

基于粗糙集的慢性病变分级方法^①

胡建强, 王 元

(厦门理工学院 计算机与信息工程学院, 厦门 361024)

通讯作者: 胡建强, E-mail: jqhucn@xmut.edu.cn

摘 要: 由于云健康生理监测数据因具有时间连续性、非精确性、模糊性等特性, 从而导致传统分类算法很难直接运用. 针对上述问题, 提出一种基于粗糙集的慢性病变分级方法. 该方法首先采用融合相关度和 Chi-merge 统计量离散化生理监测数据; 然后, 基于相容矩阵的属性约简算法去除数据冗余属性; 最后, 基于批量与增量相结合挖掘分类规则, 并基于分布计算框架 MapReduce 应用上述规则实现慢性病变智能分级. 实验验证表明, 该方法具有较高的识别准确率, 有助于个体全面认识健康风险状况.

关键词: 云健康监测; 智能分级; 生理监测数据离散化

引用格式: 胡建强, 王元. 基于粗糙集的慢性病变分级方法. 计算机系统应用, 2018, 27(12): 268-273. <http://www.c-s-a.org.cn/1003-3254/6655.html>

Classification Method of Chronic Lesions Based on Rough Sets

HU Jian-Qiang, WANG Yuan

(School of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China)

Abstract: Because of physiological monitoring data has time continuity, inaccuracy, and fuzziness, the traditional classification algorithm is difficult to be used directly. In view of the above problems, a classification method of chronic lesions based on rough sets is proposed. First, the physiological monitoring data are discretized based on fusion of correlation and Chi-merge statistics. Then, this method uses the attribute reduction algorithm based on the compatibility matrix to remove the redundant attributes of the data. Finally, classification rules are mined based on batch and incremental data, and intelligent classification of chronic diseases can be realized by applying the above rules based on MapReduce framework. Experiments show that the method has a high recognition rate, which is helpful for the individual to fully understand the health risks.

Key words: cloud health monitoring; intelligent classification; physiological monitoring data discretization

中国卫生部《健康中国 2020 战略研究报告》明确指出: 加强基层医疗体系建设, 建设基于网络体系的全民电子健康系统、推进预防慢性病及重大疾病, 大力发展民族健康产业, 投入资金总规模将达 4000 亿. 云健康监测综合运用传感器网络、云计算和大数据处理等技术, 可以实现多生理参数实时采集、海量数据

存储和大数据智能分析与处理, 是下一代家庭健康监测的主要发展方向.

目前国内外学术界围绕健康监测领域展开很多积极的探索, 主要工作包括: 美国田纳西大学设计一种实时监测人体生理 ECG 数据云 Cloud-ECG^[1], 保证云端数据存取的 ECG 数据质量. 西班牙格拉纳达大学采用

① 基金项目: 国家自然科学基金 (61872436); 福建省自然科学基金 (2018J01570); 厦门市高校科技创新项目 (3502Z20173035); 赛尔网络下一代互联网技术创新项目 (NGI20160708)

Foundation item: National Natural Science Foundation of China (61872436); Natural Science Foundation of Fujian Province (2018J01570); Xiamen Science and Technology Foundation (3502Z20173035); Next Generation of Internet Technology of CERNET Innovation Program (NGI20160708)

收稿时间: 2018-04-12; 修改时间: 2018-05-08; 采用时间: 2018-05-23; csa 在线出版时间: 2018-11-30

可穿戴式胸带传感器监测人体的心电、心率、呼吸频率、皮肤温度,生理监测数据存放在个人微云(Cloudlet);心电数据采用马拉松算法处理并以红、绿、黄颜色警示患者身体健康状况^[2].台湾国立台中科技大学基于体域网 WBANs(Wireless Body Area Networks)采集人体的血压、心率和体温等生理数据;采用 VMware 搭建的微云存储数据;采用融合粗糙集的等级评定算法^[3]协同推荐信息到相关医疗群体^[4].飞利浦保健算法研究中心提出一种 ECG 波形实时质量评估算法—ST 段心肌梗死评估算法^[5,6].东南大学采用织物电极采集 ECG 数据^[7]、北京工业大学提出采用可穿戴式多参数心脏活动监测设备采集心音、心电和心冲击图数据,在线监测心脏活动^[8];中科院提出“低成本健康海云工程”,采用无缝接入的智能医疗终端监测患者生理健康状况;利用云计算技术实现海量电子病历和电子健康档案存储^[9],但未给出生理监测数据的处理算法.

云健康监测可以有效解决现有家庭健康监测系统(终端)在数据存储容量方面的局限性,但生理监测数据仍缺乏高效、实用的智能分析算法,特别是传统的分类算法(例如 SVM, KNN, 决策树等)很难直接应用,难以满足居民慢性病智能预警和突发状况救治需要.其主要原因是:生理监测数据具有时间连续性、非精确性、模糊性等特性,实际应用存在数据遗漏或缺失.例如,体温数据相对恒定,但 24 小时略有波动;动态 ECG 信号需采用小波分析算法对 QRS 复合波实时检测和有效筛选,24 小时内波动较大^[10].

针对上述问题,本文提出一种基于粗糙集的慢性病变分级方法,帮助个体全面认识慢性病变风险,该方法包括以下步骤:首先,采用融合相关度和 Chi-merge 统计量对连续属性离散化,提高聚类能力;然后,基于相容矩阵的属性约简算法去除数据冗余属性,降低原始数据集维数;最后,基于批量和增量相结合的分类规则挖掘,确定最优分类规则集合,并基于 MapReduce 框架运用上述规则实现慢性病变智能分级.结合厦门市集美区病例数据验证算法,基于自主研发的联接厦门市“健康医疗云”的云健康监测系统 Cloud-Health^[11]平台展开上述方法的有效性验证.

1 方法原理

1.1 融合相关度和 Chi-merge 统计量的属性离散化

离散化本质上是利用选取的断点来对条件属性构成的空间进行划分,即将 m 维空间划分有限个区域.为

提高系统的聚类能力,增加对生理监测数据噪音的鲁棒性,应该采用尽可能少的断点来完成划分过程.随着属性个数和样本量的增加,候选断点数将会成倍增长^[9].断点选择对保证最小的结果断点集合至关重要.下面给出融合相关度和 Chi-merge 统计量的属性离散化的一般性方法.

定义 1^[12]. 决策表 $S = (U, R, V, f)$, $U = \{x_1, x_2, \dots, x_n\}$ 是有限的对象集合; $R = C \cup D$ 是属性集合,子集合 C 和 D 分别为条件属性集 $C = \{c_1, c_2, \dots, c_m\}$ 和决策属性集 $D = \{d\}$; m 为条件属性的个数.假设决策种类的个数为 $r(d)$, $f: U \times R$ 是一个函数,指定 U 中的每一对象 x 的属性,属性 c_i 的值域 V_{c_i} 为:

$$V_{c_i} = \{[v_{c_i}^{\min(x_1)}, v_{c_i}^{\max(x_1)}], [v_{c_i}^{\min(x_2)}, v_{c_i}^{\max(x_2)}], \dots, [v_{c_i}^{\min(x_n)}, v_{c_i}^{\max(x_n)}]\} \quad (1)$$

上式中, $\min(x_1)$ 表示 x_1^{\min} , $\max(x_1)$ 表示 x_1^{\max} , 其他同理.其中 n 为论域有限对象的个数.

定义 2. 设 $a = [a_1, a_2]$, $b = [b_1, b_2]$, 设 a^{\min} , a^{\max} , b^{\min} , b^{\max} 表示区间中最小或最大数, L 为区间 a, b 中区间长度的较大值,则区间 a, b 之间的相关度:

$$g(a, b) = \begin{cases} 0 & b^{\max} \leq a^{\min}, a^{\max} \leq b^{\min} \\ (a^{\max} - b^{\min})/L & a^{\min} < b^{\min} \leq a^{\max} \leq b^{\max} \\ 1 & a^{\min} < b^{\min} \leq b^{\max} \leq a^{\max} \\ (b^{\max} - a^{\min})/L & b^{\min} < a^{\min} \leq b^{\max} < a^{\max} \\ 1 & b^{\min} \leq a^{\min} < a^{\max} \leq b^{\max} \end{cases} \quad (2)$$

定义 3. Chi-square 是一种度量离散化数值型数据的统计量

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^p \frac{(A_{ij} - E_{ij})^2}{E_{ij}} \quad (3)$$

其中, p 表示目标属性个数, A_{ij} 表示第 i 个区间中第 j 个目标属性的实例数; E_{ij} 表示第 i 区间中,第 j 个目标属性出现频率(期望).

为克服相似区间长度不固定且区间长度相差较大问题,基于相关度离散化区间属性;本文采用 Chi-square 度量条件属性和目标属性之间关联程度,即两相邻区如果存在相似目标属性频率分布则合并为同一区间,即计算相邻区间 χ^2 值,合并 χ^2 值最小的相邻区间,迭代合并直到满足终止条件.具体描述如下:

算法 1. 融合相关度和 Chi-square 统计量的属性离散化算法

① 根据对象 x_i 在属性 c_j 的取值特点(区间数长度和相邻区间的交叉

新段落大小), 定义 c_j 相关度 g_j ;

② 对象 $U=\{x_1, x_2, \dots, x_n\}$ 属性 c_j 上的取值排序, 记为 $U^{(c_j)}=\{x_1^{(c_j)}, x_2^{(c_j)}, \dots, x_n^{(c_j)}\}$;

③ 遍历 $U^{(c_j)}$ 比较相邻对象 $x_i^{(c_j)}$ ($1 \leq i \leq n$)

若 $g(x_i^{(c_j)}, x_{i+1}^{(c_j)}) > g_j$, 继续遍历;

若 $g(x_i^{(c_j)}, x_{i+1}^{(c_j)}) \leq g_j$ 且 $f(x_i^{(c_j)}) = f(x_{i+1}^{(c_j)})$, 决策相同继续遍历;

若 $g(x_i^{(c_j)}, x_{i+1}^{(c_j)}) \leq g_j$ 且 $f(x_i^{(c_j)}) \neq f(x_{i+1}^{(c_j)})$, 添加断点 $(x_i^{(c_j)}, x_{i+1}^{(c_j)})$ 区间最小值);

④ 按遍历取断点从小到大排序实现连续数据离散化; 选取显著性水平并结合自由度 $(r(d)-1)$, 确定 c_j 统计量阈值 χ_j^2 ;

⑤ 计算连续数据离散化相邻区间 χ^2 统计量, 如果大于 χ_j^2 , 则相邻区间合并;

⑥ 按⑤循环迭代合并直至满足: 相邻区间 χ^2 统计量大于阈值 χ_j^2 或最终保留区间个数.

1.2 基于相容矩阵的属性约简算法

由于生理监测数据涉及较多属性, 且容易存在部分条件属性(如心率、血氧)取值有遗漏或缺失, 需要降低原始数据集维数, 同时避免组合爆炸和提高属性约简效率. 在保证一定的正确分类率 α 的前提下, 尽可能缩小属性约简的搜索空间.

定义 4. 对于决策表 $S=(U, R, V, f)$, 对于 $\forall \alpha \in (0.8, 1]$, 决策属性集 D 与条件属性集 C 的基于参数 α 的分类率为

$$\beta_C = \frac{\left| \bigcup_{Y_j \in U/D} C_\alpha Y_j \right|}{|U|} \quad (4)$$

其中, $C_\alpha Y_j$ 表示决策类 Y_j 相对于条件属性集 C 的 α 下的近似值, α 参数放松 $S=(U, R, V, f)$ 对近似边界的要求.

定义 5. 条件属性 C 的相容矩阵定义如下:

$$M_C = (r_{ij})_{|U| \times |U|} = \begin{cases} 0 & \exists c \in C, f(x_i, c) \neq f(x_j, c) \wedge f(x_i, c) \neq * \wedge f(x_j, c) \neq *, f(x_i, D) \neq f(x_j, D) \\ 1 & \text{其他} \end{cases} \quad (5)$$

$1 \leq i, j \leq |U|$

其中, 若条件属性的取值有遗漏或缺失, *记为遗漏属性值.

定义 6. 对于决策表 $S=(U, R, V, f)$, P, Q 是条件属性集的两个子集, 相容矩阵的交运算定义为:

$$M_P \cap M_Q = (r_{ij})_{|U| \times |U|} \cap (r_{ij})_{|U| \times |U|} = (\min(r_{ij}, r_{ij}))_{|U| \times |U|}. \quad (6)$$

下面给出基于相容矩阵的属性约简算法: 首先, 采用间接处理方法将不完备数据转化成完备数据; 其次, 将决策表中条件属性的所有组合形式用二进制编码表示, 即“1”表示属性出现, “0”表示属性不出现; 然后, 比较二进制表示的组合属性集与条件属性的分类质量,

如果相等则该组合属性作为候选集合一行; 最后, 根据准则删除那些子集已经是约简的行以及全为零的行.

算法 2. 基于相容矩阵的属性约简算法

输入: 决策表 $S=(U, R, V, f)$, $R=C \cup D$ 是属性集合, 子集 C 和 D 分别为条件属性集 $C=\{c_1, c_2, \dots, c_m\}$ 和决策属性集 $D=\{d\}$; m 为条件属性的个数; 分类率 α

输出: 最小约简属性集 C

① 采用相容矩阵并基于启发式算法^[13]将不完备决策表 $S=(U, R, V, f)$ 转化为完备数据集;

② 将条件属性 $C=\{c_1, c_2, \dots, c_m\}$ 所有组合形式通过二进制编码表示;

③ 计算编码 $j=1, 2, \dots, 2^m-1$ 所对应的属性组合 C_j 在参数 α 条件下的分类质量 β_{C_j}

$$\beta_{C_j} = \frac{\left| \bigcup_{Y_j \in U/D} C_\alpha Y_j \right|}{|U|} \quad (7)$$

④ 如果 $\beta_{C_j} = \beta_C$, 则用属性组合 C_j 代替属性集 C , 即 $C=C_j$; 重复步骤②;

⑤ 删除冗余及全为零的行, 形成最小约简属性集 C .

1.3 批量和增量相结合的分类规则挖掘

生理监测数据具有典型增量数据特点, 分类规则可以对未知数据预测, 由于分类规则强调准确性、速度、健壮性和伸缩性, 因此确定最优分类规则集合至关重要. 按照定义 1, 决策表 S 每一行对应决策规则, 存在 $r(d)$ 个决策规则, 可以表示为:

$$C(x_1), C(x_2), \dots, C(x_n) \rightarrow D_1(x_1), D_2(x_2), \dots, D_{r(d)}(x_n), \quad (8)$$

记为 $C_U \rightarrow D_U$, 可以定义规则强度、确定度和覆盖度因子.

定义 7. 决策规则 $C_U \rightarrow D_U$ 规则强度、确定度和覆盖度分别为:

$$str(C, D) = \frac{|C_U \cap D_U|}{|U|} \quad (9)$$

$$cer(C, D) = \frac{|C_U \cap D_U|}{|C_U|} \quad (10)$$

$$cov(C, D) = \frac{|C_U \cap D_U|}{|D_U|} \quad (11)$$

其中, $|\cdot|$ 运算表示集合基数.

算法 3. 基于批量和增量相结合的分类规则挖掘算法

① 根据生理监测数据确定对象的条件属性 C 和决策属性 D , 决策表 $S=(U, R, V, f)$, 选取生理监测数据集 $Data_0$;

② 对 $Data_0$ 进行预处理, 融合相关度和 Chi-merge 统计量的属性离散化连续属性值; 基于相容矩阵约简属性, 得到数据集 $Data_C$;

③ 调用基于粗糙集分类算法^[14]得到一组最优的分类规则集并保存在分类规则库中;

- ④ 读取增量数据集并进行预处理(融合相关度和 Chi-merge 统计量的属性离散化),得到 $Data_C$; 如果增量数据集为空,跳转到⑥;
- ⑤ 用当前分类规则库中的规则对 $Data_C$ 进行分类, 如果达到预期效果(规则强度、确定度和覆盖度), 则将 $Data_C$ 与 $\Delta Data_C$ 合并 $Data_C = Data_C \cup \Delta Data_C$, 跳转到④; 否则执行③;
- ⑥ 输出当前分类规则集.

算法 3 采用顺序覆盖策略, 每学习一个规则可删除该规则覆盖数据; 对其它数据重复该过程直到终止条件, 即没有训练样本/返回的规则的分类质量低于预期效果(规则强度、确定度和覆盖度).

为提高慢性病变智能分级的准确性, 特别是针对大规模生理监测数据, 需要发挥云端并行化处理优势. MapReduce 并行化框架通过对数据集的大规模操作分发给网络上的每个节点; 每个节点会周期性返回所完成的工作和最新状态. 用户只需编写 Map 函数和 Reduce 函数, 即可实现大规模数据并行处理. 慢性疾病智能分类及预测的完整过程: 实时监测获取生理数据; 经过融合相关度和 Chi-merge 统计量对连续属性离散、基于相容矩阵的属性约简算法降低原始数据集维数以及基于批量和增量相结合的分类规则挖掘; 在线生理监测数据利用上述分类规则聚类, 确定病例智能分级.

2 实验

厦门理工学院自主研制云健康监测系统 Cloud-Health, 该系统基于 Android 4.1.2, jdk1.7, 基于 Bluetooth 协议连接人体生理参数采集终端, 并支持 2G、3G、4G、WIFI 等连接云平台. Cloud-Health 实现心电、心率、血压、血氧饱和度和血糖在线监测, 并根据慢性疾病智能分类及预测的结果, 进行膳食推荐、运动推荐和就医服务推荐. Cloud-Health 支持联接厦门市“健康医疗云”, 该医疗云支持与各类健康管理终端对接, 已实现体重、体温、血压、血糖、心电、血氧、血脂等 14 项生理指标的收集, 已建立个人健康档案 330 余万份, 占 90% 常住人口, 最终满足居民慢性病变智能预警和突发状况救治的需要. 图 1 展示云健康监测系统的主界面及心电监护界面.

实验数据取自厦门市集美区 110198 条健康监测数据. 每条数据包括个人资料(年龄、性别、民族、婚姻、学历)、生活偏好(抽烟、喝酒、锻炼、睡眠时间等)、常规体检数据(身高、体重、腰围、臀围)和监测数据(心电、心率、血压、血氧饱和度和血糖), 样本平均年龄 64.9 岁. 糖尿病(Diabetes)数据共 40 个条

件属性、高血压(Hypertensive disease)数据共 35 个条件属性、冠心病(Coronary heart disease)数据共 35 个条件属性; 健康风险就是决策属性(H 高等风险、M 中等风险和 L 低风险).



图 1 云健康监测系统

以糖尿病数据为例, 运用基于粗糙集的慢性病变分级方法, 融合关联度和 Chi-merge 统计量对数据连续区间离散化, 得到初始属性 40 个、初始断点数 123; 基于相容矩阵的属性约简(正确分类率 $\alpha=0.9$), 得到剩余属性 9 个、断点数 16 个. 表 1 给出慢性病(糖尿病、高血压、冠心病)主要属性.

表 2 展示部分糖尿病数据分级预测结果. 采用批量与增量相结合的分类规则挖掘, 得到 12 条规则; 基于 MapReduce 框架运用规则库实现分级预测. 记年龄 (c_1), 身体质量指数 (c_2), 腰臀比 (c_3), 吸烟 (c_4), 甜食 (c_5), 血氧饱和度 (c_6), 心率 (c_7), 血糖 (c_8), 收缩压 (c_9), 健康风险类别 (d), 部分规则示例如下:

$$R_1 : f(x, c_1) \geq 57 \wedge f(x, c_2) \geq 26.4 \wedge f(x, c_4) \geq 0 \\ \wedge f(x, c_5) = 1 \wedge f(x, c_6) \geq 0.89 \wedge f(x, c_7) \geq (60, 90] \\ \wedge f(x, c_8) = (6.98, 9.16] \wedge f(x, c_9) \geq (115, 148] \rightarrow \\ f(x, d) = H$$

$$R_2 : f(x, c_1) \geq 57 \wedge f(x, c_2) \geq 26.4 \wedge f(x, c_3) \geq 0.83 \\ \wedge f(x, c_4) \geq 1 \wedge f(x, c_5) = 1 \wedge f(x, c_6) \geq 0.89 \\ \wedge f(x, c_7) \geq (60, 90] \wedge f(x, c_8) = (6.98, 9.16] \\ \wedge f(x, c_9) \geq (115, 148] \rightarrow f(x, d) = H$$

$$R_3 : f(x, c_1) \geq 54 \wedge f(x, c_2) \geq 23.6 \wedge f(x, c_4) = 0 \\ \wedge f(x, c_5) = 1 \wedge f(x, c_6) \geq 0.91 \wedge f(x, c_7) \geq (60, 94] \\ \wedge f(x, c_8) = (6.65, 7.12] \wedge f(x, c_9) \geq (127, 164] \rightarrow \\ f(x, d) = L$$

对 x_i 数据运用规则 R_1 : 57 岁, 男性, 身体质量指数 BMI 值为 26.9(中国标准: [24, 27.9], 偏胖), 有吸烟史、

喜欢甜食, 血氧饱和度 SpO₂ 值 92%(中国标准: <94%, 供养欠缺), 心率 (中国标准: [60, 100], 标准), 血糖 (FBG) 值范围 (6.98,7.48] (中国标准: 空腹全血血糖 FBG≥6.7 毫摩尔/升两次可诊断为糖尿病); 收缩压 (132, 136] (中国标准: ≥140 mmHg, 高血压), 经分类规则判定为高等风险 (H). 对 x_7 数据运用规则 R_3 : 54 岁, 男性, 身体质量指数 BMI 值为 24.1, 有吸烟史、喜欢甜

食, 血氧饱和度 SpO₂ 值 91%, 心率 (76,83], 血糖 (FBG) 值范围 (6.7,7.06]; 收缩压 (145,157], 经分类规则判定为低等风险 (L), 医生综合诊断后更倾向于高血压 (低等风险 L). 该方法识别准确率 89.23%、误识率 10.77%, 其根本原因在于生理监测数据具有时间连续性、非精确性、模糊性等特性.

表 1 慢性病(糖尿病、高血压、冠心病)主要属性

数据种类	主要属性
糖尿病	Age(年龄), BMI(身体质量指数), WHR(腰臀比), 吸烟, 血氧饱和度 (SpO ₂), 心率, 血糖, 收缩压/舒张压
高血压	Age(年龄), BMI(身体质量指数), 吸烟, WHR(腰臀比), 心率, 收缩压, 舒张压
冠心病	Age(年龄), BMI(身体质量指数), 血氧饱和度 (SpO ₂), 心率, 舒张压

表 2 部分糖尿病慢性病变分级结果

U	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	d
x_1	57	26.9		1	1	0.92	(75, 89]	(6.98, 7.48]	(132, 136]	H
x_2	59	27.0	0.83	1	1	0.94	(73, 84]	(7.05, 7.52]	(130, 135]	H
x_3	66	27.2		1	1	0.89	(83, 94]	(7.21, 8.54]	(129, 146]	H
x_4	68	26.4		0	1	0.89	(91, 95]	(7.52, 8.69]	(130, 138]	H
x_5	70	27.3	0.87	0	1	0.89	(60, 67]	(7.13, 8.84]	(140, 153]	H
x_6	72	29.1	0.91	0	1	0.89	(76, 94]	(8.31, 9.16]	(115, 148]	H
x_7	54	24.1		1	1	0.91	(76, 83]	(6.70, 7.06]	(145, 157]	L
x_8	62	23.6		1	1	0.92	(67, 90]	(6.65, 7.01]	(127, 145]	L
x_9	76	25.8	0.80	0	0	0.91	(73, 98]	(6.81, 7.10]	(143, 165]	L
x_{10}	65	24.9		0	1	0.93	(60, 71]	(6.74, 7.12]	(148, 164]	L

需要说明的是, 虽然对于常见的慢性疾病 (高血压、糖尿病和冠心病), 使用基于粗糙集的慢性病变分级方法虽可以得到较好的预测结果. 但限于数据等条件, 该方法目前尚未完全发挥云计算技术在大数据分析和挖掘中的优势. 如果在云端平台采用多种数据挖掘方法对生理监测数据进行协同分析, 可以提高慢性病变分级和诊断的精确度.

3 结论与展望

目前传感器网络、云计算和大数据技术正逐步应用到智慧医疗领域. 云健康监测利用附着人体的智能生理参数传感器获取人体生理数据, 并上传到健康医疗云端, 是智慧医疗领域未来发展方向之一. 为满足居民慢性病变智能预警和突发状况救治需要, 根据生理监测数据的时间连续性、非精确性和模糊性等特点, 本文提出一种基于粗糙集的慢性病变分级方法, 即首先采用融合相关度和 Chi-merge 统计量离散化生理监测数据; 然后, 基于相容矩阵的属性约简算法去除数据冗余属性; 最后, 基于批量与增量相结合挖掘分类规则,

基于分布计算框架 MapReduce 运用规则实现慢性病变智能分级. 该方法有利于帮助个体全面认识健康状况和病变风险, 有助于“健康医疗云”走入家庭慢性疾病监护领域, 有利于缓解目前医疗资源紧缺现状.

下一步工作包括: 提高云健康监测中慢性病变分级的识别率, 包括: 提高生理监测数据质量, 充分考虑降低数据的缺失值比率; 完善云健康监测系统的功能, 展开示范性应用和推广.

参考文献

- 1 Xia HN, Asif I, Zhao XP. Cloud-ECG for real time ECG monitoring and analysis. *Computer Methods and Programs in Biomedicine*, 2013, 110(3): 253–259. [doi: 10.1016/j.cmpb.2012.11.008]
- 2 Banos O, Villalonga C, Damas M, et al. PhysioDroid: Combining wearable health sensors and mobile devices for a ubiquitous, continuous, and personal monitoring. *The Scientific World Journal*, 2014, 2014: 490824.
- 3 Wang SL, Chen YL, Kuo AMH, et al. Design and evaluation of a cloud-based mobile health information recommendation

- system on wireless sensor networks. *Computers & Electrical Engineering*, 2016, 49: 221–235.
- 4 Su JH, Wang BW, Hsiao YC, *et al.* Personalized rough-set-based recommendation by integrating multiple contents and collaborative information. *Information Sciences*, 2010, 180(1): 113–131. [doi: [10.1016/j.ins.2009.08.005](https://doi.org/10.1016/j.ins.2009.08.005)]
 - 5 Firoozabadi R, Gregg RE, Babaeizadeh S. Intelligent use of advanced capabilities of diagnostic ECG algorithms in a monitoring environment. *Journal of Electrocardiology*, 2017, 50(5): 615–619. [doi: [10.1016/j.jelectrocard.2017.04.013](https://doi.org/10.1016/j.jelectrocard.2017.04.013)]
 - 6 Zhou P, Li ZC, Wang F, *et al.* Portable wireless ECG monitor with fabric electrodes. *Chinese Journal of Biomedical Engineering*, 2016, 25(4): 172–178.
 - 7 张翼, 宾光宇, 吴水才. 一种可穿戴式多参数心脏活动监测设备的设计. *中国医疗设备*, 2018, 33(3): 18–21. [doi: [10.3969/j.issn.1674-1633.2018.03.005](https://doi.org/10.3969/j.issn.1674-1633.2018.03.005)]
 - 8 Majumder S, Mondal T, Deen MJ. Wearable sensors for remote health monitoring. *Sensors*, 2017, 17(1): 130. [doi: [10.3390/s17010130](https://doi.org/10.3390/s17010130)]
 - 9 樊建平, 张元亭, 王磊. 实施海云工程 实现低成本健康. 集成技术, 2012, 1(2): 4–9.
 - 10 邹娜. 基于 Android 手机的泛在人体健康监护系统数据挖掘程序的研究与设计[硕士学位论文]. 武汉: 华中科技大学, 2012.
 - 11 胡建强. 一种联接“健康云”的家庭健康监护系统设计. *厦门大学学报(自然科学版)*, 2016, 55(1): 114–120.
 - 12 谢宏, 程浩忠, 牛东晓. 基于信息熵的粗糙集连续属性离散化算法. *计算机学报*, 2005, 28(9): 1570–1574. [doi: [10.3321/j.issn:0254-4164.2005.09.021](https://doi.org/10.3321/j.issn:0254-4164.2005.09.021)]
 - 13 陈昊, 杨俊安, 庄镇泉. 变精度粗糙集的属性核和最小属性约简算法. *计算机学报*, 2012, 35(5): 1011–1017.
 - 14 Margret AS, Clara MLJ, Jeevitha P, *et al.* Design of a diabetic diagnosis system using rough sets. *Cybernetics and Information Technologies*, 2013, 13(3): 124–139. [doi: [10.2478/cait-2013-0030](https://doi.org/10.2478/cait-2013-0030)]