

基于多项式参数拟合和支持向量机的心肌梗死识别算法

武会江, 黄启俊, 常胜, 王豪

武汉大学物理科学与技术学院, 湖北 武汉 430072

【摘要】心电图的复杂性和个人差异导致心肌梗死诊断标准的难以遵循, 本文提出了一种基于多项式参数拟合和支持向量机的心肌梗死识别算法。原始信号经过滤波、去噪等预处理之后, 使用多项式拟合的方法得到曲线拟合效果最佳的拟合系数, 并把预处理中的参数作为特征值, 支持向量机作为分类器, 实现心电信号的自动分类。该方法实现了心肌梗死信号和正常心电信号的分类, 最终识别率为82.017 9%。该方法可行性高、识别率高, 具有可扩展性。

【关键词】心肌梗死; 多项式拟合; 支持向量机; 心电图

【中图分类号】R541.7; R540.41

【文献标志码】A

【文章编号】1005-202X(2017)07-0736-03

Myocardial infarction recognition algorithm based on polynomial fitting and support vector machine

WU Huijiang, HUANG Qijun, CHANG Sheng, WANG Hao

School of Physics and Technology, Wuhan University, Wuhan 430072, China

Abstract: Based on the fact that the diagnostic criteria of myocardial infarction is hard to be followed because of the complexity and individual difference of electrocardiogram (ECG), we proposed a novel recognition method based on polynomial parameter fitting and support vector machine (SVM). The original signal is preprocessed by filtering and denoising. The optimal fitting coefficients obtained by the polynomial fitting are used as the eigenvalues, and the SVM is taken as the classifier in order to automatically classify the ECG signals. With a good feasibility, high recognition rate and extensibility, the proposed method distinguishes the myocardial infarction signals from normal ECG signals, with a final recognition rate of 82.017 9%.

Keywords: myocardial infarction; polynomial fitting; support vector machine; electrocardiogram

前言

心电图检查以其方便、迅速、准确等特点, 在心肌梗死的诊断中仍占有重要的地位^[1]。心电图对心肌梗死部位的诊断一般是以坏死型Q波出现于哪些导联为依据, 分为以下几种: 下壁心肌梗死、前间壁心肌梗死、前壁心肌梗死^[2]、广泛前壁心肌梗死、前侧壁心肌梗死、高侧壁心肌梗死、后壁心肌梗死和右室心肌梗死。

目前, 心肌梗死诊断所基于的原理主要是心电信号ST段或Q波的改变^[3-4]。但是, 因为心电图的复杂性和个人差异, 心肌梗死诊断标准难以遵循, 即使是资深的心脏病专家也只能识别心肌梗死患者中的82%的ST

段抬高^[5]。为了使一些非专业医师及民用仪器能够方便地探测心肌梗死, 很多心电特征提取的算法已经被提出, 这些算法都是基于小波算法, 并利用主成分分析和独立成分分析等进一步提取特征^[6-8]。这些特征提取算法综合考虑了时域和频域的信号特征, 但因为算法复杂度较高, 在硬件实现中要占用大量资源。

为了实现一种复杂度较低的特征提取算法, 本研究基于心电波形的多项式拟合的新型特征提取算法, 通过多项式拟合心电波形的方法, 最终得到高度拟合心电波形的多项式的系数, 并以此作为特征量。本文所采用的心电数据来自PTB数据库, 分类目标是识别正常心电信号和心肌梗死信号的二分类, 最终得到的识别率为82.017 9%。

1 多项式拟合

1.1 数据预处理

本研究所采用的心电数据为PTB心电信号数据库。PTB数据库是德国国家计量署提供, 数据来自柏

【收稿日期】2017-02-08

【基金项目】国家自然科学基金(61404094, 61574102); 湖北省自然科学基金(2014CFB694)

【作者简介】武会江, 硕士研究生, 主要研究心电信号自动识别以及分类方法, E-mail: winter_whj@whu.edu.cn

林的本杰明富兰克林医学大学的内科,包含心肌梗死数据148条,健康数据52条。每个记录包含16导联(14通道心电信号,1通道呼吸,1通道电压),采样率1 000 Hz,每段数据时长大约2 min,其目的在于心电信号算法标准的研究与教学。

本研究采用12个常规ECG导联进行有关实验,利用一个高通滤波器(截止频率0.05 Hz)来消除ECG信号基线漂移,并利用Tompkins算法检测ECG信号峰值 $R^{[9]}$,并通过截取相邻心电信号的RR间期,获得待归一化的心电信号片段。

通过对各信号片段的归一化处理,得到用来进行多项式拟合的曲线。

1.2 多项式拟合

心电信号经过滤波,归一化处理之后,得到待处理的归一化心电信号。通过多项式拟合,用一条多项式曲线函数来拟合该心电信号。通过寻找拟合效果最好的曲线,来确定合适的多项式函数最高阶数,并将得到的多项式系数作为特征,用于支持向量机(SVM)分类器分类。经预处理后得到的归一化心电波形示例见图1。

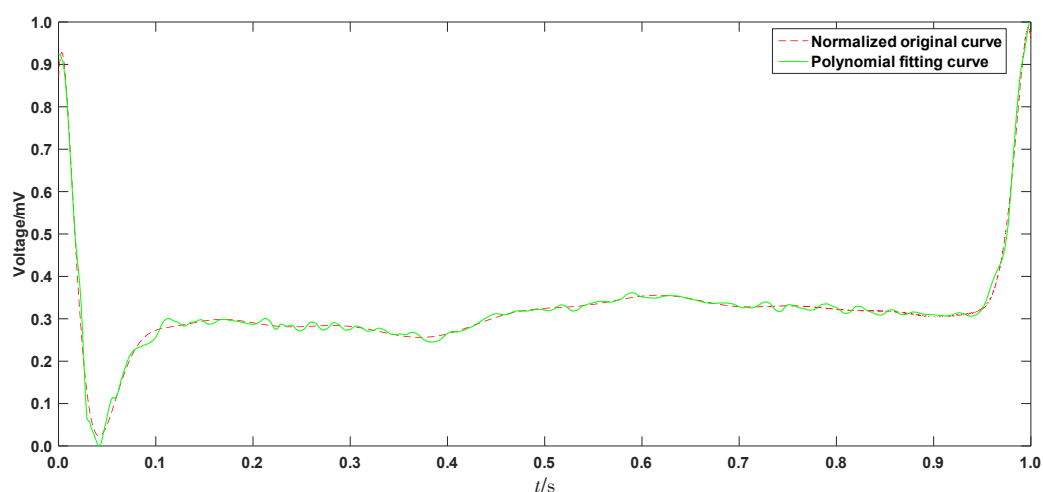


图1 预处理后得到的归一化心电波形

Fig.1 Normalized electrocardiogram (ECG) after preprocessing

拟合的多项式公式如下:

$$\text{polyFit}(x) = a_0 + a_1x + \cdots + a_nx^n \quad (1)$$

其中, a_i 表示各阶 x 的系数, n 表示多项式函数的最高阶数。

用 R^2 来衡量多项式拟合的效果:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

其中, y_i 表示原波形值, \hat{y}_i 表示拟合曲线值,

$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$, 显然 R^2 越趋近于1表明拟合效果越好。

考虑到当 n 取值过大时会出现龙格现象,即在进行高阶拟合时,边界部分会出现震荡^[10]。因此 n 取值从10到40,最终的拟合结果见图2。

由图2可知,当多项式最高次项的系数 $n=24$ 时,能得到最佳的分类结果。

因此,最终得到的特征值向量为 $[a_0, a_1, \cdots, a_{24}, C_x, C_y]$, 共含27个特征值,其中 C_x, C_y 分别为在数据预处理中,对 X, Y 坐标进行归一化操作的参数。以该特征向量作为分类器的输入,用于后续分类。

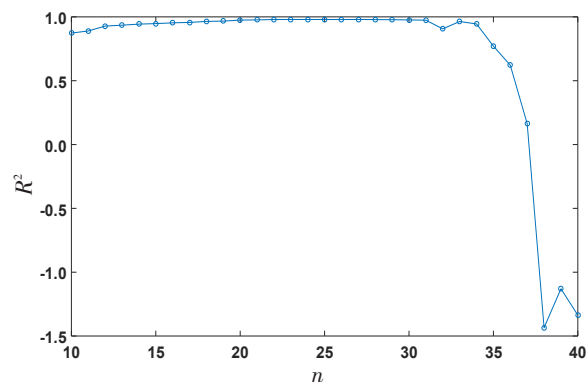


图2 n 取不同值时的拟合结果

Fig.2 Fitting results with different n

2 SVM分类

2.1 SVM介绍

SVM是一种二类分类模型,基本模型是定义在特征空间的间隔最大的线性分类器,间隔最大使SVM有别于感知机^[11]。一个分类任务实际的数据器通常包括训练集和测试集。训练集中的每一个实例包括一个目标值(标签)和多个属性(特征)。SVM的

目标是产生一个预测模型,根据测试集中实例的属性预测测试集的样本对应的目标值,即根据特征向量预测分类模型。

对于一个二类的训练集,其中元素是实例和标签集合 $(x_i, y_i), i=1, \dots, l$,其中 $x_i \in R^n$,并且 $y_i \in \{1, -1\}$,表示两个类别。如果用超平面 $\omega^T x + b = 0$ 分离,则须满足:

$$\begin{aligned} \omega^T x + b &> 0, \text{ 当 } y_i = 1 \\ \omega^T x + b &< 0, \text{ 当 } y_i = -1 \end{aligned} \quad (3)$$

得到判别函数为 $f(x) = \text{sign}(\omega^T x + b)$, x 为测试数据。通过极大化 $\omega^T x + b = \pm 1$ 的间距,从而在可能的超平面中选出有最大边际的距离 $2/\|\omega\| = \omega/\sqrt{\omega^T \omega}$ 。

对于分类问题,SVM用于找出下式中最优化问题的解:

$$\begin{aligned} \min_{\omega, b, \delta} & \frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^l \delta_i \\ \text{s.t. } & y_i(\omega^T \varphi(x_i) + b) \geq 1 - \delta_i \\ & \delta_i \geq 0; i = 1, 2, \dots, l \end{aligned} \quad (4)$$

其中, $\delta_i > 1$ 表示 x_i 在分界面错误判别的样本。 C 为错分样本的惩罚权重。

SVM常用的核函数包括线性核函数、多项式核函数、径向基核函数等核方法,本文采用的是RBF径向基核函数。

2.2 SVM分类结果

由于心肌梗死心电图表现需综合考虑各个不同导联,心肌梗死信号和正常心电信号分别从16个导联(14通道心电信号)中选取^[12]。用于SVM进行分类的训练和测试样本数如表1所示。共有训练样本25 715例,测试样本25 899例,最终得到的识别率有82.017 9%。

3 结论

表1 SVM样本数量

Tab.1 Numbers of support vector machine samples

Types of ECG	Training	Testing	Total
Myocardial infarction	20 731	20 883	41 614
Healthy control	4 984	5 016	10 000

基于心电图在心肌梗死诊断中的重要地位以及心电图的复杂性,本文提出一种基于心电波形的新型特征提取算法。采用多项式拟合的方法,即利用多项式函数来拟合原始心电波形,之后用拟合多项式的系数以及中间重要参数作为特征值,并使用SVM做为分类器,对心电特征进行分类,从而实现心肌梗死信号和正

常心电信号的自动分类,本文最终识别率达到了82.017 9%,可以用于辅助非专业医师或个人对心肌梗死的预测和诊断。但是,本文所使用的方法在一些地方还有待于提高,例如在数据处理过程中,拟合多项式的最高次项系数的选择以及SVM中重要参数的选取都会对识别结果造成很大影响,最终的分类识别率还可以通过改进拟合的方式进一步优化。

【参考文献】

- [1] 岳海. 心肌梗死心电图诊断的重要性[J]. 现代医药卫生, 2013, 29(17): 2649-2650.
YUE H. The importance of electrocardiogram diagnosis in myocardial infarction[J]. Modern Medicine & Health, 2013, 29(17): 2649-2650.
- [2] 王玉燕, 任秀娟, 郑兴锋. 心电图R波递增不良对前壁心肌梗死的诊断价值[J]. 第二军医大学学报, 2014, 35(6): 695-697.
WANG Y Y, REN X J, ZHENG X F. Diagnosis value of poor R wave progression for anterior wall myocardial infarction[J]. Academic Journal of Second Military Medical University, 2014, 35(6): 695-697.
- [3] 何艳萍, 胡晶华. 心肌梗死心电图的新进展[J]. 心脏杂志, 2014, 26(1): 115-116.
HE Y P, HU J H. New progress in myocardial infarction electrocardiogram[J]. Heart Journal, 2014, 26(1): 115-116.
- [4] 张海澄. 急性心肌梗死的心电图分类与诊断[J]. 临床心电学杂志, 2006, 15(3): 162-163.
ZHANG H C. Classification and diagnosis of acute myocardial infarction electrocardiogram[J]. Journal of Clinical Electrocardiology, 2006, 15(3): 162-163.
- [5] LIU B, LIU J K, WANG G Q, et al. A novel electrocardiogram parameterization algorithm and its application in myocardial infarction detection[J]. Comput Biol Med, 2015, 61(C): 178-184.
- [6] REMYA R S, INDIRADEVI K P, BABU K K. Classification of myocardial infarction using multi resolution wavelet analysis of ECG[J]. Procedia Technol, 2016, 24: 949-956.
- [7] 钟清华, 易子川. 基于I导联ECG的心肌梗死特征提取[J]. 生物医学工程杂志, 2013, 30(2): 260-264.
ZHONG Q H, YI Z C. Feature extraction of myocardial infarction based on lead IECG[J]. Journal of Biomedical Engineering, 2013, 30(2): 260-264.
- [8] 武扬. 心电特征提取及分类方法研究[D]. 上海: 上海交通大学, 2012.
WU Y. Research on ECG feature extraction and classification method[D]. Shanghai: Shanghai Jiao Tong University, 2012.
- [9] TOMPKINS W J. Biomedical digital signal processing[M]. Upper Saddle River: Prentice Hall, Inc, 1993: 246-253.
- [10] 张雨浓, 李名鸣, 陈锦浩, 等. 龙格现象难题破解之系数与阶次双确定方法[J]. 计算机工程与应用, 2013, 49(3): 44-49.
ZHANG Y N, LI M M, CHEN J H, et al. Solving the problem of Runge phenomenon by coefficients-and-orderdetermination method[J]. Computer Engineering and Applications, 2013, 49(3): 44-49.
- [11] 黄琼英. 支持向量机多类分类算法的研究及应用[D]. 天津: 河北工业大学, 2005.
HUANG Q Y. The study and application of support vector machine multiclass classification[D]. Tianjin: Hebei University of Technology, 2005.
- [12] 胡文瑛, 刘霞, Sophia Zhou, 等. 16导联心电图ST段改变对急性心肌梗死的临床价值[J]. 临床心电学杂志, 2008, 17(5): 340-343.
HU W Y, LIU X, ZHOU S, et al. Clinical value of ST segment change in 16-lead ECG for diagnosing acute myocardial infarction[J]. Journal of Clinical Electrocardiology, 2008, 17(5): 340-343.

(编辑:谭斯允)