

文章编号: 2095-2163(2021)07-0020-05

中图分类号: TP391

文献标志码: A

# 基于深度学习的区块链数据分片峰值聚类算法

张皓

(广东工业大学 华立学院, 广州 511325)

**摘要:** 为了提高物联网区块链数据挖掘能力,需要进行数据优化聚类处理,提出基于深度学习的区块链数据分片峰值聚类算法。采用异构有向图分析方法进行物联网区块链数据存储结构设计,结合特征空间重组技术进行物联网区块链数据结构重组,提取物联网区块链数据的关联信息特征量,采用语义相关性融合的方法进行区块链数据特征提取和自适应调度,对提取的物联网区块链数据特征量进行模糊聚类处理,采用模糊C均值聚类方法进行物联网区块链数据的网格分片峰值聚类和属性分类识别,采用深度学习方法进行数据聚类过程中的分片峰值融合和聚类分析,实现区块链数据分片峰值聚类。仿真结果表明,采用该方法进行区块链数据分片峰值聚类的收敛性较好,误分率较低,自适应学习能力较强。

**关键词:** 深度学习; 区块链; 数据; 挖掘; 聚类

## Block chain data sliced peak clustering algorithm based on deep learning

ZHANG Hao

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

**[Abstract]** In order to improve the ability of blockchain data mining in the Internet of Things, it is necessary to carry out data optimization clustering processing and propose a block chain data slice peak clustering algorithm based on deep learning. Using the heterogeneous directed graph analysis method to design the data storage structure of the Internet of Things block chain, combining the feature space reorganization technology to reorganize the data structure of the Internet of Things block chain, extracting the related information characteristic quantity of the Internet of Things block chain data, using the semantic correlation fusion method to extract the block chain data feature and adaptive scheduling, performing fuzzy clustering processing of the extracted block chain data feature quantity, using the fuzzy C-means clustering method to identify the grid segmented peak clustering and attribute classification of the block chain data of the Internet of Things, and using deep learning methods to perform the segmented peak fusion and clustering analysis in the process of data clustering, the segmented peak clustering of the block chain data is realized. The simulation results show that the convergence of block chain data segmentation peak clustering with this method is good, the error rate is low, and the adaptive learning ability is strong.

**[Key words]** deep learning; block chain; data; mining; clustering

## 0 引言

随着物联网技术的发展,需要对物联网区块链数据进行优化挖掘和存储设计,通过物联网区块链数据的聚类分析,降低数据的存储空间,提高数据的挖掘精度,相关的物联网区块链数据聚类方法研究受到人们的极大关注。对物联网区块链数据挖掘是建立在对数据的特征提取和信息识别基础上,采用模糊信息识别技术,进行物联网区块链数据的存链结构开发,结合物联网区块链数据的大数据特征分布进行异构重组<sup>[1]</sup>,建立物联网区块链数据的大数据信息分析模型,采用信息融合和资源优化调度的方法进行物联网区块链数据聚类分析。

当前,对物联网区块链数据聚类方法主要有K-Means聚类法、网格区域聚类方法、粒子群聚类方法等<sup>[2-3]</sup>,建立物联网区块链数据的特征提取和大数据分析模型,采用相关的特征分布式检测方法,实现物联网区块链数据聚类。文献[4]提出基于模糊层次结构聚类的物联网区块链数据聚类方法,采用异构有向图分析方法进行物联网区块链数据分片层次结构设计,提高数据聚类的精度,但该方法的计算开销较大,实时性不好。文献[5]提出基于特征空间重组技术的数据聚类方法,在大数据背景下进行物联网区块链数据在线聚类优化设计,结合分片重组技术进行数据聚类优化,但该方法的模糊度较大,聚类精度不高。文献[6]提出指标约束的物联网节点

**基金项目:** 2020年度广东省普通高校特色人才创新项目(自然科学类)项目(2019KTSCX223)。

**作者简介:** 张皓(1994-),男,博士研究生,主要研究方向:区块链、计算机视觉。

**通讯作者:** 张皓 Email:540218310@qq.com

**收稿日期:** 2021-03-15

部署的区块链融合技术,采用谐波谐振扰动特征分析方法,构建物联网节能降损可靠性评估模型,结合分块数据组网控制,采用数据聚类方法,实现物联网输出参数可靠性评估,但该方法的聚类性不好。

针对上述问题,本文提出基于深度学习的区块链数据分片峰值聚类算法。首先进行区块链数据的结构特征分析,然后进行区块链数据的特征提取,结合分片峰值聚类方法,实现数据聚类优化,最后进行仿真测试分析,得出有效性结论。

## 1 物联网区块链数据结构分析和特征提取

### 1.1 物联网区块链数据结构分析

采用有向图模型构建物联网区块链数据的聚类节点分布结构模型,采用相关性的统计分析方法,得到物联网区块链数据的异构存储结构模型,计算物联网区块链数据的模糊聚类特征分布集<sup>[6]</sup>,得到融合度函数定义为:

$$d_i = \sqrt{x_i^2 + y_i^2 + z_i^2} \quad (1)$$

分析物联网区块链数据集统计差异分布特性,得到每个插值点的实际值公式定义为:

$$d_{m+1}(m) = d_{k+1}(m) \pm \sqrt{(d_m(0)e^{\lambda_1} + 1)^2 - \sum_{i=1}^{m-1} [d_{m+1}(i) - d_{k+1}(i)]^2} \quad (2)$$

其中,  $d_{m+1}(m)$  为物联网区块链数据集在第  $m$  点的预测值,  $d_{k+1}(m)$  为采用第  $m$  点处采集的物联网区块链数据的模糊性特征量,根据物联网区块链数据的特征提取结果采用模糊 C 均值聚类方法进行信息处理<sup>[7]</sup>,设定全局变量,建立物联网区块链数据挖掘的模糊分割系数为:

$$Mi = L_m + \frac{N \times 0.5 - \sum f_{less}}{f_m} \times width \quad (3)$$

其中,  $Mi$  表示物联网区块链数据挖掘的中位数;  $L_m$  为物联网区块链数据挖掘的下界;  $f_m$  为物联网区块链数据的中位数;  $f_{less}$  表示各维度下物联网区块链数据的最小统计特征量。

### 1.2 物联网区块链数据特征提取

构建资源聚类的节点分布模型,在大数据背景下进行物联网区块链数据分片峰值聚类优化设计,设物联网区块链数据聚类节点图模型属性集为  $X = \{x_1, x_2, \dots, x_n\}$ ,进行物联网区块链数据聚类节点图模型设计,采用语义本体模型构造的方法,进行

物联网区块链数据处理<sup>[8]</sup>,通过自相关特征匹配方法,分析 2 组相似的物联网区块链数据的相对贴进度  $\varphi_1$  定义为:

$$(\bar{s}, \bar{a}) = \varphi_1((s_1, a_1), (s_2, a_2), \dots, (s_n, a_n)) = \Delta\left(\sum_{j=1}^n \omega_j \Delta^{-1}(s_j, a_j)\right) \quad (4)$$

其中,  $\sum_{j=1}^n \omega_j = 1, \bar{s} \in S, \bar{a} \in [-0.5, 0.5]$ 。

设  $\{(s_1, a_1), (s_2, a_2), \dots, (s_n, a_n)\}$  是一组描述物联网区块链数据聚类的语义特征分布集,建立物联网区块链数据聚类的统计分析模型,在模糊网格区域,进行物联网区块链数据聚类的统计分析和自适应调度,在模糊语义相关性融合聚类下,进行物联网区块链数据的模糊加权分析<sup>[9]</sup>,得到模糊加权的特征分布向量为  $\omega = (\omega_1, \omega_2, \dots, \omega_n)^T, \omega_j \in [0, 1]$ ,考虑等价的语义映射,进行物联网区块链数据的融合分析,建立物联网区块链数据的统计特征量,通过一个映射函数  $M$  表示,  $M: C * C' \rightarrow r$  描述物联网区块链数据的相似度信息,通过空间聚类分析的方法,进行物联网区块链数据分布式调度,建立本体模型,进行物联网区块链数据的语义相似度融合和特征聚类处理<sup>[10]</sup>。

## 2 物联网区块链数据聚类

### 2.1 区块链数据特征提取和自适应调度

采用有向图模型构建物联网区块链数据的聚类节点分布结构模型,在物联网区块链数据库中进行资源信息特征提取<sup>[11]</sup>,设  $\{(s_1, a_1), (s_2, a_2), \dots, (s_n, a_n)\}$  是物联网区块链数据的二元语义特征分量,采用异构有向图分析方法进行物联网区块链数据存储结构设计,结合特征空间重组技术进行物联网区块链数据结构重组,提取物联网区块链数据的关联信息特征量<sup>[12]</sup>,得到物联网区块链数据聚类的优化加权系数为  $\omega = ((\omega_1, a_1'), (\omega_2, a_2'), \dots, (\omega_n, a_n'))^T, \omega_j \in [0, 1]$ ,采用统计分析方法进行物联网区块链数据的聚类中心检测,得到物联网区块链数据的关联性本体结构模型为:

$$(\bar{s}, \bar{a}) = \varphi_2(((s_1, a_1), (\omega_1, a_1')), ((s_2, a_2), (\omega_2, a_2')), \dots, ((s_n, a_n), (\omega_n, a_n'))) = \Delta\left(\frac{\sum_{j=1}^n \Delta^{-1}(\omega_j, a_j') \Delta^{-1}(s_j, a_j)}{\sum_{j=1}^n \Delta^{-1}(\omega_j, a_j')}\right) = \Delta\left(\frac{\sum_{j=1}^n \beta_j \beta_j'}{\sum_{j=1}^n \beta_j'}\right) \quad (5)$$

对物联网区块链数据进行模糊处理,构建物联网区块链数据的自适应调度参数  $\nabla^2 F(x)$ , 建立物联网区块链数据聚类的模糊决策矩阵:

$$\begin{aligned} X_1 &= \begin{bmatrix} M & G & P & P \\ P & VP & M & P \\ G & M & G & EP \\ VG & P & P & G \\ EG & EP & VP & M \end{bmatrix} \\ X_2 &= \begin{bmatrix} P & M & VP & VP \\ VP & EP & G & G \\ M & G & P & VP \\ EG & VP & VP & M \\ P & VP & M & VP \end{bmatrix} \\ X_3 &= \begin{bmatrix} G & P & VP & VG \\ VP & G & P & G \\ VG & VP & G & P \\ G & VG & EG & VP \\ M & VP & M & EG \end{bmatrix} \\ \omega &= \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} = \begin{bmatrix} M & G & VP & P \\ VP & VG & P & G \\ G & P & M & G \end{bmatrix} \end{aligned} \quad (6)$$

物联网区块链数据聚类的特征提取问题转化为一个二元语义决策问题,物联网区块链数据聚类的模糊特征匹配评价指标集为  $E_k \in E(k=1,2,\dots,t)$ , 采用主体词匹配的方法,分析  $X$  的相似度函数,得到物联网区块链数据的模糊隶属度函数为:

$$\tilde{w}_k^i = \tilde{w}_{k-1}^i \frac{p(z_k/\tilde{x}_k^i)p(\tilde{x}_k^i/x_{k-1}^i)}{q(\tilde{x}_k^i/x_{k-1}^i)} \quad (7)$$

其中,  $\tilde{x}_k^i$  为物联网区块链数据的语义特征量,  $\tilde{w}_{k-1}^i$  为惯性权重。利用概念、实例和属性等实体集,得到物联网区块链数据聚类中心节点在  $k+1$  时刻的迭代函数为:

$$x_i(k+1) = x_i(k) + s \left( \frac{x_j(k) - x_i(k)}{\|x_j(k) - x_i(k)\|} \right) \quad (8)$$

采用决策树模型进行物联网区块链数据的特征重构,在物联网区块链数据的递归图模型中,采用相空间重构的方法,实现物联网区块链数据聚类的模糊特征信息采样<sup>[13]</sup>,得到物联网区块链数据的模糊信息加权权重向量  $v_i$ , 物联网区块链数据的相关特征分布矩阵表示为:

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

其中,  $c$  为物联网区块链数据聚类的搜索步数,  $\mu_{ik}$  为物联网区块链数据的语义关联度决策系数。

根据上述分析,进行物联网区块链数据的特征分析和优化调度,得到物联网区块链数据的模糊聚类中心为  $d_i$ , 采用分块特征演化方法,进行物联网区块链数据的关联特征检测,得到模糊检测向量  $v_i, v_i = ((w_1, t_1), (w_2, t_2), \dots, (w_j, t_j))$ , 结合模糊相关性融合的方法,进行物联网区块链数据的分片峰值聚类分析<sup>[14-17]</sup>。

## 2.2 数据分片峰值聚类的深度学习

采用模糊 C 均值聚类方法进行物联网区块链数据的网格分片峰值聚类 and 属性分类识别<sup>[18]</sup>, 采用深度学习方法进行数据聚类过程中的分片峰值融合和聚类分析,得到物联网区块链数据的在线聚类准则为:

$$S(i, j) = \frac{\sum_{u \in U_{ij}} (V_{u,i} - 3)(V_{u,j} - 3)}{\sqrt{\sum_{u \in U_{ij}} (V_{u,i} - \bar{V}_i)^2} \sqrt{\sum_{u \in U_{ij}} (V_{u,j} - \bar{V}_j)^2}} \quad (10)$$

物联网区块链数据聚类节点的深度学习加权系数为  $W_e = (\omega_j^{(e)}, 0)$ 。修正每个聚类自适应加权学习系数  $v_i$ , 就可得到聚类有效性评价矩阵  $R = (r_{ij}, a_{ij})_{m \times n}$  和指标权重  $W = ((\omega_1, \beta_1), (\omega_2, \beta_2), \dots, (\omega_n, \beta_n))$ , 综上分析,建立物联网区块链数据聚类的深度学习模型,构建物联网区块链数据的模糊特征分布集<sup>[19-21]</sup>, 得到物联网区块链数据的约束规划模型为:

$$\min(f) = \sum_{i=1}^m \sum_{j=1}^n C_{ij} X_{ij} \quad (11)$$

$$\text{s.t.} \begin{cases} \sum_{j=1}^n X_{ij} = a_i, i=1,2,\dots,m \\ \sum_{i=1}^m X_{ij} = b_j, j=1,2,\dots,n \\ X_{ij} \geq 0, i=1,2,\dots,m, j=1,2,\dots,n \end{cases} \quad (12)$$

由此得到物联网区块链数据聚类的最优评价集记为  $L_1, \dots, L_n$  和  $P_1^{\min}, \dots, P_n^{\min}$ , 物联网区块链数据的优化聚类模型为:

$$I_{d_{ij} \rightarrow c_x} = \left( \frac{\sum_{v=0}^{|c_x|} \cos \sin_{ij \rightarrow x}(d_{ij}, d_{xv})}{|c_x|} \right)^{-1} \quad (13)$$

其中,  $\cos \sin_{ij \rightarrow x}(d_{ij}, d_{xv})$  为物联网区块链数据的融合聚类特征集,根据上述分析,实现物联网区块链数据的优化聚类<sup>[22-23]</sup>。

## 3 仿真实验与结果分析

为了验证本文方法在实现物联网区块链数据分

片峰值聚类中的应用性能,进行实验测试分析,采用C++和Matlab 7混合编程进行物联网区块链数据聚类的算法处理,在Hadoop云平台中构建物联网区块链数据数据库结构模型,物联网区块链数据大数据分布的初始样本规模为1 200,量化分布训练集为120,深度学习的步长为2.4,自适应学习的迭代步数 $NP = 30$ ,各组数据采样集的相似度为0.68,根据上述仿真参数设定,进行物联网区块链数据峰值聚类,得到物联网区块链数据的时域分布如图1所示。

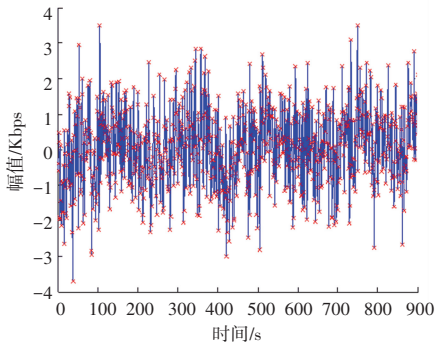


图1 数据的时域波形

Fig. 1 Time domain waveform of data

以图1所示的数据为研究对象,采用语义相关性融合的方法进行区块链数据特征提取和自适应调度,对提取的物联网区块链数据特征量进行模糊聚类处理,实现分片峰值聚类,得到聚类结果如图2所示。

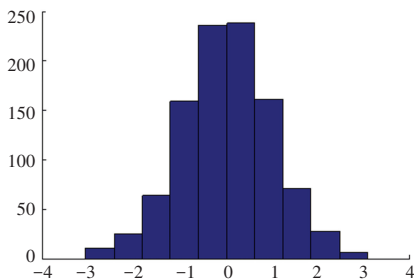


图2 数据分片峰值聚类结果

Fig. 2 Peak clustering results of data fragmentation

分析图2得知,采用本文方法进行物联网区块链数据聚类的峰值融合度较好,收敛性较强,测试误分率,得到结果见表1,分析表1得知,本文方法进行物联网区块链数据聚类的误分率较低。

表1 误分率对比

Tab. 1 Comparison of misscore rate

迭代次数	本文方法	文献[3]	文献[4]
100	0.102	0.256	0.194
200	0.034	0.145	0.173
300	0	0.104	0.163
400	0	0.083	0.088

分析图3结果得知,采用本文方法进行物联网区块链数据聚类的误分率较低,提高了数据聚类的收敛控制能力。

## 4 结束语

采用模糊信息识别技术,进行物联网区块链数据的存链结构开发,结合物联网区块链数据的大数据特征分布进行异构重组,建立物联网区块链数据的大数据信息分析模型,本文提出基于深度学习的区块链数据分片峰值聚类算法采用决策树模型进行物联网区块链数据的特征重构,在物联网区块链数据的递归图模型中,采用相空间重构的方法,实现物联网区块链数据聚类的模糊特征信息采样,采用深度学习方法进行数据聚类过程中的分片峰值融合和聚类分析,实现区块链数据分片峰值聚类。分析得知,采用本文方法进行区块链数据分片峰值聚类的收敛性较好,误分率较低,数据聚类的精度较高。

## 参考文献

- [1] 李宇帆,张会福,刘上力,等. 教育数据挖掘研究进展[J]. 计算机工程与应用, 2019, 55(14): 15-23.
- [2] 杨欢欢,李天瑞,陈馨茹. 基于螺旋图的时间序列数据可视化[J]. 计算机应用, 2017, 37(9): 2443-2448.
- [3] 李劲,岳昆,蔡娇,等. 基于距离度量的多样性图排序方法[J]. 软件学报, 2018, 29(3): 599-613.
- [4] YU Yong, WANG Ziyuan, XU Dianguo. Speed and current sensors fault detection and isolation based on adaptive observers for induction motor drivers[J]. Journal of Power Electronics, 2014, 5(14): 967-979.
- [5] BERRIRI H, NAOUAR W, BAHRI I, et al. Field programmable gate array-based fault-tolerant hysteresis current control for AC machine drives[J]. IET Electric Power Applications, 2012, 6(3): 181-189.
- [6] 李雷,王磊,赵丽,等. 基于大数据技术的配电网节能降损技术分析[J]. 电子技术与软件工程, 2020(19): 224-225.
- [7] 庞俊,于戈,许嘉,等. 基于MapReduce框架的海量数据相似性连接研究进展[J]. 计算机科学, 2015, 42(1): 1-5.
- [8] 肖文,胡娟. 基于数据集稀疏度的频繁项集挖掘算法性能分析[J]. 计算机应用, 2018, 38(4): 995-1000.
- [9] 李彦,刘军. 面向大数据的多维数据缺失特征填补仿真研究[J]. 计算机仿真, 2018, 35(10): 432-435.
- [10] 王慧健,刘峥,李云,等. 基于神经网络语言模型的时间序列趋势预测方法[J]. 计算机工程, 2019, 45(7): 13-19, 25.
- [11] 王晓雷,陈云杰,王琛,等. 基于Q-learning的虚拟网络功能调度方法[J]. 计算机工程, 2019, 45(2): 64-69.
- [12] 姚富光,钟先信,周靖超. 粒计算:一种大数据融合智能建模新方法[J]. 南京理工大学学报, 2018, 42(4): 503-510.
- [13] 王楚捷,王好贤. M-CORD下无线接入网络资源分配研究[J]. 计算机工程与应用, 2018, 54(22): 92-98.
- [14] 刘良凤,刘三阳. 基于权重差异度的动态模糊聚类算法[J]. 吉林大学学报(理学版), 2019, 57(3): 574-582.