

PhysioNet 信息资源解析及利用

庞兴梅

(山东大学齐鲁医院图书馆 济南 250012)

〔摘要〕 介绍基于 Web 的复杂生理信号和生物医学信号研究资源网站 PhysioNet 及其相互关联的组成部分: 数据库 PhysioBank, 软件库 PhysioToolkit 和网络资源平台 PhysioNet。探讨医学科研人员利用 PhysioNet 进行算法研究的过程, 并对算法的性能进行整体评价。

〔关键词〕 PhysioNet; 信息; 网络资源; 软件开发

Analysis and Application of PhysioNet Information Resources PANG Xing - mei, Library of Qilu Hospital, Shandong University, Jinan 250012, China

〔Abstract〕 The paper introduces a complex physiology and biomedicine signal study resource website: PhysioNet and related components, including a database PhysioBank, a software library PhysioToolkit and a network resource platform PhysioNet, It discusses algorithm research process with PhysioNet and overall evaluates the algorithm performance.

〔Keywords〕 PhysioNet; Information; Network resources; Software development

1 引言

PhysioNet^[1-3] 是基于 Web 的复杂生理和生物医学信号的研究资源网站, 网址为 <http://www.physionet.org>。始建于 1999 年, 由美国国立卫生研究院 (National Institutes of Health, NIH) 资助, 美国麻省理工学院健康科学与技术中心联合其他医学机构共同创建, 旨在促进研究数据和分析软件的交流, 鼓励相关研究人员之间进行广泛的协作, 促进新的分析方法和参考数据库的发展。

2 PhysioNet 的信息资源

2.1 PhysioBank

大型的、不断扩充的数据库集合, 它记录了复

杂生理信号和相关数据。根据数据的研究进展状况, 这些数据库可以分为 3 类, 即 Class 1、2、3。Class 1: 专家已经作出了标注; Class 2: 原始数据; Class3: 正在研究进展中。从专业分, PhysioBank 包括多参数、心电、神经学、影像以及综合数据库等, 如典型的 MIT - BIH 心律失常心电数据库、常用的多参数智能重症监护数据库等, 其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, 其中大部分数据带有详尽的注释。

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 3 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。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、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

〔修回日期〕 2010 - 02 - 24

〔作者简介〕 庞兴梅, 硕士, 馆员, 发表论文数篇。

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，多是能够被 PhysioToolkit 的 WFDB 库函数 `getinfo`, `putinfo` 读和写的字符串。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，标明各自的起始时间、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。结构由注释的时间、类型、子类型、通道号、数值字段和辅助信息等字段组成。

2.2 PhysioToolkit

开放源代码的软件库，它用于分析与处理、检测、交互式显示和描述、仿真生理信号，并创建数据库、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。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 WFDB (Waveform Database) 数据库开发工具包、仿真软件、绘图软件、生理信号处理软件、心率变异性分析软件、时间序列的非线性分析软件等。WFDB 工具包是用来读取和分析 PhysioBank 数据库中开放源代码的软件的集合，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 共 3 部分组成。

WFDB Library 是用来读和写信号文件、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 中的文件进行数据分析，并以新记录文件或注释文件的形式生成分析结果，缩短了应用程序的开发周期，节省了用户的时间和精力，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 (`edf2mit`, `mit2edf` 等)；信号校准 (`calsig` 等)；波形检测和定位 (`wabp`, `wqrs` 等)；心率和功率谱分析 (`hrfft`, `hrlomb` 等)；信号处理 (`fft`, `fir` 等)；信号分析

(`sigamp`, `sampfreq` 等)；注释文件操作 (`rdann`, `mrgann` 等)；绘图 (`plot2d`, `imageplt` 等) 和数据库文件操作 (`wfdbcat`, `wfdbconfig`) 等，功能极为强大。WAVE 是用于波形分析、显示和注释编辑的窗口应用程序，能实现的功能主要有：快速显示不同校准尺度下的波形和注释；快速定位到记录的任何位置；向前或向后搜索特定的注释类型；图形化编辑注释；采用不同的波形显示速度，可仿真示波器显示；高清晰打印用户选择的信号段；灵活控制与外部程序的接口；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；提供在线帮助。

2.3 PhysioNet

提供了一个开放式的自由的网络资源平台，通过网络可以免费获取 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代码软件。同时 PhysioNet 还通过在线指南对用户如何利用该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，其内容均经过严格的审查，科学性和严谨性已得到广泛的验证，享有很高的权威性。

3 应用和开发 PhysioNet 的资源

选取研究数据：根据所研究的课题，选取 PhysioBank 包含的相关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。

开发程序实现算法：在电脑中下载和安装 WFDB 后，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，参考和使用 WFDB 应用程序，能大大提高工作效率。C 语言有运行速度快、效率高的优点，Matlab 则有强大、智能的绘图功能，运算功能强大，语法规则简单。从数据库中读取数据进行分析：WFDB Library 里有许多函数能便捷地完成这个任务，如 `isigopen` 打开指定记录名的输入文件，`getvec` 读每一个可用的输入信号的采样点数据。利用 WFDB Library 函数，打开各注释文件，其中用

getann 读取注释中的各项参数, 主体程序示例如下。

```

1 #include <stdio. h>
2 #include <wfdb/wfdb. h>
3 main ( argc, argv)
4 int argc, char *argv []
5 {
6 int i, j, nsig;
7 char *name;
8 WFDB_ Anninfo a;
9 WFDB_ Annotation annot
10 if (argc < 4)
11 {
12 printf (stderr, " Usage: %s record% s annotation by 13
annotator% s \ n", argv [0], argv [1], argv [2]);
14 exit (-1);
15 }
16 name = argv [1];
17 a. name = argv [2]; a. stat = WFDB_ READ;
18 if (annopen (name, &a, 1) < 0)
19 {
20 printf (" fail to Open annotation file \ n");
21 exit (1);
22 }
23 while (getann (0, &annot) == 0)
24 printf ("%s %s %d %d %d \ n", mstimstr (an-
not. time),
annstr (annot. anntyp), annot. subtyp, annot. chan,
annot. num);
25 exit (0)
26 }

```

其中, 行 2: 指明包含 WFDB 的头文件, 提供对 WFDB 库函数的支持。行 8: 定义了 WFDB 格式的注释文件信息结构: 变量 a, 用于打开指定的注释文件进行读/写操作。行 9: 定义了 WFDB 格式的注释结构: 变量 annot, 用于保存从注释文件中读出的注释。行 12: argv [0] 是程序名, argv [1] 是记录名, argv [2] 为指定的注释文件扩展名。行 17: a. name 指定欲打开的注释文件名, a. stat 指定读写操作, 这里定义注释信息结构中的注释状态为读。行 18 - 22: 函数 annopen (记录名, 结构数组, 打开的注释文件数目) 打开指定记录的输入和输出注释文件, 返回值 = 0 时, 成功, 否则提示出错, 退出。行 23: getann (注释文件序号, * 注释结构体指针) 是从输入的注释文件序号指定的注释文件中读下一个注释, 放入由指针指向的注释结构体内, 返

回值为 0 时成功, 返回值小于 0 时出错, 退出。由 annopen 打开一个注释文件, 其注释文件序号为 0, getann 读取该文件中的注释, 写入变量 annot 里保存。行 24: 将注释结构体中的注释参数: 注释时间、类型、子类型、通道和数值打印输出。

应用 C 语言编写分析数据的程序, 并将分析结果以 WFDB 注释的形式生成注释文件保存。生成注释文件时调用 putann 函数形成 WFDB 格式的注释文件。应用 Matlab 编写程序, 读取和应用这些注释文件, 实现自己的研究算法, 例如, 使用接口程序 rd-data. m, 就可以很方便地从数据库中读出数据和注释, 应用于自己的算法。

检测算法: 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分析, 从而整体评价自己算法的性能优劣。

4 结语

PhysioNet 是分析人体生理信号的重要的数据资源, 现包括 50 多个可免费下载的生理信号数据库, 共享处理信号的开放源代码软件, 为协作进行数据分析和新算法评估提供便利, 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。本文取其中的 PhysioBank 数据库中的相关数据作为评价数据集, 利用 PhysioToolkit 软件包, 调用功能强大的库函数, 参考应用程序来开发自己的程序, 读取、检测和分析信号, 并对所研究的算法性能进行整体评价。PhysioNet 信息资源应得到充分的利用。

参考文献

- 1 <http://www.physionet.org> [EB/OL]. [2010 - 01 - 05].
- 2 Moody G B, Mark R G and Goldberger A L. PhysioNet: a web - based resource for the study of physiologic signals [J]. Engineering in Medicine and Biology, IEEE, 2001, 20 (3): 70 - 75.
- 3 Costa M, Moody G B, Henry I, et al. PhysioNet: an NIH research resource for complex signals [J]. Journal of Electrocardiology, 2003, 36 (Suppl): 139 - 144.