

# Big Data Mining Algorithm for a Heterogeneous Multi-core Platform Based on Semantic Segmentation

Zheng Long

Lanzhou University of Technology, Lanzhou, Gansu, 730050, China

## Abstract

This paper mainly studies the big data mining algorithm of heterogeneous multi-core platform based on semantic segmentation. In the process of research, the concept and characteristics of heterogeneous multi-core platform are analyzed, and on this basis, the big data clustering analysis of heterogeneous multi-core platform is conducted, and the big data mining algorithm is deeply studied, and finally the simulation experiment is conducted on the theoretical basis. The simulation experiment can draw the following conclusions: the big data data of heterogeneous multi-core platform using semantic segmentation method has high accuracy, and the feature resolution ability is relatively good, which can significantly improve the big data mining and analysis detection level of heterogeneous multi-core platform.

## Keywords

semantic segmentation; heterogeneous multi-core platform; big data; mining algorithm

## 基于语义分割的异构多核平台大数据挖掘算法

龙正

兰州理工大学, 中国·甘肃 兰州 730050

## 摘要

论文主要对基于语义分割的异构多核平台的大大数据挖掘算法进行研究。在研究过程中对异构多核平台的概念与特点等基本内容进行分析,在此基础上进行异构多核平台的大数据聚类分析,并对其大数据挖掘算法进行深入研究,最后在理论上进行仿真实验。通过仿真实验能够得出以下结论:利用语义分割方法的异构多核平台大数据具有较高的精准度,且特征分辨能力相对良好,可显著提高异构多核平台的大数据挖掘与分析检测水平。

## 关键词

语义分割; 异构多核平台; 大数据; 挖掘算法

## 1 引言

大数据时代背景下,计算机视觉等任务对计算机平台性能提出了更高的要求,需要利用更为合适的处理设备对任务运行进行优化处理。各计算机厂家为满足这一需求,纷纷开始对异构加速设备进行深入研究,由此推动了异构多核平台的快速发展,并且在现阶段已经得到良好的应用。在异构多核平台中,大数据的挖掘与检测环节具有一定的复杂性,这就需要人们使用更加科学高效的挖掘算法与检测技术,而基于语义分割能够实现上述目标,其与多种技术方法具有较高的契合性,在实际应用中可满足系统处理的多样化需求,并实现数据信息的像素级分类,进而增强对大数据的挖掘处理效果。

【作者简介】龙正(2002-),男,中国甘肃临夏人,在读本科生,从事基于语义分割的异构多核平台大数据挖掘算法研究。

## 2 异构多核平台概述

信息技术的快速发展,使得芯片与处理器相关技术也迎来了良好的发展局面。SoC技术在优化的过程中推出了多个不同类型的平台系统,其中的异构多核平台与其他平台系统相比,具有显著的优势作用。异构多核平台又叫做异构多核系统,在使用过程中并不需要绘制体积庞大的电路板,只需要利用精准严谨的计算机语言即可完成相关操作,进一步提高了操作系统的科学性与可靠性,降低了系统的运行功耗<sup>[1]</sup>。为进一步凸显SoC技术的实际应用效果,人们会利用异构多核平台的方法优化改进物联网系统的层级架构体系,以此加强物联网系统的大数据处理与控制水平。

目前,行业使用的异构平台系统有多种,如RING架构、全连接架构、mssh架构以及CORSS BAR架构等。在比较常见的全连接架构中,平台系统会以一类CPU作为主处理器,其他处理器属于协助处理器。主处理器属于平台系统的核心控制装置,协助处理器能够起到加速优化的效果。通过

集群划分的方法能够对平台系统中的处理器集群化处理，将同构处理器统一规划到一个集群系统中，并以集群为标识对异构处理器进行有效处分。

通过大数据的信息融合技术可以完成特征量化的高效化处理以及分析，并建立出大数据挖掘的优化使用模型，将其与模糊相关性的技术手段融合到一起，最终能够完成对大数据挖掘以及特征信息的全面精准提取操作。

### 3 异构多核平台大数据聚类分析

#### 3.1 大数据检测平台分析

对现阶段常见的大大数据挖掘与检测分析的方法总结分析，主要包括模糊神经网络法、关联属性特征法、PSO法以及统计信息法等<sup>[2]</sup>。但是从实际应用效果的角度分析，上述方法在应用中会存在信息融合效果低下、计算成本投入较高、可视化效果较差、模糊度较高以及抗干扰能力较弱等问题，并不利于大数据挖掘与检测工作开展。为实现对异构多核平台的大数据信息的精准全面的检测处理，论文设计了一种基于语义分割的科学的数据挖掘算法，首先建立针对模糊信息的检测分析模型。随后使用关联属性特征法，基于该方法能够得到大数据的模糊相关性信息，在此基础上能够得到数据的自动化挖掘与分析模型。利用模糊信息聚类的方法可以精准分析平台的数据统计信息的具体情况，并建立平台的信息属性链模型，从而建立决策分析的模型。利用云技术完成对大数据特征信息的全面快速检测，利用多要素联合的处理方法构建信息数据的属性链分析模型，以此得到数据挖掘的模糊决策分析模型。除此之外，采用自适应寻优的方法实现对数据挖掘的改进，能够得到数据信息的属性链表相关资源，最终得到大数据的空间分布结构。

基于融合性聚类分析的方法对数据信息进行优化挖掘处理，并将指针识别的方法应用其中，以此数据挖掘的指针分析，进而得到异构多核平台的大数据挖掘指针分布模型。其中包括线坐标指针，对应空间坐标点数据块；属性指针对应属性相关数据；左指针、右指针和下一节点指针等<sup>[3]</sup>。大数据的定位环节主要采用分块区域融合法，并建立信息存储以及查询的子系统。在对数据的特征量信息完成提取操作后，即可完成整个数据挖掘环节。

#### 3.2 数据聚类分析

本在数据的模糊信息聚类模型建立环节主要利用特征匹配的方法，将大数据模糊的特征聚类信息作为基本要素，通过语义分割的方法得到自适应加权学习模型。大数据模糊聚类分布表达公式如下：

$$z_0 = \sum_{i=1}^S Z_i(1/d_i^k) / \sum_{i=1}^S (1/d_i^k)$$

式中， $Z_0$ 代表异构多核平台的大数据关联估计值； $S$ 代表了异构多核平台的大数据的实际测量点的综合统计特征量； $d_i$ 代表了*i*点和0点之间的距离参数； $Z_i$ 代表了在

*i*点处得到的大数据的实际参数值； $k$ 代表了大数据挖掘的插值权重。

随后，利用自适应加权学习结果信息完成空间特征数据的自适应加权操作。通过此操作能够得到数据模糊加权学习公式，对数据自适应挖掘的处理操作具有显著的提升效果，最终建立出空间聚类的分析模型。具体表达公式如下：

$$\begin{cases} \dot{x} = -\alpha x + \sigma y \\ \dot{y} = -x z + r x - y \\ \dot{z} = x y - b z \end{cases}$$

式中， $x$ 、 $y$ 、 $z$ 分别代表了大数据的语义相似度在X轴坐标、Y轴坐标以及Z轴坐标中对应的特征量参数； $\sigma$ 代表了语义本体集； $r$ 代表了数据特征匹配的粗糙集； $b$ 代表了模糊度的具体系数。

### 4 异构多核平台大数据挖掘算法分析

#### 4.1 异构多核平台大数据特征提取

针对异构多核平台中的大数据进行语义关联特征量的挖掘处理，随后通过模糊属性检测的方法进行数据信息的全面检测和统计处理。通过统计分析的手段可以建立出数据的语义分割模型。具体操作方法如下：

$$\begin{bmatrix} d_1 \\ d_2 \\ \dots \\ d_N \end{bmatrix} = \begin{bmatrix} d_{11}, d_{12}, \dots, d_{1N} \\ d_{21}, d_{22}, \dots, d_{2N} \\ \dots \\ d_{N1}, d_{N2}, \dots, d_{NN} \end{bmatrix} \begin{bmatrix} w_{i1} \\ w_{i2} \\ \dots \\ w_{iN} \end{bmatrix}$$

式中， $d_{ij}$ 代表了大数据分布节点的聚类中心； $w_{ij}$ 代表第*i*个点对应的大数据挖掘的全局加权值。

利用数据语义分割模型中的信息构建对应的特征提取分析模型系统，在得到特征提取的真实数据信息后，进行数据挖掘处理。通过对STARMA(1,1)网络模型的操作处理，可以获取到可视化的数据分割处理模型系统。随后，将模糊特征聚类的操作方法与模型融合到一起，开始对数据进行统计处理与分析。通过语义分割的方法可以完成特征信息的提取。在*t*时间节点的第*i*个特征点分布的表达公式如下：

$$(w_{1j}, w_{2j}, \dots, w_{ij})$$

式中， $t$ 代表了数据的编码数值； $w_{ij}$ 代表加权系数。

通过模糊语义特征规则集得到对应的自适应加权系数，具体情况如下：

$$STDf_{ij} = \frac{Freq_{ij}}{\max I Freq_{ij}}$$

式中， $Freq_{ij}$ 代表大数据信息在挖掘寻优环节后的模糊约束特征量。

进一步分析后可以确定平台搜索信息素的浓度，具体的表达公式如下：

$$Idf_i = \log(w_{ij}N / n_i)$$

式中， $w_{ij}=tf_{ij}$ ； $N$ 代表大数据挖掘节点的维数； $n_i$ 代表了第*i*个节点对应的信息嵌入维数； $f_{ij}$ 代表了大数据中以节

点*i*和节点*j*进行采样的数据之间的关联信息；*t*代表采样时间的间隔。

### 4.2 语义动态分割与挖掘输出

进一步设计大数据信息的语义动态特征的分析处理模型。首先对大数据信息的统计特征量进行有效地提取并进行分析，以此完成对大数据信息挖掘环节的自适应寻优。具体操作流程如下：

$$D(d_i, d_j) = d_i d_j / \|d_i\| \|d_j\|$$

式中，*d<sub>i</sub>*和*d<sub>j</sub>*均代表了大数据信息在挖掘的过程中所使用的模糊规则特征量。

基于统计信息分析的方法构建大数据信息的挖掘模糊特征分布集，具体表达公式如下：

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

式中，*m*代表了样本数据的具体嵌入维数；*n*代表了分割网格的实际数量；*a<sub>mn</sub>*代表了没有进行挖掘分析操作的大数据信息中有用信息的全部统计幅值；*g<sub>mn</sub>(t)*代表了数据统计的平均值；*n(t)*代表了数据中的干扰项。

对上述分析内容综合分析，能够得到大数据信息的精准挖掘模型，具体表达公式如下：

$$F_j = \sum_{k=1}^n X_{kj}, \quad Q_j = \sum_{k=1}^n (X_{kj})^2$$

式中，*F<sub>j</sub>*代表挖掘输出的特征量；*X<sub>kj</sub>*代表语义分割的关联维数；*Q<sub>j</sub>*代表数据挖掘的语义信息分量。

在对数据挖掘环节展开自适应寻优操作后，能够完成对统计特征量相关信息数据的提取操作，而后采用FCM算法实现大数据的聚类分析与相关处理，以此完成数据挖掘的整个操作流程。

## 5 异构多核平台大数据挖掘算法的仿真实验

为验证论文提出的基于语义分割的异构多核平台大数据挖掘算法的实际应用效果，本章对其进行仿真模拟，在操作过程中使用专业的仿真模拟系统与技术手段，以此分析方法的实用性与可靠性。

在仿真模拟过程中科学设定异构多核平台大数据挖掘的各项参数，具体情况如表1所示。

表1 仿真模拟实验各项参数

项目	设置参数
大数据采样物联网节点：个	120
数据挖掘根节点：个	12
分布空间维数：个	5
数据聚类属性类别：个	10
初始频率（ <i>f<sub>1</sub></i> ）：Hz	1.4
终止频率（ <i>f<sub>2</sub></i> ）：Hz	2.4

基于上述表1各项参数的设置情况进行仿真模拟实验，以此得到数据样本对应的时域分布图，具体情况如图1所示。

以图1中的数据信息作为原始信息进行模拟研究，对

统计特征量进行提取，利用FCM算法对特征提取结果信息进行大数据的聚类处理，进而得到异构多核平台的大数据挖掘直方图。

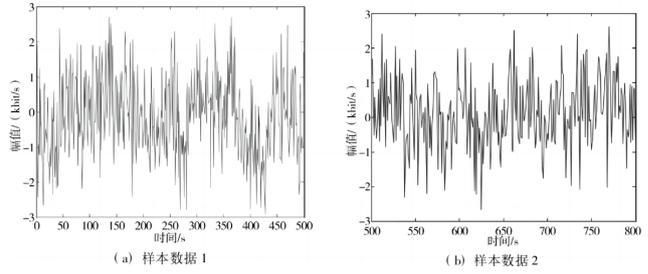


图1 异构多核平台大数据的时域分布

通过大数据挖掘直方图中的信息数据能够发现，论文提出的方法在仿真模拟实验的过程中表现出良好的特征聚敛性，能够进一步提高数据挖掘操作过程对外界因素影响的抵抗能力。

对仿真测试的大数据挖掘结果的真实性与精度进行对比分析，具体情况如下表2所示。

表2 异构多核平台大数据挖掘精度对比分析

迭代次数 使用方法	论文方法	模糊神经网络 方法	PSO 数据挖掘 方法
100	0.912	0.834	0.733
200	0.922	0.853	0.824
300	0.945	0.892	0.892
400	0.985	0.921	0.903
500	0.993	0.953	0.913

对上述各方法在不同迭代次数中的精度结果进行对比分析，能够发现，论文提出的方法具有较高的数据挖掘精度，能够为异构多核平台的大数据挖掘提供有力支持。

## 6 结论

论文在研究过程中以语义分割方法为基础，并将模糊相关性分析等方法应用其中，以此建立了异构多核平台的大数据挖掘模型，可实现对异构多核平台大数据的精准挖掘处理。利用分块区域融合的方法能够实现对大数据的精准定位，利用语义分割的方法能够实现对模糊信息特征的有效提取。构建模糊语义特征规则集完成对异构多核平台大数据挖掘的寻优控制，在决策树模型的支持下完成对大数据挖掘的优化处理。通过本次研究能够得出结论，文中提出的基于语义分割的大数据挖掘方法在实际应用中能够进行较高精准确度的数据挖掘，具有较强的抗干扰性与特征聚敛性。

### 参考文献

- [1] 周贤来.基于语义分割的异构多核平台大数据挖掘算法[J].计算机与现代化,2020(10):4.
- [2] 吴坚.自训练的域适应语义分割算法综述[J].电脑知识与技术:学术版,2023,19(17):19-22.
- [3] 郭红建,陈一飞,梅轶群.基于高维聚类的文本大数据挖掘算法仿真[J].计算机仿真,2023,40(6):499-503.