

基于模糊处理的医疗信息数据存储系统的设计

郑 健

(哈尔滨医科大学附属第一医院 信息中心, 哈尔滨 150000)

摘要: 为了提高医疗技术水平, 加快医务人员对医疗信息数据的分析, 使医疗信息数据运行更加顺畅, 减少数据存储所占空间, 需要对医疗信息数据存储系统进行设计; 当前的医疗信息数据存储系统对医疗信息数据进行存储时, 利用 FPGA 构建系统硬件, 以硬件为基础将医疗信息数据存储, 但在存储过程中, 没有将冗余数据清除, 导致存储空间易满, 容量变小, 存在正常医疗信息数据无处安放的问题; 为此, 提出一种基于模糊处理的医疗信息数据存储系统设计方法; 该方法首先对医疗信息数据的来源进行统计, 根据数据来源实现医疗数据的硬件构造, 然后利用医疗信息数据中特征相似数据间的中介点实现数据聚类, 完成对医疗信息数据的聚类存储, 最后采用高斯混合模型对冗余医疗数据进行特征分析, 以分析结果为基础, 利用分数阶 Fourier 变换对冗余医疗信息数据进行删除操作, 由此完成了基于模糊处理的医疗信息数据存储; 实验结果证明, 所提方法增加数据存储空间容量, 减少对冗余数据的存储, 加快医疗信息数据的存储速度, 提高数据存储精度, 为该领域研究发展提供强有力的依据。

关键词: 模糊处理; 医疗信息数据; 存储系统设计

Based on Fuzzy Processing of Medical Information Data Storage System Design

Zheng Jian

(First Affiliated Hospital of Harbin Medical University Information Centre, Harbin 150000, China)

Abstract: In order to improve the level of medical technology, speed up the medical staff of medical information data analysis, the medical information data run more smoothly, reduce the data storage of space, need for medical information data storage system design. The current medical information data storage system of medical information data for storage, use FPGA to build the system hardware, on the basis of the hardware to medical information data storage, but in the process of storage, there is no clear the redundant data, leading to storage space is full, smaller capacity, has the problem of the normal medical information data is placed. For this, put forward a kind of medical information data storage system based on fuzzy processing design method. This method first source of health information data statistics, according to the data source to realize the hardware structure of medical data, then use mediation between the characteristics of medical information data in the similar data points for data clustering, clustering of medical information data storage, and finally by using the gaussian mixture model to analyzing characteristics of medical data redundancy, on the basis of the analysis results, using fractional Fourier transform to delete redundant medical information data, thus complete the medical information data storage based on fuzzy processing. The experimental results show that the proposed method increases the data storage capacity, reduces the redundant data storage, to speed up the medical information data storage, improved the precision of data storage, the research in the field development provides a strong basis.

Keywords: blurred; medical information data; storage system design;

0 引言

目前, 随着互联网水平和科学技术的发展, 医疗信息数据在医疗届的应用范围广泛, 例如对医院门诊、骨科、内科、眼科等部门都有良好的辅助作用^[1]。医疗信息数据不仅从侧面提高医院医疗水平, 而且便于医务人员从中发现医疗方面的不足和漏洞, 并及时改正。随着当前医疗信息数据逐渐增加, 医疗信息数据的存储受到了社会各界的高度重视和广泛研究^[2-3]。因为医疗信息数据具有复杂性较强、灵活度较高、数量较大, 读取时间较长等特性, 所以需要对其存储

系统进行设计^[4]。大多数的数据存储系统对医疗信息数据进行存储时, 无法对其进行低损耗、低误差、高精度、高效率地存储, 导致医疗信息数据存储过程中, 经常出现恶意数据泛滥、医疗数据来路不明、存储用时较长的问题^[5]。在这种情况下, 如何提高医疗信息数据存储速度, 增加数据存储的稳定性成为了当前急需解决的问题。基于模糊处理的医疗信息数据存储系统设计方法, 对医疗信息数据进行安全可靠地存储, 是解决上述问题的有效途径^[6]。此问题受到了医疗数据存储系统研究人员的钻研, 同时也研究出了很多优秀的方法^[7]。

文献 [8] 提出了一种基于 FPGA 的医疗信息数据存储系统设计方法。该方法首先利用 Hadoop 平台, 将医疗信息数据分割成大小相同的数据块, 然后通过多副本的形式将数据块

收稿日期:2017-04-11; 修回日期:2017-04-21。

作者简介:郑 健(1988-), 男, 黑龙江哈尔滨人, 初级工程师, 主要从事网络安全方向的研究。

分布存储在不同存储系统位置，最后采用 FPGA 实现数据存储系统更好的容错机制，从而完成医疗信息数据持久化存储。该方法下的医疗信息数据存储更为持久，但是存在存储时间较长的问题。文献 [9] 提出了一种基于 CC2530 的医疗信息数据存储系统设计方法。该方法首先利用数据采集模块对医疗信息数据进行大规模的采集，然后将采集的数据经过滤波放大后，采用 CC2530 单片机实现数据传输，最后依据 MicroSD 卡的植入完成对医疗信息数据的存储。该方法虽然用时较短，但是存在存储效率偏差大的问题。文献 [10] 提出了一种基于 C51 的医疗信息数据存储系统设计方法。该方法首先利用 FPGA 作为存储系统的控制芯片，然后以数据存储芯片 NANDFlash 作为医疗信息数据的存储媒介，采用 A/D 芯片对各种医疗信息数据进行转换，最后将所有被转换的医疗信息数据统一存储至数据库中。该方法下的数据存储精度较高，但是过程过于繁琐。

针对上述产生的问题，提出一种基于模糊处理的医疗信息数据存储系统设计方法。该方法首先对医疗信息数据的来源做出了分析，以分析的结果为依据对医疗信息数据存储系统的硬件进行构造，然后以硬件构造为基础，利用 MPI 对医疗信息数据进行聚类，聚类过程中采用相似数据间的中介点将相似数据连接，从而实现数据的聚类存储，最后依据对冗余医疗信息数据特征的分析，将冗余医疗信息数据删除，实现了医疗信息数据的有序存储，完成本文所提方法下的数据存储系统设计。仿真实验证明，所提方法可以低成本、高精度地对医疗信息数据进行存储。

1 基于模糊处理的医疗信息数据存储系统设计方法

1.1 医疗信息数据存储系统设计

由于医疗信息数据中本身存在恶意数据，而在医疗信息数据存储过程中，恶意数据的来源很模糊，所以需要利用医疗信息数据来源模块对其进行具体化的分类，分类操作有利于医疗信息数据的高效存储。图 1 为医疗信息数据来源结构示意图。

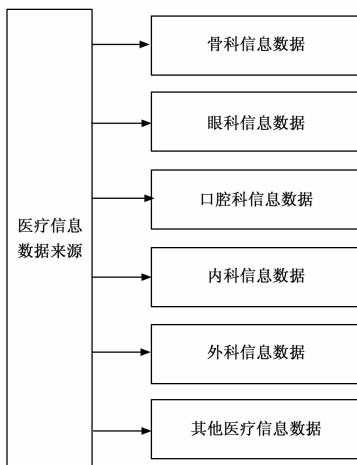


图 1 医疗信息数据来源结构

通过图 1 可以看出，医疗信息数据来源大致分为六块：骨科信息数据、眼科信息数据、口腔科信息数据、内科信息数据、外科信息数据及其他医疗信息数据。当医疗信息数据进行存储时，如果出现恶意数据，医务人员通过数据来源模块可以清晰地分辨出恶意数据来源，并及时做出处理。图 2 是医疗信息数据存储系统的构造。

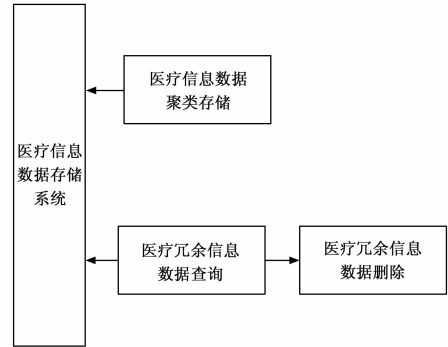


图 2 医疗信息数据存储系统设计

图 2 中医疗信息数据存储系统分为医疗信息数据聚类存储和医疗冗余信息数据查询，也就是按照冗余信息数据的特征辨别冗余数据存在，如若发现冗余数据便对其进行删除操作，有利于节省医疗信息数据存储系统的存储空间。

1.2 医疗信息数据聚类存储

以 2.1 中的各项信息数据为基础，利用 MPI 对医疗信息数据进行聚类存储，聚类存储过程中依据医疗信息数据中特征相似数据间的中介点实现数据聚类，由此完成对医疗信息数据的聚类存储。

假设，医疗信息数据网是 $M = \{E, K\}$ ，式中， E 代表医疗信息数据节点集合， K 代表数据节点之间边的集合。医疗信息数据节点代表数据存储系统中的元素，数据节点之间的边代表元素间的联系。若将整个医疗信息数据库当作由若干个社团构成，也就是由若干个特征相同的数据类构成，则社团内部的数据节点连接的比较紧凑，社团之间数据节点的连接较为稀疏。

假设 x 和 y 分别代表医疗信息数据网 M 中的两个顶点， β_x 代表顶点的结构， F 是 E 中的一个子集，且该子集不为空， α 和 δ 分别代表医疗信息数据中的聚类对象， O 代表满足 α 和 δ 聚类的条件。则医疗信息数据网 M 中的顶点结构相似度为：

$$\kappa(x, y) = \frac{|\beta(x) \cap \beta(y)|}{\sqrt{|\beta(x)| |\beta(y)|}} \quad (1)$$

其中， $\kappa(x, y)$ 代表顶点 x 和顶点 y 间的相似度。如果医疗信息数据节点与该数据邻居的顶点结构相似，则得到两者的关系式为：

$$G_\alpha(x) = \{y \in \beta(x) / \kappa(x, y) \geq \alpha\} \quad (2)$$

其中， $G_\alpha(x)$ 代表与顶点 x 相似度大于等于 α 的邻居数据集合。假设一个医疗信息数据节点有很多相似度大于等于 α 的邻居数据节点，则该数据节点是社团核心节点，则核心节点的表达方式为：

$$CORE_{\alpha, \delta}(x) \Leftrightarrow |G_\alpha(x)| \geq \delta \quad (3)$$

其中, $CORE_{\alpha,\delta}(x)$ 代表医疗信息数据节点是数据对象 α 和 δ 核心节点的值。通过核心节点值对医疗信息数据进行聚类, 聚类过程中利用特征相似数据间的中介点实现数据聚类, 其公式可表示为:

$$HUB_{\alpha,\delta}(x) \Leftrightarrow \quad (4)$$

其中, $HUB_{\alpha,\delta}(x)$ 代表医疗信息数据中特征相似数据间的中介点值, 式中, x 不属于任何团体, 假设 E 中的医疗信息数据节点至少分布在两个或者两个以上的团体时, $HUB_{\alpha,\delta}(x)$ 可以当做医疗信息数据聚类的中介点, 本文不对医疗信息数据节点分布在两个以下团体的情况做研究。由该中介点完成对医疗信息数据的聚类, 将聚类后的数据有序的存储至医疗信息数据库中。

1.3 冗余医疗信息数据的查询与删除

要完成冗余医疗信息数据删除, 首先要对其进行查询, 冗余信息数据的特征分析可以使查询效果更佳。本文利用高斯混合模型对冗余医疗数据进行特征分析, 具体过程如下。

假设, 将冗余医疗信息数据特征区间定义为:

$$\lg(i, j, f) = \{D(i, j) / (216 / f)\} \quad (5)$$

其中, i 代表冗余医疗信息数据特征数目, j 代表冗余医疗信息数据特征区间数目, f 代表对冗余医疗信息数据进行查询时的控制参数, 单位为 h, 实验证明当此参数控制在 0.07 ~ 0.08 时, 数据存储效率最高。D 代表删除冗余医疗数据中一常数单位。

为了使冗余数据查询结果更清晰, 根据冗余医疗信息数据的特征区间, 对冗余医疗信息数据特征集进行计算:

$$\lg_{uv}(i, j) = \lg_v(i, j) * s_{\sigma_f} \quad (6)$$

其中, \lg_{uv} 代表冗余医疗信息数据特征集阈值, \lg_v 代表冗余医疗信息数据特征子集, s 代表基于模糊处理存储的滤波函数, s_{σ_f} 代表冗余医疗信息数据的特征核, 由上式得到冗余医疗信息数据的特征配置为:

$$p_i(f) = q_i(f) + k_i(f) \quad (7)$$

其中, p 代表冗余医疗信息数据的特征配置, q 代表冗余数据查询质量, k 代表冗余医疗信息数据的特征配置系数, 以冗余医疗信息数据特征配置为基础, 假设冗余数据删除过程中, 删除干扰值 $z(f)$ 为 0 时, 则以此配置为中心, 得到冗余数据特征的展开结果:

$$a_{p/q,k} = a_{p,q} + \sum_{j=1}^i z(f) \quad (8)$$

其中, a 代表冗余数据特征的展开值, 由式 (8) 可知冗余医疗信息数据特征展开结果为验证医疗信息数据库内的医疗信息数据, 是否为冗余数据的验证指标, 通过对医疗信息数据特征展开结果的压缩处理, 获得冗余数据压缩特征编码。为了避免习惯性对冗余医疗信息数据删除的影响, 对该特征编码进行量化分解, 其计算方式为:

$$R_n = \sum_{j=1}^{i=1} a_{p,q} \sum_k a_{p/q,k} \quad (9)$$

其中, R_n 代表数据特征编码量化分解值, 利用量化分解的结果完成对冗余医疗信息数据的分析, 当 $R_n \geq 1$ 时, 则正在进行验证的医疗信息数据为冗余数据, 直接删除; 当 $R_n < 1$ 时, 则正在进行验证的医疗信息数据不是冗余数据, 可直接

进行存储, 由冗余医疗信息数据的特征分析过程完成冗余数据的查询。

以上述结果为依据, 采用分数阶 Fourier 变换对冗余医疗信息数据进行删除操作。具体过程如下。

设置冗余医疗信息数据的训练样本集 $S = [S_1, S_2, \dots, S_i, \dots, S_c]^T$, 在基于模糊处理的医疗信息数据存储环境中, 为了保障医疗信息数据正常存储的可靠性, 必须得到存储系统中冗余医疗信息数据流的离散分数阶 Fourier 逆变换值, 该值的表示方式为:

$$b = N_{-q} \times S \quad (10)$$

其中, b 代表冗余医疗信息数据流的离散分数阶 Fourier 逆变换值, $N_{-q} = N_{-q}^H$ 代表离散分数阶 Fourier 逆变换指标, 假设对冗余医疗信息数据删除过程中, 背景噪声具有非高斯性, 则其离散分数阶 Fourier 3 阶累积量切片 $c_3(\eta)$ 表达式为:

$$c_3(\eta) = \epsilon \sum_{j=0}^{\infty} s(j) s^2(j + \eta) \quad (11)$$

其中, ϵ 代表医院客户端点的医疗信息数据宽带, η 代表冗余医疗信息数据重构时延, 以上述结果为依据, 利用分数阶 Fourier 变换法得到冗余医疗信息数据流删除函数:

$$U_k = [u_{k1}, u_{k2}, \dots, u_{kj}, \dots, u_{ki}] \quad (12)$$

其中, U_k 代表冗余医疗信息数据流删除函数, u 代表冗余医疗信息数据流删除函数中的函数子集, 综上分析, 采用 3 阶累积量切片, 将医疗信息数据划分为若干个块, 根据冗余医疗信息数据特征的分析, 对每个数据块中的冗余数据进行彻底删除。

2 仿真实验结果与分析

为了证明基于模糊处理的医疗信息数据存储系统设计方法的可实践性, 需要进行一次仿真实验。在 Linux 的环境下搭建医疗信息数据存储实验仿真平台。实验数据取自于北京解放军医院总部的医疗信息数据存储系统, 利用本文所提方法对实验数据进行存储, 由此观察其整体有效性。表 1 为在数量相同的医疗信息数据下, 文献 [8] 所提方法、文献 [9] 所提方法和文献 [10] 所提方法与本文所提方法, 医疗信息数据存储时间 (s) 的对比。

表 1 不同方法下数据存储时间的对比

数据存储系统 设计方法	医疗信息 数据/万个	数据存储 时间/s
FPGA	100	15
CC2530	200	12
C51	300	20
模糊处理	400	6

通过对表 1 的分析, 文献 [8]、文献 [9]、文献 [10] 所提方法与本文所提方法, 在医疗信息数据量相同的情况下, 本文所提方法存储所用时间远远低于文献所提方法, 相比较之下, 文献 [10] 所提方法与本文所提方法存储时间相差最大, 这主要是因为利用本文方法进行数据存储时, 对冗余医疗信息数据进行了删除操作, 节省了存储时间, 提高了存储精度, 证明了本文所提方法的可行性较强。表 2 是文献 [8]

所提方法与本文所提方法，冗余医疗信息数据删除时间（s）的对比。

分析表 2 可知，本文所提方法在冗余医疗信息数据删除方面明显优于文献 [8] 所提方法，冗余医疗信息数据的删除，直接影响到数据的存储效率和时间，删除的越快，表示存储系统的整体性能越好，本文所提方法在删除冗余医疗信息

表 2 不同方法下冗余医疗信息数据删除时间对比

冗余数据量/ 万个	文献 [8] 方法删除 时间/s	本文方法删除 时间/s
1 000	8	4
2 000	11	6
3 000	15	7
4 000	18	9
5 000	22	11

数据时，先利用高斯混合模型对冗余数据特征进行了透彻的分析，然后采用分数阶 Fourier 变换对冗余医疗信息数据进行删除，此步骤加快了数据存储速度。图 3 为冗余医疗信息数据查询的控制参数 f 取值范围对数据存储效率（%）的影响。

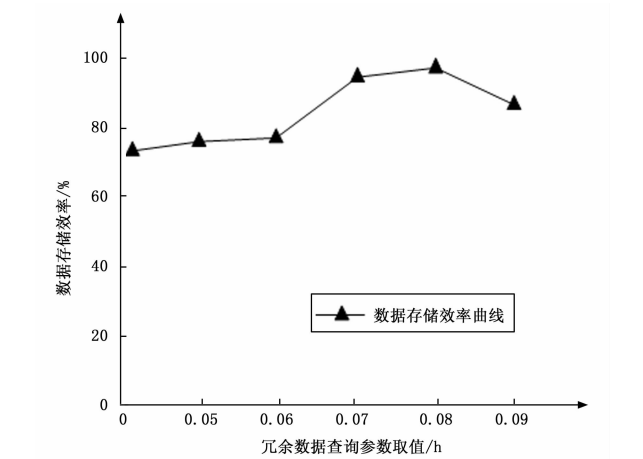


图 3 冗余数据查询的控制参数对存储效率的影响

从图 3 中可以看出，冗余医疗信息数据查询的控制参数 f 取值范围对医疗信息数据的存储效率有很大影响，当冗余数据查询控制参数在 0.05~0.06 时，医疗信息数据存储效率曲线虽然比较平稳，但存储效率并不高，在 80% 以下，与之相比当冗余数据查询控制参数在 0.07~0.08 时，数据存储效率明显升高，基本处于 97%~98%，该数据进一步证明了本文所提方法具有良好的可行性和可靠性。图 4 是文献 [8]、文献 [9]、文献 [10] 所提方法与本文所提方法数据聚类时间（s）对比。

由图 4 可知，文献 [8]、文献 [9]、文献 [10] 所提方法与本文所提方法在数据聚类时间上，都是随着数据量的增加而持续增长的，在相同数据量聚类时，本文所提方法所用时间明显较短，因为基于模糊处理的医疗信息数据存储系统设计方法中，利用了 MPI 对医疗信息数据进行聚类存储，节省了数据聚类时间，提高了医疗信息数据存储精度，为数据存储领域的后续发展指明了方向。

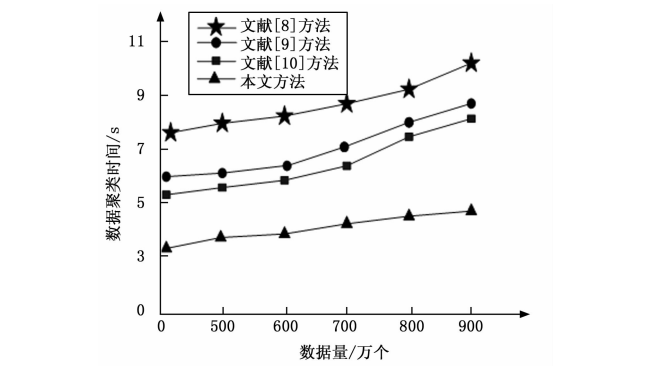


图 4 不同方法下数据聚类时间对比

仿真实验证明，所提方法可以安全可靠地对医疗信息数据进行存储，减少了医疗信息数据存储时间，提高了医务人员对医疗信息数据漏洞的分析能力，降低冗余数据对医疗信息数据存储系统的干扰，为医疗界的发展提供可靠依据，对数据存储系统设计领域有重要的借鉴意义。

3 结束语

采用当前方法对医疗信息数据进行存储时，无法对其进行高精度、低误差、稳定可靠地存储，存在数据存储空间易满，存储系统运行有延迟，数据存储效率低等问题。本文提出一种基于模糊处理的医疗信息数据存储系统设计方法。并通过实验仿真证明，所提方法可以高效率地对医疗信息数据进行存储，可行性较强，对医务人员医术的提高起到了辅助作用，为该领域的深造钻研提供支撑，成为该领域发展的重要旗帜。

参考文献：

[1] 张琳, 谭军, 白明泽. 基于 MongoDB 的蛋白质组学大数据存储系统设计 [J]. 计算机应用, 2016, 36 (S1): 232-236.

[2] 陈亨玉, 钱慧. 嵌入无损编码的海量视频数据存储系统设计 [J]. 电视技术, 2016, 40 (4): 52-55.

[3] 姜德, 马游春, 王悦凯, 等. 高速数据同步存储系统设计 [J]. 电子器件, 2016, 39 (6): 1421-1424.

[4] 姜学东, 孙海民. 大数据存储中的优化架构结构的设计与实现 [J]. 现代电子技术, 2016, 39 (24): 66-70.

[5] 王善明, 严迎建, 郭朋飞, 等. 基于国产 SOC 的数据加密存储系统设计 [J]. 电子技术应用, 2015, 41 (11): 34-37.

[6] 任莹晖, 纪会敏, 杜勇, 等. 切削刀具综合管理系统的数据采集和存储设计研究 [J]. 计算机应用研究, 2016, 33 (10): 3031-3035.

[7] 刘博伟, 黄瑞章. 基于 HBase 的金融时序数据存储系统 [J]. 中国科技论文, 2016, 11 (20): 2387-2392.

[8] 雷德龙, 郭殿升, 陈崇成, 等. 基于 MongoDB 的矢量空间数据云存储与处理系统 [J]. 地球信息科学学报, 2014, 16 (4): 507-516.

[9] 李金猛. 基于 FPGA 的数据采集存储系统的设计与实现 [J]. 电子设计工程, 2016, 24 (13): 85-87.

[10] 胡晓峰, 张亮红, 刘文怡, 等. VC 多线程流水线在数据存储系统中的设计与实现 [J]. 电子器件, 2016, 39 (4): 964-967.