

基于 GA-BP 的医疗数据访问风险评估系统设计方法

赖舒婷, 温煜

(梅州市人民医院, 广东 梅州 514000)

摘要: 为提高医疗数据访问风险评估的精准度与效率, 通过结合遗传算法 (genetic algorithm, GA) 与反向传播神经网络 (backpropagation, BP) 设计一套医疗数据访问风险评估系统。该系统采用 ZigBee 协议栈和 B/S 架构, 包含底层软件、中央控制、数据集成处理、风险状态预测及人机交互模块。利用 VIX 总线传输技术和外设总线调度方法, 系统实现了医疗数据的实时融合与数据库集成。通过 GA-BP 算法, 系统能有效提取医疗数据访问特征并进行风险分段预测。实验测试表明, 该系统显著提高了医疗数据访问的实时调度与风险控制能力, 从而增强了数据交互的准确性和风险评估的可靠性。

关键词: GA-BP; 医疗数据; 访问; 风险评估; 分段预测

中图分类号: TP183 **文献标识码:** A **文章编号:** 1003-7241(2025)09-0134-05

Design Method of Medical Data Access Risk Assessment System Based on GA-BP

LAI Shuting, WEN Yu

(Meizhou People's Hospital, Meizhou 514000, China)

Abstract: This article aims to improve the accuracy and efficiency of medical data access risk assessment. A medical data access risk assessment system is designed by combining genetic algorithm (GA) and backpropagation (BP) neural network. The system adopts the ZigBee protocol stack and B/S architecture, including underlying software, central control, data integration processing, risk state prediction, and human-machine interaction modules. By utilizing VIX bus transmission technology and peripheral bus scheduling methods, the system achieves real-time fusion of medical data and database integration. Through the GA-BP algorithm, the system can effectively extract access features of medical data and perform risk segmentation prediction. Experimental tests show that the system significantly improves the real-time scheduling and risk control capabilities of medical data access, thereby enhancing the accuracy of data interaction and the reliability of risk assessment.

Keywords: GA-BP; medical data; access; risk assessment; segmented prediction

0 引言

随着智能化医疗信息管理系统的发展, 利用人工智能技术对医疗数据的访问风险进行评估成为可能。医疗数据访问风险主要指在医疗信息管理系统中, 由于各种因素 (如黑客攻击、系统漏洞、人为误操作等) 导致医疗数据可能被非法访问、篡改、泄露的风险。这些风险不仅威胁着患者的隐私安全, 还可能影响医疗机构的正常运行和医疗决策的准确性。

通过构建医疗信息管理大数据分析模型, 并结合集成信息调度和融合聚类分析方法, 可以建立有效的医疗数据访问风险评估模型^[1-2]。在日常医疗实践中, 医疗数据的访问风险可能发生在多个环节。例如, 在远程医疗活动中, 患者的检验报告、诊断结果、用药信息等涉及个人隐私的健康医疗信息需要通过互联网进行传输和存储, 若网络被不明身份人员接入或存在安全漏洞, 则数据面临着被窃取、篡改的风险。此外, 医疗中心涉及跨机构

数据汇聚, 集中存储着大量敏感信息, 如果没有建立相应的存储安全保障机制, 则可能面临非法登录、越权访问等风险。许多学者进行了研究, 赵东升研究基于物元可拓模型的入侵风险评估^[3]。依据实际的网络运行情况, 选取入侵风险评估指标, 建立指标体系, 建立风险评估物元可拓模型。以确定最终的入侵风险。李依霖研究隐私信息入侵风险评估^[4]。根据风险度来评估复杂网络大规模隐私信息被入侵的风险等级, 采用熵权系数法对评判集中的向量进行计算, 以获得风险因素的综合权值。实现对复杂网络大数据信息入侵风险评估。但是上述方法局限性较大, 不能适用于医疗数据研究。

对此, 本文设计了基于 GA-BP 的医疗数据访问风险评估系统, 以实现医疗数据访问风险评估系统在风险评估和数据访问查准能力方面的提升。

1 系统总体构架和大数据分析

1.1 医疗数据访问风险评估信息系统总体构架

采用语义抽取方法进行医疗数据访问风险评估的信

息融合和调度,采用RFID射频标签进行医疗数据访问风险评估信息的二维码标签识别。采用XML模型和Zig-bee技术进行软件开发设计和组网设计^[5]。总体结构如图1所示。

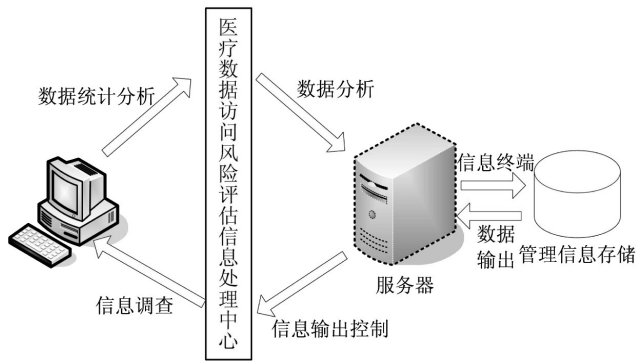


图1 系统总体设计构架

根据图1所示的医疗数据访问风险评估信息处理模型,提取医疗数据访问风险评估信息的语义相关性特征量。

1.2 系统功能模块组件分析

首先采用ZigBee协议栈和B/S构架模型实现对医疗数据访问风险评估的底层架构设计。医疗数据访问风险评估包括底层软件模块、中央控制模块、数据集成处理模块、风险状态预测模块及人机交互模型^[6]。将系统分为设备层、数据层、展现层、应用层,得到医疗数据访问风险评估系统的功能结构组件如图2所示。

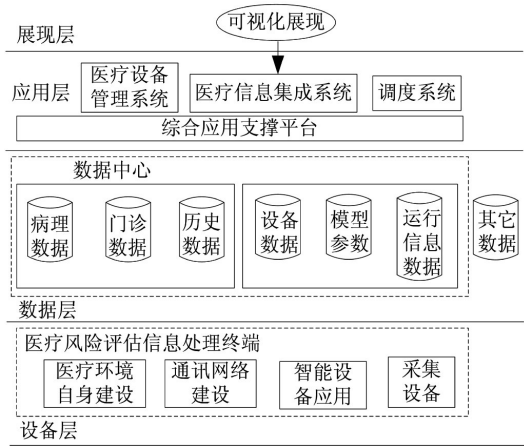


图2 医疗数据访问风险评估系统的功能结构组件

根据图2的医疗数据访问风险评估系统的功能结构组件,利用计算机、通信技术和传感器RFID标签识别技术进行医疗数据访问的过程控制和模块化的开发设计^[7]。

2 医疗数据访问风险评估的算法设计

2.1 医疗数据访问的信息融合处理

在上述构建的医疗数据访问风险评估系统的总体结构模型的基础上,建立医疗数据访问风险评估信息管理系统的IETF的会话协议(session initiation protocol, SIP)^[8],并进行医疗数据访问风险评估信息融合,得到的医疗数据访问风险评估信息管理字结构见表1。

表1 医疗数据访问风险评估信息管理字结构

表结构字段	标识符	字结构长度	数据最大集	表结构字段	标识符	字结构长度	数据最大集
姓名	XM-001	944	328	表头流水号	ID-010	3 910	1 835
集成信息	IN-002	1 058	393	医疗信息分布权重	ST-011	3 096	1 549
教学信息	BL-003	998	360	日程任务	KR-012	3 484	1 721
医疗数据项目	XY-004	1 605	590	模糊特征匹配	RW-013	5 438	2 569
簇聚类	SH-005	1 812	786	关联度	TH-014	4 578	2 323
语义	CT-006	1 564	655	在线辅助治疗状态	SG-015	5324	2 581
文本	SS-007	2 819	1 311	输出	SD-016	944	328
标签运行参数	PO-008	2 107	1 032	状态管理	REV-017	1 058	393
医疗资源关联	ZS-009	2 224	1 075	-	-	-	-

根据对医疗数据访问风险离散变量的集合的概率分布式重组,计算不同时刻医疗数据访问风险的出力聚类参数,以衡量聚类质量^[9]。提取序列隐藏关系,并采用二乘拟合方法,得到医疗数据访问风险评估的统计学模型为

$$f(V) = \frac{1}{\sqrt{2\pi}\sigma_v} \exp\left(-\frac{V^2}{2\sigma_v^2}\right) \quad (1)$$

式中, σ_v 为 C_i 和 C_j 医疗数据访问风险评估的聚类簇。根据医疗数据访问风险评估状态分布的差异性融合标记,按照最佳聚类数N展开医疗数据访问风险聚类的空间维数,采用相似的聚类簇重标记方法,建立医疗数据访问风

险异构特征检测,得到医疗数据访问风险评估动态修正的波动参数为

$$s = \frac{c-v}{c+v} = \left(1 - \frac{v}{c}\right) \left(1 - \frac{v}{c} + \left(\frac{v}{c}\right)^2 - \dots\right) = 1 - \frac{2v}{c} + 2\left(\frac{v}{c}\right)^2 + \dots \quad (2)$$

式中, c, v 分别表示医疗数据访问风险筛选聚类簇参数。结合任务要求和数据特征分配,分析医疗数据访问风险评估时序数据的规则集,基于数据动态修正和自适应反馈调节,得到风险评估的约束图谱特征量为

$$\int \left| \langle g, U(x)g \rangle \right|^2 d\mu(x) < \infty \quad (3)$$

采用三层Stacking模型建立医疗风险数据访问的潮流模型,并采用二值化图谱特征挖掘方法提取医疗数据

访问风险关联特征量^[10],知识图谱扰动分布参数为

$$S_T(x, x') = S_n(x, x') *_{G_1} \dots *_{G_n} S_2(x, x') *_{G_2} S_1(x, x') \quad (4)$$

式中, $*_{G_i}$ 表示医疗数据可视化知识图谱的三维空间分布集。

结合小波域特征分解方法实现对医疗数据访问风险的可视化量化重构,结合自相关特征分析方法进行医疗风险数据的特征挖掘。

2.2 医疗数据的可视化特征分析

通过投影法实现医疗数据访问风险评估的相空间重构,结合风险容量参数识别方法进行医疗数据访问风险评估的可视化数据融合。假设医疗数据访问风险评估的通道数为 N_{WT} ,评估对象的模糊度分布特征量为 N_{sub} ,风险监测动态范围为 $n \in [1, N_{sub}]$,FCM算法的目的是将 N_{WT} 个向量数据集划分为 n 个子集。取FCM的价值函数为

$$J = \sum_{i=1}^n \sum_{j=1}^{N_{WT}} u_{ij}^m d_{ij}^2 \quad (5)$$

式中, m 为加权系数,基于重标记法和投影法合并与校正的医疗数据访问风险融合知识图谱,得到校正输出的规范化正交范围为 $[1, \infty)$; d_{ij} 为第 i 个数据异构融合的图谱与第 j 个欧几里得距离,取 $d_{ij} = \|O_i - X_j\|$; X_j 为第 j 个稀疏散点图的分段线性化节点位置 ($j=1, 2, \dots, N_{WT}$); u_{ij}^m 为医疗数据访问风险特征分布的隶属度权重,介于 $0 \sim 1$ 之间; O_i 为医疗数据访问风险评估的自适应跟踪模板参数为 d, q 轴分量^[11]。 O_i 的更新计算公式为

$$O_i(k+1) = \sum_{j=1}^n u_{ij}^m(k) X_j / \sum_{j=1}^n u_{ij}^m(k) \quad (6)$$

在第 k 次迭代中,若 $\forall j, r, d_{ij}(k) > 0$,引入状态变量 M_{di} 和 M_{qi} ,得到医疗数据访问风险分析的动态方程线性化处理模型为

$$u_{ij} = \frac{1}{\sum_{r=1}^c \left(\frac{d_{ij}(k)}{d_{rj}(k)} \right)^{\frac{2}{m-1}}} \quad (7)$$

若 $\forall j, r, d_{ij}(k) = 0$,则对医疗数据访问风险分析融合的GA-BP控制模型为

$$u_{ij}(k) = \begin{cases} 0, & i \neq r \\ 1, & \text{其他} \end{cases} \quad (8)$$

当 $\|O(k+1) - O(k)\|$ 的矩阵范数小于给定阈值时迭代结束,便可得出 n 个GA-BP状态特征分布位置,建立医疗数据访问风险的子空间寻优模型,因此优化函数目标 F 可表示为

$$F = \lambda_1 C_t + \lambda_2 C_r \quad (9)$$

式中, C_t 为统计学模型分析的动态成本信息及参数; λ_1 和 λ_2 分别为调节控制的速度和精度。引入医疗数据访问风险参数分布的状态变量,得到联合自相关系数 C_t 可以表示为

$$C_t = C_0 + C_C + C_S \quad (10)$$

式中, C_0 为医疗数据分析特征组成部分,含技术风险、诊疗风险等; C_C 为医疗数据风险动态访问成本; C_S 为医疗数据访问风险可视化融合特征量。

2.3 GA-BP算法特征提取和风险分段预测

医疗数据访问监测的风险点主要包括:非授权访问、数据泄露、数据篡改、越权访问^[12]。基于此采用遗传算法-BP(GA-BP)神经网络算法进行风险分段预测。GA-BP算法是一种结合遗传算法和BP神经网络的优化算法。其通过遗传算法来优化BP神经网络的权重和阈值,以提高训练的效果和结果。其具体过程如下:

(1) 初始化种群。使用遗传算法,初始化一组随机的神经网络权重和阈值。这些被称为父代个体;

(2) 适应度评估。使用BP神经网络对每个父代个体进行训练,并计算其适应度值。使用误差平方和作为适应度函数;

(3) 选择操作。根据适应度值,选择一部分优秀的父代个体作为下一代的种群,得到经优化的BP神经网络模型,即GA-BP模型。一般采用轮盘赌选择方法进行选择;

(4) 交叉操作。对被选择的父代个体进行交叉操作,生成子代个体;

(5) 变异操作。对子代个体进行变异操作,引入一定程度的随机性。这可以增加种群的多样性和探索性;

(6) 更新种群。将交叉和变异后得到的子代个体与原始的父代个体合并,形成新的种群;

(7) 重复步骤2到步骤6,直到达到预定的迭代次数或满足终止条件为止。

3 系统软件开发实现

在上述算法设计的基础上,进行风险评估系统的软件开发和设计,软件实现结构流程如图3所示。

在LOG-CONTROL-BLOCK model(记录控制块模型)中,建立医疗数据访问风险评估系统的数据加载模型和风险预测模型。通过视觉引导技术和人机交互技术实现对医疗数据访问风险评估过程中的信息交互和传输控制。采用标准化的IEC 61850标准控制协议实现医疗数据访问风险评估控制和串行数据处理。在融合层中通过B/S模块和I/O数据输出模块实现医疗数据访问风险评估的软件开发和数据总线调度,将数据连接SQLServer数据库,在ADO.NET组件库中读取数据源的数据。

4 实验测试

为了测试本文方法在实现医疗数据访问风险评估的应用性能,进行仿真测试分析。操作系统选择Ubuntu 20.04 LTS,Python版本为3.8.5,使用Jupyter Notebook作

为编程环境。机器学习库和工具包括 Keras 2.4.3 和 TensorFlow 2.3.0 用于建立、训练和评估神经网络模型,以及 NumPy 1.19.1 和 Pandas 1.1.2 用于数值计算和数据处理。神经网络结构为 3 层,选定 ReLU 作为激活函数,遗传算法参数设置种群大小为 50 个个体,迭代次数为 100 次,选择、交叉和变异的概率分别为 0.5、0.8 和 0.1。BP 神经网络训练参数设定学习率为 0.001。给出医疗数据访问风险数据分布长度为 1 024,频谱分布的带宽为 126 bps,医疗数据访问风险数据聚类的动态因子为 0.38,数据传输速率 2 400 KB/s,信息传输方式为异步方式,医疗数据访问风险融合的关联信息分布见表 2。

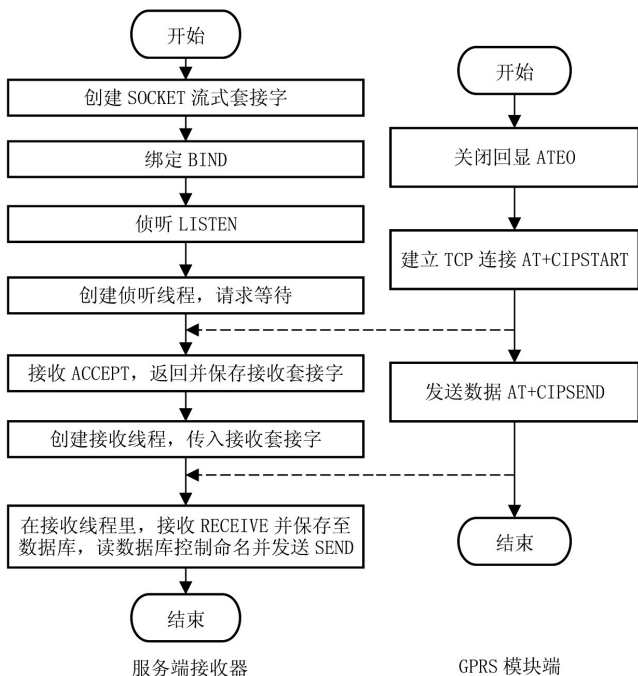


图 3 医疗数据访问风险评估软件实现流程

表 2 医疗数据访问风险融合的关联信息分布

综合关联度	C11	C12	C21	C22
风险监测点 1	0.546	0.553	0.482	0.283
风险监测点 2	0.448	0.436	0.522	0.304
风险监测点 3	0.120	0.374	0.419	0.467
风险监测点 4	0.249	0.255	0.295	0.497
风险监测点 5	0.557	0.129	0.487	0.222
风险监测点 6	0.099	0.549	0.210	0.194
风险监测点 7	0.099	0.235	0.121	0.292
风险监测点 8	0.255	0.216	0.378	0.474
风险监测点 9	0.339	0.498	0.569	0.333

根据上述参数设定,得到医疗数据访问风险样本数据时间序列如图 4 所示。

在该实验中,仿真流程包括数据准备、特征提取、训练模型、风险分段预测和评估分析。以图 5 的医疗数据访问风险数据为测试对象,采用本文的信息融合聚类技术进行医疗访问数据聚类处理,结果如图 5 所示。

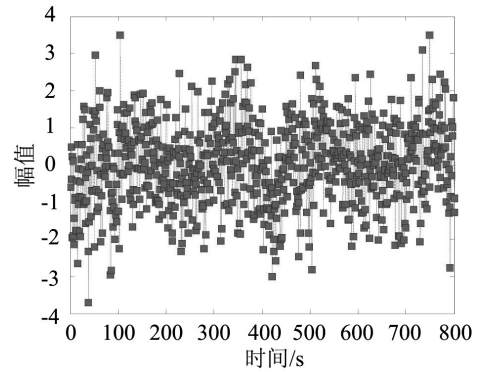
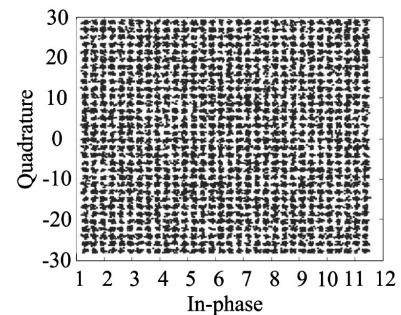
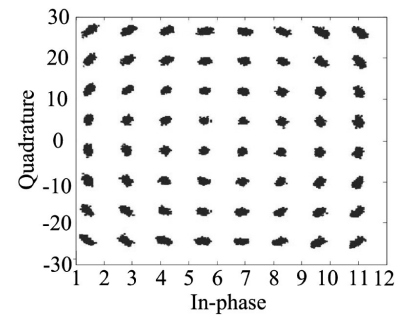


图 4 医疗数据访问风险样本序列



(a) GA-BP 聚类前



(b) GA-BP 聚类后

图 5 医疗访问数据聚类结果

根据上述数据处理结果,实现对医疗数据的访问风险评估。以文献[3]方法与文献[4]方法为对比方法。测试评估的预测精度,得到对比结果如图 6 所示。

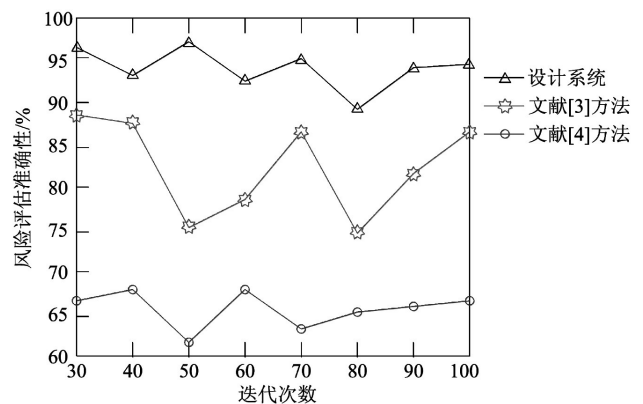


图 6 风险评估准确性对比

(下转第 148 页)