

# 基于量子计算的医疗数据敏感度度量

李晓峰<sup>1</sup>, 焦洪双<sup>1</sup>, 王妍玮<sup>2</sup>

(1. 黑龙江外国语学院 信息工程系, 黑龙江 哈尔滨 150025;  
2. 普度大学 机械工程系, 印第安纳州 西拉法叶市 IN47906)

**摘要:**针对传统的医疗数据敏感度度量方法存在的度量开销大、数据查准率和查全率低的问题,提出基于量子计算的医疗数据敏感度度量方法。首先采用分布式样本重构方法重组医疗数据的分布式结构,建立医疗数据敏感度度量的统计分析模型;其次采用量化回归分析方法进行医疗数据的模糊融合和聚类分析,建立其定量递归分析模型;最后采用量子计算方法进行医疗数据敏感度度量过程中的自适应寻优控制,建立量子寻优约束进化模型,采用动态全局规划方法实现对医疗数据敏感度的度量。实验结果表明,采用该方法进行医疗数据敏感度度量的统计分析能力较强,且度量时间开销较小、对敏感数据的查准率和查全率较高,提高了对医疗数据的检索和特征分辨能力,有效实现了对医疗数据的统计分析和检测度量。

**关键词:**量子计算;医疗数据;敏感度;度量;统计分析;模糊融合;聚类分析;全局规划

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2021)01-0187-05

doi:10.3969/j.issn.1673-629X.2021.01.033

## Sensitivity Measurement of Medical Data Based on Quantum Computing

LI Xiao-feng<sup>1</sup>, JIAO Hong-shuang<sup>1</sup>, WANG Yan-wei<sup>2</sup>

(1. Department of Information Engineering, Heilongjiang International University, Harbin 150025, China;  
2. Department of Mechanical Engineering, Purdue University, West Lafayette IN47906, US)

**Abstract:** In order to address the problems existing in the traditional medical data sensitivity measurement methods, such as high measurement cost, low accuracy and low recall rate, a medical data sensitivity measurement method based on quantum computing is proposed. First, the distributed sample reconstruction method is used to reconstruct the distributed structure of medical data and establish the statistical analysis model of medical data sensitivity measurement. Secondly, the fuzzy fusion and clustering analysis of medical data are carried out by quantitative regression analysis method, and the quantitative recursive analysis model is established. Finally, the quantum computing method is used for adaptive optimization control in the process of medical data sensitivity measurement, the quantum optimization constraint evolution model is established, and the dynamic global programming method is used to measure the sensitivity of medical data. Experiment shows that the proposed method has a strong statistical analysis ability to measure the sensitivity of medical data, a low cost of measuring time, and a high precision and recall rate of sensitive data, which improves the retrieval and feature resolution ability of medical data and effectively realizes the statistical analysis and detection measurement of medical data.

**Key words:** quantum computing; medical data; sensitivity; measurement; statistical analysis; fuzzy fusion; clustering analysis; global planning

## 0 引言

利用大数据分析和处理方法对医疗信息进行管控、建立医疗大数据的融合调度和敏感度表征模型,有利于提高对医疗大数据的信息检测和统计分析能力。通常来说,对医疗大数据进行处理的方法是建立在

医疗数据的统计分析和大数据融合采样基础上的<sup>[1]</sup>,建立医疗数据的量化分析模型,结合模糊信息调度方法,有利于提高医疗信息的表征能力<sup>[2]</sup>。由于很多医疗应用程序都包含敏感信息的输入过程,为确保敏感信息的安全性,使患者的个人权益不受侵害,研究敏感

收稿日期:2020-03-18

修回日期:2020-07-22

基金项目:国家自然科学基金资助项目(61803117);教育部科技发展中心产学研创新基金(2018A01002);国家科技部创新方法专项(2017IM010500)

作者简介:李晓峰(1978-),男,博士研究生,教授,CCF高级会员(16287S),研究方向为数据挖掘、机器学习、计算机视觉、智慧医疗、智能交通、体育工程学。

信息的安全保护机制尤为重要,而这一过程,需建立在对医疗数据敏感度进行度量的基础上。因此,相关的医疗数据敏感度度量和特征分析方法研究,在医疗数据的信息诊断和检索等领域中具有很好的应用价值。

目前,已有很多专家学者在该领域进行了研究,所得到的医疗数据的敏感度度量方法主要有统计特征分析方法、三维特征重构方法、C均值重构度量方法等。总的来说,对医疗数据敏感度进行度量的方法是建立在医疗数据的谱特征分析和特征提取结果的基础上的,从中提取医疗数据的敏感度特征分量,再结合高阶统计信息融合方法对医疗数据进行三维重构。在此基础上,根据医疗数据的融合聚类分析结果结合医疗数据的表面重建,实现医疗数据的三维重构和特征分析,完成对敏感度的度量<sup>[3]</sup>。文献[4]中提出了一种基于改进全卷积神经网络的大数据表面重建和敏感度度量方法,在该方法中,采用无线射频识别技术进行大数据采样和敏感度特征分析,结合卷积神经网络分析方法,实现对数据的敏感度度量。但该方法进行医疗数据敏感度度量的自适应性不好,统计分析能力不强。文献[5]中提出了一种基于关联规则特征检测的医疗数据敏感度度量方法,在该方法中,首先对医疗信息管理系统中的医疗数据的存储结构进行分析,采用支持向量机算法区分医疗数据的属性类别,再在同类属性数据中筛选出敏感度数据并对其敏感度进行计算。然而在利用该方法进行医疗数据敏感度度量的时间开销较大,度量过程过于繁琐。文献[6]中提出了一种基于秩约束密度敏感距离的自适应聚类方法,该方法首先引入密度敏感距离相似度量方法扩大不同类数据间的距离,并将秩约束施加于拉普拉斯矩阵,使相似矩阵的连通区域数量等于聚类数量,将数据划分至相应的类别中,在聚类的基础上实现对数据敏感度的度量。然而该方法对敏感数据的查准率较低,对敏感数据的采集和获取结果不理想。

量子算法是指利用量子计算的并行性和纠缠性等特征、将量子理论与计算机技术相结合的新型计算模式。由于量子的独特性质,使得量子算法能够适应大数据量的处理,计算成本也大大减少。为此,针对当前方法中存在的度量过程自适应性差、度量开销大、对敏感数据查准率较低的问题,该文提出一种基于量子计算的医疗数据敏感度度量方法。整体思路如下:首先采用分布式样本重构方法对医疗数据的分布式结构进行重组,然后采用量化回归分析方法对医疗数据进行模糊融合和聚类分析,根据融合分析结果建立定量递归分析模型,在此基础上,结合量子计算对度量过程进行寻优约束,并采用动态全局规划方法实现对医疗数据敏感度的度量。最后通过仿真实验结果证明了该方

法在提高医疗数据敏感度度量性能方面的优越性能。

## 1 医疗数据的分布式结构重组及统计分析模型

### 1.1 医疗数据的分布式结构重组

为了实现对医疗数据敏感度的准确度量,首先构建医疗数据的分布式结构重组模型,采用高阶统计特征分析方法,进行医疗数据的分布式结构重组过程中的特征提取和分布式特征检测<sup>[7]</sup>,继而建立医疗数据度量的模糊关联规则特征检测模型,采用一条 NURBS 曲线进行医疗数据的分布式结构重组<sup>[8-9]</sup>,这一过程表述如下:

$$P = \frac{\sum_{i=0}^n NW_i C_i}{\sum_{i=0}^n NW_i} \quad (1)$$

其中,  $N$  表示采集到的用于进行敏感度度量的医疗数据;  $P$  表示重组结果;  $C_i$  ( $i = 0, 1, \dots, n$ ) 表示医疗数据的分布式度量的控制顶点;  $W_i$  ( $i = 0, 1, \dots, n$ ) 表示自适应学习的权因。在此基础上,对医疗数据的分布式结构重组的权因子进行量化寻优,当分布式权值满足  $W_0 > 0, W_n > 0$  时,采用模糊加权学习方法,假设  $N_{i,k}$  表示第  $k$  次寻优规范的样条函数,则由递推公式计算医疗数据的分布式结构重组模型为:

$$N_{i,k} = \frac{u \times N_{i,k-1}}{u_{i+k} - u_i} + \frac{u_{i+k+1} - u}{u_{i+k+1} - u_{k+1}} \quad (2)$$

其中,  $U = (u_0, u_1, \dots, u_{i+k+1})$  表示医疗数据敏感度度量节点矢量,  $u$  表示 NURBS 曲线的自变量。根据上述重组结果,结合关联规则挖掘方法进行医疗数据敏感度度量的自适应寻优,从中提取出医疗数据的关联特征量,再采用关联特征检测方法进行医疗数据敏感度度量过程中的模糊加权学习。在这一过程中,得到的统计特征量为:

$$X = x_i (P - N_{i,k} \times \alpha) \quad (3)$$

其中,  $x_i$  表示医疗数据敏感度度量的状态矢量,  $\alpha$  表示模糊加权系数。对于所得的统计特征量,结合自适应学习方法对医疗数据进行三维特征重建。设置  $r$  个不同的聚类中心中医疗数据结构重组的状态因子,得到医疗数据敏感度度量的动态增量函数  $h_1, h_2, \dots, h_r, \dots, h_r$ , 每一个函数满足  $h_i: \{0, 1\}^* \rightarrow [1, m]$ 。采用线性映射方法,建立医疗数据敏感度度量的模糊度检测模型,结合结构重组方法进行医疗数据敏感度度量和统计特征分析,所得的模糊度检测结果为:

$$Z = j \times \sum_{j=1}^n (X + A_j h_r) + mh_r \quad (4)$$

其中,  $m$  表示医疗数据三维特征动态重构的模糊度分

布特征集,  $A_j$  表示不同模糊度下医疗数据敏感度度量的聚类中心, 其中  $j (j = 1, 2, \dots, n)$  表示模糊度。

## 1.2 医疗数据度量的统计分析模型

在上述研究的基础上, 根据模糊度检测结果建立医疗数据敏感度度量的统计分析模型, 采用量化回归分析方法进行医疗数据敏感度度量的模糊融合和聚类分析<sup>[10]</sup>。首先, 采用样本回归分析方法进行医疗数据度量的统计特征分析, 得到的统计特征量表示如下:

$$V = \frac{\sum_{i=0}^n \sum_{j=0}^n NW_i d_{i,j}}{Z} \quad (5)$$

其中,  $d_{i,j} (i, j = 0, 1, \dots, n)$  表示医疗数据敏感度度量的控制顶点。采用插补算法进行医疗数据敏感度度量过程中的二维插值运算, 以等弧长为度量尺度, 进行医疗数据敏感度度量曲线分割<sup>[11-12]</sup>。使用  $f$  表示医疗数据敏感度度量的样条曲线, 根据时间  $t$  的变化, 利用一阶泰勒级数展开上述统计特征量, 得到医疗数据统计特征量的时间尺度分解式为:

$$V = Z + \frac{dU}{dt} + H. O. T |_{t=t_i} (t_{i+1} - t_i) \quad (6)$$

其中,  $H. O. T$  表示高阶微量。对于式(6)中的  $t_i$ , 采用  $i$  次插补方法, 进行医疗数据敏感度度量的量化回归分析, 建立统计分析模型, 对应的插补时刻, 得到医疗数据的敏感度度量的量子计算微量<sup>[13]</sup>, 可定义为:

$$V(t) = \left\| \frac{dV}{dt} \right\| + H. O. T |_{t=t_i} (t_{i+1} - t_i) \quad (7)$$

忽略医疗数据量子计算的高阶微量  $H. O. T$ , 根据量子计算方法<sup>[14]</sup> 可得到医疗数据统计分析的参数增量  $\Delta V$  如下:

$$\Delta V = T_s \times (V(t) - V) \quad (8)$$

其中,  $T_s$  为曲线插补周期。在此基础上, 采用支持向量机模型, 进行医疗数据敏感度度量的动态增量控制, 得到控制误差性能曲线为:

$$l = \Delta V(M + c)e \quad (9)$$

其中,  $M$  表示医疗数据敏感度特征分布的正定值;  $e$  表示医疗数据统计分析的模糊度函数;  $c$  表示为聚类误差。则根据误差控制结果对医疗数据进行统计分析, 结果如下:

$$G = l(\Delta V \times M - c \times Z) \quad (10)$$

通过得到的医疗数据的统计分析模型, 建立医疗数据敏感度度量的定量递归分析模型, 采用量子计算方法进行医疗数据敏感度度量过程中的自适应寻优控制。

## 2 医疗数据敏感度度量设计

### 2.1 量子计算过程

在上述采用分布式样本重构方法进行医疗数据的

分布式结构重组, 并建立医疗数据敏感度度量的统计分析模型的基础上, 进行医疗数据敏感度度量模型的设计。该文提出了基于量子计算的医疗数据敏感度度量方法。采用量子计算方法进行医疗数据敏感度度量过程中的自适应寻优控制, 采用模网格分区域聚类分析方法, 建立医疗数据敏感度度量的关联规则特征分布集, 结合分簇融合方法进行医疗数据的敏感度度量。分布集的簇模型描述为:

$$E = \text{diag}\{\max\{|\rho_1^+|, |\rho_1^-|\}, \dots, \max\{|\rho_n^+|, |\rho_n^-|\}\} = \text{diag}\{\rho_1, \dots, \rho_n\} \quad (11)$$

其中,  $\rho_1, \dots, \rho_n$  为一组医疗数据敏感度特征分布的关联特征量。根据量子隐形传态原理可得到共享的量子纠缠特征值  $O$ 。对医疗数据的敏感度的关联进行映射, 根据映射结果, 得到在邻域空间  $q$  内, 医疗数据敏感度度量的量子计算统计分布集为:

$$F = O \times (G \times q + E) \quad (12)$$

采用量子计算方法, 得到医疗数据敏感度度量的模糊关联度, 输出为  $s$ , 得到量子寻优进化模型为:

$$\delta = [s \times \sum_{i=1}^n (F + G)] + E \quad (13)$$

### 2.2 医疗数据敏感度度量实现

在上述获得医疗数据的模糊相关性特征量的基础上进行医疗数据的敏感度度量, 在高维相空间中进行医疗数据的敏感度度量, 得到医疗数据敏感度度量的模糊聚类中心为  $A_\lambda$ , 其中,  $\lambda$  表示空间维数<sup>[15]</sup>。当医疗数据敏感度度量的模糊度特征量满足  $\sum_{\lambda=1}^n A_\lambda = 1$  时, 采用自相关特征检测方法进行医疗数据敏感度度量, 输出结果为:

$$Y = \frac{\sum_{\lambda=1}^n A_\lambda \delta}{\lambda(\lambda - 1)/2} \quad (14)$$

综上所述, 该文结合量子计算方法实现了对医疗数据敏感度度量。首先初始化医疗数据, 在建立统计分析模型的基础上, 采用样本重构方法重组医疗数据的分布式结构, 然后采用量化回归分析方法对医疗数据进行模糊融合和聚类分析, 建立其定量递归分析模型, 继而采用量子计算进行医疗数据敏感度度量过程中的自适应寻优控制, 通过全局动态规划方法实现对医疗数据敏感度的度量。其实现过程如图1所示。

## 3 实验与结果分析

为了测试所提的基于量子计算的医疗数据敏感度度量方法的实际应用性能, 设计如下仿真实验进行验证。

### 3.1 实验环境和数据集

实验环境设置情况如下: 实验所有医疗数据来自

于 ADNI 数据库(adni. loni. usc. edu), 医疗数据敏感特征分布样本长度为 1 200, 对医疗数据的敏感属性分布的维数为 12, 对医疗数据网格聚类的大小 40 \* 40, 对医疗数据的统计特征分析的样本训练集为 60, 关联度特征分布系数为 0. 12。硬件环境为: Windows7 系统, Visual Studio2010 操作平台。

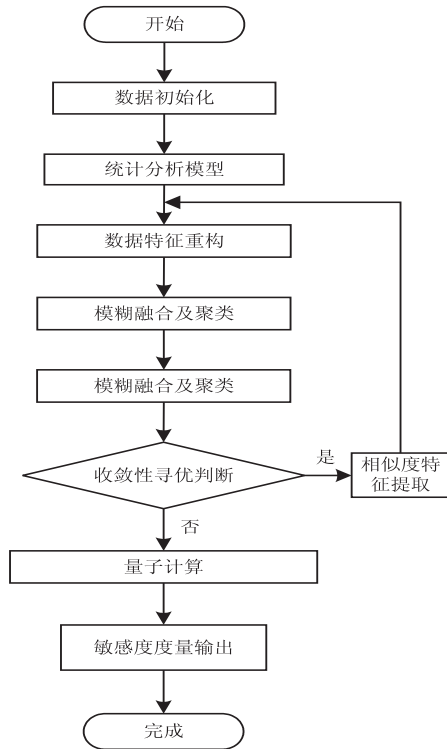


图 1 基于量子计算的医疗数据敏感度量实现

为使实验结果具有说明性, 将所提的基于量子计算的医疗数据敏感度量方法与文献[4]中的基于改进全卷积神经网络的大数据表面重建和敏感度量方法、文献[5]中的基于关联规则特征检测的医疗数据敏感度量方法、文献[6]中的基于秩约束密度敏感距离的自适应聚类方法作对比。

### 3.2 实验指标

(1) 数据查准率。

查准率是一种衡量检索过程的准确度的指标, 通过查准率的对比, 可以判断不同方法对敏感医疗数据的检索能力, 其计算过程如下:

$$\text{敏感数据查准率} = \frac{\text{检索出的敏感数据量}}{\text{数据总量}} \times 100\% \quad (15)$$

(2) 度量时间开销。

负载开销指在数据敏感度量过程中所花费的时间, 可以判断不同方法的时间消耗情况。度量开销结果由 Visual Studio2010 操作平台自动统计。

(3) 查全率。

查全率是指由度量过程检索出的相关数据量与数

据总量的比率, 由数据内容、数量和运行环境的平稳性来决定, 是衡量度量成功度和自适应性的一项指标, 其计算过程如下:

$$\text{查全率} = \frac{\text{度量数据量}}{\text{数据总量}} \times 100\% \quad (16)$$

### 3.3 实验结果与分析

根据上述实验条件和指标的设定情况, 进行医疗数据的敏感度量实验。首先对医疗数据进行采集, 在此基础上, 建立医疗数据敏感度量的定量递归分析模型, 采用量子计算方法进行医疗数据敏感度量过程中的自适应寻优控制, 实现敏感表征。得到的医疗数据敏感表征结果如图 2 所示。

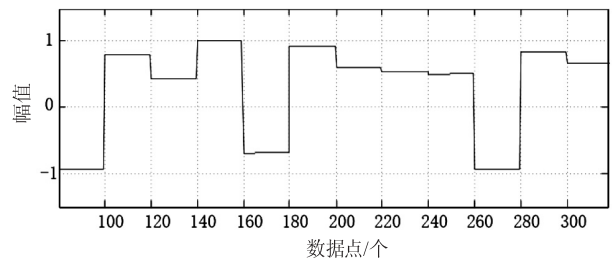


图 2 医疗数据敏感表征结果

分析图 2 可知, 采用基于量子计算的医疗数据敏感度量方法后, 医疗数据敏感度的波动幅值始终保持在[-1, 1]之间, 波动情况较为稳定, 证明利用基于量子计算的医疗数据敏感度量方法进行医疗数据敏感度量过程的敏感特征辨识能力较好, 信息反馈能力较强、自适应优势明显。

测试不同数据敏感度量方法的数据查准率, 对比结果如表 1 所示。

表 1 不同医疗数据敏感度量方法的查准率对比结果

迭代次数/次	文中方法	文献[4]方法	文献[5]方法	文献[6]方法	%
100	94.5	88.4	82.4	81.4	
200	96.8	90.3	88.4	84.3	
300	96.9	93.3	89.3	85.7	
400	97.6	92.5	90.2	86.4	
500	97.2	93.7	90.6	86.1	

分析表 1 可知, 随着实验迭代次数的不断增加, 不同方法的度量查准率也在不断发生变化, 整体表现出上升态势。其中, 文献[5]方法的度量查准率的上升幅度最大, 但其度量查准率值低于文中方法、文献[4]方法和文献[6]方法。文中方法的度量查准率上升幅度虽小, 但度量查准率值更高, 证明采用文中方法在医疗数据敏感度量过程中, 对医疗敏感数据的捕获能力较强, 检索误差较小, 能够有效实现敏感医疗数据的查准。

为进一步对度量方法的有效性进行检验,测试不同数据敏感度度量方法的度量开销和查全率,结果分别如图 3 和图 4 所示。

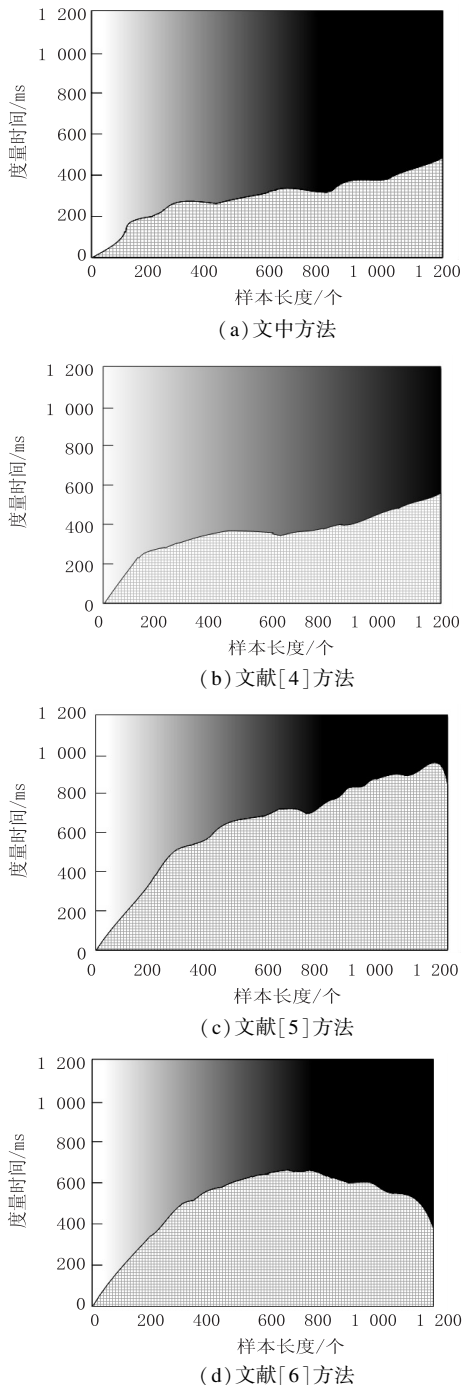


图 3 不同医疗数据敏感度度量方法度量所需时间对比  
分析图 3 可知,随着数据样本数量的增加,同医疗数据敏感度度量方法度量所需的时间也在不断变化。文中方法和文献[4]、文献[5]方法的时间开销均呈现出上升态势,而文献[6]方法的时间开销先上升后下降。但四种方法中,文献[5]方法的时间开销最大,文中方法和文献[4]方法的时间开销较接近,但文中方法的时间开销更小,证明采用文中方法在医疗数据敏感度度量方法的时效性更强。

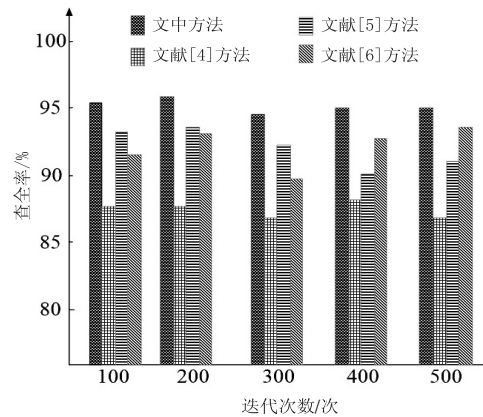


图 4 不同医疗数据敏感度度量方法查全率对比

分析图 4 可知,随着实验迭代次数的不断增加,不同医疗数据敏感度度量方法的查全率也在不断发生变化。但文中的基于量子计算的医疗数据敏感度度量方法的查全率始终在 4 种方法中保持最高,维持在 95% 左右,证明该方法具有较强的自适应性,对数据的统计分析能力较强。

### 4 结束语

为对医疗数据的敏感度进行准确度量,提出基于量子计算的医疗数据敏感度度量方法。采用分布式样本重构方法进行医疗数据的分布式结构重组,对医疗数据进行统计分析,结合量化回归分析方法进行医疗数据敏感度度量的模糊融合和聚类分析,建立医疗数据敏感度度量的定量递归分析模型,采用量子计算方法建立医疗数据敏感度度量的量子寻优约束进化模型,根据动态全局规划结果完成医疗数据敏感度的度量。经实验研究得知,利用该方法进行医疗数据敏感度度量的辨识能力较好、统计分析能力较强,且自适应性较强,为保证医疗信息的安全性奠定了基础。在今后的研究中,将进一步对该方法进行优化,以期使所提的数据敏感度度量方法在度量时效和应用范围两个方面有效突破。

### 参考文献:

- [1] 夏宇航,高大启,阮 彤,等. 基于知识图谱的医疗病历数据存储服务研究[J]. 计算机工程,2019,45(1):9-16.
- [2] GREENGARD S. Finding a healthier approach to managing medical data [J]. Communications of the ACM, 2018, 61(5):31-33.
- [3] SHAIKHINA T, KHOVANOVA N A. Handling limited datasets with neural networks in medical applications; a small-data approach[J]. Artificial Intelligence in Medicine, 2017, 75(11):51-63.
- [4] 师彦文,王宏杰. 基于新型不纯度度量的代价敏感随机森林分类器[J]. 计算机科学,2017,44(11A):98-101.

(下转第 197 页)