



计算机科学

COMPUTER SCIENCE

基于标签共现和特征局部相关的心电异常检测方法

韩京宇, 钱龙, 葛康, 毛毅

引用本文

韩京宇, 钱龙, 葛康, 毛毅. 基于标签共现和特征局部相关的心电异常检测方法[J]. 计算机科学, 2023, 50(3): 139-146.

HAN Jingyu, QIAN Long, GE Kang, MAO Yi. [ECG Abnormality Detection Based on Label Co-occurrence and Feature Local Pertinence](#) [J]. Computer Science, 2023, 50(3): 139-146.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[一种基于超网络的多目标回归方法](#)

Multi-target Regression Method Based on Hypernetwork

计算机科学, 2022, 49(11A): 211000205-9. <https://doi.org/10.11896/jsjcx.211000205>

[监督和半监督学习下的多标签分类综述](#)

Survey of Multi-label Classification Based on Supervised and Semi-supervised Learning

计算机科学, 2022, 49(8): 12-25. <https://doi.org/10.11896/jsjcx.210700111>

[路网上基于时空锚点的移动对象群体和个体运动监测方法](#)

Detecting Group-and-individual Movements of Moving Objects Based on Spatial-Temporal Anchors of Road-network

计算机科学, 2020, 47(11): 294-303. <https://doi.org/10.11896/jsjcx.191100083>

[Web数据仓库研究综述](#)

计算机科学, 2004, 31(11): 19-23.

[网络受限移动对象不确定性轨迹的索引](#)

计算机科学, 2008, 35(3): 79-83.

基于标签共现和特征局部相关的心电异常检测方法

韩京宇 钱龙 葛康 毛毅

南京邮电大学计算机学院 南京 210023

江苏省大数据安全与智能处理重点实验室 南京 210023

摘要 自动的心电异常识别是一个多标签分类问题,多通过对每个标签训练一个二分类器来实现异常识别。由于异常数目多,特征和异常间以及不同异常间的相关性复杂,自动检测的效果并不理想。为了充分利用异常和特征间的依存关系,提出了一种基于异常标签共现和特征局部相关(Label Co-occurrence and Feature's local Pertinence,LCFP)的心电异常识别方法。首先,根据标签共现性和特征局部相关性,为标签构建包含宏特征和微特征的联合特征空间。宏特征采用狄利克雷过程混合模型聚类构建,以区分不同的共现标签集;微特征是原始特征空间的一个子集,用于区分共现标签集中的各个标签。进而,在联合特征空间为每个异常训练一个一对多(One-Versus-All)的概率分类器。其次,为充分利用异常的关联,提出在概率分类器排序基础上区分相关和非相关标签,采用Beta分布自适应地学习锚阈值和相关度阈值,以确定实例的相关标签集。LCFP是一种检测多种心电异常的通用方法,提高了心电异常识别的精度。在两个真实数据集上,F1指标分别提高了4%和22.4%,验证了所提方法的有效性。

关键词:心电异常;多标签分类;标签共现;狄利克雷过程混合模型;Beta分布;锚阈值

中图法分类号 TP311.132

ECG Abnormality Detection Based on Label Co-occurrence and Feature Local Pertinence

HAN Jingyu, QIAN Long, GE Kang and MAO Yi

School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Jiangsu Key Laboratory of Big Data Security & Intelligent Processing, Nanjing 210023, China

Abstract Automatic electrocardiogram(ECG) abnormality detection is a multi-label classification problem, which is commonly solved by training a binary-relevance classifier for each abnormality. Due to the large number of abnormalities, the complex correlations between features and abnormalities, and those among different abnormalities, existing methods' performance is not satisfying. To make full use of the dependencies between features and abnormalities, this paper proposes a novel abnormality detection method based on label co-occurrence and feature local pertinence(LCFP). Firstly, we set up a consolidated feature space consisting of both the macro-features and micro-features based on the label co-occurrence and features' pertinence. The macro-features are constructed with a clustering approach based on Dirichlet process mixture model(DPMM), thus distinguishing different co-occurrence label sets. The micro-features are a subset of primitive features, which serves to distinguish between the labels in the same labelset. Next, we train a one-versus-all classifier which returns a relevance probability. Secondly, to make use of the different correlation degrees among different abnormalities, we propose to differ the relevant labels from the irrelevant ones based on the sorting according to the probabilities given by the classifiers. In particular, we propose to exploit the Beta distribution to adaptively learn the anchor thresholds and correlation thresholds, thus determining the relevant labels of an instance. Our LCFP method is a universal way to detect every possible ECG abnormalities, which effectively improves the detection accuracy. The results on two real datasets show that our method can achieve an improvement of 4% and 22.4%, respectively, in terms of F1, which proves the effectiveness of our method.

Keywords Electrocardiogram abnormality, Multi-label classification, Label co-occurrence, Dirichlet process mixture model, Beta distribution, Anchor thresholds

到稿日期:2022-01-29 返修日期:2022-05-12

基金项目:国家自然科学基金(62002174)

This work was supported by the National Natural Science Foundation of China(62002174).

通信作者:韩京宇(jyhan@njupt.edu.cn)

1 引言

根据世界卫生组织的数据,心脏病是人们健康的头号杀手^[1]。心电信号记录了人们的心脏活动,作为一种非介入治疗手段,在临床上被广泛应用^[2]。人工识别异常不仅需要专业的医生,而且费时,因此自动的心电异常识别成为了大家关注的焦点^[3-6]。由于心电异常固有的关联性,一个人(实例)会同时呈现若干种异常(标签),因此自动的心电异常检测在机器学习中是一个典型的多标签分类问题^[7-8]。

自动心电异常检测技术可以分成两大类,即基于特征预抽取的方法和非基于特征预抽取的方法。前者首先显式地抽取一系列特征,然后基于特征训练分类器^[3-4];后者直接将心电信号作为输入,自动分析识别,不需要预先抽取特征^[9-10]。例如,深度学习中将心电图分解成一系列的图像片段,直接学习识别。最近,基于知识规则进行心电异常识别也受到了关注^[11],但这类方法需要大量的人工参与和专业知识,代价较高。

目前心电异常识别工作多关注于若干异常的识别^[3-4, 9, 12-13],而同时识别多种心电异常的效果则不够理想,主要挑战在于:1)心电信号作为生物电信号,具有相同异常的人们,表现的心电信号特征不尽相同,容易混淆;2)多标签分类方法直接应用到心电异常检测,效果并不理想,一方面,心电异常呈现不同的共现标签组,不同的标签组需要不同的特征组合区分,另一方面,每个异常又与特定的特征值相关联,因此,区分不同病症,不仅需要识别能区分共现标签组的特征组合,还需要根据组合内部的特征值差异来进行进一步的区分;3)由于异常标记成本高,可用的训练样本有限,如何自适应地根据样本寻找最合适的参数仍是亟待解决的问题。

我们注意到心电异常呈现出不同的共现性,有些异常频繁共同出现,而有些异常则很少共同出现。此外,特征对不同的异常具有不同的相关性^[2],如前间壁心肌梗死(Anteroseptal myocardial Infarction)经常和完全性左束支阻滞(Complete Left Bundle Branch Block)同时出现,并在ST段和QRS波群呈现共同特征,又如,前壁心肌梗死(Anterior Myocardial Infarction)经常和前间壁心肌梗死(Anteroseptal Myocardial Infarction)共同出现,并且QRS波群常呈现QR或rS波形^[2]。鉴于此,本文提出利用异常共现和特征局部相关(LCFP)的方法来提高心电异常检测精度。LCFP方法包括两个阶段,第一阶段通过构建区分能力更强的联合特征空间,为每个异常训练一对多的概率分类器。第二阶段对所有可能的异常根据概率排序,根据异常相关性,学习锚阈值和相关度阈值,识别每个实体的相关异常。

具体地,在第一阶段,结合宏特征和微特征为每个异常训练二分类器,宏特征主要用于区分一组共现异常和另外一组共现异常,微特征用于在一个共现异常组内区分不同的异常。每个二分类器返回每个异常属于实例的概率。在第二阶段,在所有异常标签排序的基础上,构建阈值最小生成树来确定一个实例的所有相关异常。本文提出采用Beta分布来学习两类阈值,即锚阈值和相关度阈值。锚阈值用于确定以高置信度属于实例的异常;相关阈值用于基于已经确认的相关

异常来识别其他相关异常。

本文的贡献包括:

(1)提出了根据心电异常的共现性和特征局部相关性来构建宏特征和微特征的联合空间,两类特征相互补充。

(2)针对宏特征构建,提出了基于狄利克雷过程的混合模型(Dirichlet Process Mixture Model, DPMM)对心电样本进行聚类,从而识别宏特征空间中最合适的支撑轴。

(3)在学习分割相关和非相关异常阈值的过程中,提出采用Beta分布来刻画阈值分布,从而自适应地根据可用样本拟合阈值。下文中异常和标签、人和实例指代相同。

2 相关研究

2.1 心电图异常检测

心电图异常检测的方法大致分为两类:基于显式特征抽取的方法和基于深度学习的方法。基于显式特征抽取的方法在训练分类模型前需要先进行特征抽取,然后训练模型。文献[3]提出采用两层神经网络来检测4种类型的心电节拍。第一层网络使用统计特征作为输入,实现心电节拍分类;第二层网络使用第一层网络的输出作为输入进行训练。文献[12]采用3类神经网络分类器,即反向传播网络、前馈网络和多层感知器,将心电信号分为正常类或异常类。文献[4]通过提取形态特征来检测心律失常。文献[14]比较了采用支持向量机(Support Vector Machine, SVM)、Adaboost(Ada)、神经网络(Artificial Neural Network, ANN)和朴素贝叶斯分类器来区分正常和异常心拍的性能。文献[15]提出采用知识迁移,用加权支持向量机对心电信号进行分类,避免标记大量的训练样本。文献[11]提出基于规则,使用抽象特征识别节拍异常。这类基于特征预抽取的方法通常针对特定异常进行检测,并且其性能易受限于前期的数据预处理。

基于深度学习的方法把心电图像作为输入,训练深度模型,无须显式地提取特征。文献[9]提出使用深度神经网络识别室性早搏。文献[10]将每个心电节拍转换为二维灰度图像作为输入,使用深度二维卷积神经网络检测心律失常。文献[16]使用卷积神经网络和长短期记忆网络构建深度模型,根据正常心电信号预测后续的正常心电信号。这类深度学习方法需要大量的训练数据,在样本有限的情况下,性能难以保证。

2.2 多标签分类

多标签分类的挑战在于分类器的输出空间大小与类标签的数量呈指数关系。通常通过问题转换和算法适应^[7-8, 17]来解决。问题转换方法可以分为3类,即二分类、标签排序和多分类。代表性的二分类包括二元相关^[18]和分类器链^[19]两类方法。前者将多标签分类问题分解为一组独立的二分类问题;后者将多标签分类转化为二分类链,链中的后续二分类器建立在前面的分类器预测的基础上。标签排序的常见方法是校准标签排序(Calibrated Label Ranking),它将多标签分类问题转化为标签排序问题来解决^[20]。多类别分类算法,如随机K标签集(Random K-Labelsets, RKL)等^[21],将多标签分类问题转化为若干个多类分类问题来解决。

算法适应类方法通过改造其他算法来解决多标签分类

问题^[17,22-24]。文献[22]提出的多标签 K 近邻(Multi-Label K -nearest Neighbour, MLK)方法,采用 K 近邻技术来处理多标签数据,根据最大后验规则预测标签。多标签决策树采用决策树技术来处理多标签数据,根据多标签熵的信息增益递归地构建决策树^[23]。文献[24]提出的排序支持向量机(Ranking SVM)采用最大边距策略进行多标签分类,通过最小化经验排序损失,训练一组线性分类器来预测标签。

2.3 标签相关性和特征选取

有些工作利用标签相关性来提高多标签分类的效果。根据标签关联程度,分成一阶、二阶和高阶算法^[18-21]。一阶策略不考虑标签间的相关性,把多标签分解成多个独立的二分类问题^[18,22]来解决,有时误差会较大。二阶方法考虑了标签对之间的关联性,如校准标签排序通过训练两两标签的分类器,来获得相关标签和不相关标签的排序^[20],集体多标签分类器(Collective Multi-Label Classifier)^[25]利用条件随机场来捕捉标签对的相关性。高阶方法考虑多个标签之间的关联,如分类器链和随机 K 标签集等^[19,21]。相比一阶方法,二阶和高阶方法能获得更好的泛化性能,但计算复杂度更高。

一些研究通过寻找标签合适的特征集来提高分类效果,主要分成过滤式、包裹式和空间嵌入 3 类方法。过滤式利用训练集自身的特点筛选出特征子集后再将其送入分类器进行学习,特征选择的过程与后续学习器无关^[26];包裹式直接把最终要使用的学习器的性能作为特征子集的评价标准^[27];空间嵌入法通过变换特征空间来发现标签的最优特征,以提高分类效果,如 LIFT 通过聚类来构建每个标签的最优特征^[28]。本文与上述工作的主要区别在于:宏特征是多个标签共享的特征,用于区分不同的标签集;而微观特征是对共现标签进行进一步区分的有效特征;宏特征和微特征互相补充。

3 问题描述和一对多分类器训练

心电信号常被基线漂移、肌电和电工噪声等干扰^[29-30]。本文工作的原型系统采用小波变换过滤电工和肌电噪声^[30],然后采用 6 阶 Butterworth 带通滤波器消除基线漂移^[31]。进而,基于信号幅度、形态、横向时长和波的方向识别 3 种最重要的波形,即 P 波、QRS 波群和 T 波^[32-33]。最后,基于 3 种波形识别 6 类原始特征,如表 1 所列,构造原始特征空间 $S^d = \{f_1, \dots, f_d\}$,其中 d 代表特征数目。

表 1 6 类原始特征

Table 1 Six categories of primitive features

Feature category	Example
Horizontal interval	Duration of PR segment is 0.11 s
Vertical amplitude	Amplitude of R peak is 0.25 mV
Wave direction	P waveform is upward
Deviation of wave axis	QRS axis is between[90,150]
Waveform Morphology	QRS complex takes rsR' form
Consecutive wave pattern	QRS complex comes earlier 0.05 seconds

给定实例 $st \in S^d$ 和所有标签集合 $B = \{r_1, \dots, r_i, \dots, r_u\}$, u 代表所有标签的数目。寻找一个多标签分类模型 $h: S^d \rightarrow 2^B$,为每个实例分配一个相关标签子集。表 2 列出了实验数据集中的部分异常标签。

表 2 部分异常标签示例

Table 2 Example of some abnormality labels

Complete left bundle branch block	Anteroseptal myocardial infarction
Anterior myocardial infarction	Inferior myocardial\ \ infarction
Atrial fibrillation	Sinus tachycardia
Left anterior fascicular block	Right atrial hypertrophy

首先,对于每个标签 r ,构建包含宏特征和微特征的联合特征 $MF(r) = M^{\text{mac}} \cup M^{\text{mic}}(r)$ 。宏特征 $M^{\text{mac}} = [f_1^{\text{mac}}, \dots, f_v^{\text{mac}}]$ 用于区分不同的共现标签集;微特征用于区分共现标签集中的不同标签,是原始特征的子集合,记为 $M^{\text{mic}}(r) = [f_1^{\text{mic}}, \dots, f_w^{\text{mic}}]$ 。然后,在宏特征和微特征组成的联合特征集上为每个标签训练一对多的二分类器。给定实例 st 和标签 r ,二分类器返回一个介于 0 和 1 之间的值,指示 r 属于 st 的概率。下面分别论述如何构建宏特征和微特征。

3.1 基于狄利克雷过程混合模型聚类构建宏特征

本文分两步构建宏特征:第一步,识别包含所有标签且重叠最小的覆盖集;第二步,通过狄利克雷过程混合模型聚类,在每个标签集的对应实例上构建支撑轴,进而联合所有支撑轴形成宏特征空间。

3.1.1 基于频繁标签集挖掘构建标签覆盖集

首先,通过频繁模式增长(Frequent-pattern Growth)算法^[34]识别所有频繁 k 项标签集。

定义 1(频繁 k 项标签集) 给定含 $k(k > 0)$ 个标签的标签集 bs 和支持度阈值 η ,如果 bs 出现在数据集 D 上的频率超过 η ,则 bs 是一个频繁 k 项标签集。

挖掘出所有的频繁 k 项标签集后,剔除是其他频繁标签集真子集的 1 项标签集,形成一个包括所有标签的覆盖集 $\beta = \{bs_1, \dots, bs_n\}$,满足:

$$\bigcup_{i=1}^{|\beta T|} bs_i = B \quad (1)$$

定义 2(重叠率) 给定一个覆盖集 β ,其重叠率是 $lr(\beta) =$

$$\sum_{i=1}^{|\beta|} \sum_{i < j} |bs_i \cap bs_j|, \text{ 这里 } bs_i (bs_j) \text{ 是 } \beta \text{ 中的一个频繁 } k \text{ 项标签集。}$$

重叠率越小,覆盖集越好。最优的覆盖集 βT 是 β 的一个子集,须满足条件(2):

$$\beta T = \text{argmin } lr(\beta) \quad (2)$$

这是组合优化问题,采用遗传算法求解,即获得一个最优的覆盖集 βT 。

3.1.2 狄利克雷过程混合模型聚类构建支撑轴

宏特征空间由 βT 中所有标签集对应的支撑轴表征。为构建各个标签集的支撑轴,将标签集对应的实例划分成若干类簇,每个类簇用一个轴表征。为合理划分类簇,提出通过狄利克雷过程混合模型聚类实现,不需要指定类簇个数,不受数据维数的限制^[35]。

不失一般性,设实例的原始特征符合多元正态分布,通过中国餐馆过程(Chinese Restaurant Process, CRP)来描述狄利克雷过程,每个实例 $st_i (1 \leq i \leq N)$ 由以下过程产生:

$$(u, \Sigma)_{1, \dots, \infty} \sim NIW(\mu_0, k_0, \Psi_0, v_0) \quad (3)$$

$$Z_{1, \dots, i, \dots, N} \sim CRP(\gamma) \quad (4)$$

$$st_{1, \dots, i, \dots, N} \sim MN((u, \Sigma)_{z_i}) \quad (5)$$

该生成模型中,实例由多元正态分布 MN 产生,类簇

分配由中国餐馆过程 $CRP(\gamma)$ 决定, 其中 γ 是聚焦参数。 Z_i 是实例 st_i 对应的类簇; 作为狄利克雷过程的基分布, 逆威沙特分布 (Normal-Inverse-Wishart, NIW) 是多元正态分布 MN 的共轭先验分布。逆威沙特分布中, \mathbf{u}_0 是 N 维向量, 代表最初的平均值; k_0 用作平滑因子, 控制 Ψ_0 中各个元素的放缩比例; v_0 是自由度, 初始化为原始特征数目; Ψ_0 是成对偏差积, 初始化为 $N \times N$ 的常数矩阵。

为确定一个实例所属的类簇, 采用吉布斯取样实现, 如算法 1 所示。

算法 1 Gibbs Sampling For Assignments

输入: 实例集 $D = \{st_1, \dots, st_N\}$, 狄利克雷过程混合模型参数

输出: 每个实例所属的类簇

1. Initialize $\{Z_i^{(0)} \mid 1 \leq i \leq N\}$;
2. $t \leftarrow 1$;
3. While t do
4. For $i \leftarrow 1$ to N do
5. $Z_i^{(t)} \sim P(Z_i \mid st_1, Z_1^{(t)}, \dots, Z_{i-1}^{(t)}, \dots, Z_{i+1}^{(t-1)}, \dots, Z_N^{(t-1)})$;
6. End
7. If $(Z_1^{(t)}, \dots, Z_N^{(t)})$ is the same as $(Z_1^{(t-1)}, \dots, Z_N^{(t-1)})$ then
8. $t \leftarrow 0$
9. End
10. End
11. Return (Z_1, \dots, Z_N)

该算法不断迭代, 直至每个样本所属的类簇不再变化。在迭代过程中, 每个实例所属类簇的概率按照式(6)修改。

$$P(Z_i = m \mid st_i, Z_{-i}) \propto P(Z_i = m \mid Z_{-i}) P(st_i \mid Z_i = m, Z_{-i}) \quad (6)$$

这里, Z_{-i} 是除 st_i 之外的实例的类簇分配。式(6)的证明过程如下。

证明:

$$P(Z_i = m \mid st_i, Z_{-i}) = P(Z_i = m \mid st_{1:i}, Z_{-i}) = \frac{P(Z_i = m, Z_{-i}, st_{1:i})}{P(Z_{-i}, st_{1:i})} \quad (7)$$

分子和分母按照概率乘法分解, 由 $p(st_{-i} \mid Z_i = m, Z_{-i}) = p(st_{-i} \mid Z_{-i})$, 且 $p(st_i \mid st_{-i}, Z_i = m, Z_{-i}) = p(st_i \mid Z_i = m, Z_{-i})$, 推得式(7)等于:

$$\frac{P(Z_i = m \mid Z_{-i}) P(st_i \mid Z_i = m, Z_{-i})}{P(st_i \mid st_{-i}, Z_{-i})} \quad (8)$$

联合式(6)和式(8), 可得:

$$P(Z_i = m \mid st_i, Z_{-i}) \propto P(Z_i = m \mid Z_{-i}) P(st_i \mid Z_i = m, Z_{-i})$$

证毕。

式(6)的第一项是在 st_i 外所有实例的类簇分配概率给定条件下, st_i 属于类簇 m 的概率。根据中国餐馆过程, 计算为:

$$P(Z_i = m \mid Z_{-i}) = \begin{cases} \frac{n_{m,-i}}{i + \gamma - 1}, & m \text{ is an existing cluster id} \\ \frac{\gamma}{i + \gamma - 1}, & m \text{ is a new cluster id} \end{cases}$$

其中, $n_{m,-i}$ 是类簇 m 中除 st_i 外的实例数目。式(6)中的第二项是所有类簇分配确定条件下, st_i 的出现概率, 即:

$$P(st_i \mid Z_i = m, Z_{-i}) \sim MN(u_{m,-i}, \Sigma_{m,-i}) \quad (9)$$

其中, $u_{m,-i}$ 和 $\Sigma_{m,-i}$ 是类簇 m 不包含 st_i 情况下的均值和方差。

在迭代过程中, 类簇 m 的参数修改为:

$$\mu_m = \frac{k_0 \mu_0 + n_m \bar{x}}{k_0 + n_m} \quad (10)$$

$$k_m = k_0 + n_m \quad (11)$$

$$\Psi_m = \Psi_0 + \sum_{i=1}^{n_m} (x_i - \bar{x})(x_i - \bar{x})^T + \frac{k_0 n_m}{k_0 + n_m} (\bar{x} - \mu_0)(\bar{x} - \mu_0)^T \quad (12)$$

$$v_m = v_0 + v_m \quad (13)$$

其中, n_m 是类簇 m 中的实例数, \bar{x} 是该类簇中所有实例的平均值。

假定标签集 bs_i 对应的实例被划分成 $n_{tw}(i)$ 个互不相交的类簇, 所有类簇的中心 $ct(bs_i) = \{c^1, \dots, c^j, \dots, c^{n_{tw}(i)}\}$, 每个中心是宏特征空间的一个支撑轴。所有的支撑轴确定后, 给定一个实例, 计算它和各个支撑轴的距离, 从而将实例映射到宏特征空间, 实现有效区分。

例 1 假设原始特征空间含 6 个维度, 覆盖集中含两个标签集 $r_1 r_2$ 和 $r_3 r_4$, $r_1 r_2$ 对应的实例被划分为两个类簇, 两个类簇中心即支撑轴 c^1 和 c^2 是两个长度为 6 的向量。 $r_3 r_4$ 对应的实例被划分为两个类簇, 两个类簇中心即支撑轴分别是 c^3 和 c^4 , 也是两个长度为 6 的向量。给定一个实例 st_1 , 设它与 4 个支撑轴的欧几里得距离分别是 0.63, 0.7, 0.65, 0.78, 则 st_1 在宏特征空间中的坐标是 (0.63, 0.7, 0.65, 0.78)。

3.2 抽取单个标签的微特征

给定一个标签, 微特征用于区分它和其他共现标签。为了识别标签 r 的微特征, 首先识别它的所有共现标签, 记为 $OS(r)$; 然后, 求得 r 的所有共现标签的正例和反例的并集 $ST(r)$, 即 $ST(r) = (\bigcup_{x \in OS(r)} ST_x^p) \cup (\bigcup_{x \in OS(r)} ST_x^N)$, 其中 ST_x^p 是同时包含标签 r 和 x 的实例集合, ST_x^N 是含标签 r 但无标签 x 的实例集合。最后, 在 $ST(r)$ 上通过特征选取, 选取原始特征的一个子集, 作为 r 的微特征 $M^{mic}(r)$ ^[34]。

4 基于阈值最小生成树识别相关异常

为了基于所有分类器的返回概率确定一个实例的相关标签, 提出训练一个阈值最小生成树来区分相关标签和非相关标签。下面分述如何构建生成树, 如何学习阈值和预测相关标签。

4.1 基于相关度构建最小生成树

定义 3(相关度) 给定数据集 D 和两个标签 r_1 和 r_2 , 其相关度 $cd(r_1, r_2) = \frac{cnt(r_1 \cap r_2)}{cnt(r_1 \cup r_2)}$, 其中 $cnt(r_1 \cap r_2)$ 是 D 中同时包含 r_1 和 r_2 的实例数目, $cnt(r_1 \cup r_2)$ 是 D 中包含 r_1 或 r_2 的实例数目。

根据克鲁思卡尔(Kruskal)算法, 按照如下步骤创建一棵最小生成树, 树的每个顶点代表一个标签, 树的边代表两个标签的相关度。

(1) 计算所有标签对的相关度, 并对所有的标签对进行降序排序。

(2) 每次抛出排序列表头部的标签对, 根据标签对在最小生成树中加一条边和相应顶点, 检查树中是否形成环。如果

形成环,则抛弃新加入的边和顶点。

(3)重复步骤(2),直到树中仅有 $u-1$ 条边。

直观地,最小生成树代表了连接所有标签的最频繁路径。为了基于最小生成树识别一个实例的相关标签,要学习两种阈值,即锚阈值和相关度阈值。每个标签 r 对应一个锚阈值 $ah(r)$;同时,标签 r 和每个近邻标签 r^n 有一个相关度阈值 chr, r^n 。给定实例 st 和标签 r ,如果分类器的返回概率 $pr(st, r) \geq ah(r)$,则 r 是 st 的锚标签,也是相关标签。尽管 $pr(st, r)$ 小于锚阈值,但若 r^n 是 st 的相关标签,并且 $pr(st, r^n) \geq cd(r, r^n) \geq ch(r, r^n)$,则 r 也是 st 的相关标签。

定义 4(阈值最小生成树) 一个阈值最小生成树包含 u 个节点和 $u-1$ 条边,记为 $TT=(V, E)$ 。其中,每个 $v \in V$ 是一个三元组 $(r, ah(r), ch)$, r 是一个标签, $ah(r)$ 是 r 的锚阈值, ch 是相关度阈值集合,每个元素是 r 和一个近邻标签 r^n 的相关度阈值,记为 $ch(r, r^n)$ 。

例 2 图 1 给出了一个阈值最小生成树,它包含 6 个顶点、5 条边,每个顶点 v_i 对应标签 r_i ,每条边标注了相关度。

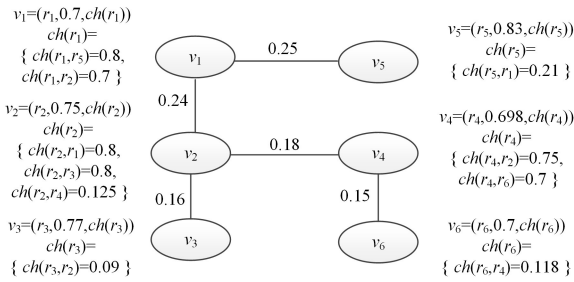


图 1 阈值最小生成树示例

Fig. 1 Illustration of threshold-based minimum spanning tree

4.2 学习锚阈值和相关度阈值

锚阈值用于区分锚标签和非锚标签,相关度阈值用于关联新的相关标签。由于这两个阈值都是介于 0 和 1 之间,设其服从 Beta 分布,其参数可采用闭合形式推断获得。以标签 r 的锚阈值为例,其概率密度是:

$$f(ah(r); \alpha, \beta) = \frac{1}{\beta \alpha^\alpha} \int ah(r)^{\alpha-1} (1-ah(r))^{\beta-1} \quad (14)$$

其中, α 和 β 两个参数影响密度函数的形状。根据 Beta 分布的性质, $ah(r)$ 可以用分布的期望来表征,如式(15)所示:

$$ah(r) = \frac{\alpha}{\alpha + \beta} \quad (15)$$

对于一个实例,对 B 中的所有标签按照分类器返回的概率降序排序,锚标签集的定义如下。

定义 5(锚标签集) 给定一个实例 st 的相关标签集 $RS(st)$ 和 B 中所有标签的降序排列 $drs(st)$,若 r 是 $drs(st)$ 第一个不在 $RS(st)$ 中的标签,则 st 的锚标签集是排在 r 之前的标签构成的集合,记为 $AS(r)$ 。

然后,在含有 r 作为锚标签的所有实例上学习 $ah(r)$ 的 Beta 分布。对于每个实例,将 $pr(st, r)$ 作为分割点,对于第 i 个样本, Beta 分布更新如下:

$$f(ah(r); \alpha^{i-1} + 100 \cdot pr(st, r), \beta^{i-1} + 100 \cdot (1 -$$

$$pr(st, r)) \quad (16)$$

其中, α^{i-1} 和 β^{i-1} 是上次迭代中的两个参数。最后:

$$ah(r) = \frac{\alpha^{(N)}}{\alpha^{(N)} + \beta^{(N)}} \quad (17)$$

例 3 说明如何学习锚阈值。假设 $RS(st) = \{r_1, r_4, r_5\}$, $pr(st, r_1) = 0.8$, $pr(st, r_2) = 0.4$, $pr(st, r_3) = 0.5$, $pr(st, r_4) = 0.7$, $pr(st, r_5) = 0.6$, $pr(st, r_6) = 0.62$, 因此 $drs(st) = \langle r_1, r_4, r_6, r_5, r_3, r_2 \rangle$ 。由于锚标签集 $AS(r) = \{r_1, r_4\}$, 因此 r_4 对应分割点。假设 r_4 的初始 Beta 分布为 $\alpha^{(0)} = 1$ 和 $\beta^{(0)} = 1$, 根据式(16) r_4 对应的 Beta 分布更新为 $\alpha^{(1)} = 1 + 100 \cdot 0.7 = 71$ 和 $\beta^{(1)} = 1 + 100 \cdot 0.3 = 31$ 。

进一步,在最小生成树上学习相关度阈值。 $nb(r)$ 代表 r 在最小生成树上的所有邻居标签。针对每个标签对 (r, r^n) 的相关度阈值,在同时拥有这两个标签并且满足 $pr(st, r) < ah(r)$ 的实例上学习获得,如算法 2 所示。

算法 2 learnCorrelationThreshold

输入: 标签集 $B = \{r_1, \dots, r_u\}$, 带标签接受概率的训练集 D , 锚阈值 $\{ah(r_1), \dots, ah(r_u)\}$, 最小生成树 MST

输出: 最小生成树上标签对的相关度阈值

1. $NLP \leftarrow$ extract all neighbour label pairs on MST and initialize the Beta distributions;
2. For each $st \in D$ do
3. For each label pair $(r, r^n) \in NLP$ do
4. If $(r^n \in RS(st) \wedge r \in RS(st) \wedge pr(st, r) < ah(r))$ then
5. $\zeta \leftarrow pr(st, r^n) \cdot cd(r, r^n)$;
6. $\alpha(r, r^n) \leftarrow \alpha(r, r^n) + \zeta \cdot 100$;
7. $\beta(r, r^n) \leftarrow \beta(r, r^n) + 100 - \zeta \cdot 100$;
8. End
9. End
10. End
11. Return the means of all Beta distributions

该算法中,若实例上的标签 (r, r^n) 满足学习条件,则通过计算 $pr(st, r^n) \cdot cd(r, r^n)$ 的乘积修改对应的 Beta 分布。最后,取 Beta 分布的平均值作为标签对的相关度阈值。该算法的时间复杂度为 $O(|D| \cdot u \cdot sn)$, sn 是最小生成树上标签的最大邻居数。

4.3 预测相关标签

给定一个实例,若一个标签与该实例已经识别的相关标签密切相关,则该标签也可能是该实例的相关标签。根据这个思想,采用算法 3 确定一个实例的相关标签。

算法 3 predRelevantLabels

输入: 实例 st , 标签集 $B = \{r_1, \dots, r_u\}$, 所有标签的概率分类器, 阈值最小生成树 TT

输出: st 的相关标签

1. $rel \leftarrow \phi$;
2. For each label $r \in B$ do
3. //判定初始相关标签集
4. If $pr(st, r) \geq ah(r)$ then
5. $rel \leftarrow rel \cup r$;
6. visited[r] \leftarrow true;
7. End

```

7. End
8. If rel= $\phi$  then
9.   Put the first ranked label into rel;
10. End
    //宽度优先搜索相关标签的近邻标签
11. LL $\leftarrow$ rel, LLnw $\leftarrow$  $\phi$ ;
12. For each r  $\in$  LL do
13.   NB $\leftarrow$ find r's unvisited neighbors on TT;
14.   For each nl $\in$  NB do
15.     If pr(st, r) * cd(r, nl)  $\geq$  ch(r, rn) then
16.       LLnw $\leftarrow$ LLnw  $\cup$  {nl};
17.       visit[nl]  $\leftarrow$  true;
18.     End
19.   End
20. rel $\leftarrow$ rel  $\cup$  LLnw;
21. LL $\leftarrow$ (LL \ {r})  $\cup$  LLnw;
22. End
23. Return rel;

```

最初,算法检测每个标签对应的分类器返回的接受概率,如果该概率大于锚阈值,则该标签是一个相关标签。然后,根据宽度优先搜索来检查相关标签的近邻标签。若标签 r 的一个近邻标签 r_n 是 st 的相关标签,并且满足 $pr(st, r^n) \cdot d(r, r^n) \geq ch(r, r^n)$, 则 r 就是一个相关标签。依此类推,直到所有的标签都被检查,或者剩余的未检查标签没有邻居是相关标签,则算法停止。算法的时间复杂度主要由确定初始标签集和通过相关度识别新的相关标签的时间决定。设实例的最大相关标签数为 R , 剪枝树上一个标签的最大邻居数为 H , 则算法 3 的时间复杂度为 $O(u + R \cdot H)$ 。

5 实验分析

实验在 Windows 10 平台运行, CPU 3.4 GB, 内存 16 GB。实验在两个实际数据集上进行: 第一个数据集记为 dcom, 包含 6000 条社区体检心电图数据^[36], 取样频率是 500 Hz。每条数据包含 12 个导联的 10 s 的信号, 共有 20 个标签。第二个数据集是 MIT-BIH 的公共数据集^[37], 记为 dmit, 取样频率是 360 Hz, 因数据稀疏性, 提取了 8 个异常标签。原型系统用 Python 实现。除特殊说明, 只汇报 dcom 上的结果。

实验中采用 3 个指标, 即正确率 (precision, PN)、召回率 (recall, RL) 和 F1 值。

$$PN(r) = \frac{TP(r)}{TP(r) + FP(r)} \quad (18)$$

$$RL(r) = \frac{TP(r)}{TP(r) + FN(r)} \quad (19)$$

$$F1(r) = \frac{2 * PN(r) * RL(r)}{PN(r) + RL(r)} \quad (20)$$

涉及的 TP, FP, FN 的含义如表 3 所列。为保证结果的客观性, 采用 6 折交叉验证: 整个数据集分成 6 份, 每类标签的样本在 6 份数据集中平分。依次进行 6 轮, 每轮取 4 份作为训练集训练多个概率分类器, 1 份作为验证集训练阈值最小生成树, 1 份作为测试集。最后汇报 6 轮的平均值。概率分类器用随机森林实现。

表 3 TP, FP, FN 的含义

Table 3 Meaning of TP, FP, and FN

notation	meaning
TP	# labels correctly detected as the ground-truth ones
FP	# labels incorrectly detected as the ground-truth ones
FN	# of labels incorrectly detected not as the true ones

5.1 DPMM 聚类效果分析

为了衡量聚类效果, 采用轮廓系数 (Silhouette Coefficient, SC) 来度量^[34]。

(1) k_0 对聚类效果的影响

由图 2 可见, 随着 k_0 的增大, 类簇数目会有微小的波动, 但幅度很小。图 3 给出了把 Ψ_0 的元素值设置为 20 时, 随着 k_0 从 1 上升到 50, 平均轮廓系数基本没有大的波动。综合图 2、图 3 可见, 狄利克雷过程混合模型聚类对 k_0 不敏感。

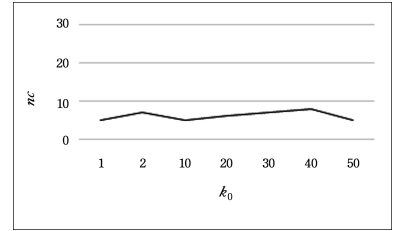


图 2 k_0 对平均类簇数目的影响 (Ψ_0 元素值为 20)

Fig. 2 Impact of k_0 on average number of class clusters (element value in Ψ_0 is 20)

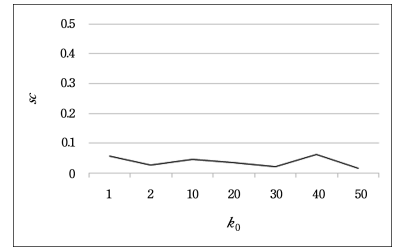


图 3 k_0 对聚类轮廓系数的影响 (Ψ_0 元素值为 20)

Fig. 3 Impact of k_0 on clustering silhouette coefficients (element value in Ψ_0 is 20)

设定 $\Psi_0 = 20$, k_0 从 1 增加到 50, 图 2 给出了类簇个数 nc 随 k_0 的变化。

(2) Ψ_0 的元素值对聚类结果的影响

设定 k_0 为 2, 图 4 给出了类簇数目随 Ψ_0 元素值的变化, 随着 Ψ_0 元素值从 1 上升到 150, 类簇数目快速下降, 后趋于平稳。图 5 给出了平均轮廓系数随着 Ψ_0 元素值的变化情况, 随着 Ψ_0 元素值从 1 上升到 150, 平均轮廓系数上升, 然后趋于平稳。在下文的实验中选取 $\Psi_0 = 90$ 。

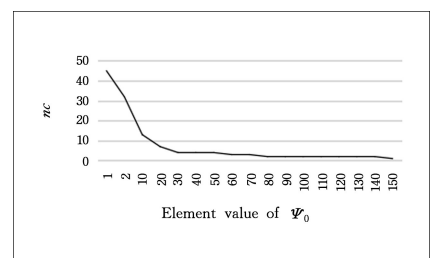


图 4 Ψ_0 元素值对类簇数目的影响 ($k_0 = 2$)

Fig. 4 Impact of Ψ_0 on average number of clusters ($k_0 = 2$)

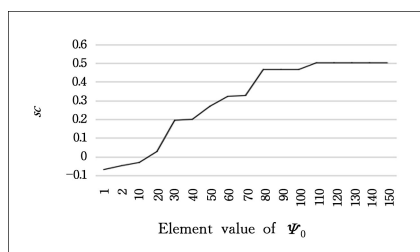


图5 Ψ_0 元素值对聚类轮廓系数的影响($k_0=2$)

Fig. 5 Impact of Ψ_0 on clustering silhouette coefficient($k_0=2$)

5.2 联合特征空间效果分析

为评估联合特征空间的效果,在两个数据集上比较了两种策略的效果,即直接在原始特征空间进行识别和在联合特征空间(宏特征和微特征)进行识别的效果。图6列出了在dcom数据集上3个指标的结果比较,图7列出了在dmit数据集上3个指标的结果比较。从图6、图7可知,通过结合微特征和宏特征,相比原始特征空间,在dcom数据集上F1提升了7.23%,在dmit上F1提升了5.72%。这主要是因为,宏特征和微特征可以互相补充,提高了异常标签的区分能力。

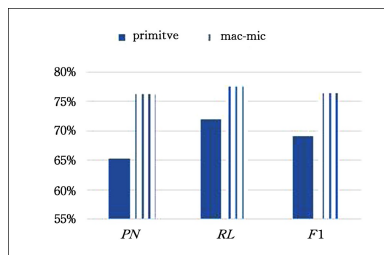


图6 dcom上联合特征空间和原始特征空间的效果比较

Fig. 6 Effect comparison of consolidated space and primitive space on dcom dataset

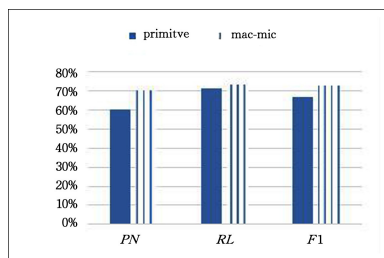


图7 dmit上联合特征空间和原始特征空间的效果比较

Fig. 7 Effect comparison of consolidated space and primitive space on dmit dataset

5.3 锚阈值和相关度阈值效果分析

为了评估锚阈值和相关度阈值的效果,表4列出了不设定相关标签分割阈值、单独采用锚阈值和同时设定锚阈值和相关度阈值3种策略的准确率、召回率和F1值。对于不设定相关标签分割阈值的方法,在标签排序后,取前 nk 个标签, nk 从1变化到6。从表4可见,双阈值策略的F1指标比锚阈值策略高4.82%,高出最好的无阈值策略($nk=1$)5.57%,证实了基于最小生成树的双阈值方法可以有效地区分相关标签和非相关标签。这是因为双阈值方法充分考虑了标签间的关联性,并且通过Beta分布,根据可得样本准确地学习区分阈值。

表4 各种不同策略的识别效果

Table 4 Detection effects of different strategies

(单位:%)

		PN	RL	F1
without threshold	$nk=1$	82.85	63.56	70.80
	$nk=2$	54.46	84.24	64.65
	$nk=3$	40.92	92.43	54.38
	$nk=4$	32.86	95.68	46.30
	$nk=5$	27.52	97.10	40.29
	$nk=6$	23.86	98.16	35.93
anchor threshold only		78.16	65.67	71.55
both thresholds		76.28	77.52	76.37

5.4 相关工作比较

本文将LCFB方法和文献[14]中采用的表现最好的支持向量机(SVM)、Adaboost(Ada)和人工神经网络(ANN)进行了比较,同时与多标签K近邻(MLK)^[22]、随机k标签集(RKL)^[21]两种多标签分类方法进行了比较,它们分别代表高阶和一阶的多标签分类方法。文献[14]用SVM,Ada,ANN来区分异常和非异常标签。本文由于每个样本对应多个病症,是多标签分类问题,因此为每个标签训练一对多的分类器。为进行合理比较,5种相关方法训练分类器时没有构建宏特征和微特征,也没有构建最小生成树。篇幅所限,不再叙述。

结束语 在智慧医疗中,采用机器学习实现自动的心电异常检测具有广泛的临床应用价值,不仅可以降低人力成本,且可提供实时的异常预警。本文根据心电异常数据的特点,提出利用异常共现性和特征的局部相关性进行异常检测,设计了特征构建方法和基于阈值学习区分相关标签和非相关标签的具体方法,在实际数据集上分析了影响因素并验证了方法的有效性。在未来的工作中,将进一步探究根据病症的演化特征来提高异常检测效果。

参考文献

- [1] WORLD H O. Cardio-vascular diseases (CVDs) [EB/OL]. (2021-06-11) [2021-06-11]. <https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-cvds>.
- [2] YANG H. Course book of Electrocardiogram Specialty[M]. Beijing: Beijing University Medical Press, 2005: 18-34.
- [3] INAN G, ELIF D U. ECG beat classifier designed by combined neural network model[J]. Pattern Recognition, 2005, 38(2): 199-2008.
- [4] ZHU W L, CHEN X H, WANG Y, et al. Arrhythmia Recognition and Classification Using ECG Morphology and Segment Feature Analysis[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2019, 16(1): 131-138.
- [5] YANG C D, JIA Z, LI X W. Study on ECG Signal Recognition and Classification Based on U-Net++[J]. Computer Science, 2021, 48(10): 121-126.
- [6] JIN L P, DONG J. Deep learning research on clinical electrocardiogram analysis [J]. SCIENTIA SINICA Informationis, 2015, 45(3): 398-416.
- [7] ZHANG M L, ZHOU Z H. A Review on Multi-Label Learning Algorithms[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(8): 1819-1837.
- [8] EVA G, SEBASTIAN V. A Tutorial on Multilabel Learning

- [J]. *ACM Computing Surveys*, 2015, 47(3): 1-38.
- [9] TAE J J, HYUN J P, YOUNG-HAK K. Premature ventricular contraction beat detection with deep neural networks [C] // Proceedings of 15th IEEE International Conference on Machine Learning and Applications. New York: IEEE Press, 2016: 49-56.
- [10] TAE J J, HOANG M N, DAEYOUN K, et al. ECG arrhythmia classification using a 2-D convolutional neural network [EB/OL]. (2018-03-24) [2018-04-25]. <https://dblp.org/rec/bib/journals/corr/abs-1804-06812>.
- [11] TOMÁS T, PAULO F, JESÚS P. Heartbeat Classification Using Abstract Features From the Abductive interpretation of the ECG [J]. *IEEE Journal of Biomedical and Health Informatics*, 2018, 22(2): 409-420.
- [12] HARI M R, ANURAG T, SHAILJA S. ECG signal processing for abnormalities detection using multi-resolution wavelet transform and Artificial Neural Network classifier [J]. *Measurement*, 2013, 46(9): 3238-3246.
- [13] WANG Z, LI X F, ZHU H J, et al. Impact of Left Bundle Branch Area Pacing on Clinical and Electrocardiogram Characteristics Among Bradycardia Patients With Right Bundle Branch Block [J]. *Chinese Circulation Journal*, 2021, 36(1): 22-27.
- [14] CELIN S, VASANTH K. ECG Signal Classification Using Various Machine Learning Techniques [J]. *Journal of Medical Systems*, 2018, 42(2018): 240-250.
- [15] LI K, DU N, ZHANG A D. Detecting ECG Abnormalities via Transductive Transfer learning [C] // Proceedings of ACM-BCB. New York: IEEE Press, 2012: 35-74.
- [16] KEIJI G, SHOTA H, HIDEH-IRO O, et al. Building Normal ECG Models to Detect Any Arrhythmias Using Deep Learning [C] // Proceedings of the 47th International Conference on Computing in Cardiology. New York: IEEE Press, 2020: 45-51.
- [17] LI F, MIAO D Q, ZHANG Z F, et al. Mutual information-based granular feature-weighted multi-label learning k-nearest neighbor algorithm [J]. *Computer Research and Development*, 2017, 54(5): 1024-1035.
- [18] MATTHEW R, BOUTELL, LUO J B, et al. Learning multi-label scene classification [J]. *Pattern Recognition*, 2004, 37(9): 1757-1771.
- [19] JESSE R, BERNHARD P, GEOFF H, et al. Classifier chains for multi-label classification [C] // ECML PKDD 2009. Berlin: Springer, 2009: 254-269.
- [20] JOHANNES F, EYKE H, ENELDO L M, et al. Multi-label classification via calibrated label ranking [J]. *Machine Learning*, 2008, 73: 133-153.
- [21] GRIGORIOS T, IOANNIS K, IOANNIS V. Random k-labelsets for multi-label classification [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2011, 23(7): 1079-1089.
- [22] ZHANG M L, ZHOU Z H. ML-KNN: A Lazy Learning Approach to Multi-Label Learning [J]. *Pattern Recognition*, 2007, 40(7): 2038-2048.
- [23] AMANDA C, ROSS D. KING. Knowledge discovery in multi-label phenotype data [C] // Proceedings of International Conference on Principles of Data Mining and Knowledge Discovery. 2001, 42-53.
- [24] ELISSEEFF A, WESTON J. A kernel method for multi-labelled classification [C] // Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic. New York: IEEE Press, 2001: 681-687.
- [25] NADIA G, ANDREW M C. Collective Multi-Label Classification [C] // Proceedings of the 14th ACM International Conference on Information and Knowledge Management. Berlin: Springer Press, 2005: 93-99.
- [26] HALL M. Correlation-Based feature selection for discrete and numeric class machine learning [C] // Proceedings of the 17th International Conference on Machine Learning. Berlin: Springer Press, 2000: 359-366.
- [27] RON K, GEORGE H. John. Wrappers for Feature Subset Selection [J]. *Artificial Intelligence*, 1997, 97(1): 273-324.
- [28] ZHANG M L, WU L. LIFT: Multi-Label Learning with Label-Specific Features [J]. *IEEE Transactions on Pattern Analysis and Machine Learning*, 2015, 37(1): 107-120.
- [29] HAMED D H, MARYAM M. A Multi Rate Marginalized Particle Extended Kalman Filter for P and T Wave Segmentation in ECG Signals [J]. *IEEE Journal of Biomedical and Health Informatics*, 2019, 23(1): 112-122.
- [30] LI C W, ZHENG C X, TAI C F. Detection of ECG Characteristic Points Using Wavelet Transforms [J]. *IEEE Transactions on Biomedical Engineering*, 1995, 42(1): 21-28.
- [31] SMIT H W, VERTON K, GRIMBERGEN C A. A Low-Cost Multichannel Preamplifier for Physiological Signals [J]. *IEEE Transactions on Biomedical Engineering*, 1987, 34(4): 307-310.
- [32] LAGUNA P, THAKOR N V. New algorithm for QT interval analysis in 24 hour Holter ECG: Performance and applications [J]. *Medical Biological Engineering and Computing*, 1990, 28: 67-73.
- [33] SALAH H, ASMA B A, MOHAMED H B. A robust QRS complex detection using regular grammar and deterministic automata [J]. *Biomedical Signal Processing and Control*, 2018, 40: 263-274.
- [34] HAN J W, MICHELINE K, PEI J. *Data Mining: Concepts and Techniques* [M]. AMSTERDAM: Morgan Kaufmann, 2012: 443-474.
- [35] FERGUSON T S. A Bayesian Analysis of Some Non-Parametric Problems [J]. *The Annals of Statistics*, 1973, 1(2): 209-230.
- [36] Masked ECG DATASET [EB/OL]. (2020-03-01) [2021-10-27]. <https://github.com/hjyresearch228/PAC>.
- [37] MIT-BIH ARRHYTHMIAS [EB/OL]. (2005-03-01) [2005-03-27]. <https://physionet.org/content/mitdb/1.0.0/>.



HAN Jingyu, born in 1976, Ph. D, professor, is a member of China Computer Federation. His main research interests include biomedical data processing, machine learning and spatio-temporal database.