

The Mobile Medical Service Based on the Analysis of Automatic Diagnosis of Holter Electrocardiogram Data

Donghai Huang¹, Xiaoya Li¹, Changjian Tu²

¹Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing

²Microcardio Technology Co., Ltd., Beijing

Email: huangdhchn@yeah.net, xyli@amss.ac.cn, Tu.changjian@microcardio.com

Received: Jan. 7th, 2016; accepted: Jan. 24th, 2016; published: Jan. 27th, 2016

Copyright © 2016 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this paper, we consider the mobile medical service from the perspective of diagnosis of electrocardiogram (ECG) signal, and expect to promote the development of mobile medical service through the implementation of automatic diagnosis of ECG. We focus on the main problem during the procedure of implementing the automatic diagnosis, which is illustrated with ECG clinical data. And we propose a hierarchical clustering method based on tree structure to improve the results of ECG pattern classification. The simulation results on clinical data collected by smart device such as Holter, reveal that the tree-like hierarchical clustering method can effectively detect abnormal heart beat from ECG data set.

Keywords

Electrocardiogram, Pattern Classification, Automatic Diagnosis, Mobile Medical Service

基于Holter心电数据自动诊断的移动医疗服务

黄东海¹, 李晓亚¹, 涂昌建²

¹中国科学院数学与系统科学研究院应用数学研究所, 北京

²北京微心百源科技发展有限公司, 北京

Email: huangdhchn@yeah.net, xyli@amss.ac.cn, Tu.changjian@microcardio.com

收稿日期: 2016年1月7日; 录用日期: 2016年1月24日; 发布日期: 2016年1月27日

摘要

本文从心电图(ECG)信号自动诊断的角度切入移动医疗服务, 期望以实现ECG自动诊断来带动移动医疗服务的发展。文中重点描述了实现自动诊断过程中面临的主要问题, 并结合临床数据进行验证, 提出以树结构层次聚类的方式提高ECG模式分类效果。在Holter临床数据上的实验结果表明, 树形层次聚类法可以很好的找到数据集中的异常心拍。

关键词

心电图, 模式分类, 自动诊断, 移动医疗服务

1. 引言

移动医疗将医疗服务与互联网和移动互联网技术融合, 是现代医疗与健康管理服务的新形态。在移动医疗服务架构下, 人们改变过去只能前往医院问诊的就医方式, 而是可以实时实地获得医生的建议, 以及与自身健康相关的诊断报告, 甚至可以直接获得远程医疗, 这在很大程度上节省了人们的就医时间和就医成本。

移动医疗由于对传统医疗服务的颠覆性改变, 发展十分迅速。在资本、技术和政策的多重因素推动下, 如春雨掌上医生、大姨吗、康康血压等行业内的先发企业如雨后春笋班涌现, 以BAT(百度 Baidu、阿里巴巴 Alibaba、腾讯 Tencent)为代表的互联网公司背靠自身具有的技术优势也在积极布局整个行业生态。然而在如此百花齐放的盛景背后, 更多的是细分领域内密集的同质化竞争, 而医疗的核心却触及甚少。目前, 经过互联网技术改造后的在线预约以及在线问诊等服务, 只是将传统的就医模式进行了信息化优化和升级, 只能做到单纯的线上沟通或者医疗建议, 而无法实现真正意义上的诊断, 自动诊断之路就显得更加遥远。图1展示了患者的完整就医行为链, 将整个过程分为院前、院中、院后三个环节, 从图中可以看出, 目前的移动医疗服务仅仅处于对当下就医模式的信息化改造, 很多环节并没有充分信息化和数字化。

穿戴式医疗仪器、无线通讯技术与网络技术的结合, 以及大数据分析技术的发展成熟, 使得实现真正意义上的自动诊断成为可能。无线传感技术和无创连续检测技术的发展, 使得患者产生的数据可以方便地得到实时采集, 并透过穿戴设备传输到诊断云平台, 通过利用大数据分析技术, 最终给患者提供诊断报告, 如图2所示。史丹佛大学将电子医疗记录、全基因组序列、保险和医药记录、可穿戴式感测器和社会环境数据等作为数据分析对象, 辅助医生和研究人员更好地预测个人罹患特定疾病的几率, 以此制定出早期检查和预防的方案。北大医信的临床数据中心(CDR)将患者症状、诊疗情况、医嘱资讯、用药效果进行分析和利用, 辅助医生进行相关疾病诊断。

在移动医疗服务的框架中, 疾病与医疗诊断的整套分类系统, 都需要重新改写。不再像如今的简化模式, 十分粗略地将个体分配到极为宽泛的疾病类别中, 而是借助患者的生理数据和疾病表征等因素, 实现精确划分、个性化医疗。在未来的移动医疗平台中, 大量的患者数据将被组合并为一体。在数据收集的基础之上, 移动服务中的自动诊断, 以及从临床取样中获取信息的能力, 都将得到很大提升。人类

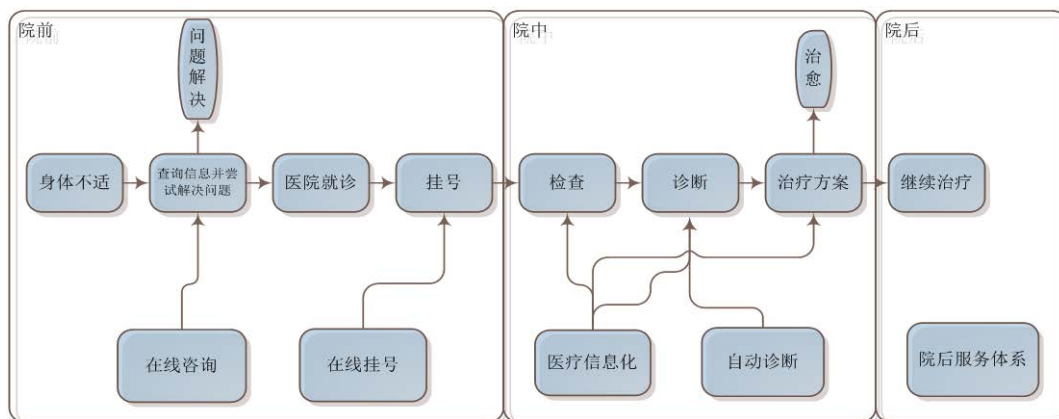


Figure 1. The patient behavior chain

图 1. 患者就诊行为链

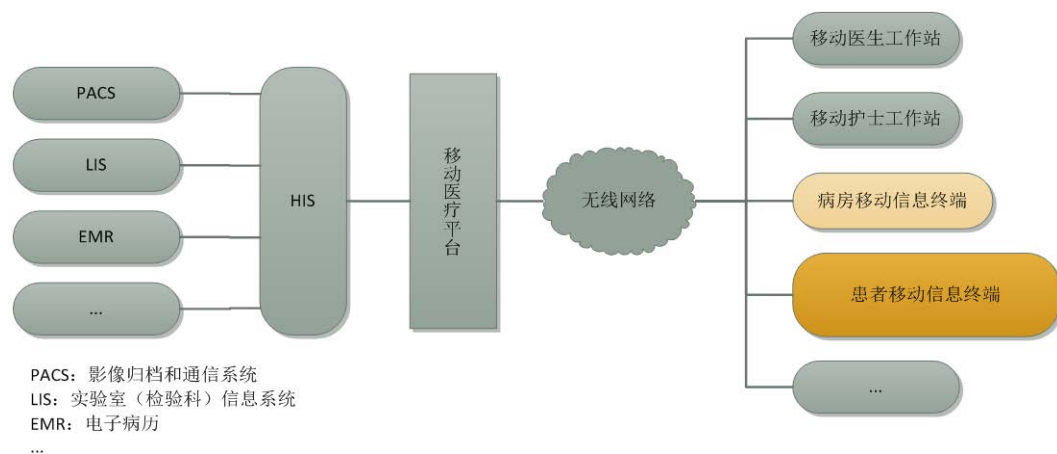


Figure 2. The framework of mobile medical service

图 2. 移动医疗框架

的疾病信息呈现多类别多变异和信息繁杂的特点,要做到对所有疾病实现一概而论的自动诊断极为困难,有针对性的各点突破才是目前的可能。正如埃里克·托普(Eric Topol)的观点,未来移动端的诊断工具一定是多样化地针对各类疾病的,如无线传感器与基因组学相结合的心脏病检测工具、依靠嵌入式纳米传感器检测血液中分子标签的新型智能手机应用等[1]。

心电图(electrocardiogram, ECG)是记录人体心脏电活动的可视时间序列,由于其采集简便可靠,对病人的无创性,已被广泛应用于心脏相关疾病的诊断工作。荷兰生理学家爱因托芬(William Einthoven)由于在 ECG 领域的杰出贡献,荣获 1924 年生理学及医学诺贝尔奖[2]。一批实时监测患者心电波动的穿戴设备,在移动医疗服务的浪潮下应运而生,为人们实现智慧医疗提供了一种新手段。本文以 ECG 临床数据(动态心电图数据, Holter)为研究背景,通过数据分析建立以 ECG 分类和诊断为基础的自动诊断模式。

2. ECG 数据特点及分析现状

临床数据中,病态 ECG 种类繁多、变异性极大,同种病理不同患者的心电图、甚至同一患者不同时刻的心电图都存在着很大的差异[3] [4]。这使得在人工识别各类 ECG 信号时,实时性和准确都不能得到很好的保证。

为了提高医疗诊断效率、缩短诊断时间,20 世纪 60 年代 Pipberger 等提出计算机辅助 ECG 分析的设

想[5]，之后大量学者着手 ECG 模式识别和分类问题的研究工作[6]。董军等将 ECG 模式分类总结为四个阶段：知识推理阶段、结构(语法)模式识别、统计模式识别以及神经网络分类[7]。其中，统计分析方法和神经网络模型取得了很好的效果，例如用于 ECG 特征提取和选择的主成分分析(principal component analysis, PCA)、独立成分分析(independent component analysis, ICA)、小波分析(wavelet analysis, WA)，用于 ECG 分类的 K 最近邻(KNN, k-nearest neighbor)、贝叶斯概率模型(Bayes)、支持向量机(SVM)，以及自组织神经网络(self-organizing map, SOM)和 BP 算法等。近几年，随着深度学习技术的兴起和发展，已有越来越多的学者将 CNN 等深度学习神经网络模型用于 ECG 模式分类工作[8]。

经过几十年的长足发展，用于 ECG 模式分类的方法越来越多，在细节处理上的技巧也越来越成熟。很多模型和算法在 MIT-BIT 等标准数据库上测试的性能都取得了很好的效果，但 ECG 模式分类成功应用于临床诊断的案例却极少。其中，最主要的原因是很多分类算法都是在如 MIT-BIT 上的标准数据库上做的测试，而 MIT-BIT 包含的患者记录数量非常少，且大多是 2 导联数据，噪声少干扰小。反观当前临床数据，采用 12 导联数据，数据体量大、噪声多，而采集于移动设备上的数据干扰项就更多。这就导致算法在临床数据集上的性能表现得不尽如人意。通过对 Holter 临床数据的分析，总结出导致此问题的主要原因有：

1) 算法的泛化能力很弱，对数据源的鲁棒性不够。很多算法在固定数据集上表现出令人满意的结果，但当数据源发生变化，算法的准确率往往也会随之出现较大幅度的下降。而在临床上数据集上，尤其是动态 ECG (ambulatory electrocardiogram, AECG 或 Holter)诊断方面，数据集的改变是不可避免的，病人的切换就意味着数据集的变化。并且在处理实时情况下的异常心拍识别时，类别的可分性会随着数据量的增加而变弱，这也导致了模型的泛化能力下降。

2) 没有很好的特征和方法来表示和区分异常心拍。实现自动诊断的必经之路是要识别病人表现疾病特征的那些心拍，这就需要在 ECG 模式分类的基础上剔除正常心拍以及干扰心拍。而对于 ECG 的疾病特征提取与表示就显得尤其重要，就目前而言，考虑融合医学专家诊断过程中所关注的特征或许是可行途径之一。

3) 提取异常心拍的算法时间复杂度太高，很难做到实时性。目前对异常心拍的处理，多在事后进行，采用静态数据分析的方式，通过模式匹配的方法，提取异常心拍。由于 ECG 数据分布的复杂性，此类模型通常需要较长的时间才能训练出适应数据集的算法。在移动端的 ECG 诊断中，数据量大、实时性要求高，传统的模型和算法不再适用，需要设计出新的算法和模式。

3. 树形结构层次聚类分析 Holter 数据

以典型的树形结构构造适用于 Holter 数据的分类模型。首先，运用自顶向下的多层结构，在不同的层次训练不同的聚类模型。在聚类的结果基础上递归再聚类，根据聚类效果判定是否需要继续进行聚类分析。其次，在多属性类别聚类分析的基础上，训练分类器识别未知心拍中的噪声和异常心拍。图 3 展示了此方法的主要分析框架，其核心之处在于根据不同的样本选择不同的聚类分析模型以及设定合适的阈值判定是否需要进一步聚类。

图 3 的分析框架使得在 ECG 数据量较大时，能够很好降低类别间的交叉度，提高聚类效果；使得类别数目可以根据数据本身具有的结构进行调节。因此，在噪声干扰和异常心拍较多时，同一水平的聚类效果往往意味着更多的类别数目。并且在聚类分析和训练分类器的各个环节上，根据不同的效率指标，可以选择不同的聚类算法和分类算法，这就将多种算法实现了有机融合。

图 4 显示了在 Holter 二十四小时临床数据集上的部分聚类分析结果，这里采用了 SOM 神经网络模型和学习矢量量化神经网络 LVQ (Learning Vector Quantization)。图 3 的框架仅对不符合聚类阈值标准的

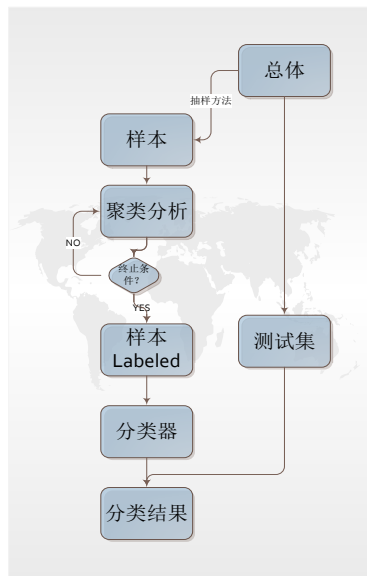


Figure 3. The analytical framework
图 3. 分析框架

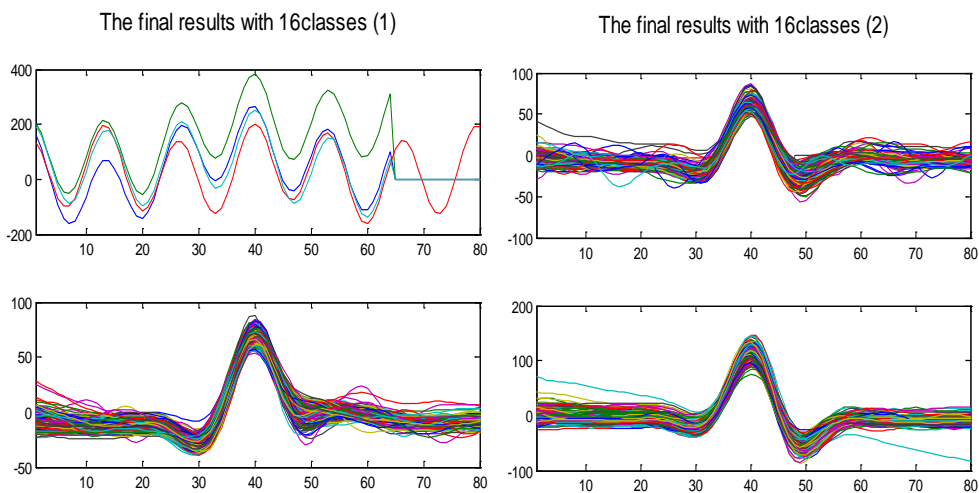


Figure 4. The illustration for clustering
图 4. 聚类结果展示

类别进行再聚类，因此随着算法深入，输入算法的数据集会随之减少。在 2000 个 Holter 临床心拍组成的数据集上，采用树形结构的聚类方法显示了很好的效果。数据集被聚成了 16 类，从形态特征上看，每类之间都呈现较大的差异，聚类效果很好，如图 4 所示。在此基础上，将数据纪录多的类别记入“大分类”，数据纪录少的类别记入“小分类”，分别采用 BP 神经网络模型和 SVM 支持向量机模型进行训练。

树形结构是一种递归实现，它保证了 Holter 数据做聚类分析时的效果。然而由于递归算法的特性，其无法直接应用于新数据分类，识别新数据异常心拍。为实现这个目的，我们利用监督算法给数据加上标签，并输入监督学习算法训练分类器，将分类器用于识别新数据异常心拍。由于树形聚类算法将类别间交叉度都降到了很低的程度，所以在这里我们尝试的 SVM 和 BP 神经网络的分类算法都达到了比较好的效果，SVM 的准确率为 92.7%，BP 神经网络的准确率为 93.01%。图 5 展示了 BP 神经网络算法在分类处理时的收敛情况。

可以看到，到目前为止这种树形聚类算法的表现都非常好，这是因为分类算法最大程度地利用树形聚类带来的优势。但是当我们将分类算法应用在新的数据源上时，该算法的泛化能力却极其低下。我们采用不同于原 2000 个心拍数据的临床数据，最后分类算法的识别结果如图 6 (这里只展示了部分结果)。在 ECG 模式分类的研究中，算法在临床数据上泛化能力差一直是学者们十分头痛的问题，也是阻碍 ECG 自动诊断在临床上应用的主要原因之一。本文尽管采取了树形聚类降低了类别交叉度，但是树形结构的算法本身是递归算法，尽管通常显得很简洁，但需要很大的内存消耗和时间复杂度，尤其在数据量增大的情况下。

树形聚类的设计目的是为了更准确的捕捉异常心拍，将其作为疾病诊断特征之一。因此，先天心拍数据的加入除了增加算法的时间和空间复杂度外，并没有其它的作用。基于此考虑，为了弥补树形聚类的不足，我们在聚类之前首先采用模式匹配的方法将原数据集进行先天心拍剔除处理。根据大数定律，以均匀分布的方式抽样患者心拍数据；均值化处理，得到先天心拍模板；再根据先天心拍模板过滤整个数据集；在过滤的过程中，增加的先天心拍作为更新先天心拍模板的数据来源。在 Holter 数据集上的实验结果显示，经过模式匹配处理，可以过滤掉 90% 以上的 ECG 数据。

经过处理后的数据集，体量明显缩小，树形聚类的效果也得到了很大提升。图 7 展示了修改后的聚类算法识别出的异常心拍。结果显示，树形聚类可以有效提取剩余 10% 数据中的异常心拍。然而，树形聚类的递归实现，使得算法本身具有了过拟合的特点，因此在训练分类算法进行实时异常心拍预警处理时，这个方法仍然遇到了困难。树形聚类是基于经验阈值进行递归的，经验值的确定极其重要，太高会过拟合，太低会使后来的分类算法难以收敛。这是在以后的工作中有待提高的地方。

4. 结论

各种用于心电监测的移动设备，使得 ECG 数据采集的规模日益增长，医生和研究人员能够以更细微的粒度去挖掘和分析 ECG 中隐藏的形态特征和疾病信息。目前来看，ECG 分类以及预测模型的建立并不是主要困难，大数据分析技术的日趋成熟，使得可供选择的技术和策略都大大改善，建立的模型和算

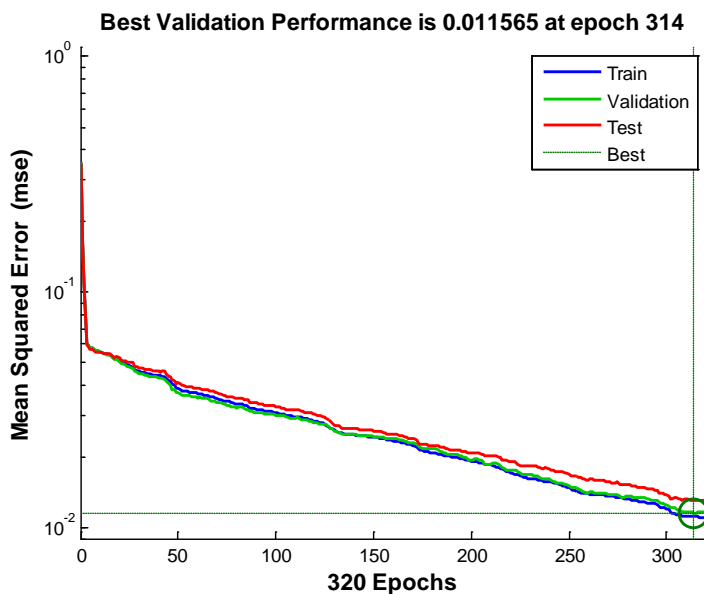


Figure 5. Schematic diagram of convergence for BP neural network algorithm

图 5. BP 神经网络算法收敛示意图

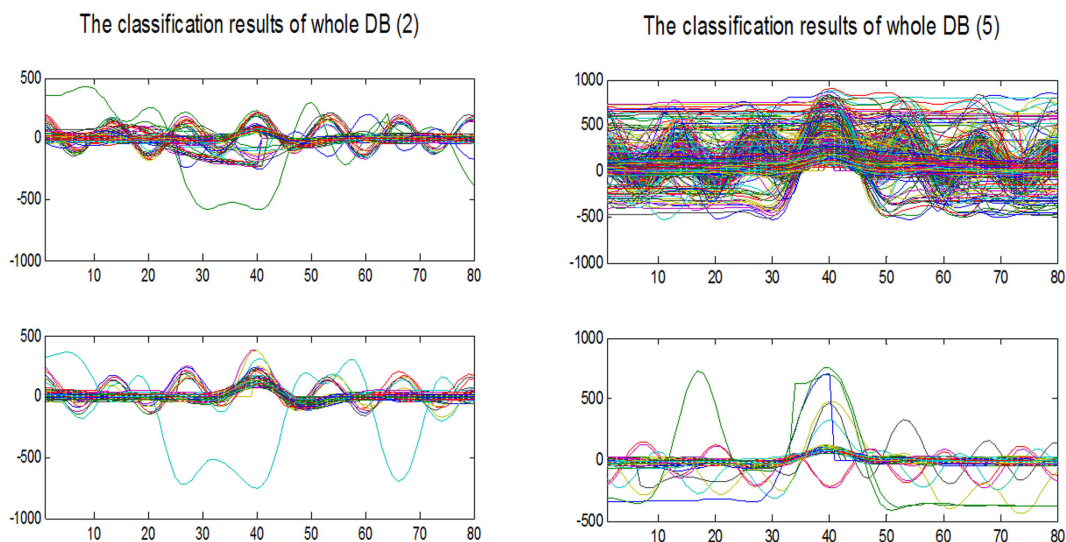


Figure 6. Generalization ability test for classification algorithm

图 6. 分类算法泛化能力测试效果图

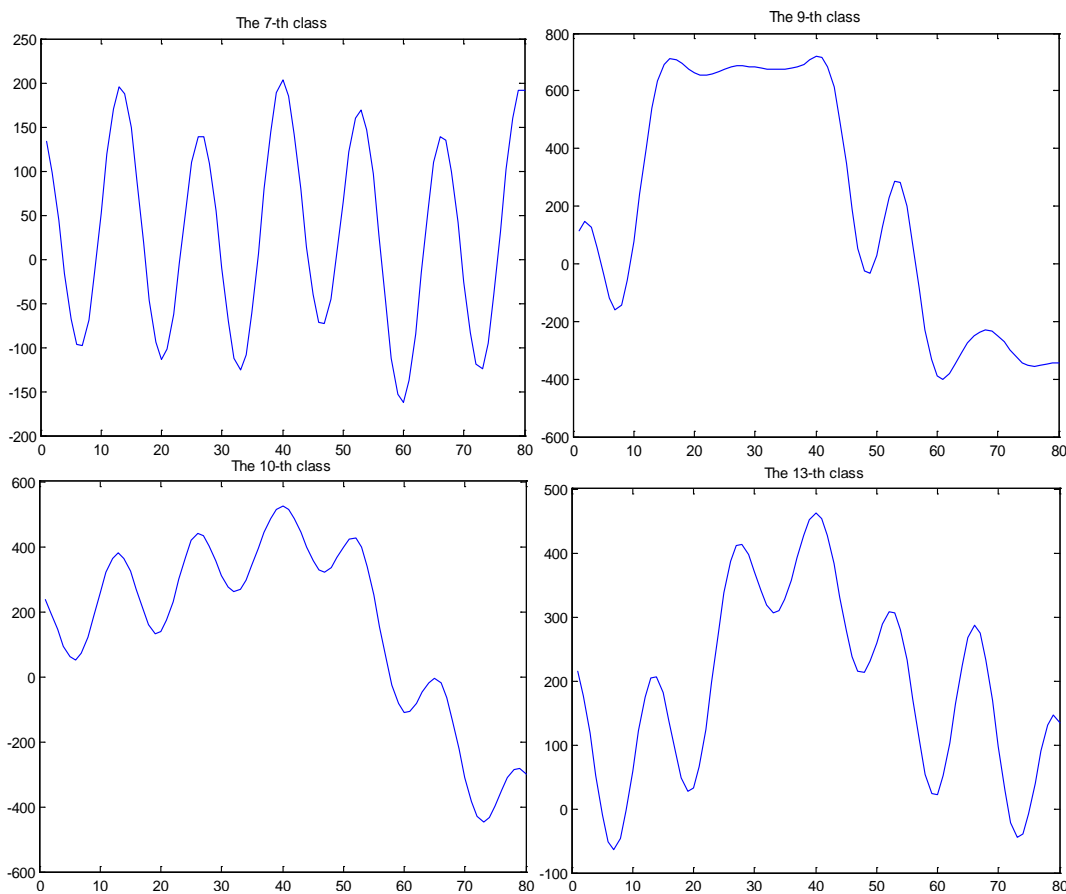


Figure 7. The abnormal heart beat identified by modified tree-like clustering method

图 7. 修正后的树形聚类识别出的异常心拍

法在标准数据库上的测试结果都能得到比较满意的结果。主要的困难点是建立合适的模型及算法，包括与临床应用相匹配的模型，能够辅助临床正确决策、实现自动诊断的算法。

本文提出的模型虽然克服了很多困难,但是在临床应用中仍然有不足。如前面所提到的,训练得到的聚类模型无法在每个树节点中保存下来。因此,聚类算法的结果只能是给训练数据贴上标签,而为了捕捉新数据的异常心拍,需要进一步训练适用的分类器。虽然我们通过剔除先天心拍极大的降低了数据体量,通过树形层次聚类很好的捕捉了异常心拍,但是这同样也会带来由于数据量减少分类器泛化能力降低的困扰。综合考虑计算时间和泛化能力,以识别动态心拍中的异常心拍为目的的算法,应该是基于足够数据集的稳健模型。该模型从在大规模数据集上长时间训练得来,具有相当的泛化能力,并且可以通过捕捉到的实时异常数据对模型进行更新。采用异常数据更新的原因是异常数据体量小,计算时间短,并且由于模型在大数据集上训练的稳健性,其对先天心拍数据已有非常好的识别能力。基于以上考虑,未来的工作是采用深层神经网络模型对本文分析框架进行修正和补充,弥补由于模型无法保存而带来的重复训练。

未来的 ECG 自动诊断技术还依赖于数据采集、数据分析和临床经验的多方面结合。目前的 ECG 临床数据受干扰因素多,很大部分都没有如标准数据库中明确标注心拍信息,这给后来的数据分析工作带来很大困难。未来 ECG 数据的采集以及检测工作,需要依赖于多工具的交叉,如使用血压与心拍波峰交叉检验[9]。从目前的研究来看,ECG 数据分析工作采用的方法主要是,基于机器学习的方法和统计分析的方法。鉴于 ECG 数据本身的多维性(文本、数字、图像)、异质性和时空动态等特征,已有很多学者将多种不同模型和算法进行融合,用于 ECG 数据分析。就目前来看,其临床效果仍然不令人满意。下一个可能的技术突破可能是引入更适宜的多学科方法,比如数据包络分析(DEA)技术与数据挖掘技术的结合[10]。更多适用的算法需要研究者不断的尝试和比较。

参考文献 (References)

- [1] Eric, T. (2013) *The Creative Destruction of Medicine*. Basic Books AZ.
- [2] 李志平, 张福利, 马学博. 心电研究与现代心电图监测法建立的历史回顾[J]. 中华医史杂志, 1999, 29(4): 215-219.
- [3] Engin, M., Fedakar, M., Engin, E.Z., *et al.* (2006) Feature Measurements of ECG Beats Based on Statistical Classifiers. *Measurement*, **40**, 904-912. <http://dx.doi.org/10.1016/j.measurement.2006.10.012>
- [4] Inan, O.T., Giovangrandi, L. and Kovacs, G.T.A. (2006) Robust Neural-Network Based Classification of Premature Ventricular Contractions Using Wavelet Transform and Timing Interval Features. *IEEE Transactions on Biomedical Engineering*, **53**, 2507-2515. <http://dx.doi.org/10.1109/TBME.2006.880879>
- [5] Friedemann, W., Stallmann, D., Hubert, V., *et al.* (1961) Automatic Recognition of Electrocardiographic Waves by Digital Computer. *Proceedings of the Fifth Annual Meeting of the Biophysical Society*, The Biophysical Society, St. Louis, 1138-1143.
- [6] Clifford, G.D., Azuaje, F. and McSharry, P.E. (2006) *Advance Methods and Tools for ECG Data Analysis*. Artech House, London.
- [7] 王丽萍, 董军. 心电图模式分类方法研究进展与分析[J]. 中国生物医学工程学报. 2010(6): 916-925.
- [8] 金林鹏, 董军. 面向临床心电图分析的深度学习算法研究[J]. 中国科学: 信息科学, 2015(3): 398-416.
- [9] Bo, Y., Soo, K.T., Bart, H., Christopher, M. and Yi, S. (2014) Robust Identification of Heartbeats and Blood Pressure Signals and Noise Detection. *Computing in Cardiology*, **41**, 565-568.
- [10] Wu, D. (2009) Supplier Selection: A Hybrid Model Using DEA, Decision Tree and Neural Network. *Expert Systems with Applications*, **36**, 9105-9112. <http://dx.doi.org/10.1016/j.eswa.2008.12.039>