

文章编号: 2095-2163(2021)06-0144-05

中图分类号: TP391

文献标志码: A

基于多种群协同进化算法的数据并行聚类算法

孙柳

(广东工业大学 华立学院, 广州 511325)

摘要: 为了提高云存储空间多维资源数据挖掘能力,需要进行数据并行聚类处理,提出基于多种群协同进化算法的数据并行聚类算法。构建云存储空间多维资源数据的参数采集模型,对采集的云存储空间多维资源数据进行模糊并行特征分布式重组,提取云存储空间多维资源数据聚类特征参数集,采用关联粗糙集特征分析方法进行云存储空间多维资源数据的多尺度小波结构分解,结合多种群协同控制的方法,建立云存储空间多维资源数据的数据并行聚类模型,通过关联协同滤波检测方法,进行云存储空间多维资源数据的分组特征检测和融合聚类处理,根据差分进化方法进行云存储空间多维资源数据的聚类中心寻优,遍历云存储空间多维资源数据聚类区域的候选目标集,实现对云存储空间多维资源数据的数据并行关联规则聚类和可靠性挖掘。仿真结果表明,采用该方法进行云存储空间多维资源数据挖掘的精度较高,聚类性较好,收敛性较强。

关键词: 多种群; 协同进化; 云存储; 数据; 并行聚类

Data parallel clustering algorithm based on multi-group co-evolution algorithm

SUN Liu

(Huali College Guangdong University of Technology, Guangzhou 511325, China)

[Abstract] In order to improve the data mining ability of multi-dimensional resources in cloud storage space, data parallel clustering processing is needed, and a data parallel clustering algorithm based on multi-group co-evolution algorithm is proposed. Construct a parameter collection model of cloud storage space multi-dimensional resource data, perform fuzzy parallel feature distributed reorganization of the collected cloud storage space multi-dimensional resource data, extract cloud storage space multi-dimensional resource data clustering feature parameter set, and use the associated rough set feature analysis method Multi-scale wavelet structure decomposition of cloud storage space multi-dimensional resource data, combined with multiple group collaborative control methods, establishes a parallel clustering model of cloud storage space multi-dimensional resource data, and grouping cloud storage space multi-dimensional resource data through associated collaborative filtering detection methods Feature detection and fusion clustering processing, according to the differential evolution method to optimize the clustering center of cloud storage space multi-dimensional resource data, traverse the candidate target set of the cloud storage space multi-dimensional resource data clustering area, and realize the cloud storage space multi-dimensional resource data Parallel association rule clustering and reliability mining. The simulation results show that this method has higher precision, better clustering and stronger convergence for multi-dimensional resource data mining in cloud storage space.

[Key words] multi-group; co-evolution; cloud storage; data; parallel clustering

0 引言

随着云存储空间多维资源分布数据库存储和信息传输技术的发展,云存储空间多维资源分布数据库的数据信息维数越来越多,需要结合大数据和云信息处理技术,构建云存储空间多维资源分布数据库的大数据并行聚类模型,提高云存储空间多维资源分布数据库数据的检测和识别能力。通过云存储空间多维资源数据并行聚类和特征分析,构建云存储空间多维资源数据聚类分析模型^[1],提高云存储空间多维资源分布数据库的信息管理能力。相关的并行聚类方法研究,在云存储和资源分布数据库的组网设计和大数据信息管理中具有重要意义^[2]。

对云存储空间多维资源数据并行聚类是建立在对数据的候选特征分析基础上,通过贝叶斯关联规则分析,进行云存储空间多维资源数据并行聚类^[3]。传统方法中,对云存储空间多维资源数据并行聚类方法主要有:基于模糊信息检测的云存储空间多维资源数据并行聚类方法^[4]、基于统计分析的云存储空间多维资源数据并行聚类方法^[5]、基于粗糙集特征匹配的云存储空间多维资源数据并行聚类方法^[6]等。由于传统方法进行云存储空间多维资源数据并行聚类存在适应度水平不高,抗干扰性不好等问题。对此,本文提出基于多种群协同进化算法的云存储空间多维资源数据并行聚类方法。首先构建云存储空间多维资源数据的参数采集模型,对

作者简介: 孙柳(1988-),男,硕士,助教,主要研究方向:软件工程。

收稿日期: 2021-03-26

采集的云存储空间多维资源数据进行模糊并行特征分布式重组,提取云存储空间多维资源数据聚类特征参数集,采用关联粗糙集特征分析方法进行云存储空间多维资源数据的多尺度小波结构分解,然后采用多种群协同控制的方法,建立云存储空间多维资源数据的并行聚类模型。通过关联协同滤波检测方法,进行云存储空间多维资源数据的分组特征检测和融合聚类处理,利用差分进化方法进行云存储空间多维资源数据的聚类中心寻优,遍历云存储空间多维资源数据聚类区域的候选目标集,实现对云存储空间多维资源数据的并行关联规则聚类和可靠性挖掘。经仿真测试分析,展示了本文方法在提高云存储空间多维资源数据并行聚类能力方面的优越性能。

1 云存储空间多维资源数据存储特征分析

1.1 结构分析

为了实现基于多种群协同进化算法的云存储空间多维资源数据并行聚类,构建云存储空间多维资源数据的参数采集和优化存储结构模型,采用多维特征空间融合和匹配调度的方法,进行云存储空间多维资源数据的传输结构分析,通过信道转换和均衡配置,进行云存储空间多维资源数据融合^[7],得到云存储空间多维资源数据存储结构模型如图1所示。



图1 云存储空间多维资源数据存储结构模型

Fig. 1 Cloud storage space multi-dimensional resource data storage structure model

在云存储空间多维资源数据存储结构模型中,采用演化贝叶斯准参数估计方法,构造云存储空间多维资源数据的分类存储器,通过多维信息重组和分块区域重构,进行云存储空间多维资源数据的网格分块区域调度^[8]。在临近区域中,边缘特征融合测度作为云存储空间多维资源数据挖掘的候选区域,遍历这些区域获得云存储空间多维资源数据的聚类中心子集,在候选目标集中,得到云存储空间多维资源数据聚类信息熵为:

$$q_i(t) = \text{slerp}(q_i(t_1), q_i(t_2), t) = \frac{\sin(1-t)\theta}{\sin\theta} q_i(t_1) + \frac{\sin t\theta}{\sin\theta} q_i(t_2). \quad (1)$$

采用熵函数聚类方法,进行云存储空间多维资源数据分布式概率密度重组,得到云存储空间多维资源数据聚类的随机概率密度条件 $p(v_i | y = 1)$ 、 $p(v_i | y = 0)$,其满足高斯分布:

$$u_m = \cos[2\pi f_0 \tau_m(\theta)], v_m = \sin[2\pi f_0 \tau_m(\theta)]. \quad (2)$$

式中, μ_1, σ_1 和 μ_0, σ_0 分别为云存储空间多维资源数据的目标样本数据和标准信息差。

采用多维特征分解方法,进行云存储空间多维资源数据信息特征重构,得到云存储空间多维资源数据的模糊信息聚类样本分布为:

$$S(\varphi) = \sum_{t=0}^T \frac{\int_0^1 |h_{1(k-1)}(t) - h_{1k}(t)|^2 dt}{\int_0^1 h_{1(k-1)}^2(t) dt}, \quad (3)$$

$$p\left(\sum_{i=1}^L c_{4s_i} e^{2\varphi_i(m-n)} = r_i\right) = p \int \frac{x(\tau)}{t-\tau} + \prod T. \quad (4)$$

式中, $\alpha < \zeta < \beta$, $l(z)$ 为云存储空间多维资源数据样本位置; l_i 为云存储空间多维资源数据聚类区域位置; D^α 和 $D^{\zeta\beta}$ 分别为正样本和负样本。根据云存储空间多维资源数据的结构参数分析,进行云存储空间多维资源数据的优化聚类 and 挖掘^[9]。

1.2 特征分析

采用关联粗糙集特征分析方法进行云存储空间多维资源数据的多尺度小波结构分解,结合特征收敛性控制的方法,通过云存储空间多维资源分布数据库多属性样本重组^[10],得到云存储空间多维资源数据的模糊相关系数:

$$F_j = \sum_{i=1}^m P_{ij} u_i(k-1). \quad (5)$$

结合灰度特征重组和语义分布式融合方法,得到云存储空间多维资源数据聚类的随机概率密度分布集。云存储空间多维资源数据的多维概率密度函数为:

$$u_{i/j}(k-1/k-1) = P(m_i(k-1)/m_j(k), z^{k-1}) = \frac{1}{c_j} P_{ij} u_i(k-1). \quad (6)$$

采用类间平衡聚类分析方法,聚类过程以 $\frac{1}{2s}$ 的概率实现收敛,收敛结果为1,得到云存储空间多维资源数据的类间分离的特征提取公式为:

$$\{\delta x_{i+1}(j_k) = x_{j_{k+1}} - x_{i+1} | k \in 1, \dots, N_b\}. \quad (7)$$

采用机器学习的分类学习方法,得到云存储空

间多维资源数据的联合特征分布参数 φ 和 θ 。采用重采样策略,得到云存储空间多维资源分布数据库的特征分配概率 $P(z_i = j | z = i, w_i)$ 的算式为:

$$\begin{aligned} x(k+1) &= \Phi_i(k)x(k) + w_i(k), \quad i = 1, 2, \dots, m; \\ z(k) &= H_i(k)x(k) + v_i(k), \quad i = 1, 2, \dots, m. \end{aligned} \quad (8)$$

其中, $n_{-i,j}^{(w_i)}$ 表示云存储空间多维资源分布数据库多属性联合分布权重; w_i 表示被分配在主题词聚类集 j 中的次数; $n_{-i,j}^{(z)}$ 表示云存储空间多维资源数据特征分布主题 j 中的样本数; $n_{-i,j}^{(d)}$ 表示联合概率密度分布集。采用关联粗糙集特征分析方法进行云存储空间多维资源数据的多尺度小波结构分解,表示为:

$$x^i(0) = \hat{x}^i(0), \quad (9)$$

$$u_i(0) = P(m_i(0)/z(0)). \quad (10)$$

综上所述,构建了云存储空间的多维资源数据融合模型,结合特征检测方法,实现数据并行聚类分析^[11]。

2 云存储空间多维资源数据聚类

2.1 云存储空间多维资源分布数据库多属性数据特征融合

采用多种群协同控制的方法,建立云存储空间多维资源数据的并行聚类模型,通过关联协同滤波检测方法^[12],得到云存储空间多维资源数据聚类的更新规则约束参数的解:

$$M = \{m_i | i = 1, 2, \dots, m\}, \quad (11)$$

根据云存储空间多维资源数据的属性分布进行模糊聚类,得到云存储空间多维资源数据的差分进化约束的相关性因子为:

$$K(t) = \left(\sum_{m=1}^M u_m \right) \cos(2\pi f_0 t) - \left(\sum_{m=1}^M v_m \right) \sin(2\pi f_0 t). \quad (12)$$

其中,云存储空间多维资源数据融合的特征分布矩阵为 $\mathbf{R} = (r_{ij}, a_{ij})_{m \times n}$, 基于数据层面构建大数据分类模型,得到云存储空间多维资源数据分类的联合特征解为:

$$\Delta P = P^r - P^n, \quad (13)$$

获取原始数据集,引入云存储空间多维资源数据的互信息熵,即:

$$P^n = F(a, d, \alpha, \theta), \quad (14)$$

以 P 为云存储空间多维资源分布数据库多属性分布的概率密度为:

$$P^r = F(a + \Delta a, d + \Delta d, \alpha + \Delta \alpha, \theta + \Delta \theta).$$

根据云存储空间多维资源数据的融合参数应满

足:

$$T_{i,j}(t) = \frac{|p_{i,j}(t) - \Delta p(t)|}{p_{i,j}(t)}. \quad (15)$$

用 $U_{i,j}(t)$ 表示的云存储空间多维资源数据动态特征分布信息熵。

基于决策边界的类样本分析方法^[13],得到共享的通道数为 P ,构建云存储空间多维资源数据聚类的联合关联决策函数为:

$$Dp(om, rm) = \frac{1}{n} \sum_{i=1}^n d(om^i, rm^i). \quad (16)$$

其中, $d(om^i, rm^i)$ 表示联合度评估系数。充分利用数据空间的类间指数分布,采用差分进化方法,基于高斯概率分布方法,云存储空间多维资源分布数据库多属性特征融合输出为 $E_k \in E(k = 1, 2, \dots, t)$ 。根据类别的不同属性,得到云存储空间多维资源分布数据库多属性数据特征融合模型为 $P_i \in P(i = 1, 2, \dots, m)$ 。

综上所述,采用差分进化方法,进行云存储空间多维资源分布数据库多属性参数识别和聚类^[14]。

2.2 云存储空间多维资源数据并行聚类

计算描述云存储空间多维资源数据的互信息特征向量 $\mathbf{v}_i, \mathbf{om}_j^i, \mathbf{rm}_j^i$, 通过差分进化方法进行云存储空间多维资源数据的聚类中心寻优,遍历云存储空间多维资源数据聚类区域的候选目标集,得到多属性大数据聚类的模糊集为:

$$\begin{aligned} q_i(t_1) &= [w_1, x_1, y_1, z_1], \\ q_i(t_2) &= [w_2, x_2, y_2, z_2] + V_i. \end{aligned} \quad (17)$$

其中, V_i 为云存储空间多维资源数据的关联分析度量值,使用联合特征分布式进化方法,得到云存储空间多维资源数据并行聚类的联合公式为:

$$Dv(om, rm) = \frac{1}{n} \sum_{i=1}^{n-1} |d(om^{i+1}, om^i) - d(rm^{i+1}, rm^i)|. \quad (18)$$

在非线性可分的数据集中,得到云存储空间多维资源分布数据库多属性并行聚类输出的相似系数为:

$$u_{AB} = \frac{s}{n} + \frac{f}{n} i + \frac{p}{n} j. \quad (19)$$

其中: p 为云存储空间多维资源数据的分布集, f 为云存储空间多维资源数据分布的联合特征参数分布集。用 4 元组 (E_i, E_j, d, t) 来表示云存储空间多维资源数据的主特征量,采用决策树调度和多属性差分进化方法,得到并行聚类输出的联合特征量:

$$ck = \begin{cases} x_k & \text{if } m_k + U(-1,1) * (m_k - lk) < x_k, \\ \bar{x}_k & \text{if } m_k + U(-1,1) * (m_k - l_k) > \bar{x}_k, \\ m_k + U(-1,1) * (m_k - l_k) & \text{otherwise.} \end{cases} \quad (20)$$

式中, m 为云存储空间多维资源数据并行聚类的进化维数, $(d_{ik})^2$ 为非线性数据集。

综上所述,通过差分进化方法进行云存储空间多维资源数据的聚类中心寻优,实现对云存储空间多维资源数据的并行关联规则聚类和可靠性挖掘。

3 仿真测试

对云存储空间多维资源数据采集的样本长度为 1024,云存储空间的特征分布维数为 3,嵌入维数为 125,数据分类的属性为 6,多种群迭代的部署为 24,差分进化的迭代数为 100。根据上述参数设定,得到云存储空间多维资源数据统计特征量分布如图 2 所示。

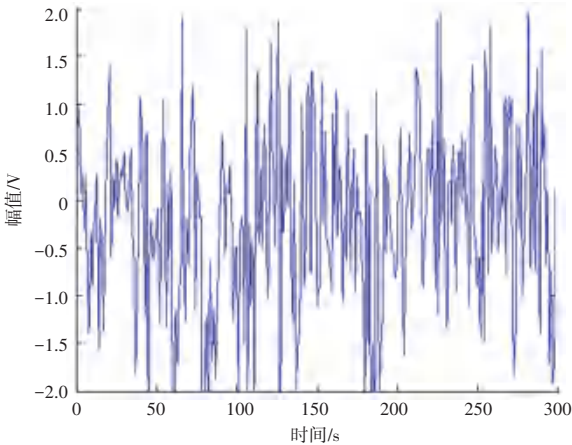


图 2 云存储空间多维资源数据统计特征量分布

Fig. 2 Distribution of statistical characteristics of multi-dimensional resource data in cloud storage space

根据图 2 大数据检测结果,实现云存储空间多维资源数据聚类,得到并行聚类预测值如图 3 所示。

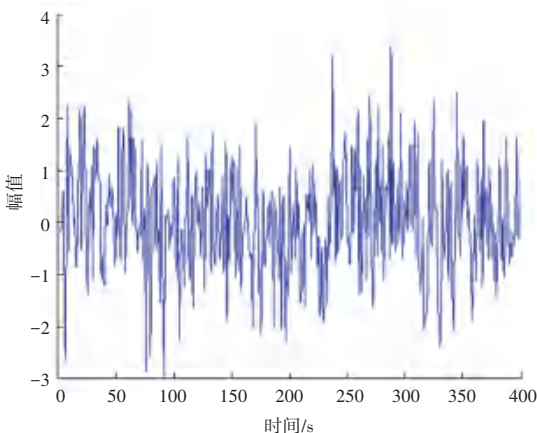


图 3 数据并行聚类预测值

Fig. 3 Data parallel clustering predicted value

分析图 3 得知,本文方法进行云存储空间多维资源数据的特征并行聚类的聚敛度水平较高,数据聚类融合性较好。测试数据分类的准确率,得到聚类误差收敛结果如图 4 所示。

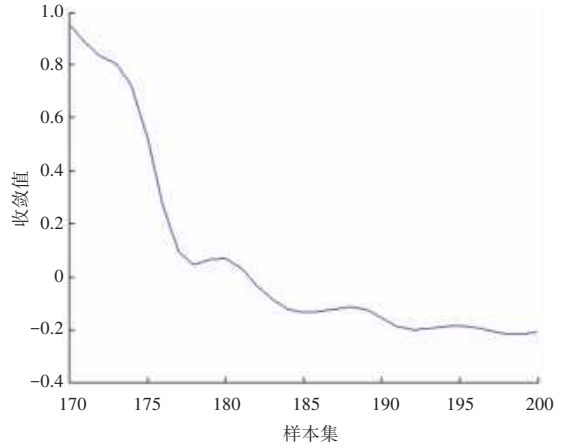


图 4 数据聚类收敛曲线

Fig. 4 Data clustering convergence curve

分析图 4 得知,本文方法对云存储空间多维资源数据分类的正确率较高。在不同的数据聚类中心,测试云存储空间多维资源数据挖掘的识别率,得到测试结果如图 5 所示。

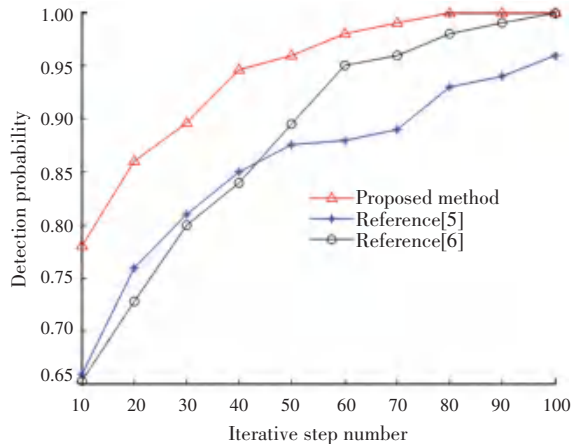


图 5 数据并行聚类的识别率

Fig. 5 Recognition rate of data parallel clustering

根据图 5 仿真结果得知,本文方法进行云存储空间多维资源数据并行聚类处理,提高了数据的识别率。

4 结束语

本文提出基于多种群协同进化算法的云存储空间多维资源数据并行聚类方法,采用多维特征空间融合和匹配调度,进行云存储空间多维资源数据的传输结构分析,结合灰度特征重组和语义分布式融

(下转第 152 页)