

引入时频聚集交叉项干扰抑制的大数据聚类算法

胡先兵¹ 赵国庆²

(国家计算机网络应急技术处理协调中心 北京 100029)¹ (北京交通大学电子信息工程学院 北京 100044)²

摘要 面向大数据集管理的数据聚类方法研究在模式识别、故障诊断和数据挖掘等领域具有重要的研究意义。传统的大数据聚类算法采用混合差分进化的粒子群算法,因数据信息流分量之间的交叉作用而出现的类间交叉项干扰影响了聚类分量的正确判断,聚类效果不好。提出了一种基于时频聚集交叉项干扰抑制的大数据聚类算法。在面向传播学视域下物联网大数据库中生成大数据聚类信息特征向量,对任意两个分簇矢量进行近邻样本的隶属度训练,在时间滑动窗口模型中进行信息调度,采用高频分量抑制方法实现时频聚集交叉项的干扰抑制,通过频域卷积相似性融合处理,采用粒子群优化算法进行聚类适应度计算,以实现大数据聚类算法改进。仿真结果表明,采用该算法进行大数据聚类,具有较好的抗干扰性和自适应性,聚类准确度较高。

关键词 大数据,聚类,粒子群,干扰抑制

中图分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2016.4.040

Large Data Clustering Algorithm Introducing Time and Frequency Clustering Interference Suppression

HU Xian-bing¹ ZHAO Guo-qing²

(National Computer Network Emergency Coordination Center, Beijing 100029, China)¹

(School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China)²

Abstract The data clustering method of large data sets has important research significance in pattern recognition, fault diagnosis, data mining and so on. Traditional data clustering algorithms use hybrid differential evolution particle swarm optimization algorithm, but because the cross term interference caused by interactions between components of data flow has an impact on the correct judgement, the clustering effect is not good. A large data clustering algorithm was proposed based on the time-frequency clustering algorithm. In the large database of Internet of things from the perspective of communication, big data clustering feature vector is generated. Arbitrary two cluster vectors are trained of membership grade about neighbor sample, and information is scheduled in time sliding window model. High frequency component suppression method is used to conduct interference suppression of time-frequency clustering interactions. By similarity fusion processing frequency domain convolution and using particle swarm optimization algorithm to do cluster fitness calculation, we improved the data clustering algorithm. The simulation results show that the algorithm has good anti-interference and self-adaptive performance, and it has high accuracy.

Keywords Large data, Clustering, Particle swarm, Interference suppression

1 引言

近些年,随着计算机与网络技术的不断发展,服务器中数据库数据越来越多,面对这种大数据网络时代,与数据分析相关的数据挖掘和数据聚类技术成了大数据管理的重要研究工作。在大数据管理技术的研究中,首先需要对数据进行分析,这是计算机应用的基础。数据分析主要分为两方面:探索性分析与验证性分析,而探索性分析最常用的方法为聚类分析技术。聚类算法是通过寻找相同类别、相同典型数据,来简化数据集分析的复杂度,提高运行速度。聚类分析是大数据管理的重要过程,它既可以单独运作,也可在数据库中进行处理。因此,面向大数据集管理的数据聚类方法研究在模式识

别、故障诊断和数据挖掘等领域具有重要的研究价值和意义,相关算法研究受到人们的重视^[1]。

数据聚类是对大量无规律的数据进行静态特征分析的一门技术,通过一定算法和手段对采集的海量大数据进行特征量提取,构建分类器以实现数据的分类输出和管理,典型的数据聚类算法主要有 K-means 算法、模糊 c 均值(Fuzzy c-means, FCM)聚类算法、层次聚类算法、分割聚类算法、粒子群算法和智能仿生群聚类算法等^[2],相关算法的研究取得了一定的成果。其中文献[3]提出一种基于混沌差分优化聚类的机械故障特征数据聚类算法,其利用 Logistics 混沌映射进行数据特征分类,把混沌扰动量引入到进化种群中,提高了故障特征的聚类性能,但该算法存在计算复杂、特征空间维数较

到稿日期:2015-08-28 返修日期:2015-10-30

胡先兵(1981—),男,工程师,主要研究方向为计算机应用、网络安全, E-mail: anyphu@163.com; 赵国庆(1967—),男,博士,副教授,主要研究方向为计算机应用、网络安全。

高的问题;文献[4]提出一种基于模糊c均值融合法的数据聚类算法,该算法采用相似度信息参量自适应地调整聚类中心,以适应动态的数据集变化的聚类需求,但该算法对初始值具有较强的敏感性,容易陷入局部极值点收敛;文献[5]提出一种基于网格模糊划分的大数据聚类算法,反复调整聚类的网格中心点进行聚类优化,但该算法对初始聚类中心的依赖性较大,不容易得到全局最优解;文献[6]提出一种基于K均值聚类和粒子滤波的Web数据聚类算法,对群体中的个体进行最优解划分,但该算法中由于海量数据存在大量干扰选项、对噪声数据敏感,影响了聚类的准确性;文献[7]将蛙跳群算法引入到数据聚类中,提出一种基于模糊因组融合的混合蛙跳群算法进行数据聚类,实现对海量多模态数据的优化聚类,但该算法的缺点是寻优进化过程中对模糊边缘的局部最优解搜索性能相对较差,随着迭代次数的增加容易陷入局部收敛。此外,文献[8]对聚类算法进行改进,采用混合差分进化的粒子群算法,由于数据信息流分量之间的交叉作用出现类间交叉项干扰,影响了聚类分量的正确判断,聚类效果不好。

针对上述问题,本文提出一种基于时频聚集交叉项干扰抑制的大数据聚类算法。首先进行了基础知识描述和相关定义介绍,生成大数据聚类的信息特征向量,然后进行时频聚集交叉项干扰抑制算法设计,最后实现数据聚类算法改进。通过仿真实验进行了性能验证,展示了本文算法在提高大数据聚类性能方面的优越性,展示了较好的应用前景。

2 基础知识描述及大数据聚类的信息特征向量生成

2.1 大数据聚类算法思想及基础知识

本文研究大数据聚类的目标是面向传播学视域下物联网大数据数据库,提高大数据数据库的优化管理和数据调度能力。因此需要进行大数据优化聚类算法设计,在数据库中选择一个小样本^[9],对面向传播学视域下物联网大数据数据库中的海量数据进行数据库模块化处理,设大数据数据库的执行交叉项概念格结点 $G_1 = (M^1, Y_1)$, $G_2 = (M^2, Y_2)$, 则 $G_1 \subseteq G_2 \Leftrightarrow Y_1 \subseteq Y_2$, 令 $A = \{a_1, a_2, \dots, a_n\}$ 为面向传播学视域下物联网大数据信息流预处理特征序列训练集的属性集, $B = \{b_1, b_2, \dots, b_m\}$ 为查询属性类别集, a_i 的属性值为 $\{c_1, c_2, \dots, c_k\}$ 。现在把有限数据集 X 分为 c 类,将训练集与它们所属的类别进行关联,得到第 i 个节点的多层自回归空间并行特征搜索矩阵为:

$$s(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} a_{mn} g_{mn}(t) \quad (1)$$

其中, a_{mn} 称作 Gabor 展开系数,而聚类粒度为:

$$g_{mn}(t) = g(t - mT) e^{j2\pi(nF)t}, m, n = 0, \pm 1, \pm 2, \dots \quad (2)$$

其中, $\gamma_{mn}(t)$ 是 Gabor 基函数的对偶函数,对数据属性集合进行时频对偶重排,假设有有限数据集:

$$X = \{x_1, x_2, \dots, x_n\} \subset R^c \quad (3)$$

设任意两个聚类簇 M_i 与 M_j 的中心距离为 $Clustdist(M_i, M_j)$, 其中, $i \neq j, 1 \leq i \leq q, 1 \leq j \leq q$, 大数据聚类类别总数为 m , 定义每个分支节点中的数据聚类中心矩阵为:

$$V = \{v_{ij} | i = 1, 2, \dots, c, j = 1, 2, \dots, s\} \quad (4)$$

定义 1(大数据聚类时间窗口快照) 设 t 时刻的样本的类别时间窗口跨度为 p , 则在数据采样区间 $[t - p, p]$ 内类别判断数据流 DB_i 为时间窗口 B 的中心邻域, 记为 $b_i(size) = |b_i|, b_i(span) = p$ 。

定义 2(数据聚类时频聚集交叉项) 若将数据流在聚类过程中的时频平面任意点 (t, f) 的值记为 SB , 则令聚类特征时序 i 时的邻近点 $(t - t', f - f')$ 上的加权为 $SB_i = \{b_{i-w+1}, b_{i-w+2}, \dots, b_i\}$ 。原时频坐标 (t, f) 的函数为时频聚类交叉项。若聚类窗口基本数为 n , 则数据信号分量扩散的时间跨度为 $sb_i(span) = n * p$ 。

根据上述面向传播学视域下物联网大数据聚类基本思想的描述, 结合上述基本定义, 可见, 大数据聚类的首要要素是进行特征提取, 通过数据预处理和特征训练、认证、测试和评估, 筛选出有用的数据特征, 采用适应性计算方法(本文采用粒子群优化算法)实现数据聚类^[6], 面向传播学视域下物联网大数据数据库中的数据优化聚类流程如图 1 所示。

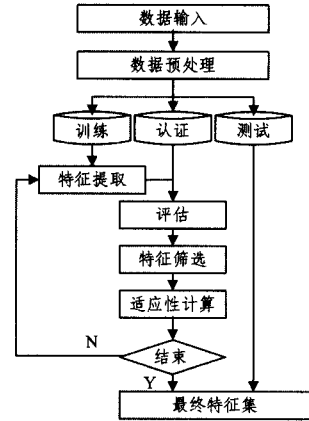


图1 数据优化聚类流程

2.2 大数据聚类的信息特征向量生成

在上述基本定义的基础上, 根据图 1 可知, 数据聚类的关键是进行特征提取和特征筛选。在数据流微簇中进行特征分析, 给出数据流微簇的定义, 基于滑动窗口的数据流聚类方法, 设传播学视域下物联网大数据在时频聚类分布空间的微簇为 CF , 记为 $CF = \langle F, Q, n, RT_1, RT_2, RW \rangle$, 数据流滑动窗口总数为 n , 为离群点新建一个簇, 定义 3 种离群点相似度系数为 (RT_1, RT_2) , 当前数据聚类的时间窗口大小为 RW , 则 $F_j = \sum_{k=1}^n X_{kj}, Q_j = \sum_{k=1}^n (X_{kj})^2$ 。定义欧氏距离 $dis(w_1, w_2)$ 为连接两个数据分簇聚类中心的最短路径上 n 条边的权值之和, 即:

$$dis(w_1, w_2) = \sum_{i=1}^n weight(e_i) \quad (5)$$

对频率求平均, 设时刻 t 时的时序窗口平均测度 ϵ 满足 $2^{-\lambda} < \epsilon, \lambda > 0$, 则数据流的瞬时谱的平均频率 $W(t)$ 为:

$$W(t) = v * \sum_{i=0}^{t-t_c} 2^{-\lambda} = \frac{v}{1-2^{-\lambda}} \quad (6)$$

将数据流微簇理解为一组解析信号, 定义数据类群适应度值为簇中心到簇内点的平均瞬时频率时间, 即 $\langle f \rangle = \langle f_i \rangle$, 其中:

$$\langle f \rangle = \frac{\int_{-\infty}^{\infty} f |Z(f)|^2 df}{\int_{-\infty}^{\infty} |Z(f)|^2 df}, \langle f_i \rangle = \frac{\int_{-\infty}^{\infty} f_i |z(t)|^2 dt}{\int_{-\infty}^{\infty} |z(t)|^2 dt} \quad (7)$$

进一步将数据导入调度控制模块中, 由于大数据聚类的信息特征向量的相似度与概念深度之间存在正比的关系, 数据流微簇的时域特征 $z(t)$ 与瞬时特征相对应, 频域特征 $Z(f)$ 的群延迟 $\tau_g(f)$ 为一个瞬时聚类特征参数, 它表示频谱 $Z(f)$

中频率为 f 的各个聚类簇的类间分类, 定义为:

$$\tau_g(f) = -\frac{1}{2\pi df} \arg[Z(f)] \quad (8)$$

对于每个待识别的样本 Y , V_i 为聚类中心的第 i 个矢量, 在时间滑动窗口模型中对大数据聚类信息特征向量进行信息调度, 在滑动窗口内数据特征划分约束函数表示为:

$$U = \{\mu_{ik} | i=1, 2, \dots, c, k=1, 2, \dots, n\} \quad (9)$$

其中, i 为当前样本总数, k 为近邻样本总数, 数据流为 SB , 引入一种数据流聚类相似性函数, 得到聚类目标函数为:

$$J_m(U, V) = \sum_{k=1}^n \sum_{i=1}^c \mu_{ik}^m (d_{ik})^2 \quad (10)$$

第 i 个聚类中心矢量的聚类权重系数 $l(j)$ 为:

$$l(j) = \frac{d_i^{-\alpha}}{\sum_{j=1}^k d_j^{-\alpha}} \quad (11)$$

为了提高数据分类精度, 对任意两个分簇矢量 X, Y 进行近邻样本的隶属度训练, 近邻样本隶属于类别 ω_i 的类判别函数 $G(Y_j, \omega_i)$ 为:

$$G(Y_j, \omega_i) = [p_1 * \sum_{j=1}^k l(x) \frac{\sum_{i=1}^c N(V_i)}{k}] + p_2 * N + p_3 * \frac{\sum_{i=1}^n (V_i)}{N_j} \quad (12)$$

通过定义可知, $\{p_1, p_2, p_3\}$ 为相似度特征向量的识别参数, $p_1 + p_2 + p_3 = 1$ 。通过上述处理生成大数据聚类信息特征向量, 以此为基础进行特征训练、认证、测试和评估, 为数据优化聚类提供准确的数据基础。

3 改进的大数据聚类算法实现

3.1 时频聚集交叉项干扰抑制算法

在上述进行了大数据聚类信息特征向量生产和特征提取的基础上, 采用时频分析方法构建数据流微簇时频聚集分布空间, 在时间滑动窗口模型中对大数据进行时频聚集分析, 以达到数据聚类的目的。但分析上述特征提取过程可见, 在两个数据分簇聚类中心出现类间交叉项干扰, 影响了聚类分量的正确判断, 聚类效果不好。为了克服传统方法的缺陷, 本文提出一种基于时频聚集交叉项干扰抑制的大数据聚类算法。进行时频聚集交叉项干扰滤波算法设计, 通过窗函数宽度自适应调整数据流微簇滑动窗口模型中的聚类特征中心, 进行自适应调度, 得到的聚类输出目标端信息分量为:

$$r_1 = x(t) - c_1 \quad (13)$$

自适应调整的滑动时间窗口权重为:

$$w_{ij} = \beta \times w(e_p k_q), \beta > 1 \quad (14)$$

在数据聚类过程中, 多分量大数据聚类信息特征中出现不同时频强度的交叉作用, 产生干扰, 其中干扰项为:

$$n(t) = F_a \cdot X(t) \quad (15)$$

式中:

$$X = [X_a(0), X_a(1), \dots, X_a(N-1)]^T \quad (16)$$

其中, F_a 是 $N \times N$ 维矩阵, 设计一原始实值时间函数 $h(t)$, 满足:

$$\int_{-\infty}^{\infty} h(t) dt = 1, h(t) = h(-t), h(t) = 0, |t| > 1/2 \quad (17)$$

上式说明, 数据聚类分簇矢量的时间函数是一个定义在区间 $[-1/2, 1/2]$ 内的时限信号, 为一个对称时间函数, 可以采用高频分量抑制方法实现对时频聚集交叉项干扰的抑制,

高频分量为:

$$\phi(\tau, v) = H(\tau v) \quad (18)$$

在 $H(v)$ 中用 τv 代替 v , 求 $h(t)$ 的 Fourier 变换得到:

$$H(v) = \int_{-\infty}^{\infty} h(t) e^{-j2\pi vt} dt \quad (19)$$

在 Fourier 时频聚集空间中, 需要更多的高频分量来处理多分量信号的频域干扰, 设 $D = \{T_1, T_2, \dots, T_n\}$ 是一个时间延迟为 τ 的加窗函数大数据集, 该大数据集有 n 个区间特征项集, 并包含 m 个不同的项, 即 $I = \{i_1, i_2, \dots, i_m\}$ 。通常情况下, 多普勒频移带类间距离为 r , 采用频域可变方法对频率分辨率进行调整, 只对 τ 加窗函数 $h(\tau)$ 以达到减小交叉项的目的。此时, 时频聚集时延 $\Delta t = r/c$ 。通过频域卷积相似度融合处理, 面向传播学视域下物联网大数据一共聚成 q 类, 通过时频聚集交叉项干扰抑制提高数据聚类的精度, 时频聚集交叉项干扰抑制聚类算法流程如图 2 所示。

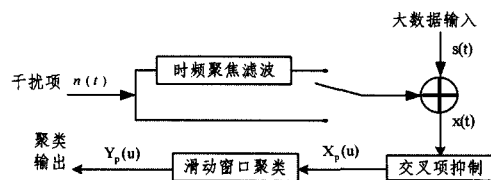


图2 时频聚集交叉项干扰抑制聚类算法流程

3.2 聚类算法实现

在上述进行交叉项干扰抑制处理的基础上, 通过特征提取结果, 采用粒子群优化算法进行聚类适应性计算, 实现数据聚类, 算法实现描述如下。

对每一批到达的数据利用时频聚集交叉项干扰抑制进行滤波处理, 创建多个线程的信息流特征, 通过对变量 u 和 τ 加窗函数 $h(\frac{u}{\tau})$ 实现交叉项抑制, 若设 B_i 到达时 $B_{i-j} \in A[j]$, $1 \leq j \leq n+1, i-j > 0$, 针对大多数数据集属性的度量标准, 找到粒子群的 k 个聚类中心 $\{a_1, a_2, \dots, a_k\}$, 通过先设定好密度参数 (MP 和 ϵ), 计算区域 D 内的任意一个数据点, 设微簇表示为 num , 初始时令 $num=0$, 对第 d 维空间中的粒子群按照聚类中心的窗口函数密度来划分成 k 类 $\{C_1, C_2, \dots, C_k\}$, 划分原则按照粒子种群的全局最优位置集合 $P_g = \{p_{g1}, p_{g2}, \dots, p_{gd}\}$ 进行位置更新, 全局寻优时, 在孤立核心簇中, 若 $w \geq \beta \mu$, 则大数据聚类的微簇特征向量为:

$$(\overline{CF}^2 * 2^{-\lambda \Delta t}, \overline{CF} * 2^{-\lambda \Delta t}, n * 2^{-\lambda \Delta t}, t_l, \tau) \quad (20)$$

计算新的聚类中心, 按照, $c_j' = \frac{1}{|C_j|} \sum_{X \in C_j} X (j=1, 2, \dots, k)$, 对粒子的全局寻优能力进行调节, 得到聚类中心矩阵各元素为:

$$F_a(m, n) = A_a e^{(j/2) \cos \alpha \cdot m^2 \Delta t^2} \cdot e^{(j/2) \cos \alpha \cdot n^2 \Delta t^2} \cdot e^{-j \frac{\sin(\alpha)}{N} \cdot 2\pi \cdot m \cdot n} \quad (21)$$

计算样本集中余下各点与各聚类中心的时间间隔, 执行数据聚类的自适应计算粒子更新机制为:

$$v_{id}^{(t+1)} = w * v_{id}^t + c_1 * r_1 (p_{id}^t - x_{id}^t) + c_2 * r_2 (p_{gd}^t - x_{id}^t) \quad (22)$$

其映射策略为线性递减时, 进一步进行时频聚集交叉项干扰抑制, 随机地选择 3 个不同的个体 $\{x_{r1}, x_{r2}, x_{r3}\}$, 以 x_{r1} 为基向量, 以 $\{x_{r2}, x_{r3}\}$ 为差分向量, 通过降低惯性系数, 提高聚类的收敛性, 得到聚敛特征函数:

$$z_{(i,d)} = x_{r1} + F * (x_{r2} - x_{r3}) \quad (23)$$

其中 F 为缩放因子。根据个体的自适应度值比较,输出最终的聚类集合,即:

$$\begin{cases} v_{id}^{(t+1)} = w * v_{id}^t + c_1 * r_1 (p_{id}^t - x_{id}^t) + c_2 * r_2 (p_{id}^t - x_{id}^t) \\ x_{id}^{(t+1)} = x_{id}^t + v_{id}^{(t+1)} \end{cases} \quad (24)$$

其中微簇数据的权重阈值 $u_i^{(k+1)} = \{u_{i1}^{(k+1)}, u_{i2}^{(k+1)}, \dots, u_{id}^{(k+1)}\}$, 粒子群聚类的种群 d 维分量 $x_i^k = \{x_{i1}^k, x_{i2}^k, \dots, x_{id}^k\}$, m_i 是随机整数, C_R 为平均粒距,选取最优值保存为新的种群解,实现数据聚类。

4 仿真实验与结果分析

为了测试本文算法在实现面向传播学视域下物联网大数据数据库的大数据优化聚类时的性能,进行仿真实验。仿真实验的硬件环境配置为:Windows 7 系统的 PC 机,2.89GHz 双核 Core 四处理器,1GB 内存。仿真软件为 Matlab 7,首先读取物联网大数据库内存存储的大数据,数据集来自物理网大数据库 B400C20D40,数据集采样共包含 8766539 条记录。数据集的个体适应度特征满足高斯均匀分布,在簇半径内进行时间窗口模型构建,采用微簇随机分类识别方法进行聚类参量初始化设置。实验初始的参数为 $InitN=1000$,个体的适应度值为 $v=1000$,离线阶段聚类的时间滑动窗口调控因子 $\lambda=0.25$,微簇半径设为 $R=16$,粒距聚集度设为 $\mu=10$,聚类中心的测量误差系数设为 $\beta=0.001$,类间距的调整权重 $(\alpha, \beta) = \{1, 3, 2, 6\}$,并令两个子种群的采样时间间隔 $T=5$,候选解的更新周期 $T_d=9s$,时频聚集交叉项干扰强度 $n=5dB$,平均适应度值阈值为 $[0.6, 0.8]$,加速分布因子为 2。

根据上述仿真环境和参量设计,进行大数据聚类仿真和性能分析,在数据库中选择一个小样本,对样本进行分析采集,得到物联网大数据库中的海量数据模块化随机分布结果,如图 3 所示。

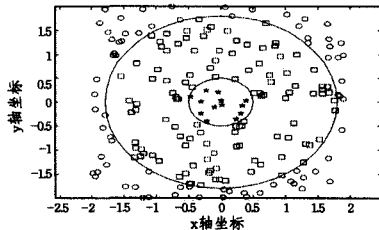


图3 大数据库中的海量数据模块化随机分布结果

从图 3 可见,原始的大数据分布具有随机性,如果不通过特征信息提取和数据聚类提取,将无法有效进行分类识别和数据分析。采用本文方法对大数据进行特征提取,使得聚类特征在时频空间中聚焦,得到时频特征聚焦结果,如图 4 所示。

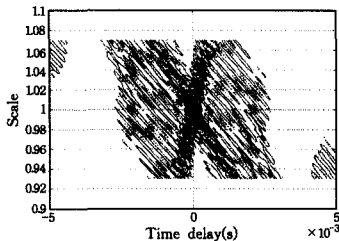


图4 大数据聚类时频特征聚焦结果

从图 4 可见,原始的大数据聚类时频特征聚焦过程中,由于数据信息流分量之间的交叉作用出现类间交叉项干扰,影响了聚类分量的正确判断,数据聚类性能不好。采用本文方

法,引入时频聚集交叉项干扰抑制算法,在 Fourier 时频聚集空间中,通过频域卷积相似度融合处理,得到干扰抑制后输出的特征聚焦结果如图 5 所示。

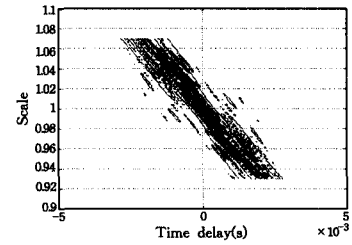


图5 时频聚集交叉项干扰抑制输出

从图 5 可见,采用本文算法能有效抑制大数据聚类的交叉项类间干扰,提高了聚类性能,由此实现数据聚类。为了对比算法性能,采用本文算法和传统算法,通过 10000 次蒙特卡洛实验,分析大数据在不同数据聚类长度下的聚类性能,得到数据聚类准确度对比结果如图 6 所示。从图 6 可见,采用本文算法能有效提高数据聚类的准确度,展示了较好的性能。

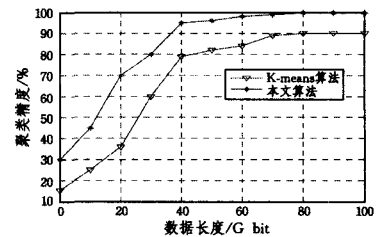


图6 聚类性能对比

结束语 数据聚类是对大量无规律的数据进行静态特征分析,对采集的海量大数据进行特征量提取,构建分类器实现对数据的分类输出和管理,数据聚类是模式智能识别和故障诊断等应用技术的基础,本文提出一种基于时频聚集交叉项干扰抑制的大数据聚类算法。在时间滑动窗口模型中对大数据聚类的信息特征向量进行信息调度,采用高频分量抑制方法实现时频聚集交叉项干扰抑制,在 Fourier 时频聚集空间中,通过频域卷积相似度融合处理,提高数据聚类的精度。研究表明,采用本文算法进行大数据聚类,具有较好的抗干扰能力,聚类准确度较高。

参考文献

- [1] Wen Tian-zhu, Xu Ai-qiang, Cheng Gong. Multi-fault diagnosis method based on improved ENN2 clustering algorithm[J]. Control and Decision, 2015, 30(6): 1021-1026 (in Chinese)
文天柱, 许爱强, 程恭. 基于改进 ENN2 聚类算法的多故障诊断方法[J]. 控制与决策, 2015, 30(6): 1021-1026
- [2] Xing Chang-zheng, Liu Jian. Evolutionary data stream clustering algorithm based on integration of affinity propagation and density[J]. Journal of Computer Applications, 2015, 35(7): 1927-1932 (in Chinese)
邢长征, 刘剑. 基于近邻传播与密度相融合的进化数据流聚类算法[J]. 计算机应用, 2015, 35(7): 1927-1932
- [3] Tao Xin-min, Song Shao-yu, Cao Pan-dong, et al. A Spectral Clustering Algorithm Based on Manifold Distance Kernel[J]. Information and Control, 2012(3): 307-313 (in Chinese)
陶新民, 宋少宇, 曹盼东, 等. 一种基于流形距离核的谱聚类算法[J]. 信息与控制, 2012(3): 307-313
- [4] Tian Gang, He Ke-qing, Wang Jian, et al. Domain-Oriented and

Tag-Aided Web Service Clustering Method[J]. Chinese Journal of Electronics, 2015, 43(7): 1266-1274 (in Chinese)

田刚, 何克清, 王健, 等. 面向领域标签辅助的服务聚类方法[J]. 电子学报, 2015, 43(7): 1266-1274

[5] Wu Tao, Chen Li-fei, Guo Gong-de. High-dimensional data clustering algorithm with subspace optimization[J]. Journal of Computer Applications, 2014, 34(8): 2279-2284 (in Chinese)

吴涛, 陈黎飞, 郭躬德. 优化子空间的高维聚类算法[J]. 计算机应用, 2014, 34(8): 2279-2284

[6] Xin Yu, Yang Jing, Tang Chu-heng, et al. An Overlapping Semantic Community Detection Algorithm Based on Local Semantic Cluster[J]. Journal of Computer Research and Development, 2015, 52(7): 1510-1521 (in Chinese)

辛宇, 杨静, 汤楚衡, 等. 基于局部语义聚类的语义重叠社区发现算法[J]. 计算机研究与发展, 2015, 52(7): 1510-1521

[7] Liao Lü-chao, Jiang Xin-hua, Zou Fu-min, et al. A Spectral Clustering Method for Big Trajectory Data Mining with Latent Semantic Correlation[J]. Chinese Journal of Electronics, 2015, 43(5): 956-964

[8] Yu Xiao-dong, Lei Ying-jie, Yue Shao-hua, et al. Research on PSO-based intuitionistic fuzzy kernel clustering algorithm[J]. Journal of Communication, 2015(5): 74-80 (in Chinese)

余晓东, 雷英杰, 岳韶华, 等. 基于粒子群优化的直觉模糊核聚类算法研究[J]. 通信学报, 2015(5): 74-80

[9] Liang Hua-dong, Han Jiang-hong. Clustering and Sorting Radar Signals Based on Multi-wavelet Packets Characteristics of Bispectrum[J]. Acta Photonica Sinica, 2014, 43(3): 152-159 (in Chinese)

梁华东, 韩江洪. 采用双谱多类小波包特征的雷达信号聚类分选[J]. 光子学报, 2014, 43(3): 152-159

[10] Sun Yan-wei, Peng Zhi-ming, Li Jian-bo. Community detection algorithm based on particle swarm optimization and fuzzy clustering[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2015, 27(5): 660-666 (in Chinese)

孙延维, 彭智明, 李健波. 基于粒子群优化与模糊聚类的社区发现算法[J]. 重庆邮电大学学报(自然科学版), 2015, 27(5): 660-666

(上接第 176 页)

行。经过多次实验,测得该实时内核的主要时间参数及第一个调度周期内任务的执行结果,分别如表 1、图 5 所示。

表 1 混合实时系统内核主要的时间参数

内核时间参数	时间(μ s)
TT 任务切换时间	2.32
ET 任务切换时间	2.35
TT/ET 切换时间	2.44
中断响应时间	0.66
时钟中断总时间	10.92

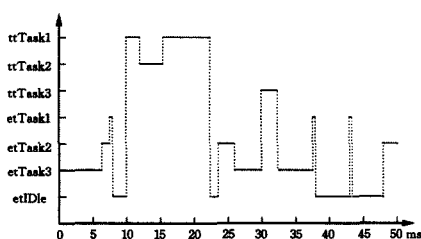


图 6 混合任务调度实验结果

由表 1 可知,该系统内核的时间特性满足大多数实时应用的需要。其中,时钟中断中由于要完成系统与时间有关的一些更新并进行 TT/ET 任务间的调度与切换,因此时间较长。

由图 6 所示的任务执行结果可知,TT 任务 ttTask1、ttTask2、ttTask3 在预定的时间点立即执行,体现出内核的实时性和确定性。0 时刻 TT 任务空闲,同时就绪的 ET 任务 etTask3、etTask2 和 etTask1 按照优先级依次执行,系统具有良好的确定性与可靠性。中断级及任务级 TT 任务与 ET 任务的灵活切换,使得系统调度更加灵活,系统利用率高。

结束语 本文根据 OSEK/VDX 标准,基于 μ C/OS-II 系统内核,扩展了其对于时间触发任务调度的支持,实现了系统对于时间/事件触发混合任务调度的支持,并给出了时间触发任务的可调度性测试算法。实验证明,该实时内核具有良好的实时性、确定性与可靠性;中断级与任务级的时间/事件触发任务调度使得内核更具灵活性,并能提高系统的利用率;该内核具有较大的理论研究与应用价值。

参考文献

[1] Van Den Heuvel M M H P, Bril R J, Lukkien J J, et al. RTOS support for mixed time-triggered and event-triggered task sets [C]//Proceedings of the 2012 IEEE 15th International Conference on Computational Science and Engineering. IEEE Computer Society, 2012: 578-585

[2] Kopetz H. Event-triggered versus time-triggered real-time systems[M]//Operating Systems of the 90s and Beyond. Springer Berlin Heidelberg, 1991: 86-101

[3] Liu C L, Layland J W. Scheduling algorithms for multiprogramming in a hard-real-time environment[J]. Journal of the ACM (JACM), 1973, 20(1): 46-61

[4] Baruah S, Fohler G. Certification-cognizant time-triggered scheduling of mixed-criticality systems[C]//2011 IEEE 32nd Real-Time Systems Symposium (RTSS). IEEE, 2011: 3-12

[5] Itami Y, Ishigooka T, Yokoyama T. A Distributed Computing Environment for Embedded Control Systems with Time-Triggered and Event-Triggered Processing[C]//14th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications, 2008(RTCSA'08). IEEE, 2008: 45-54

[6] Chen Xi, Lv Wei-jie, Liu Lu-yuan. Design and implementation of real time kernel supporting event/time mixed trigger[J]. Computer Engineering and Applications, 2008, 44(16): 87-89 (in Chinese)

陈曦, 吕伟杰, 刘鲁源. 事件/时间触发嵌入式操作系统内核的设计[J]. 计算机工程与应用, 2008, 44(16): 87-89

[7] Dan Tu-nan, Zhu Li-ping, Yan Ji-xun. A Mixed Trigger Schedule Method Based on Time-Triggered Safety Critical Operating System[C]//China Aviation Science and Technology Conference. 2013: 1-5 (in Chinese)

谈图南, 朱立平, 颜纪迅. 一种基于时间触发的安全关键操作系统混合调度策略[C]//2013 首届中国航空科学技术大会论文集. 2013: 1-5

[8] OSEK Group. OSEK/VDX Operating System Specification[S/OL]. [2005-02-17]. <http://www.osek-idx.org>

[9] OSEK Group. OSEK/VDX Time-triggered Operating System Specification, Version 1.0[S/OL]. [2005-02-17]. <http://www.osek-idx.org>