基于数据场的无线传感网数据双重聚类算法

石美红,赵 辉,贾郑磊,雷 燕,张祥俊

(西安工程大学 计算机科学学院,陕西 西安 710048)

摘 要:为了解决大规模高维 WSN 数据的双重聚类问题,提出一种基于数据场的 WSN 数据聚类方法.借用数据场的思想,将 WSN 数据空间非线性映射到势能空间;结合 WSN 数据的分布特点,采用概率熵度量数据的质量;根据数据场形成的等势线分布特征,采用极小判定法得到数据聚类结果.通过人工合成数据集实验测试,结果表明,对于随机分布的大规模数据集的聚类效果和聚类精确度,此算法优于 ICC 和 DFCM 的双重聚类算法,且具有较低的时间复杂度和良好的可扩展性.

关键词:无线传感器网;大规模高维数据;双重聚类;数据场;势能

An Dual Clustering Algorithm for Data in Wireless Sensor Networks Based on Data-field Model

SHI Mei-hong, ZHAO Hui, JIA Zheng-lei, LEI Yan, ZHANG Xiang-jun

(School of Computer Science, Xi'an Polytechnic University, Xi'an 710048, China)

Abstract: The paper has proposed a novel self-organizing-mapping algorithm based on data-field model for dual clustering of large-scale and high dimensional data in Wireless Sensor Networks (WSNs) in order to has good performance in cluster problems. The method maps the WSNs data from data space to the appropriate potential space in data field, which measures the interactions of the elements in large-scale and high dimensional data by taking probabilistic entropy of data distribution in the WSNs as the mass of data field, thus generating a two-dimension data field. Then, by employing distribution features of the potential center and the equipotential lines, without significantly increasing the time complexity, the good clustering result is obtained by minimum potential difference determination method. The comparing experiments on the synthetic datasets demonstrate the effectiveness of the algorithm. Experimental results show that the proposed method improves the clustering effect and has exact clustering result compared with other dual clustering algorithm, i. e. ICC and DFCM, and it has good scalability.

Key words: wireless sensor networks; large-scale and high dimensional data; dual clustering; data-filed; potential

作者简介:

石美红 女,(1956-),教授.研究方向为智能信息处理、图形图像处理与模式识别等.E-mail: meihong_shi@163.com.

赵 辉 男,(1990-),硕士研究生.研究方向为智能信息处理.

贾郑磊 男,(1993-).研究方向为智能信息处理.

雷 燕 女,(1994-).研究方向为智能信息处理.

张祥俊 男,(1993-).研究方向为智能信息处理.