

## Y-30-0018

## 生物医学信号研究资源网站—PHYSIONET

庞兴梅\*

山东大学齐鲁医院 250012

## 引言

自 1999 年起, 由美国 National Institutes of Health (NIH) 资助, 美国麻省理工学院 (MIT) 联合波士顿 Beth Israel Deaconess 医学中心、波士顿大学、McGill 大学等单位建立了基于 Web 的复杂生理信号和生物医学信号研究资源网站 PhysioNet ([www.physionet.org](http://www.physionet.org)), 用于促进研究数据和分析软件的交流, 激励相关研究人员之间的广泛协作, 同时促进新的分析方法和参考数据库的发展。

PhysioNet 包括三个相互关联的部分: (1) PhysioBank 是一个大型的、不断扩充的、由生物医学研究机构共享的、记录复杂生理信号和相关数据的数据文档集合, 其中大部分数据带有详尽的注释, 数据来源于正常人及常见多发病种的病人, 如心脏猝死、心力衰竭、心律失常、癫痫、睡眠呼吸暂停综合症等, 收录了包括 MIT-BIH 心律失常心电数据库、多参数智能重症监护数据库 (the Multi-Parameter Intelligent Monitoring for Intensive Care II, MIMIC II) 等在内的典型生理信号数据库; (2) PhysioToolkit 是一个开放源码的软件库, 它用于分析与处理生理信号、应用传统理论与技术和基于统计物理学和非线性动力学的新方法检测生理信号、交互式显示和描述信号、创建数据库、仿真生理信号、定量地评价和比较各种分析方法、以及分析不平衡和不稳定的过程等; (3) PhysioNet 同时提供了一个开放式的网络资源平台, 可以自由地通过网络访问, 免费获取 PhysioBank 和 PhysioToolkit 的原始数据和分析工具, 发布和交换生物医学信号以及处理信号的开放源码软件, 为协作进行数据分析和新算法评估提供便利; 同时还通过在线指南对用户如何利用该网络资源提供服务和训练, 并设置著作论文、最新进展、讨论、挑战、常见问题等专题。PhysioNet 为研究人员提供了一个方便获得临床数据、开发数据分析算法、分享研究成果的平台, 为临床教育提供重要的病例资料, 内容经过严格的审查, 其科学性和严谨性已得到广泛的验证, 享有很高的权威性。

## 1. PHYSIOBANK 数据和文件

PhysioBank 包括多参数数据库、心电数据库、神经学数据库、影像数据库、合成数据库等。数据库中的每一个病例记录的所有相关文件都保存在以该记录命名的子目录下, 每一条记录包括至少三类文件, 头文件 (.hea)、信号文件 (.dat)、和注释文件 (.atr, .al, .alM 等), 分别以记录名后加独特的扩展名来标示。如 a40001 记录, 包括信号文件 a40001.dat、头文件 a40001.hea 和注释文件 a40001.al。

头文件是描述信号属性的文本文件, 内容包括记录名、信号数目、信号文件的统一资源定位地址 (Uniform Resource Locator, URL)、贮存格式、信号数量和类型、信号的片断数、采样频率、校验和、数字化特征、记录的持续时间和起始时间等信息。内容用换行符分隔成多行, 至少包

括一条记录行, 指定记录的名字, 节段数量、信号数量、采样频率、基础计数值、每个信号的采样点数目、基础时间和日期; 信号详细说明行指定保存信号的采样数据的文件名和信号格式; 其最后一行紧接着是内容行, 是能被 WFDB 库函数 getinfo、putinfo 读和写的字符串; 多节段记录行包括对从 0 开始排序的节段的详细说明; 节段详细说明行包括记录名和每一信号的采样点数量, 字段间以空格分开。

信号文件是存储信号的数字化采样点的二进制文件, 一个组的信号以相同的格式存储, 最常用的是 8 位和 16 位数据格式。

注释文件记录了对信号特征的注释信息, 如 .atr 注释文件包含对此记录中的每一个心电 QRS 波群和相应节律的注释, 表明各自的起始时间、类型和心律等信息; .al 注释文件记录监护仪报警信息; .alM 注释文件记录了专家组对报警进行人工审定的结果。注释结构由长整型的注释时间、字符型的注释类型、子类型、通道号、数值字段和辅助信息等字段组成。

## 2. PhysioToolkit 软件包

PhysioToolkit 是一个庞大且不断扩充的软件库, 包含了多种软件包, 主要有:

1. WFDB (waveform database) 数据库开发工具包, 用于读取和分析 PhysioBank 数据库。
2. 仿真软件, 人工合成的各种动力学波形, 用于算法分析和模型评估。
3. 绘图软件, 将文本或二进制数据绘成高品质的图形。
4. 生理信号处理软件, 提供了一些通用的信号处理算法。
5. 心率变异性 (Heart Rate Variability, HRV) 分析软件。
6. 时间序列的非线性分析软件。

WFDB 软件包是用来读取和分析 PhysioBank 中数据库的开放源码的软件, 可在 Linux、Unix、Mac、Windows 等操作系统下运行。WFDB 软件包由 WFDB Library、WFDB 应用程序和 WAVE 三部分组成。

WFDB Library 是用来读、写信号文件、注释文件、头文件的函数集合。这些函数可在用户以 C、Fortran、Java、Matlab 等语言编写的程序中调用, 从而读取 PhysioBank 文件进行数据分析、以新记录文件或注释文件的形式生成分析结果, 这不仅为用户节省了大量的时间和精力, 而且可随时为数据库增加新的内容。

WFDB 应用程序是一个经过测试的命令工具集, 用于信号分析和处理, 如不同数据格式间的转换 (a2m、ahaconvert、m2a、edf2mit、mit2edf)、信号校准 (calsig)、波形检测和定位 (ecgpuwave、sqrs、wabp、wqrs)、心率和功率谱分析 (ihr、hrfft、hrlomb、hrmem)、信号处理 (fft、fir、mfilt、dfa)、信号分析 (sigamp、sigavg、signame、sampfreq)、注释文件操作 (rdann、wrann、sumann、ann2rr、rr2ann、mrgann)、绘图 (plot2d、plot3d、plt、pltf、imageplt) 和数据库文件操作 (wfdbcats、wfdbcollate、wfdbconfig、wfdbdesc) 等, 功能强大。

例如: 命令 rdann, 功能是读取并显示指定记录的注释文件, 命令格式为: "rdann -r 记录名 -a 注释名 [选项]", 以文本格式输出注释的内容, 输出包括注释时间 (格

\* 庞兴梅, 馆员, 生物医学工程学硕士, 方向: 医学信息, 医学信号检测分析。

式为小时:分:秒:微秒),采样点序列值、注释类型、子类型、通道号、数值字段以及辅助信息串。如“rdann -a atr -r 200 -f 0 -t 5:0 -p v”显示第 200 号记录的参考注释. atr 文件的第一个 5 分钟内所有室性早搏的注释。

又如:命令 rdsamp, 功能是读取指定记录的信号文件, 命令格式为:“rdsamp -r 记录名 [选项]”, 将采样点值转换成十进制数据输出, 当无选项时从指定记录的起点开始输出所有采样点, 每一行输出包括采样点编号和每一个通道的采样点数值。

Wave 是用于波形分析、显示和注释编辑的窗口应用程序, 主要功能有:

1. 快速显示不同校准尺度下的波形和注释;
2. 快速定位到记录的任何一位置, 通过存贮历史数据、提前预读等方式, 提高程序运行效率和缩短网络传输时间;
3. 向前和向后搜索特定的注释类型;
4. 使用标准的或用户自定义的注释来编辑图形化注释;
5. 采用不同的波形显示速度, 可仿真示波器显示, 能够前进或后退;
5. 高清晰打印用户选择的信号段;
7. 灵活控制与外部程序的接口来处理和分析信号;
8. 提供远程操作模式, 通过浏览器外部程序进行 wave 的波形显示;
9. 提供在线帮助。

Wave 的命令格式:“wave -r record[+record...][options...]”, 如使用选项“-a 注释文件类型”, 将显示指定记录的波形和相应注释。

### 3. 应用开发

医学科研人员可将 physionet 应用到自己的研究中, 来验证和评估各种实用算法。

首先, 选取 PhysioBank 包含的各个数据库中的研究数据, 这些数据库样本量大, 且提供了专家组的金标准注释, 分析结果具有较高的科学性和统计学意义; 其次, 开发程序实现自己研究的算法, 在电脑中下载和安装 WFDB 后, 研究者用 C 语言和 MATLAB 等软件编写数据的读取、检测、分析、比较等应用程序, 其中可以直接调用 WFDB 库函数, 读取数据库中的数据, 对信号进行分析, 生成分析结果, 大大提高了工作效率; 并可以将分析结果与专家注释金标准对照, 得出比较结果并加以统计计算, 从而整体评价自己算法的性能优劣。

### 4. 总结

PhysioNet 是分析人体生理信号的重要的数据资源和研究开发平台。我们取 PhysioBank 数据库中的数据作为评价数据集, 利用 PhysioTool kit 软件包开发程序, 对研究算法的性能进行整体评价