

统计学习课程心电数据库的简要说明

2020-11-20

1. 训练集

- 1.1 训练数据放在ECGTrainData文件夹中，一共有10个子文件夹，每个子文件夹的名字代表其中包含心电数据的Label。正常心电图文件夹中包含50条心电记录，其余各文件夹中包含10条心电记录，每条心电记录中包含20个左右心跳。
- 1.2 训练集中的每条心电记录都是已经按照不同心跳切分好。将每条记录load进matlab的工作空间后，会得到一个变量名为Beats的结构体。Beats结构体中有4个域：beatNumber, label, rPeak, beatData。下面的表格对4个域进行了说明：
beatNumber：该条记录中包含的心跳数。
label：与其所在文件夹的名称对应，是该记录中各心跳的共有标签。
beatData：一个cell数组，每个cell是一个 $N \times 12$ 的矩阵，代表了一个心跳的具体数据，其中 N 是该心跳的采样点数，12代表有12导的数据，每导数据代表空间心电向量环在不同方向的投影。
rPeak：一个数组，该数组的第 i 个数据代表第 i 个心跳R波波峰所在的位置。

2. 测试集

测试数据放在ECGTestData文件夹中，共计70条心电记录，与训练集数据不同的是，测试集中的数据并没有按照心跳切分好，每条记录都包含了若干条心跳。将每条记录load进matlab的工作空间后，会得到一个变量名为data，大小为 $N \times 13$ 的数据矩阵，其中 N 为采样点数，第1列的数值从0开始一直到 $N-1$ ，只是一个index，没有特殊含义，从第2列开始一直到13列代表的是12导心电数据。