S(c_j)}{d(c_k, c_j)} \right] \quad (4)$$

式中： K 表示聚类的类别数量， $S(c_k)$ 表示聚类簇内数据之间的平均距离， $S(c_j)$ 表示聚类簇内数据到该聚类簇中心的平均距离， $d(c_k, c_j)$ 表示不同分簇聚类之间的距离。

结合式（4）所示的计算方式，DB 指数越低，意味着对应方法的聚类效果越好，即反映了聚类簇内高凝聚性和簇间高分离性的理想状态。本文利用这种评估方法，客观比较不同聚类方案的有效性。

2.3 测试结果与分析

2.3.1 有效性分析

在进行实现聚类的过程中，其数据特征挖掘的全面性则直接影响后续聚类的效果。因此，为验证所提方法的有效性，先针对数据特征提取的完整度对所提方法进行分析。利用所提方法对 10 组数据样本进行挖掘，每组包含 100 个数据，其结果如图 4 所示。

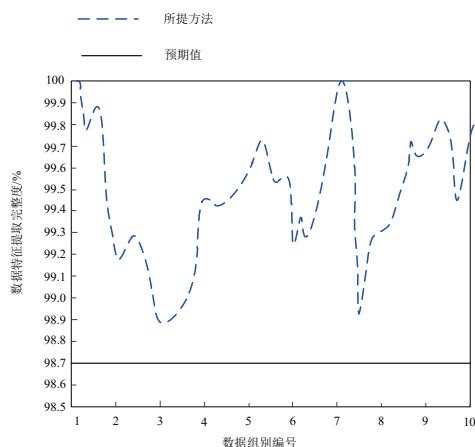


图4 数据特征提取完整度结果

根据图4可知,采用所提方法进行数据特征挖掘,其可充分考虑数据的时空相关性,其数据特征挖掘的完整度较高,最低在98.9%。针对各数据样本组别的特征挖掘完整度均在预期以上。由此基于上述分析可说明,所提方法具有有效性,可为后续聚类提供可靠的数据特征支持。

2.3.2 对比分析

针对DB指数,对不同方法的测试结果进行统计,得到的数据结果如表1所示。

表1 不同方法测试结果对比表

节点数量 / 个	粒子群聚类方法	改进 PSO 聚类方法	本文设计聚类方法
100	0.168 4	0.162 2	0.162 4
200	0.169 1	0.168 9	0.163 3
300	0.169 9	0.172 0	0.164 5
400	0.170 7	0.177 4	0.164 6
500	0.171 2	0.183 9	0.163 7
平均值	0.169 9	0.172 9	0.163 7

结合表1所示的测试结果对三种不同聚类方法的性能进行分析,其中,在粒子群聚类方法下,DB指数整体呈现出随着节点数量的增加逐渐增大的发展趋势,虽然具体的增幅相对较小,但是具体的DB指数始终在0.168 0~0.172 0区间,处于相对偏高的状态。在改进 PSO 聚类方法下,DB指数整体随着节点数量的增加逐渐增大的发展趋势更加明显,涨幅与粒子群聚类方法相比明显增大,当节点数量不超过200个时,DB指数在0.17以下;当节点数量达到200个以上时,DB指数最大值达到了0.183 9(节点数量为500个时)。在本文设计聚类方法下,DB指数与节点数量之间并未表现出明显的相关性,始终稳定在0.165以下。从整体角度分析,本文设计聚类方法的DB指数均值分别低于粒子群聚类方法0.006 2,低于改进 PSO 聚类方法0.009 2。

由此可以得出结论,本文设计的无线传感器网络数据分簇聚类方法可以适应不同无线传感器网络场景,实现对数据的有效分簇聚类。

3 结语

在无线传感器网络(WSN)中,数据之间的相关性,特别是从时空角度分析,具有深刻的含义。这种相关性主要体现在节点之间的空间位置关系和时间序列数据的变化规律方面,是反馈无线传感器网络数据特性的重要参数之一。本文提出考虑时空相关性的无线传感器网络数据分簇聚类方法研究,切实实现了对无线传感器网络数据的有效聚类,大大提高了聚类效果。借助本文的研究与设计,也希望能够为实际的无线传感器网络数据分析和管理工作提供参考。

参考文献:

- [1] 陈功平,王红.基于大数据聚类的通信网络安全态势预测技术[J].淮阴师范学院学报(自然科学版),2024,23(1):20-26.
- [2] 黄俊萍.基于改进K-means数据聚类算法的网络入侵检测[J].成都工业学院学报,2024,27(2):58-62+97.
- [3] 李红艳,徐寅森,张子栋.蜂窝移动网络大数据聚类异常挖掘方法仿真[J].计算机仿真,2024,41(2):406-409+414.
- [4] 程宁,李超.基于粒子群算法的无线传感网络大数据聚类优化方法[J].传感技术学报,2023,36(8):1316-1322.
- [5] 原大明.基于改进PSO的无线传感器网络数据自适应聚类算法[J].现代电子技术,2023,46(11):99-102.
- [6] 张小庆.基于改进GSA的数据聚类机制[J].计算机应用与软件,2021,38(2):27-32+84.
- [7] 郑真真,李婧彬.基于数学模型的高效通信网络入侵检测方法[J].长江信息通信,2024,37(2):117-119.
- [8] 胡红彬,李忠,刘建华,等.基于数据聚类的无线网络通信数据加密系统[J].电子设计工程,2024,32(3):125-128+133.
- [9] 胡艳华,张春玉,崔亚楠,等.云环境下改进自回归模型的网络数据去重仿真[J].计算机仿真,2024,41(1):443-446+536.
- [10] 李叶飞,马昊燕,荆树君,等.基于数据聚类的共享电源无线网络通信数据加密系统[J].电子设计工程,2024,32(1):19-23.
- [11] 潘玉恒,赵豪,鲁维佳,等.一种基于测量数据相似性的WSN分簇路由方法[J].传感技术学报,2023,36(3):481-489.
- [12] 陈勇,郭云柱,王威,等.模糊K-Harmonic-Kohonen网络的FTIR光谱数据聚类分析[J].光谱学与光谱分析,2023,43(1):268-272.
- [13] 葛维静,冯圆圆,刘宗洋,等.一种基于网络流量风险数据聚类的APT攻击溯源方法[J].通信技术,2022,55(10):1354-1362.

【作者简介】

杨志杰(1982—),男,江苏南京人,本科,讲师,研究方向:信息安全、计算机通信。

(收稿日期:2024-07-23)