

# 基于混合算法的医院信息化大数据自动存储方法

董托阳

(哈尔滨医科大学附属第二医院, 黑龙江 哈尔滨 150081)

**摘要:** 考虑到医院信息数据较为庞杂, 导致自动存储效果不佳等问题, 提出基于混合算法的医院信息化大数据自动存储方法。通过混合算法对医院信息大数据中的异常数据进行识别, 以完成医疗数据的预处理过程。再对医疗大数据进行智能化聚类, 实现对医院不同部门的信息需求特征分类要求, 以此更加快速便捷地进行信息的分类存储。对自动分类存储的策略进行论述, 而对每一项存储任务数据资源进行格式规范, 并在建立了多维存储空间之后, 将其通过不同路径分类存入到总数据库中。实验结果表明, 所提方法对于医院信息数据的自动存储过程相比传统方法而言, 数据存储速率更高, 具有较好的并发处理性能, 能够满足研究需求。

**关键词:** 混合算法; 信息化; 大数据; 数据存储

中图分类号: TP311.13 文献标识码: A 文章编号: 1003-7241(2025)04-0066-05

## Method for Automatic Storage of Big Data in Hospital Informatization Based on Hybrid Algorithms

DONG Tuo-yang

(Department Unit of the Second Affiliated Hospital of Harbin Medical University, Harbin 150081 China)

**Abstract:** Considering that the hospital information data is relatively complex, which leads to poor automatic storage effect, this paper proposes a hybrid algorithm based automatic storage method of hospital information big data. The abnormal data in the hospital information big data are identified by the hybrid algorithm to complete the preprocessing process of medical data. Then, the intelligent clustering of medical big data is carried out to realize the classification requirements of the information demand characteristics of different departments of the hospital, so as to make the classification and storage of information faster and more convenient. The strategy of automatic classification storage is discussed, so that the format of each storage task data resource is standardized, and after the establishment of multi-dimensional storage space, it is classified into the total database through different paths. The experimental results show that the proposed method has higher data storage rate than the traditional method for the automatic storage process of hospital information data, indicating that the method in this paper has better concurrent processing performance in the practical application process, and can fully meet the initial research needs.

**Keywords:** hybrid algorithm; information; big data; data storage

### 0 引言

目前, 大多数医院已经开始引入信息技术, 在此基础上, 逐步实现了从门诊到住院全过程的自动化, 极大地简化了医院就诊看诊步骤。但随着我国信息化程度的提高, 数据的数量也在不断增加, 根据医院内部不同的部门不同的数据存储要求, 基于此需要进行不同的分类存储, 而这种储存要求会耗费很大的资源 and 时间。在此背景下, 如何有效对医院信息化大数据进行自动化、高效地分类存储, 是当前研究的热点<sup>[1]</sup>。目前, 我国在大数据的存储技术方面已经有了大量的研究结果, 当前的主流存储方式都基于云计算发挥其自身的特性, 从而有效地避免

了由于关系数据库不能进行横向扩展而造成的数据过载问题。本文以面向大数据的横向、纵向两个尺度为研究对象, 研究了当前主流的并行聚类算法。其中, 面向多核处理器, 图形处理器, 以及可编程逻辑门阵列等。另外, 还对目前流行的几种基于聚类验证的标准算法进行了对比, 但这种方法在应用时, 没有考虑到多维数据的特性。在多维数据处理过程中, 突然增大的任务负荷使得系统的压力难以支持存储任务的正常执行。本项目提出的以 Spark 为基础的数据存储方式, 将其应用于实践中, 以提高其并行性<sup>[2]</sup>。然而, 在实际应用中, 由于其所面临的负荷较大, 所占用的时间也较长, 因此, 在实际应用中, 其压缩性能有待提高。为此, 本项目拟在存储方式中引入混合计算技术, 研究新型的医疗类别大数据自动存储方法。

\*基金项目: 黑龙江医药卫生科技计划项目 (2007-454, 2006-139)

收稿日期: 2023-12-28

## 1 医院信息化大数据自动存储方法

为了实现对医院的信息化管理,本文研究医院信息大数据自动存储方法,整体流程如图1所示。

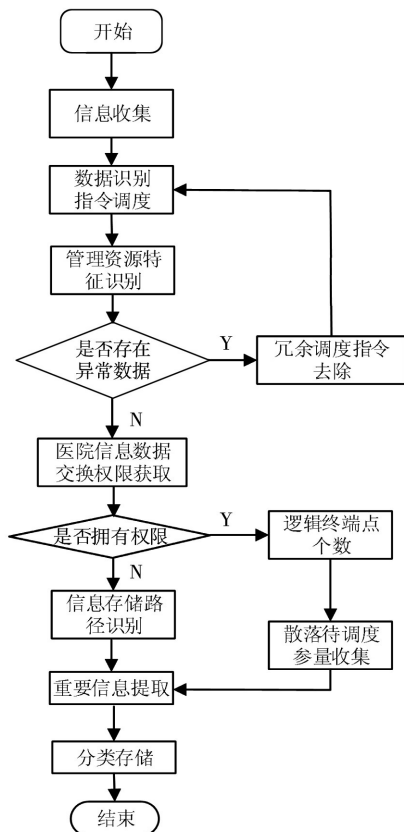


图1 医院信息数据分类存储流程图

本文首先对医院的信息数据进行采集,再对数据进行数据识别和指令调度,判断其中是否存在异常数据。若存在异常数据,则需采用冗余调度指令将其去除;若不存在异常数据,则直接获取医院信息数据的交换权限。在此基础上,对信息存储路径进行识别,提取其中的重要信息,实现分类存储。下面将对此过程展开具体分析。

### 1.1 医院信息异常数据识别算法

在对医院信息数据进行分类存储之前,首先对其中的异常数据进行识别。通过对大数据参数的上、下界分割关系的分析,来判断所存储的信息格式的正确性。在这个基础上,将网格环中的信息参数,快速地传递到数据库中<sup>[3]</sup>。在不考虑过度传输的前提下,基于网格作用系数和分布式信息量化等多种因素共同作用下的混合计算模型依赖度的计算效果。网格作用系数常表示为 $s$ ,其具有显著的时序特征,即在给定的存贮期间 $t$ ,随着总的信息量的增大,数据特征的物理参数将呈现出逐渐增大的趋势<sup>[4]</sup>。若医院信息化条件常表示为 $c$ ,在大数据存储空间 $[D_1, D_2]$ 内,则对医院信息数据混合存储格式特征的依赖程度进行计算,算法为:

$$\varepsilon = 2t|c - s| \int_{i=1}^{j-1,2,3,\dots} \frac{f(n)}{(l - \Delta l) \ln \frac{D_2}{D_1} - 1} \quad (1)$$

在给定了被控医疗资源的时间序列后,该算法中的属性函数和模糊规则等参数条件总是保持不变的。对于一个标准的医疗管理资源数据集,两个邻近的数据始终以同样的计划频率运行,但是,因为数据本身占用的内存不一样,所以各个频率周期的波动峰和波谷也是不一样的。为了更好地确定医院管理资源的最优控制参数,定义 $\omega$ 代表医院数据提取频度, $\mu$ 表示提取数据的波峰起伏的平均值, $\bar{\kappa}$ 表示数据提取波动谷。通过以上几个变量,可以得到最优的医疗数据控制参数的计算结果如下:

$$u = \varepsilon \int_{i=1}^{k=1,2,3,\dots} [l(\omega - 1) + \Delta l \ln \mu] - \bar{\kappa} \quad (2)$$

式中, $A$ 代表模糊调度数据隶属度。

在此基础上对医疗资源的动态调度与传输,该方法既支持现实中的数据传输,又支持虚拟的资源传输。在关联业务单元存在清晰的调度需求时,根据网络的动态变化,将各不忙务资源按可用频谱利用率进行分区,从而实现医疗管理系统中的数据进行初始定位,从而实现一次动态的调度转移<sup>[5]</sup>。针对切换过程中医院信息数据发生的状态变动,将可信度因素引入到实时监测技术中。该方法能较好地解决因虚假数据造成的监测结果不一致的问题。设 $\gamma$ 为医院系统中的正常的可信度; $\gamma'$ 为传输分类过程中检测到的异常时的可信度; $G(x)$ 为数据可信度,则可以求出医院信息数据安全折合系数,算法如下:

$$S(c) = u \sum \frac{G(x)}{g - g'} + 1 \quad (3)$$

基于安全折合系数进行异常数据的识别和消除,其中当 $S(c) \leq 1$ 时说明数据存在异常,基于此需要进一步数据传输安全进行优化,若医院信息数据所占用的频谱率始终保持为 $\eta$ ,信息数据传输量上限也为 $K_{max}$ ,则将医院信息数据传输标准计算过程表示为:

$$H = S(c) - K_{max} \prod_{i=1}^{j=1,2,3,\dots} \eta(\varphi + \theta)^{\frac{1}{2}} \quad (4)$$

式中, $\varphi$ 代表医院管理资源的动态调度传输标准, $l$ 代表虚拟资源传输量的最小取值条件, $\theta$ 代表不同特征数据的动态传输系数。

根据上述算法,进行数据传输可以更好地保证医院信息数据的传输安全,实现对不同类别数据的安全管理和划分,方便后续进行存储操作。

### 1.2 医院信息化大数据特征分类

针对医疗大数据,本项目拟利用混合计算方法对其进行分类,并对其进行分类,以减少对数据进行压缩所需的时间和资源<sup>[6]</sup>。假设医疗机构评级资料集 $Q$ 代表该样

本集合中的样本点数,并求出该样本点 $a$ 与 $b$ 的欧氏距离。公式如下:

$$\lambda = all \left( q_k \sum_{i,j=1,2,\dots,n}^{k-1,2,\dots} (a_i - b_j)^{k-1} \right) \quad (5)$$

式中: $\lambda$ 表示 $a, b$ 间的欧氏距离; $k$ 表示数据存储空间维度, $i$ 表示第 $i$ 维坐标, $j$ 表示类别个数, $q_k$ 表示第 $k$ 维坐标中的特征数据。

基于上述算法进一步设置不同类别数据的初始化特征聚类,设特征聚类的中心集合为 $R$ ,在进行数据特征识别的过程中存在些随机噪声 $s_i$ ,基于此,在噪声环境下的数据类别特征识别函数为:

$$f(n) = \left( \lambda - q_k \sum_{i=1,2,\dots,n}^{k=1,2,\dots} R : k(i-j)(a-b) \right) - s_i \quad (6)$$

大数据的混存形式可分为四大类:并列型、联合型、递进型和交互型<sup>[7]</sup>。其中,并置大数据存储格式是指这样一种信息结构,它的开始和结束节点的类型是完全一样的,在分布式网格环境下,满足此类应用要求的云计算数据信息传输速度非常高,能够将散点信息转换为无需互感节点协作的簇结构。设起始混合信息标度参量常表示为 $n$ ,基于此对特征数据存储路径进行建模,使其与初始分布授权系数具有相同的物理作用<sup>[8]</sup>。设终止分布式权限常表示为 $r_2$ ,在给定的存贮期条件下,可以减小大数据结构中的平均电导率,逆向提升云参数的平均存贮数量。若特征信息标度参量常表示为 $m$ ,为使其与终端分布授权系数具有同样的物理作用,则可以用如下方式来表达大数据的存储关系:

$$\begin{cases} l = g \int_0^{\infty} \frac{\lambda^2 |m+n-1|}{f(n) - e} \\ \Delta l = w \int_0^{\infty} \frac{f(n)}{\lambda^2 |m-n| e} \end{cases} \quad (7)$$

式中, $e$ 代表数据信息初始传输速率, $g$ 代表网格数据信息的平均存储条件, $w$ 代表特征数据信息的平均转化条件。

在不考虑分布式信息网格干扰的情况下,计算聚类数据的最大容量(即存储极限)受初始分布授权和初始混合信息尺度参数共同影响。在给定的数据存储期时间内,可以加速大数据结构中的平均传输率,从而有效地提升存储极限条件<sup>[9]</sup>。数据存储格式是一种信息结构,针对不同特征数据的存储需求需要对存储格式进行分类定义。在分布式网格环境下,数据传输速率需要一直维持在2 000 bps—4 000 bps之间。数据的存储形式是指具有不同的起始和结束两种不同的信息结构,在分布式网格环境下,为满足这种应用要求的数据传输速度非常缓慢,并且需要互连节点协同工作,才能从散点信息向簇结构转变<sup>[10]</sup>。表1列出了混合存储格式定义原则。

表1 混合存储格式定义原理

网格节点特征	信息传输速率/bps	是否需要互感节点配合
起始节点与终止节点类型完全相同	$\geq 6\ 000$	否
起始节点质量明显高于终止节点质量	4 000-6 000	是
终止节点质量明显高于起始节点质量	2 000-4 000	无具体限制
起始节点与终止节点类型完全不相同	0-2 000	是

在此基础上,进一步提出了一种新的混合计算模型,即以此为基础,结合数据定义特征,构建分布式大数据的混合型混合传输管理方法。

### 1.3 医院信息化大数据自动存储的实现

在对医院信息化大数据进行特征分类和异常数据识别后,针对得到的数据进行自动存储处理。根据医院信息化大数据的特征分类结果,可以将不同类型的数据存储在不同的存储介质或存储策略中。

对于关键数据,如病人的实时生命体征数据、重要的医疗影像等,选择SSD磁盘这样的高速存储设备。以此确保数据的快速读取和写入,从而支持实时的医疗决策和操作<sup>[11-12]</sup>。对于访问频率较低、数据量大的非关键数据,如历史医疗记录、日常行政数据等,使用分布式文件系统(hadoop distributed file system, HDFS)进行存储。这样的系统能够横向扩展,为大量数据提供经济高效的存储解决方案。对于很少访问的历史数据进行数据归档操作,将数据从在线存储设备迁移到磁带库等更低成本的存储介质上,进一步降低成本并确保数据的长期保存<sup>[13]</sup>。随着医院业务的变化和数据的增长,之前设定的数据存储策略就需要及时调整。因此,应定期评估现有存储策略的有效性,并根据需要进行调整。例如,某些“非关键”数据可能在某段时间内变得“关键”,这时就需要将其迁移到更高速的存储设备上。不论是关键数据还是非关键数据,都需要确保数据的安全和合规性。对所有数据进行加密存储,并严格管理访问权限,确保只有授权人员能够访问敏感数据。同时,要定期备份数据,防止数据丢失。

考虑到数据的增长速率对于选择数据存储方法至关重要,快速增长的数据可能需要可扩展的存储解决方案,而较慢增长的数据可能更适合静态的存储方法。数据增长速率 $\eta$ 的计算过程可表示为:

$$\eta = \frac{D_m - D_0}{\Delta t} \times 100\% \quad (8)$$

式中, $D_m$ 代表评估周期内的最终数据量, $D_0$ 代表评估周期开始的数据量, $\Delta t$ 代表评估周期的时间长度。

实现医院信息数据的自动存储过程,具体步骤如下:

步骤一:确定评估周期,并记录初始数据量和最终数据量;

步骤二:使用式(8)计算数据增长速率。将初始数据

量、最终数据量以及评估周期时间长度代入公式中进行计算；

步骤三：根据计算得到的数据增长速率，判断数据的增长类型。当数据增长速率超过50%时，需要选择可扩展的存储解决方案，如分布式文件系统或云存储等，以适应数据的快速增长和可伸缩性。当数据增长速率低于50%时，则需选择静态的存储方法，如专用服务器或传统存储设备等，以满足数据的稳定存储需求；

步骤四：自动执行存储策略，将数据按照选定的存储解决方案进行自动存储。确保关键数据存储在高速设备中，非关键数据按照访问频率和容量需求存储在合适的介质中，历史数据归档到低成本存储介质上；

步骤五：定期评估存储策略的有效性。根据医院业务的变化和数据的增长情况，及时调整存储策略。例如，根据需求将数据从静态存储方法迁移到可扩展存储解决

方案，或者根据访问频率变化调整数据的存储位置；

步骤六：确保数据的安全和合规性。对所有数据进行加密存储，并严格管理访问权限，确保敏感数据只能由授权人员访问。同时，定期备份数据以防止数据丢失。

通过以上明确、具体的步骤，医院可以根据数据增长速率的结果，实现医院信息数据的自动存储过程，并确保数据的安全、高效和合规性。

## 2 实验结果分析

为了检验所提出的方法的实用性，设计基于传统神经网络的数据存储方法进行对比试验。在同样的实验条件下，将实验组和对照组控制开关都关闭，并在规定的实验时间内，对实验组和对照组的数值指标进行记录。在实验组的主机上安装了基于全息数字水印技术的医院信息管存系统，而对照组的主机则安装了传统的医院信息管理系统。

表2 实验参数设定

实验组	实验参数	对照组	实验参数
管存应用主机	SR258 1U机架式服务器主机	管存应用主机	SR258 1U机架式服务器主机
内、外机存储条件	8 GB内存、1 TB硬盘	内、外机存储条件	8 GB内存、1 TB硬盘
系统单位管存时长	15 min	系统单位管存时长	15 min
单位时间内的数据存储总量	512 MB	单位时间内的数据存储总量	512 MB
是否允许双向数据传输行为的同时存在	是	是否允许双向数据传输行为的同时存在	是

在进行存储效果监测的过程中，数据存储量、相邻存储节点之间的UDI限制是衡量当前医疗数据自动存储状况的重要指标，通常，当数据容量较大、UDI限制指数较低时，系统主机能够更好地存储海量医院数据，反之，则会降低。表3记录了试验期间实验组和对照组数据存储的真实变化。

表3 医院数据存储量对比表

实验时间/ min	数据存储量/MB		数据容纳总量/ $\times 10^3$ TB	
	实验组	对照组	实验组	对照组
2	261	260	6.8	4.3
4	284	261	7.2	4.3
6	349	269	7.7	4.3
8	375	269	8.0	4.3
10	408	228	8.5	4.1
12	433	207	8.9	3.6
14	456	198	8.9	3.2
16	467	201	8.9	2.8

从表3可以看出，实验组的数据存储随试验时间的增加而呈现出先升高后平稳的态势，而对比的数据存储经过一段时间的平稳后，便出现了持续的衰减，离理想的存储极限值还有很大的差距，一直都在试验组的极值水平之下。进一步对比了相同环境下两种方法在5 000 GB数据的读写速率，将检测数据绘制成图，具体如图2所示。

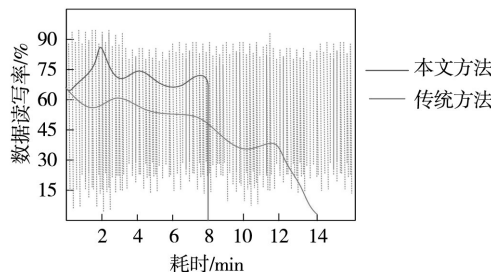


图2 对比实验检测结果

基于图2可知在相同的干扰环境下本文方法对海量医院信息数据的读写速率优于传统方法，且在进行大量信息数据存储的过程中，本分方法耗时明显更低，由此说明基于混合运算的医院信息数据智能存储方法在实际应用过程中具有更高的实用性，可以更好地为医院各个部门的工作开展提供支持，充分满足当前研究要求。

## 3 结束语

医院信息系统的建立依赖于对数据的处理和存储，因此，对医院分级数据的自动化存储进行研究，有助于实现医院的信息化。本项目以医疗机构数据的自动化存储为研究对象，通过对海量数据的分析，提出一种基于混合算法的医院信息数据自动存储方法。在此基础上，针对所提算法性能要求进行了仿真实验，实验表明，基于混合

算法的医院信息数据自动存储方法具有较好的性能,能够较好地解决传统的数据存储方式所遇到的一些问题。然而,受限于研究时间及个体能力,该方法在某些细节上仍有亟待解决的问题,例如,数据的安全问题、数据空间的自动扩充等。在今后的研究中,将对上述具体问题进行深入的分析 and 探讨,并对其进行改进。

#### 参考文献:

- [1] 马小明,刘永涛,陈晓楠.基于智能聚类的医院分类大数据自动存储方法[J].自动化技术与应用,2022,41(11):84-87.
- [2] 戴飞,程颖,焦攀科.基于无线传感器网络的远程会诊信息自动存储方法研究[J].自动化技术与应用,2022,41(9):128-131.
- [3] 吴丹,李鹏,杜明超.基于云计算的医疗大数据自动存储与共享系统[J].自动化技术与应用,2022,41(7):54-57.
- [4] 任浩.医院备份数据库自动恢复和监测的研究与应用[J].电子技术与软件工程,2021(20):178-181.
- [5] 马海峰,王俊华,薛庆水,等.用户可实时撤销的云存储数据完整性验证方案[J].计算机工程与设计,2023,44(10):2943-2949.
- [6] 任云龙,马新江,许效源.医院供配电设备在线状态监测与预警系统研究[J].中国设备工程,2023(19):189-191.

[7] 胡圣,秦昆,喻雪松,等.全球地理多元流网络数据存储与可视化系统的设计与实现[J].测绘地理信息,2023,48(5):155-158.

[8] 王意.面向混合存储的时空大数据存储策略研究[J].地理空间信息,2023,21(9):10-13.

[9] 张益兵.基于区块链技术的海量Web数据关联存储系统设计[J].电子设计工程,2023,31(19):157-161.

[10] 刘飞飞.基于区块链的超高压变电站大数据存储研究[J].智能城市,2023,9(9):36-38.

[11] 张彦芳.基于大数据和区块链技术的农机远程监测系统[J].农机化研究,2023,45(9):197-200.

[12] 邹文景,唐良运,甘莹,等.基于Hadoop技术的物联网大数据同步存储系统设计[J].电子设计工程,2023,31(18):114-117,122.

[13] 危华明,廖剑平.海量数据存储中云服务器性能加速方法仿真[J].计算机仿真,2023,40(5):515-519.

作者简介:董托阳(1987-),男,硕士,研究实习员,研究方向:医院信息化。

(上接第56页)

伺服控制系统在实际应用过程中,根据图中曲线分析,支撑力随时间线在1 070 kN和1 100 kN之间的变化,保持了一个平稳控制,结果表明:系统稳定,能够保证基坑的安全。

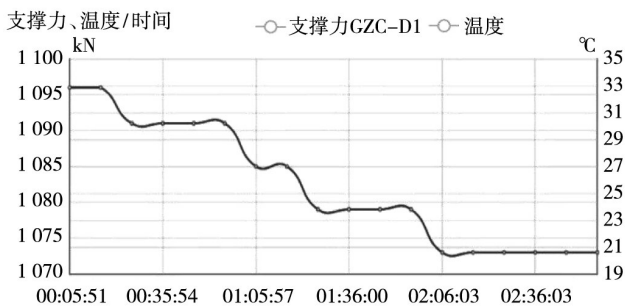


图9 现场监控

## 4 结束语

在基坑工作中,传统的钢支撑固定支撑而不能随意施压泄压,导致工程中需要考虑诸多因素且施工困难,而基于PLC的钢支撑智能伺服系统可以实时监控钢支撑的受力情况,利用可靠性高、适应性强的PLC硬件以及功能强大的编程和运算命令,可以对设备进行伺服控制、冗余控制等。基于PLC的钢支撑智能伺服系统为技术人员开展基坑工作提供便利,对钢支撑的信息、状态实现高效管理,解决了过往钢支撑存在的问题,在一定程度上提高了工作效率。

#### 参考文献:

- [1] 刘成. IPS预应力鱼腹梁钢支撑在深基坑工程中的受力机理与变形控制研究[D].合肥:安徽建筑大学,2018.
- [2] 唐玲玉.钢支撑对深基坑围护桩变形影响及优化研究[D].青岛:山东科技大学,2018.
- [3] 张忠苗,房凯,刘兴旺,等.粉砂土地铁深基坑支撑轴力监测分析[J].岩土工程学报,2010,32(S1):426-429.
- [4] 卞国强.钢支撑伺服系统应用的若干问题及对策措施研究[J].山西建筑,2018,44(15):70-71.
- [5] 王年朋.基于PLC的顶驱智能防碰控制系统的研究与开发[D].重庆:重庆大学,2019.
- [6] 侯倩祯.基于PLC三相异步电机控制系统设计[J].矿业装备,2021(5):240-241.
- [7] 梅幼亚,姚毅,于洋,等.基于西门子S7-300与威纶通触摸屏的换热站自控系统[J].山东工业技术,2015(13):220-221.
- [8] 杨振波.基于PLC的自动化包装生产线集成控制系统设计[J].自动化技术与应用,2022,41(4):23-27.

作者简介:彭昊(1998-),男,在读硕士,研究方向:控制工程。

通信作者:刘显晖(1972-),男,教授级高级工程师,研究方向:预应力技术,智能建造及装备。