



(12) 发明专利

(10) 授权公告号 CN 112837799 B

(45) 授权公告日 2021.10.15

(21) 申请号 202110125345.X

(22) 申请日 2021.01.29

(65) 同一申请的已公布的文献号
申请公布号 CN 112837799 A

(43) 申请公布日 2021.05.25

(73) 专利权人 广东德澳智慧医疗科技有限公司
地址 523000 广东省东莞市松山湖高新技术
产业开发区阿里山路19号产业化中
心10栋5层501室、502室、503室、504
室、505室

(72) 发明人 曹茂诚 王洪平

(74) 专利代理机构 南昌合达信知识产权代理事
务所(普通合伙) 36142
代理人 李旦

(51) Int.Cl.

G16H 40/60 (2018.01)

G16H 50/20 (2018.01)

G16H 50/70 (2018.01)

(56) 对比文件

CN 110890146 A, 2020.03.17

CN 111863232 A, 2020.10.30

CN 108806791 A, 2018.11.13

US 2017300646 A1, 2017.10.19

审查员 白露霜

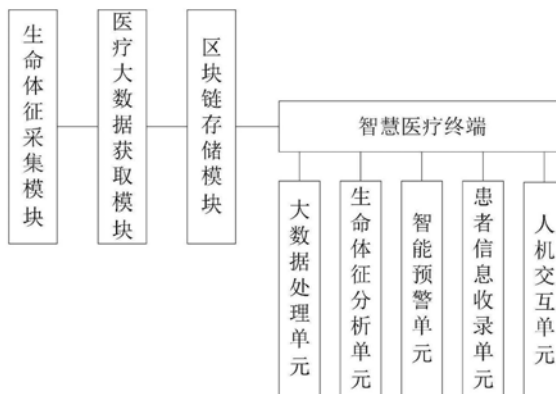
权利要求书4页 说明书10页 附图1页

(54) 发明名称

基于区块链的远程互联网大数据智慧医疗系统

(57) 摘要

基于区块链的远程互联网大数据智慧医疗系统,包括生命体征采集模块、医疗大数据获取模块、区块链存储模块和智慧医疗终端,根据医疗大数据获取模块收集的生命体征大数据建立根据生命体征数据对身体状态进行评估的健康评估模型,将生命体征采集模块采集到的患者的生命体征数据输入到健康评估模型中,从而评估患者的身体状态是健康还是危险,并在危险时进行预警。本发明的有益效果:实现了远程对患者的身体状态进行统一监护,并当患者的身体状态出现危险时,能够及时的进行预警。



1. 基于区块链的远程互联网大数据智慧医疗系统,其特征是,包括生命体征采集模块、医疗大数据获取模块、区块链存储模块和智慧医疗终端,所述生命体征采集模块和患者的床位号进行绑定,用于采集患者的生命体征数据,并将采集的生命体征数据和患者的床位号通过互联网传输至智慧医疗终端,所述医疗大数据获取模块用于收集生命体征大数据,并将收集的生命体征大数据传输至区块链存储模块进行存储,所述智慧医疗终端包括大数据处理单元、生命体征分析单元、智能预警单元、患者信息收录单元和人机交互单元,智慧医疗终端从区块链存储模块中调取所述生命体征大数据,并将调取的生命体征大数据输入至大数据处理单元进行处理,所述生命体征分析单元根据处理后的生命体征大数据建立根据生命体征数据对患者的身体状态进行评估的健康评估模型,将接收到的患者的生命体征数据输入到健康评估模型中,从而评估患者的身体状态是健康还是危险,当评估患者的身体状态为危险时令智能预警单元进行预警,所述患者信息收录单元用于记录患者的基本信息、患者的生命体征数据和评估所得的患者的身体状态,医护人员可以通过在人机交互单元中输入患者的基本信息查询患者的生命体征数据和患者的身体状态;大数据处理单元用于对生命体征大数据进行聚类,并在聚类的过程中去除生命体征大数据中的噪声数据,具体包括:

(1) 在生命体征大数据中选取类中心;

(2) 根据选取的类中心对生命体征大数据中的数据进行聚类,并在聚类的过程中去除生命体征大数据中的噪声数据;

采用下列方式在所述生命体征大数据中选取类中心:

设Y表示生命体征大数据组成的集合, y_i 表示集合Y中的第i个数据,定义 $s(y_i)$ 表示数据 y_i 在集合Y中的全局相似系数,且 $s(y_i) = \exp\left(-\frac{\sum_{j=1}^{M(Y)} |y_i - y_j|}{M(Y)-1}\right)$,其中, y_j 表示集合Y中的第j个数据, $M(Y)$ 表示集合Y中的数据个数;设 $U(y_i)$ 表示数据 y_i 的邻域数据集合,给定正整数M,采用下列步骤在集合Y中选取M个数据加入到集合 $U(y_i)$ 中:

步骤1: 设 $y_i(1)$ 表示集合Y中距离数据 y_i 最近的数据, $\omega(y_i(1))$ 表示数据 $y_i(1)$ 的局部区域,且 $\omega(y_i(1))$ 为以数据 $\frac{y_i(1)+y_i}{2}$ 为中心、以 $|y_i(1) - y_i|$ 为边长的正方体区域,将数据 $y_i(1)$ 加入到集合 $U(y_i)$ 中;

步骤2: 在局部区域 $\omega(y_i(1))$ 外的数据中确定距离数据 y_i 最近的数据,将该数据记为 $y_i(2)$,设 $\omega(y_i(2))$ 表示数据 $y_i(2)$ 的局部区域,且 $\omega(y_i(2))$ 为以数据 $\frac{y_i(2)+y_i}{2}$ 为中心、以 $|y_i(2) - y_i|$ 为边长的正方体区域,将数据 $y_i(2)$ 加入到集合 $U(y_i)$ 中;

步骤3: 继续在局部区域 $\omega(y_i(1))$ 和局部区域 $\omega(y_i(2))$ 外的数据中确定距离数据 y_i 最近的数据,将该数据记为 $y_i(3)$,设 $\omega(y_i(3))$ 表示数据 $y_i(3)$ 的局部区域,且 $\omega(y_i(3))$ 为以数据 $\frac{y_i(3)+y_i}{2}$ 为中心、以 $|y_i(3) - y_i|$ 为边长的正方体区域,将数据 $y_i(3)$ 加入到集合 $U(y_i)$ 中;

步骤4: 继续按照上述步骤1、步骤2和步骤3中的方法确定数据加入到集合 $U(y_i)$ 中,直到集合 $U(y_i)$ 中的数据个数等于M时,则停止向集合 $U(y_i)$ 中加入数据;

采用下列公式在集合Y中筛选出可作为类中心的候选数据:

$$f(y_i) = \left| s(y_i) - \frac{\sum_{l=1}^M \rho(y_i(l), y_i) * s(y_i(l))}{\sum_{l=1}^M \rho(y_i(l), y_i)} \right|$$

式中, $f(y_i)$ 表示数据 y_i 在集合 Y 中的类中心属性值, $y_i(1)$ 表示集合 $U(y_i)$ 中的第1个数据, $s(y_i(1))$ 表示数据 $y_i(1)$ 在集合 Y 中的全局相似系数, $\rho(y_i(1), y_i)$ 表示数据 $y_i(1)$ 相较于数据 y_i 的距离权值, 且 $\rho(y_i(l), y_i) = \exp\left(-\frac{|y_i - y_i(l)|}{\max_{1 \leq l \leq M} |y_i - y_i(l)|}\right)$;

当数据 y_i 的类中心属性值 $f(y_i) < \frac{s(y_i)}{10}$ 时, 则将数据 y_i 视为类中心的候选数据, 且判定所述数据 y_i 为非噪声数据; 当数据 y_i 的类中心属性值 $f(y_i) \geq \frac{s(y_i)}{10}$ 时, 则将数据 y_i 视为未聚类数据;

设 $L(Y)$ 表示集合 Y 中视为类中心的候选数据组成的集合, 在集合 $L(Y)$ 中选取类中心, 并根据选取的类中心对集合 $L(Y)$ 中的候选数据进行聚类, 具体包括:

Step(1): 选取集合 $L(Y)$ 中具有最大全局相似系数的候选数据为第一个类中心, 将所述第一个类中心记为 c_1 , 类中心 c_1 所在的类集合记为 C_1 , 将类中心 c_1 在集合 $L(Y)$ 中删除, 并采用下列步骤在当前集合 $L(Y)$ 中筛选出属于类集合 C_1 的其他候选数据, 具体为:

步骤(1): 设 $l_k(1)$ 表示第1次筛选时集合 $L(Y)$ 中的第 k 个候选数据, 定义 $G(l_k(1), C_1)$ 表示候选数据 $l_k(1)$ 和类集合 C_1 之间的聚类函数, 且 $G(l_k(1), C_1)$ 的表达式为:

$$G(l_k(1), C_1) = \theta(l_k(1), C_1) * |s(l_k(1)) - s(c_1)|$$

式中, $\theta(l_k(1), C_1)$ 表示判断函数, 设 $N(l_k(1))$ 表示在集合 Y 中选取的距离候选数据 $l_k(1)$ 最近的 M 个数据组成的邻域集合, $m(l_k(1), C_1)$ 表示邻域集合 $N(l_k(1))$ 中存在的属于类集合 C_1 的数据个数, 当 $m(l_k(1), C_1) \neq 0$ 时, $\theta(l_k(1), C_1) = 1$, 当 $m(l_k(1), C_1) = 0$ 时, $\theta(l_k(1), C_1) = 0$, $s(c_1)$ 表示类中心 c_1 在集合 Y 中的全局相似系数, $s(l_k(1))$ 表示候选数据 $l_k(1)$ 在集合 Y 中的全局相似系数;

当 $0 < G(l_k(1), C_1) < \frac{s(c_1)}{10}$ 时, 则判定候选数据 $l_k(1)$ 为类集合 C_1 中的数据, 将候选数据 $l_k(1)$ 加入到类集合 C_1 中, 并将候选数据 $l_k(1)$ 在集合 $L(Y)$ 中删除, 当 $G(l_k(1), C_1) = 0$ 或者 $G(l_k(1), C_1) \geq \frac{s(c_1)}{10}$ 时, 则将候选数据 $l_k(1)$ 在集合 $L(Y)$ 中保留;

步骤(2), 设 $l_k(2)$ 表示第2次筛选时当前集合 $L(Y)$ 中的第 K 个候选数据, 定义 $G(l_k(2), C_1)$ 表示候选数据 $l_k(2)$ 和类集合 C_1 之间的聚类函数, 且 $G(l_k(2), C_1)$ 的表达式为:

$$G(l_k(2), C_1) = \theta(l_k(2), C_1) * \left| s(l_k(2)) - \frac{\sum_{y_{1,z} \in C_1} \rho(y_{1,z}, l_k(2)) * s(y_{1,z})}{\sum_{y_{1,z} \in C_1} \rho(y_{1,z}, l_k(2))} \right|$$

其中, $\theta(l_k(2), C_1)$ 表示判断函数, 设 $N(l_k(2))$ 表示在集合 Y 中选取距离候选数据 $l_k(2)$ 最近的 M 个数据组成的邻域集合, $m(l_k(2), C_1)$ 表示邻域集合 $N(l_k(2))$ 中存在的属于类集合 C_1 的数据个数, 当 $m(l_k(2), C_1) \neq 0$ 时, $\theta(l_k(2), C_1) = 1$, 当 $m(l_k(2), C_1) = 0$ 时, $\theta(l_k(2), C_1) = 0$, $s(l_k(2))$ 表示候选数据 $l_k(2)$ 在集合 Y 中的全局相似系数, $y_{1,z}$ 表示类集合 C_1 中的第 z 个数据, $\rho(y_{1,z}, l_k(2))$ 表示数据 $y_{1,z}$ 相较于候选数据 $l_k(2)$ 的距离权值, 且 $\rho(y_{1,z}, l_k(2)) =$

$\exp\left(-\frac{|y_{1,z}-l_K(2)|}{\max_{y_{1,z}\in C_1}|y_{1,z}-l_K(2)|}\right)$, $s(y_{1,z})$ 表示数据 $y_{1,z}$ 在集合 Y 中的全局相似系数;

当 $0 < G(l_K(2), C_1) < \frac{s(l_K(2))}{10}$ 时, 则判定候选数据 $l_K(2)$ 为类集合 C_1 中的数据, 将候选数据 $l_K(2)$ 加入到类集合 C_1 中, 并将候选数据 $l_K(2)$ 在集合 $L(Y)$ 中删除, 当 $G(l_K(2), C_1) = 0$ 或者 $G(l_K(2), C_1) \geq \frac{s(l_K(2))}{10}$ 时, 则将候选数据 $l_K(2)$ 在集合 $L(Y)$ 中保留;

当第二次筛选时在集合 $L(Y)$ 中筛选出属于类集合 C_1 的数据时, 则继续按照步骤 (2) 中的方法在集合 $L(Y)$ 中进行第三次筛选, 直到在当前筛选次数时没有在集合 $L(Y)$ 中筛选出属于类集合 C_1 的数据时即停止在集合 $L(Y)$ 中进行下一次的筛选;

Step (2): 继续选取当前集合 $L(Y)$ 中具有最大全局相似系数的候选数据为第二个类中心, 将所述第二个类中心记为 c_2 , 所述类中心 c_2 所在的类集合记为 C_2 , 将类中心 c_2 在集合 $L(Y)$ 中删除, 并采用上述步骤 (1) 和步骤 (2) 中的方法在当前集合 $L(Y)$ 中筛选出属于类集合 C_2 的其他候选数据; 筛选完成后, 将类集合 C_2 中的候选数据在当前集合 $L(Y)$ 中删除;

重复上述 Step (1) 和 Step (2) 中的方法直到当前集合 $L(Y)$ 中剩余的候选数据个数为 0 时, 即完成了在生命体征大数据中选取类中心, 并完成了对所述生命体征大数据的初步聚类。

2. 根据权利要求 1 所述的基于区块链的远程互联网大数据智慧医疗系统, 其特征是, 所述生命体征分析单元采用支持向量机建立根据生命体征数据进行身体状态评估的健康评估模型, 采用大数据处理单元处理后的生命体征大数据作为训练和测试所述支持向量机的样本集。

3. 根据权利要求 2 所述的基于区块链的远程互联网大数据智慧医疗系统, 其特征是, 患者的基础信息包括患者的姓名、年龄和床位号。

4. 根据权利要求 3 所述的基于区块链的远程互联网大数据智慧医疗系统, 其特征是, 大数据处理单元用于对生命体征大数据进行聚类, 并在聚类的过程中去除所述生命体征大数据中的噪声数据, 确定对生命体征大数据聚类所得的各个类集合对应的身体状态的标签, 在支持向量机的训练过程中, 将生命体征大数据的类集合作为支持向量机的输入值, 将类集合所对应的身体状态的标签作为支持向量机的输出值。

5. 根据权利要求 4 所述的基于区块链的远程互联网大数据智慧医疗系统, 其特征是, 所述身体状态的标签包括健康和危险。

6. 根据权利要求 1 所述的基于区块链的远程互联网大数据智慧医疗系统, 其特征是, 根据选取的类中心对所述生命体征大数据中的数据进行聚类, 并在聚类的过程中去除生命体征大数据中的噪声数据, 具体为:

根据选取的类中心和初步聚类结果对集合 Y 中剩余的未聚类数据进行聚类, 设 $D(Y)$ 表示集合 Y 中未聚类数据组成的集合, d_a 表示集合 $D(Y)$ 中的第 a 个未聚类数据, $N(d_a)$ 表示集合 Y 中距离未聚类数据 d_a 最近的 M 个数据组成的邻域集合, 定义 $h(d_a)$ 表示未聚类数据 d_a 在集合 $D(Y)$ 中的聚类优先级, 且 $h(d_a) = \frac{m(d_a)}{M} * s(d_a)$, 其中, $m(d_a)$ 表示邻域集合 $N(d_a)$ 中已聚类数据的个数, $s(d_a)$ 表示未聚类数据 d_a 在集合 Y 中的全局相似系数;

Step1: 对此时集合 $D(Y)$ 中具有最大聚类优先级的未聚类数据进行优先聚类, 设 d_e 表示

集合 $D(Y)$ 中的第 e 个未聚类数据,且 $d_e = \operatorname{argmax}_{d_a \in D(Y)} h(d_a)$, $N(d_e)$ 表示集合 Y 中距离未聚类数据 d_e 最近的 M 个数据组成的邻域集合, $m(d_e)$ 表示集合 $N(d_e)$ 中已聚类数据的个数;

当 $m(d_e) = 0$ 时,则判定集合 $D(Y)$ 中的未聚类数据均为噪声数据,并将所述噪声数据从 $D(Y)$ 中都删除;

当 $m(d_e) \neq 0$ 时,设 $J_{e,p}$ 表示集合 $N(d_e)$ 中的第 p 个已聚类数据,将已聚类数据 $J_{e,p}$ 所在的类集合表示为 $C_{e,p}$,定义 $J(d_e, C_{e,p})$ 为未聚类数据 d_e 和类集合 $C_{e,p}$ 之间的分布检测系数,则 $J(d_e, C_{e,p})$ 的计算公式为:

$$J(d_e, C_{e,p}) = \begin{cases} \left| \frac{\sum_{J_{e,p,q} \in N'(J_{e,p})} |J_{e,p} - J_{e,p,q}|}{M} - |d_e - J_{e,p}| \right|, & M_{e,p} = 1 \\ \left| \frac{\sum_{J_{e,v} \in N(d_e)} \sum_{J_{e,v,b} \in N'(J_{e,v})} |J_{e,v} - J_{e,v,b}|}{M_{e,p} * M} - \frac{\sum_{J_{e,v} \in N(d_e)} |J_{e,v} - d_e|}{M_{e,p}} \right|, & M_{e,p} > 1 \end{cases}$$

式中, $M_{e,p}$ 表示集合 $N(d_e)$ 中存在的属于类集合 $C_{e,p}$ 的已聚类数据的个数, $N'(J_{e,p})$ 表示类集合 $C_{e,p}$ 中距离已聚类数据 $J_{e,p}$ 最近的 M 个已聚类数据组成的集合, $J_{e,p,q}$ 表示集合 $N'(J_{e,p})$ 中的第 q 个已聚类数据, $J_{e,v}$ 表示集合 $N(d_e)$ 中的第 v 个已聚类数据,且 $J_{e,v}$ 为类集合 $C_{e,p}$ 中的数据, $N'(J_{e,v})$ 表示类集合 $C_{e,p}$ 中距离已聚类数据 $J_{e,v}$ 最近的 M 个已聚类数据组成的集合, $J_{e,v,b}$ 表示集合 $N'(J_{e,v})$ 中的第 b 个已聚类数据;

设 $M(d_e)$ 表示集合 $N(d_e)$ 中的已聚类数据所处的不同类集合的个数, $C_{e,n}$ 表示数据 d_e 和所述 $M(d_e)$ 个类集合之间拥有最小分布检测系数的类集合,即 $C_{e,n} = \operatorname{argmin}_{1 \leq p \leq M(d_e)} J(d_e, C_{e,p})$;

当未聚类数据 d_e 和类集合 $C_{e,n}$ 之间满足: $\left| s(d_e) - \frac{\sum_{y_{e,n,r} \in C_{e,n}} \rho(y_{e,n,r}, d_e) * s(y_{e,n,r})}{\sum_{y_{e,n,r} \in C_{e,n}} \rho(y_{e,n,r}, d_e)} \right| < \frac{s(d_e)}{10}$ 时,则将未聚类数据 d_e 加入到类集合 $C_{e,n}$ 中,并将未聚类数据 d_e 在集合 $D(Y)$ 中删除,判定未聚类数据 d_e 为非噪声数据,当未聚类数据 d_e 和类集合 $C_{e,n}$ 之间满足: $\left| s(d_e) - \frac{\sum_{y_{e,n,r} \in C_{e,n}} \rho(y_{e,n,r}, d_e) * s(y_{e,n,r})}{\sum_{y_{e,n,r} \in C_{e,n}} \rho(y_{e,n,r}, d_e)} \right| \geq \frac{s(d_e)}{10}$ 时,则认定未聚类数据 d_e 为噪声数据,将未聚类数据 d_e 在集合 $D(Y)$ 中删除,其中, $s(d_e)$ 表示未聚类数据 d_e 在集合 Y 中的全局相似系数, $y_{e,n,r}$ 表示类集合 $C_{e,n}$ 中的第 r 个数据, $s(y_{e,n,r})$ 表示数据 $y_{e,n,r}$ 在集合 Y 中的全局相似系数, $\rho(y_{e,n,r}, d_e)$ 表示数据 $y_{e,n,r}$ 相较于未聚类数据 d_e 的距离权值,且 $\rho(y_{e,n,r}, d_e) = \exp\left(-\frac{|y_{e,n,r} - d_e|}{\max_{y_{e,n,r} \in C_{e,n}} |y_{e,n,r} - d_e|}\right)$;

Step2:按照上述Step1中的方法重新在当前集合 $D(Y)$ 中选取具有最大聚类优先级的数据进行优先聚类,直到集合 $D(Y)$ 中的未聚类数据个数为0时,则停止聚类。

基于区块链的远程互联网大数据智慧医疗系统

技术领域

[0001] 本发明创造涉及智慧医疗领域,具体涉及一种基于区块链的远程互联网大数据智慧医疗系统。

背景技术

[0002] 随着人们健康意识的增强和社会老龄化,医院的患者越来越多,加之人们维权意识的增强,护士普遍感到工作压力大和工作任务繁重,一个病人从入院到出院所需的各项处理90%由护士完成;90.42%的护士每周工作时间超过40小时,74.2%的护士有夜班情况,由于护理院护工缺口大,导致招工门槛不断降低,目前普遍存在学历低,缺乏专业护理技能的情况,在繁重的工作压力下,责任心欠缺或职业懈怠导致单纯依靠人力24小时连续看护的难度非常大。

[0003] 针对上述情况,为减少护士的工作量,本发明提供一种基于区块链的远程互联网大数据智慧医疗系统,根据生命体征大数据建立根据生命体征数据对患者的健康状态进行评估的健康评估模型,能够实现远程对患者进行统一监护,从而减少了医护人员的工作量的同时,在患者身体状态出现危险时,能够及时发现并进行预警。

发明内容

[0004] 针对上述问题,本发明旨在提供一种基于区块链的远程互联网大数据智慧医疗系统。

[0005] 本发明创造的目的通过以下技术方案实现:

[0006] 基于区块链的远程互联网大数据智慧医疗系统,包括生命体征采集模块、医疗大数据获取模块、区块链存储模块和智慧医疗终端,所述生命体征采集模块和患者的床位号进行绑定,用于采集患者的生命体征数据,并将采集的生命体征数据和患者的床位号通过互联网传输至智慧医疗终端,所述医疗大数据获取模块用于收集生命体征大数据,并将收集的生命体征大数据传输至区块链存储模块进行存储,所述智慧医疗终端包括大数据处理单元、生命体征分析单元、智能预警单元、患者信息收录单元和人机交互单元,智慧医疗终端从区块链存储模块中调取所述生命体征大数据,并将调取的生命体征大数据输入至大数据处理单元进行处理,所述生命体征分析单元根据处理后的生命体征大数据建立根据生命体征数据对患者的身体状态进行评估的健康评估模型,将接收到的患者的生命体征数据输入到健康评估模型中,从而评估患者的身体状态是健康还是危险,当评估患者的身体状态为危险时令智能预警单元进行预警,所述患者信息收录单元用于记录患者的基础信息、患者的生命体征数据和评估所得的患者的身体状态,医护人员可以通过在人机交互单元中输入患者的基础信息查询患者的生命体征数据和患者的身体状态。

[0007] 优选地,所述生命体征分析单元采用支持向量机建立根据生命体征数据进行身体状态评估的健康评估模型,采用大数据处理单元处理后的生命体征大数据作为训练和测试所述支持向量机的样本集。

[0008] 优选地,所述患者的基础信息包括患者的姓名、年龄和床位号。

[0009] 优选地,所述大数据处理单元用于对生命体征大数据进行聚类,并在聚类的过程中去除所述生命体征大数据中的噪声数据,确定生命体征大数据聚类所得的各个类集合对应的身体状态的标签,在支持向量机的训练过程中,将生命体征大数据的类集合作为支持向量机的输入值,将类集合所对应的身体状态的标签作为支持向量机的输出值。

[0010] 优选地,所述身体状态的标签包括健康和危险。

[0011] 优选地,所述大数据处理单元用于对生命体征大数据进行聚类,并在聚类的过程中去除生命体征大数据中的噪声数据,具体包括:

[0012] (1) 在生命体征大数据中选取类中心;

[0013] (2) 根据选取的类中心对生命体征大数据中的数据进行聚类,并在聚类的过程中去除生命体征大数据中的噪声数据。

[0014] 优选地,采用下列方式在所述生命体征大数据中选取类中心:

[0015] 设 Y 表示生命体征大数据组成的集合, y_i 表示集合 Y 中的第 i 个数据,定义 $s(y_i)$ 表示数据 y_i 在集合 Y 中的全局相似系数,且 $s(y_i) = \exp\left(-\frac{\sum_{j=1}^{M(Y)} |y_i - y_j|}{M(Y)-1}\right)$,其中, y_j 表示集合 Y 中的第

j 个数据, $M(Y)$ 表示集合 Y 中的数据个数;设 $U(y_i)$ 表示数据 y_i 的邻域数据集合,给定正整数 M ,采用下列步骤在集合 Y 中选取 M 个数据加入到集合 $U(y_i)$ 中:

[0016] 步骤1:设 $y_i(1)$ 表示集合 Y 中距离数据 y_i 最近的数据, $\omega(y_i(1))$ 表示数据 $y_i(1)$ 的局部区域,且 $\omega(y_i(1))$ 为以数据 $\frac{y_i(1)+y_i}{2}$ 为中心、以 $|y_i(1) - y_i|$ 为边长的正方体区域,将数据 $y_i(1)$ 加入到集合 $U(y_i)$ 中;

[0017] 步骤2:在局部区域 $\omega(y_i(1))$ 外的数据中确定距离数据 y_i 最近的数据,将该数据记为 $y_i(2)$,设 $\omega(y_i(2))$ 表示数据 $y_i(2)$ 的局部区域,且 $\omega(y_i(2))$ 为以数据 $\frac{y_i(2)+y_i}{2}$ 为中心、以 $|y_i(2) - y_i|$ 为边长的正方体区域,将数据 $y_i(2)$ 加入到集合 $U(y_i)$ 中;

[0018] 步骤3:继续在局部区域 $\omega(y_i(1))$ 和局部区域 $\omega(y_i(2))$ 外的数据中确定距离数据 y_i 最近的数据,将该数据记为 $y_i(3)$,设 $\omega(y_i(3))$ 表示数据 $y_i(3)$ 的局部区域,且 $\omega(y_i(3))$ 为以数据 $\frac{y_i(3)+y_i}{2}$ 为中心、以 $|y_i(3) - y_i|$ 为边长的正方体区域,将数据 $y_i(3)$ 加入到集合 $U(y_i)$ 中;

[0019] 步骤4:继续按照上述步骤1、步骤2和步骤3中的方法确定数据加入到集合 $U(y_i)$ 中,直到集合 $U(y_i)$ 中的数据个数等于 M 时,则停止向集合 $U(y_i)$ 中加入数据;

[0020] 采用下列公式在集合 Y 中筛选出可作为类中心的候选数据:

$$[0021] \quad f(y_i) = \left| s(y_i) - \frac{\sum_{l=1}^M \rho(y_i(l), y_i) * s(y_i(l))}{\sum_{l=1}^M \rho(y_i(l), y_i)} \right|$$

[0022] 式中, $f(y_i)$ 表示数据 y_i 在集合 Y 中的类中心属性值, $y_i(1)$ 表示集合 $U(y_i)$ 中的第1个数据, $s(y_i(1))$ 表示数据 $y_i(1)$ 在集合 Y 中的全局相似系数, $\rho(y_i(1), y_i)$ 表示数据 $y_i(1)$ 相较于数据 y_i 的距离权值,且 $\rho(y_i(l), y_i) = \exp\left(-\frac{|y_i - y_i(l)|}{\max_{1 \leq l \leq M} |y_i - y_i(l)|}\right)$;

[0023] 当数据 y_i 的类中心属性值 $f(y_i) < \frac{s(y_i)}{10}$ 时,则将数据 y_i 视为类中心的候选数据,且判定所述类中心的候选数据为非噪声数据;当数据 y_i 的类中心属性值 $f(y_i) \geq \frac{s(y_i)}{10}$ 时,则将数据 y_i 视为未聚类数据;

[0024] 设 $L(Y)$ 表示集合 Y 中视为类中心的候选数据组成的集合,在集合 $L(Y)$ 中选取类中心,并根据选取的类中心对集合 $L(Y)$ 中的候选数据进行聚类,具体包括:

[0025] Step(1):选取集合 $L(Y)$ 中具有最大全局相似系数的候选数据为第一个类中心,将所述第一个类中心记为 c_1 ,类中心 c_1 所在的类集合记为 C_1 ,将类中心 c_1 在集合 $L(Y)$ 中删除,并采用下列步骤在当前集合 $L(Y)$ 中筛选出属于类集合 C_1 的其他候选数据,具体为:

[0026] 步骤(1):设 $l_k(1)$ 表示第1次筛选时集合 $L(Y)$ 中的第 k 个候选数据,定义 $G(l_k(1), C_1)$ 表示候选数据 $l_k(1)$ 和类集合 C_1 之间的聚类函数,且 $G(l_k(1), C_1)$ 的表达式为:

$$[0027] \quad G(l_k(1), C_1) = \theta(l_k(1), C_1) * |s(l_k(1)) - s(c_1)|$$

[0028] 式中, $\theta(l_k(1), C_1)$ 表示判断函数,设 $N(l_k(1))$ 表示在集合 Y 中选取的距离候选数据 $l_k(1)$ 最近的 M 个数据组成的邻域集合, $m(l_k(1), C_1)$ 表示邻域集合 $N(l_k(1))$ 中存在的属于类集合 C_1 的数据个数,当 $m(l_k(1), C_1) \neq 0$ 时, $\theta(l_k(1), C_1) = 1$,当 $m(l_k(1), C_1) = 0$ 时, $\theta(l_k(1), C_1) = 0$, $s(c_1)$ 表示类中心 c_1 在集合 Y 中的全局相似系数, $s(l_k(1))$ 表示候选数据 $l_k(1)$ 在集合 Y 中的全局相似系数;

[0029] 当 $0 < G(l_k(1), C_1) < \frac{s(c_1)}{10}$ 时,则判定候选数据 $l_k(1)$ 为类集合 C_1 中的数据,将候选数据 $l_k(1)$ 加入到类集合 C_1 中,并将候选数据 $l_k(1)$ 在集合 $L(Y)$ 中删除,当 $G(l_k(1), C_1) = 0$ 或者 $G(l_k(1), C_1) \geq \frac{s(c_1)}{10}$ 时,则将候选数据 $l_k(1)$ 在集合 $L(Y)$ 中保留;

[0030] 步骤(2),设 $l_k(2)$ 表示第2次筛选时当前集合 $L(Y)$ 中的第 K 个候选数据,定义 $G(l_k(2), C_1)$ 表示候选数据 $l_k(2)$ 和类集合 C_1 之间的聚类函数,且 $G(l_k(2), C_1)$ 的表达式为:

$$[0031] \quad G(l_k(2), C_1) = \theta(l_k(2), C_1) * \left| s(l_k(2)) - \frac{\sum_{y_{1,z} \in C_1} \rho(y_{1,z}, l_k(2)) * s(y_{1,z})}{\sum_{y_{1,z} \in C_1} \rho(y_{1,z}, l_k(2))} \right|$$

[0032] 其中, $\theta(l_k(2), C_1)$ 表示判断函数,设 $N(l_k(2))$ 表示在集合 Y 中选取距离候选数据 $l_k(2)$ 最近的 M 个数据组成的邻域集合, $m(l_k(2), C_1)$ 表示邻域集合 $N(l_k(2))$ 中存在的属于类集合 C_1 的数据个数,当 $m(l_k(2), C_1) \neq 0$ 时, $\theta(l_k(2), C_1) = 1$,当 $m(l_k(2), C_1) = 0$ 时, $\theta(l_k(2), C_1) = 0$, $s(l_k(2))$ 表示候选数据 $l_k(2)$ 在集合 Y 中的全局相似系数, $y_{1,z}$ 表示类集合 C_1 中的第 z 个数据, $\rho(y_{1,z}, l_k(2))$ 表示数据 $y_{1,z}$ 相较于候选数据 $l_k(2)$ 的距离权值,且 $\rho(y_{1,z}, l_k(2)) =$

$$\exp\left(-\frac{|y_{1,z} - l_k(2)|}{\max_{y_{1,z} \in C_1} |y_{1,z} - l_k(2)|}\right), s(y_{1,z})$$
表示数据 $y_{1,z}$ 在集合 Y 中的全局相似系数;

[0033] 当 $0 < G(l_k(2), C_1) < \frac{s(l_k(2))}{10}$ 时,则判定候选数据 $l_k(2)$ 为类集合 C_1 中的数据,将候选数据 $l_k(2)$ 加入到类集合 C_1 中,并将候选数据 $l_k(2)$ 在集合 $L(Y)$ 中删除,当 $G(l_k(2), C_1) = 0$ 或者 $G(l_k(2), C_1) \geq \frac{s(l_k(2))}{10}$ 时,则将候选数据 $l_k(2)$ 在集合 $L(Y)$ 中保留;

[0034] 当第二次筛选时在集合L(Y)中筛选出属于类集合 C_1 的数据时,则继续按照步骤(2)中的方法在集合L(Y)中进行第三次筛选,直到在当前筛选次数时没有在集合L(Y)中筛选出属于类集合 C_1 的数据时即停止在集合L(Y)中进行下一次的筛选;

[0035] Step(2):继续选取当前集合L(Y)中具有最大全局相似系数的候选数据为第二个类中心,将所述第二个类中心记为 c_2 ,所述类中心 c_2 所在的类集合记为 C_2 ,将类中心 c_2 在集合L(Y)中删除,并采用上述步骤(1)和步骤(2)中的方法在当前集合L(Y)中筛选出属于类集合 C_2 的其他候选数据;筛选完成后,将类集合 C_2 中的候选数据在当前集合L(Y)中删除;

[0036] 重复上述Step(1)和Step(2)中的方法直到当前集合L(Y)中剩余的候选数据个数为0时,即完成了在生命体征大数据中选取类中心,并完成了对所述生命体征大数据的初步聚类。

[0037] 优选地,根据选取的类中心对所述生命体征大数据中的数据进行聚类,并在聚类的过程中去除生命体征大数据中的噪声数据,具体为:

[0038] 根据选取的类中心和初步聚类结果对集合Y中剩余的未聚类数据进行聚类,设D(Y)表示集合Y中未聚类数据组成的集合, d_a 表示集合D(Y)中的第a个未聚类数据, $N(d_a)$ 表示集合Y中距离未聚类数据 d_a 最近的M个数据组成的邻域集合,定义 $h(d_a)$ 表示未聚类数据 d_a 在集合D(Y)中的聚类优先级,且 $h(d_a) = \frac{m(d_a)}{M} * s(d_a)$,其中, $m(d_a)$ 表示邻域集合 $N(d_a)$ 中已聚类数据的个数, $s(d_a)$ 表示未聚类数据 d_a 在集合Y中的全局相似系数;

[0039] Step1:对此时集合D(Y)中具有最大聚类优先级的未聚类数据进行优先聚类,设 d_e 表示集合D(Y)中的第e个未聚类数据,且 $d_e = \operatorname{argmax}_{d_a \in D(Y)} h(d_a)$, $N(d_e)$ 表示集合Y中距离未聚类数据 d_e 最近的M个数据组成的邻域集合, $m(d_e)$ 表示集合 $N(d_e)$ 中已聚类数据的个数;

[0040] 当 $m(d_e) = 0$ 时,则判定集合D(Y)中的未聚类数据均为噪声数据,并将所述噪声数据从D(Y)中都删除;

[0041] 当 $m(d_e) \neq 0$ 时,设 $J_{e,p}$ 表示集合 $N(d_e)$ 中的第p个已聚类数据,将已聚类数据 $J_{e,p}$ 所在的类集合表示为 $C_{e,p}$,定义 $J(d_e, C_{e,p})$ 为未聚类数据 d_e 和类集合 $C_{e,p}$ 之间的分布检测系数,则 $J(d_e, C_{e,p})$ 的计算公式为:

$$[0042] \quad J(d_e, C_{e,p}) = \begin{cases} \left| \frac{\sum_{J_{e,p,q} \in N'(J_{e,p})} |J_{e,p} - J_{e,p,q}|}{M} - |d_e - J_{e,p}| \right|, & M_{e,p} = 1 \\ \left| \frac{\sum_{J_{e,v} \in N(d_e)} \sum_{J_{e,v,b} \in N'(J_{e,v})} |J_{e,v} - J_{e,v,b}|}{M_{e,p} * M} - \frac{\sum_{J_{e,v} \in N(d_e)} |J_{e,v} - d_e|}{M_{e,p}} \right|, & M_{e,p} > 1 \end{cases}$$

[0043] 式中, $M_{e,p}$ 表示集合 $N(d_e)$ 中存在的属于类集合 $C_{e,p}$ 的已聚类数据的个数, $N'(J_{e,p})$ 表示类集合 $C_{e,p}$ 中距离已聚类数据 $J_{e,p}$ 最近的M个已聚类数据组成的集合, $J_{e,p,q}$ 表示集合 $N'(J_{e,p})$ 中的第q个已聚类数据, $J_{e,v}$ 表示集合 $N(d_e)$ 中的第v个已聚类数据,且 $J_{e,v}$ 为类集合 $C_{e,p}$ 中的数据, $N'(J_{e,v})$ 表示类集合 $C_{e,p}$ 中距离已聚类数据 $J_{e,v}$ 最近的M个已聚类数据组成的集合, $J_{e,v,b}$ 表示集合 $N'(J_{e,v})$ 中的第b个已聚类数据;

[0044] 设 $M(d_e)$ 表示集合 $N(d_e)$ 中的已聚类数据所处的不同类集合的个数, $C_{e,n}$ 表示数据 d_e 和所述 $M(d_e)$ 个类集合之间拥有最小分布检测系数的类集合,即 $C_{e,n} = \operatorname{argmin}_{1 \leq p \leq M(d_e)} J(d_e, C_{e,p})$;

[0045] 当未聚类数据 d_e 和类集合 $C_{e,n}$ 之间满足： $\left|s(d_e) - \frac{\sum_{y_{e,n,r} \in C_{e,n}} \rho(y_{e,n,r}, d_e) * s(y_{e,n,r})}{\sum_{y_{e,n,r} \in C_{e,n}} \rho(y_{e,n,r}, d_e)}\right| < \frac{s(d_e)}{10}$ 时，则将未聚类数据 d_e 加入到类集合 $C_{e,n}$ 中，并将未聚类数据 d_e 在集合D(Y)中删除，判定未聚类数据 d_e 为非噪声数据，当未聚类数据 d_e 和类集合 $C_{e,n}$ 之间满足： $\left|s(d_e) - \frac{\sum_{y_{e,n,r} \in C_{e,n}} \rho(y_{e,n,r}, d_e) * s(y_{e,n,r})}{\sum_{y_{e,n,r} \in C_{e,n}} \rho(y_{e,n,r}, d_e)}\right| \geq \frac{s(d_e)}{10}$ 时，则认定未聚类数据 d_e 为噪声数据，将未聚类数据 d_e 在集合D(Y)中删除，其中， $s(d_e)$ 表示未聚类数据 d_e 在集合Y中的全局相似系数， $y_{e,n,r}$ 表示类集合 $C_{e,n}$ 中的第r个数据， $s(y_{e,n,r})$ 表示数据 $y_{e,n,r}$ 在集合Y中的全局相似系数， $\rho(y_{e,n,r}, d_e)$ 表示数据 $y_{e,n,r}$ 相较于未聚类数据 d_e 的距离权值，且 $\rho(y_{e,n,r}, d_e) = \exp\left(-\frac{|y_{e,n,r} - d_e|}{\max_{y_{e,n,r} \in C_{e,n}} |y_{e,n,r} - d_e|}\right)$ ；

[0046] Step2:按照上述Step1中的方法重新在当前集合D(Y)中选取具有最大聚类优先级的数据进行优先聚类，直到集合D(Y)中的未聚类数据个数为0时，则停止聚类。

[0047] 本发明创造的有益效果：根据生命体征大数据建立根据患者的生命体征数据对患者的身体状态进行评估的健康评估模型，实现了远程对患者的身体状态进行统一监护，从而减轻了医护人员的工作量，并且能够在患者的身体状态处于危险时及时发现，从而提高了施救效率；采用大数据处理单元对生命体征大数据进行处理，并利用处理后的生命体征大数据对支持向量机进行训练，从而建立根据生命体征数据对患者的身体状态进行评估的健康评估模型，在利用生命体征大数据对支持向量机进行训练之前，先对生命体征大数据进行聚类处理，并在聚类处理的过程中去除生命体征大数据中的噪声数据，从而避免噪声数据对支持向量机的评估准确度的影响，采用聚类所得的类集合作为训练支持向量机的输入值，能够显著减少训练所需的时间并且提高了支持向量的性能；在对生命体征大数据的聚类过程中，提出一种新的类中心的选取方式，能够适应不同密度类和不同尺寸类的类中心的选取，具有较高的类中心的选取精度；提出了一种根据选取的类中心对生命体征大数据进行聚类的方法，能够在对大数据进行有效聚类的同时，避免噪声数据对聚类结果的影响，使得聚类结果具有较高的准确度。

附图说明

[0048] 利用附图对发明创造作进一步说明，但附图中的实施例不构成对本发明创造的任何限制，对于本领域的普通技术人员，在不付出创造性劳动的前提下，还可以根据以下附图获得其它的附图。

[0049] 图1是本发明结构示意图。

具体实施方式

[0050] 结合以下实施例对本发明作进一步描述。

[0051] 参见图1，本实施例的基于区块链的远程互联网大数据智慧医疗系统，包括生命体征采集模块、医疗大数据获取模块、区块链存储模块和智慧医疗终端，所述生命体征采集模块和患者的床位号进行绑定，用于采集患者的生命体征数据，并将采集的生命体征数据和患者的床位号通过互联网传输至智慧医疗终端，所述医疗大数据获取模块用于收集生命体

征大数据,并将收集的生命体征大数据传输至区块链存储模块进行存储,所述智慧医疗终端包括大数据处理单元、生命体征分析单元、智能预警单元、患者信息收录单元和人机交互单元,智慧医疗终端从区块链存储模块中调取所述生命体征大数据,并将调取的生命体征大数据输入至大数据处理单元进行处理,所述生命体征分析单元根据处理后的生命体征大数据建立根据生命体征数据对患者的身体状态进行评估的健康评估模型,将接收到的患者的生命体征数据输入到健康评估模型中,从而评估患者的身体状态是健康还是危险,当评估患者的身体状态为危险时令智能预警单元进行预警,所述患者信息收录单元用于记录患者的基础信息、患者的生命体征数据和评估所得的患者的身体状态,医护人员可以通过在人机交互单元中输入患者的基础信息查询患者的生命体征数据和患者的身体状态。

[0052] 优选地,所述生命体征分析单元采用支持向量机建立根据生命体征数据进行身体状态评估的健康评估模型,采用大数据处理单元处理后的生命体征大数据作为训练和测试所述支持向量机的样本集。

[0053] 优选地,所述患者的基础信息包括患者的姓名、年龄和床位号。

[0054] 本优选实施例提供一种远程的智慧医疗系统,根据生命体征大数据建立根据患者的生命体征数据对患者的身体状态进行评估的健康评估模型,实现了远程对患者的身体状态进行统一监护,从而减轻了医护人员的工作量,并且能够在患者的身体状态处于危险时及时发现,从而提高了施救效率。

[0055] 优选地,所述大数据处理单元用于对生命体征大数据进行聚类,并在聚类的过程中去除所述生命体征大数据中的噪声数据,确定生命体征大数据聚类所得的各个类集合对应的身体状态的标签,在支持向量机的训练过程中,将生命体征大数据的类集合作为支持向量机的输入值,将类集合所对应的身体状态的标签作为支持向量机的输出值。

[0056] 优选地,所述身体状态的标签包括健康和危险。

[0057] 本优选实施例采用大数据处理单元对生命体征大数据进行处理,并利用处理后的生命体征大数据对支持向量机进行训练,从而建立根据生命体征数据对患者的身体状态进行评估的健康评估模型,在利用生命体征大数据对支持向量机进行训练之前,先对生命体征大数据进行聚类处理,并在聚类处理的过程中去除生命体征大数据中的噪声数据,从而避免噪声数据对支持向量机的评估准确度的影响,采用聚类所得的类集合作为训练支持向量机的输入值,能够显著减少训练所需的时间并且提高了支持向量的性能。

[0058] 优选地,所述大数据处理单元用于对生命体征大数据进行聚类,并在聚类的过程中去除生命体征大数据中的噪声数据,具体包括:

[0059] (1) 在生命体征大数据中选取类中心;

[0060] (2) 根据选取的类中心对生命体征大数据中的数据进行聚类,并在聚类的过程中去除生命体征大数据中的噪声数据。

[0061] 优选地,采用下列方式在所述生命体征大数据中选取类中心:

[0062] 设Y表示生命体征大数据组成的集合, y_i 表示集合Y中的第i个数据,定义 $s(y_i)$ 表示数据 y_i 在集合Y中的全局相似系数,且 $s(y_i) = \exp\left(-\frac{\sum_{j=1}^{M(Y)} |y_i - y_j|}{M(Y)-1}\right)$,其中, y_j 表示集合Y中的

第j个数据, $m(Y)$ 表示集合Y中的数据个数;设 $U(y_i)$ 表示数据 y_i 的邻域数据集合,给定正整数M,M的值可以取5,采用下列步骤在集合Y中选取M个数据加入到集合 $U(y_i)$ 中:

[0063] 步骤1: 设 $y_i(1)$ 表示集合 Y 中距离数据 y_i 最近的数据, $\omega(y_i(1))$ 表示数据 $y_i(1)$ 的局部区域, 且 $\omega(y_i(1))$ 为以数据 $\frac{y_i(1)+y_i}{2}$ 为中心、以 $|y_i(1)-y_i|$ 为边长的正方体区域, 将数据 $y_i(1)$ 加入到集合 $U(y_i)$ 中;

[0064] 步骤2: 在局部区域 $\omega(y_i(1))$ 外的数据中确定距离数据 y_i 最近的数据, 将该数据记为 $y_i(2)$, 设 $\omega(y_i(2))$ 表示数据 $y_i(2)$ 的局部区域, 且 $\omega(y_i(2))$ 为以数据 $\frac{y_i(2)+y_i}{2}$ 为中心、以 $|y_i(2)-y_i|$ 为边长的正方体区域, 将数据 $y_i(2)$ 加入到集合 $U(y_i)$ 中;

[0065] 步骤3: 继续在局部区域 $\omega(y_i(1))$ 和局部区域 $\omega(y_i(2))$ 外的数据中确定距离数据 y_i 最近的数据, 将该数据记为 $y_i(3)$, 设 $\omega(y_i(3))$ 表示数据 $y_i(3)$ 的局部区域, 且 $\omega(y_i(3))$ 为以数据 $\frac{y_i(3)+y_i}{2}$ 为中心、以 $|y_i(3)-y_i|$ 为边长的正方体区域, 将数据 $y_i(3)$ 加入到集合 $U(y_i)$ 中;

[0066] 步骤4: 继续按照上述步骤1、步骤2和步骤3中的方法确定数据加入到集合 $U(y_i)$ 中, 直到集合 $U(y_i)$ 中的数据个数等于 M 时, 则停止向集合 $U(y_i)$ 中加入数据;

[0067] 采用下列公式在集合 Y 中筛选出可作为类中心的候选数据:

$$[0068] \quad f(y_i) = \left| s(y_i) - \frac{\sum_{l=1}^M \rho(y_i(l), y_i) * s(y_i(l))}{\sum_{l=1}^M \rho(y_i(l), y_i)} \right|$$

[0069] 式中, $f(y_i)$ 表示数据 y_i 在集合 Y 中的类中心属性值, $y_i(1)$ 表示集合 $U(y_i)$ 中的第1个数据, $s(y_i(1))$ 表示数据 $y_i(1)$ 在集合 Y 中的全局相似系数, $\rho(y_i(1), y_i)$ 表示数据 $y_i(1)$ 相较于数据 y_i 的距离权值, 且 $\rho(y_i(l), y_i) = \exp\left(-\frac{|y_i-y_i(l)|}{\max_{1 \leq l \leq M} |y_i-y_i(l)|}\right)$;

[0070] 当数据 y_i 的类中心属性值 $f(y_i) < \frac{s(y_i)}{10}$ 时, 则将数据 y_i 视为类中心的候选数据, 且判定所述类中心的候选数据为非噪声数据; 当数据 y_i 的类中心属性值 $f(y_i) \geq \frac{s(y_i)}{10}$ 时, 则将数据 y_i 视为未聚类数据;

[0071] 设 $L(Y)$ 表示集合 Y 中视为类中心的候选数据组成的集合, 在集合 $L(Y)$ 中选取类中心, 并根据选取的类中心对集合 $L(Y)$ 中的数据进行聚类, 具体包括:

[0072] Step(1): 选取集合 $L(Y)$ 中具有最大全局相似系数的候选数据为第一个类中心, 将所述第一个类中心记为 c_1 , 类中心 c_1 所在的类集合记为 C_1 , 将类中心 c_1 在集合 $L(Y)$ 中删除, 并采用下列步骤在当前集合 $L(Y)$ 中筛选出属于类集合 C_1 中的候选数据, 具体为:

[0073] 步骤(1): 设 $l_k(1)$ 表示第1次筛选时集合 $L(Y)$ 中的第 k 个候选数据, 定义 $G(l_k(1), C_1)$ 表示候选数据 $l_k(1)$ 和类集合 C_1 之间的聚类函数, 且 $G(l_k(1), C_1)$ 的表达式为:

$$[0074] \quad G(l_k(1), C_1) = \theta(l_k(1), C_1) * |s(c_1) - s(l_k(1))|$$

[0075] 式中, $\theta(l_k(1), C_1)$ 表示判断函数, 设 $N(l_k(1))$ 表示在集合 Y 中选取的距离候选数据 $l_k(1)$ 最近的 M 个数据组成的邻域集合, $m(l_k(1), C_1)$ 表示邻域集合 $N(l_k(1))$ 中存在的属于类集合 C_1 的数据个数, 当 $m(l_k(1), C_1) \neq 0$ 时, $\theta(l_k(1), C_1) = 1$, 当 $m(l_k(1), C_1) = 0$ 时, $\theta(l_k(1), C_1) = 0$, $s(c_1)$ 表示类中心 c_1 在集合 Y 中的全局相似系数, $s(l_k(1))$ 表示候选数据 $l_k(1)$ 在集合 Y 中的全局相似系数;

[0076] 当 $0 < G(l_k(1), C_1) < \frac{s(c_1)}{10}$ 时, 则判定候选数据 $l_k(1)$ 为类集合 C_1 中的数据, 将候选数据 $l_k(1)$ 加入到类集合 C_1 中, 并将候选数据 $l_k(1)$ 在集合 $L(Y)$ 中删除, 当 $G(l_k(1), C_1) = 0$ 或者 $G(l_k(1), C_1) \geq \frac{s(c_1)}{10}$ 时, 则将候选数据 $l_k(1)$ 在集合 $L(Y)$ 中保留;

[0077] 步骤 (2), 设 $l_k(2)$ 表示第 2 次筛选时当前集合 $L(Y)$ 中的第 K 个候选数据, 定义 $G(l_k(2), C_1)$ 表示候选数据 $l_k(2)$ 和类集合 C_1 之间的聚类函数, 且 $G(l_k(2), C_1)$ 的表达式为:

$$[0078] \quad G(l_k(2), C_1) = \vartheta(l_k(2), C_1) * \left| s(l_k(2)) - \frac{\sum_{y_{1,z} \in C_1} \rho(y_{1,z}, l_k(2)) * s(y_{1,z})}{\sum_{y_{1,z} \in C_1} \rho(y_{1,z}, l_k(2))} \right|$$

[0079] 其中, $\vartheta(l_k(2), C_1)$ 表示判断函数, 设 $N(l_k(2))$ 表示在集合 Y 中选取距离候选数据 $l_k(2)$ 最近的 M 个数据组成的邻域集合, $m(l_k(2), C_1)$ 表示邻域集合 $N(l_k(2))$ 中存在的属于类集合 C_1 的数据个数, 当 $m(l_k(2), C_1) \neq 0$ 时, $\vartheta(l_k(2), C_1) = 1$, 当 $m(l_k(2), C_1) = 0$ 时, $\vartheta(l_k(2), C_1) = 0$, $s(l_k(2))$ 表示候选数据 $l_k(2)$ 在集合 Y 中的全局相似系数, $y_{1,z}$ 表示类集合 C_1 中的第 z 个数据, $\rho(y_{1,z}, l_k(2))$ 表示数据 $y_{1,z}$ 相较于候选数据 $l_k(2)$ 的距离权值, 且 $\rho(y_{1,z}, l_k(2)) =$

$$\exp\left(-\frac{|y_{1,z}-l_k(2)|}{\max_{y_{1,z} \in C_1} |y_{1,z}-l_k(2)|}\right), s(y_{1,z}) \text{ 表示数据 } y_{1,z} \text{ 在集合 } Y \text{ 中的全局相似系数};$$

[0080] 当 $0 < G(l_k(2), C_1) < \frac{s(l_k(2))}{10}$ 时, 则判定候选数据 $l_k(2)$ 为类集合 C_1 中的数据, 将候选数据 $l_k(2)$ 加入到类集合 C_1 中, 并将候选数据 $l_k(2)$ 在集合 $L(Y)$ 中删除, 当 $G(l_k(2), C_1) = 0$ 或者 $G(l_k(2), C_1) \geq \frac{s(l_k(2))}{10}$ 时, 则将候选数据 $l_k(2)$ 在集合 $L(Y)$ 中保留;

[0081] 当第二次筛选时在集合 $L(Y)$ 中筛选出属于类集合 C_1 的数据时, 则继续按照步骤 (2) 中的方法在集合 $L(Y)$ 中进行第三次筛选, 直到在当前筛选次数时没有在集合 $L(Y)$ 中筛选出属于类集合 C_1 的数据时即停止在集合 $L(Y)$ 中进行下一次的筛选;

[0082] Step (2): 继续选取当前集合 $L(Y)$ 中具有最大全局相似系数的候选数据为第二个类中心, 将所述第二个类中心记为 c_2 , 所述类中心 c_2 所在的类集合记为 C_2 , 将类中心 c_2 在集合 $L(Y)$ 中删除, 并采用上述步骤 (1) 和步骤 (2) 中的方法在当前集合 $L(Y)$ 中筛选出属于类集合 C_2 中的候选数据; 筛选完成后, 将类集合 C_2 中的候选数据在当前集合 $L(Y)$ 中删除;

[0083] 重复上述 Step (1) 和 Step (2) 中的方法直到当前集合 $L(Y)$ 中剩余的候选数据个数为 0 时, 即完成了在生命体征大数据中选取类中心, 并完成了对所述生命体征大数据的初步聚类。

[0084] 本优选实施例用于在生命体征大数据中选取类中心, 从而根据选取的类中心对生命体征大数据进行聚类。在对大数据进行聚类时, 类中心的选取直接影响着后期聚类结果的准确度和聚类的效率, 也决定着噪声数据检测的准确度; 传统的类中心的选取方式多数容易受到类密度、类尺寸的影响, 从而容易选取出高密度类、具有较小尺寸类的类中心而忽略低密度类或者具有较大尺寸类的类中心, 从而影响最终的聚类效果, 针对上述现象, 本优选实施例提出的在生命体征大数据中筛选出可作为类中心的候选数据的方式能够有效的筛选出不同密度、不同尺寸类的类中心, 即提出的类中心的筛选方式不受类密度和类尺寸的影响, 对于低密度类或者具有较大尺寸类的类具有相同的类中心的检测精度, 因为本优选

实施例的类中心的筛选方式是通过计算所述数据的全局相似系数和其邻域数据的全局相似系数的加权均值之间的绝对差值来衡量所述数据的中心属性,数据的全局相似系数能够有效的衡量所述数据在生命体征大数据中的分布特性,邻域数据集合中邻域数据的选取方式能够保证所述数据在选取的邻域数据中的中心性,避免选取的邻域数据都处于所述数据一边的现象,再根据所述数据的全局相似系数和其邻域数据的全局相似系数的加权均值之间的绝对差值来衡量所述数据的中心属性,当所述数据处于类中心或类中心附近时,无论其处于何种密度或尺寸的类中,其全局相似系数和选取的邻域数据的全局相似系数都具有较大的相似性,因此,通过计算所述数据和其邻域数据之间全局相似系数的相似性能够有效的判断所述数据的中心属性,并且不受类密度或者类尺寸的影响,从而提高较小密度或者较大尺寸类的类中心的检测精度;按照上述方法能够有效的筛选出生命体征大数据中处于类中心或者类中心附近的数据,继续采用本优选实施例提出的类中心的选取方式,能够有效的选取类中心的同时,将类中心附近的数据聚类到其所对应的类集合中,即完成了生命体征大数据的初步聚类,为接下来的聚类和噪声检测奠定了基础。

[0085] 优选地,根据选取的类中心对所述生命体征大数据中的数据进行聚类,并在聚类的过程中去除生命体征大数据中的噪声数据,具体为:

[0086] 根据选取的类中心和初步聚类结果对集合Y中剩余的未聚类数据进行聚类,设D(Y)表示集合Y中未聚类数据组成的集合, d_a 表示集合D(Y)中的第a个未聚类数据, $N(d_a)$ 表示集合Y中距离未聚类数据 d_a 最近的M个数据组成的邻域集合,定义 $h(d_a)$ 表示未聚类数据 d_a 在集合D(Y)中的聚类优先级,且 $h(d_a) = \frac{m(d_a)}{M} * s(d_a)$,其中, $m(d_a)$ 表示邻域集合 $N(d_a)$ 中已聚类数据的个数, $s(d_a)$ 表示未聚类数据 d_a 在集合Y中的全局相似系数;

[0087] Step1:对此时集合D(Y)中具有最大聚类优先级的未聚类数据进行优先聚类,设 d_e 表示集合D(Y)中的第e个未聚类数据,且 $d_e = \operatorname{argmax}_{d_a \in D(Y)} h(d_a)$, $N(d_e)$ 表示集合Y中距离未聚类数据 d_e 最近的M个数据组成的邻域集合, $m(d_e)$ 表示集合 $N(d_e)$ 中已聚类数据的个数;

[0088] 当 $m(d_e) = 0$ 时,则判定集合D(Y)中的未聚类数据均为噪声数据,并将所述噪声数据从D(Y)中都删除;

[0089] 当 $m(d_e) \neq 0$ 时,设 $J_{e,p}$ 表示集合 $N(d_e)$ 中的第p个已聚类数据,将已聚类数据 $J_{e,p}$ 所在的类集合表示为 $C_{e,p}$,定义 $J(d_e, C_{e,p})$ 为未聚类数据 d_e 和类集合 $C_{e,p}$ 之间的分布检测系数,则 $J(d_e, C_{e,p})$ 的计算公式为:

$$[0090] \quad J(d_e, C_{e,p}) = \begin{cases} \left| \frac{\sum_{J_{e,p,q} \in N'(J_{e,p})} |J_{e,p} - J_{e,p,q}|}{M} - |d_e - J_{e,p}| \right|, & M_{e,p} = 1 \\ \left| \frac{\sum_{J_{e,v} \in N(d_e)} \sum_{J_{e,v,b} \in N'(J_{e,v})} |J_{e,v} - J_{e,v,b}|}{M_{e,p} * M} - \frac{\sum_{J_{e,v} \in N(d_e)} |J_{e,v} - d_e|}{M_{e,p}} \right|, & M_{e,p} > 1 \end{cases}$$

[0091] 式中, $M_{e,p}$ 表示集合 $N(d_e)$ 中存在的属于类集合 $C_{e,p}$ 的已聚类数据的个数, $N'(J_{e,p})$ 表示类集合 $C_{e,p}$ 中距离已聚类数据 $J_{e,p}$ 最近的M个已聚类数据组成的集合, $J_{e,p,q}$ 表示集合 $N'(J_{e,p})$ 中的第q个已聚类数据, $J_{e,v}$ 表示集合 $N(d_e)$ 中的第v个已聚类数据,且 $J_{e,v}$ 为类集合 $C_{e,p}$ 中的数据, $N'(J_{e,v})$ 表示类集合 $C_{e,p}$ 中距离已聚类数据 $J_{e,v}$ 最近的M个已聚类数据组成的集合, $J_{e,v,b}$ 表示集合 $N'(J_{e,v})$ 中的第b个已聚类数据;

[0092] 设 $M(d_e)$ 表示集合 $N(d_e)$ 中的已聚类数据所处的不同类集合的个数, $C_{e,n}$ 表示数据 d_e 和所述 $M(d_e)$ 个类集合之间拥有最小分布检测系数的类集合,即 $C_{e,n} = \operatorname{argmin}_{1 \leq p \leq M(d_e)} J(d_e, C_{e,p})$;

[0093] 当未聚类数据 d_e 和类集合 $C_{e,n}$ 之间满足: $\left|s(d_e) - \frac{\sum_{y_{e,n,r} \in C_{e,n}} \rho(y_{e,n,r}, d_e) * s(y_{e,n,r})}{\sum_{y_{e,n,r} \in C_{e,n}} \rho(y_{e,n,r}, d_e)}\right| < \frac{s(d_e)}{10}$

时,则将未聚类数据 d_e 加入到类集合 $C_{e,n}$ 中,并将未聚类数据 d_e 在集合 $D(Y)$ 中删除,判定未聚类数据 d_e 为非噪声数据,当未聚类数据 d_e 和类集合 $C_{e,n}$ 之间满足: $\left|s(d_e) -$

$\frac{\sum_{y_{e,n,r} \in C_{e,n}} \rho(y_{e,n,r}, d_e) * s(y_{e,n,r})}{\sum_{y_{e,n,r} \in C_{e,n}} \rho(y_{e,n,r}, d_e)}\right| \geq \frac{s(d_e)}{10}$ 时,则认定未聚类数据 d_e 为噪声数据,将未聚类数据 d_e 在

集合 $D(Y)$ 中删除,其中, $s(d_e)$ 表示未聚类数据 d_e 在集合 Y 中的全局相似系数, $y_{e,n,r}$ 表示类集合 $C_{e,n}$ 中的第 r 个数据, $s(y_{e,n,r})$ 表示数据 $y_{e,n,r}$ 在集合 Y 中的全局相似系数, $\rho(y_{e,n,r}, d_e)$ 表示

数据 $y_{e,n,r}$ 相较于未聚类数据 d_e 的距离权值,且 $\rho(y_{e,n,r}, d_e) = \exp\left(-\frac{|y_{e,n,r} - d_e|}{\max_{y_{e,n,r} \in C_{e,n}} |y_{e,n,r} - d_e|}\right)$;

[0094] Step2:按照上述Step1中的方法重新在当前集合 $D(Y)$ 中选取具有最大聚类优先级的数据进行优先聚类,直到集合 $D(Y)$ 中的未聚类数据个数为0时,则停止聚类。

[0095] 本优选实施例用于根据选取的类中心和初步聚类结果对生命体征大数据中的未聚类数据进行聚类,并去除所述生命体征大数据中的噪声数据,对未聚类数据定义聚类优先级,所述聚类优先级综合考虑了未聚类数据的全局相似系数和其邻域集合中的已聚类数据个数,当所述未聚类数据拥有较大的全局相似系数并且其邻域集合中拥有较多的已聚类数据时,其具有较大概率为类集合中的数据,因此采用迭代的方式在未聚类数据集合中选取具有最大聚类优先级的未聚类数据进行聚类,能够保证优先对未聚类数据中的非噪声数据进行聚类,并且为接下来的未聚类数据的聚类奠定基础的同时,能够避免噪声数据对聚类结果的影响;在对具有最大聚类优先级的未聚类数据进行聚类时,当此时的最大聚类优先级为0时,则判定生命体征大数据中剩余的未聚类数据均为噪声数据,当此时的最大聚类优先级不为0时,则定义分布检测系数用于衡量所述未聚类数据和其邻域集合中已聚类数据所处类集合之间的分布相似性,定义的分布检测系数通过计算邻域集合中已聚类数据和其所在类集合中 M 个较近的已聚类数据之间的平均距离来衡量待检测类集合中数据的分布特性,通过计算所述未聚类数据和其邻域集合中已聚类数据之间的距离来衡量所述数据和待检测类集合中数据之间的分布特性,最后将这两者之间的分布特性进行比较,和未聚类数据的分布特性最为相似的类集合具有最大概率为未聚类数据所属于的类集合,因此,选取和未聚类数据之间具有最小分布检测系数值的类集合进行检测,在检测过程中,通过比较未聚类数据和待检测类集合中的数据的全局相似系数之间的相似性来判断所述未聚类数据是否为该类集合的数据,从而能够有效的对未聚类数据进行聚类,并且能够有效的检测出未聚类数据中的噪声数据。

[0096] 最后应当说明的是,以上实施例仅用以说明本发明的技术方案,而非对本发明保护范围的限制,尽管参照较佳实施例对本发明作了详细地说明,本领域的普通技术人员应当理解,可以对本发明的技术方案进行修改或者等同替换,而不脱离本发明技术方案的实质和范围。

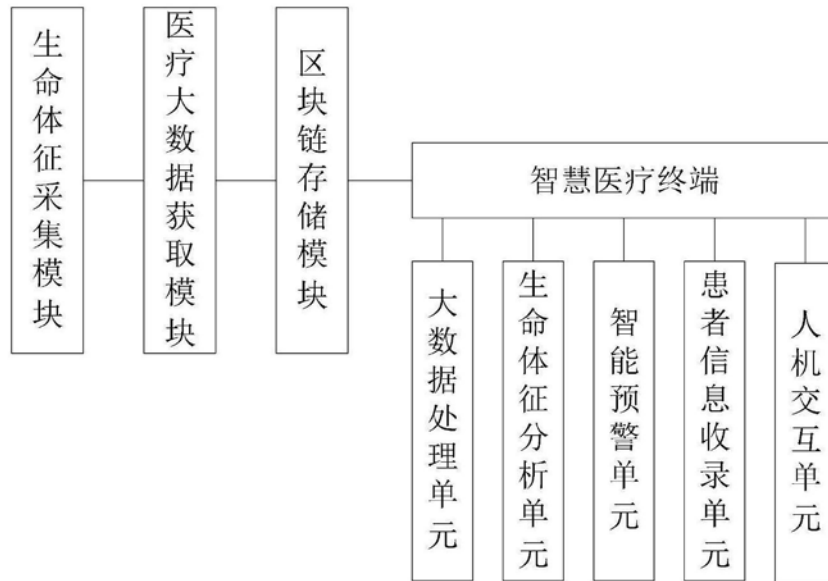


图1