201703文件夹说明

该文件夹下程序实现的功能如下：

1. 原始数据下载，mingw
2. 将下载的原始波形数据文件（.hea和.dat）转换为.mat格式（convert\_wavedata.sh）
3. 将第一步生成的.mat数据根据头文件中的基线及增益，转换成正常范围内的生理参数值，并按统一的变量顺序进行存储；（loaddata.m）
4. 根据急性低血压的定义，从第二步中按顺序存储的数据中筛选发生急性低血压及未发生急性低血压的样本
5. 原始数据下载

波形数据通过rsync命令下载，该命令将波形数据库复制到本地,且下载过程可以中断，续传。

下载步骤：

打开MinGW/msys/1.0/msys.bat,输入如下命令

mkdir -p /usr/database/mimic2wdb/matched

rsync -CaLvz physionet.org::mimic2wdb-matched /usr/database/mimic2wdb/matched

下载的数据保存在 /usr/database/mimic2wdb/matched文件夹中，即/MinGW/msys/1.0/database/mimic2wdb/matched文件夹中。

1. 数据格式转换（convert\_wavedata.sh）

下载到的数据以头文件(.hea)和数据文件(.dat)的形式表示，且因为单个波形数据及趋势数据的数据量比较大，所以通常将波形数据分成多个部分进行存储，在进行格式转换过程中，只需要根据波形数据(\*.hea)及趋势数据(\*n.hea)进行转换，转换过程中会自动调用各个部分的数据，其中\*的命名规则为：‘病人编号-年-月-日-时-分’。

格式转换流程为：

1. 切换至存储下载数据的文件夹
2. 对每个病人编号对应的文件夹进行循环，找到总的趋势数据头文件，利用wfdb2mat命令生成对应的.mat文件

wfdb2mat -r filename

1. wfdb2mat命令运行后会自动生成’filename’nm.mat和’filename’nm.hea文件，将生成的这两个文件，移动到L：/Available中，病人编号对应的文件夹。
2. 去除基线和增益（loaddata.m）

wfdb2mat转换得到的.mat文件中的数据按行存储，且包含基线和增益，并不是正常的生理参数值。因此，需要根据正常值=（测量值-基线值）/增益，将测量值转换成常规形式的生理参数值。

处理过程：

1. 每一个病人编号对应的文件夹中，找到wfdb2mat自动生成的\*nm.hea文件，从中提取出每个生理参数对应的基线和增益，并将文件名称、基线、增益存储到’\*nm\_hea.mat’文件中，
2. 按照提取出的基线、增益转换数据，并按照指定的顺序存储变量：心率hr、动脉收缩压（ABPSys）、动脉舒张压（ABPDias）、动脉平均压（ABPMean）、脉搏（PULSE）、呼吸（RESP）、血氧（SPO2）、无创收缩压（NBPSys）、无创舒张压（NBPDias）、无创平均压（NBPMean）。
3. 将去除了基线、增益并按顺序存储的数据文件以’\*nm\_select.mat’的名称进行存储，后续从该数据文件中筛选急性低血压病例。
4. 样本筛选(findAHE.m和AHEEpisode.m)

3.1急性低血压样本的筛选

急性低血压的定义：预测窗口内，任意30min或更长时间内，有超过90%的平均动脉压数据小于60mmHg时，认为发生了急性低血压。

在按照上述定义编写筛选程序后，利用challenge项目中给的110个已知样本对程序进行测试，发现会出现误判的情况，为使挑战项目中的110个样本能够全部正确分类，对定义中的有关参数进行调整，调整后的定义为：预测窗口内，任意21min或更长时间内，有超过90%以上的平均动脉压数据小于62mmHg时，认为发生了急性低血压。

筛选步骤：

1. findAHE.m文件对’\*nm\_select.mat’文件中第4列，即平均动脉压数据进行处理，筛选长度为11小时，该函数对全部的平均动脉压数据，以600个数据点为单位，1个数据点为步长，进行移动。每移动一次，提取出600个点之后的1个小时数据，即预测窗口，用’AHEEpisode.m’函数判断预测窗口内的数据是否发生了急性低血压，如果发生了急性低血压，记录4个参数：低血压判别窗的起始位置、判别窗内发生急性低血压的数据段的起始位置、判别窗内发生急性低血压的数据段长度、急性低血压点数占整个数据段的百分比。对于每个’\*nm\_select.mat’文件，记录所有筛选出的急性低血压样本的上述四个参数，选择最后1个记录作为该病人编号对应的急性低血压样本，保存660个点数的数据段，命名为’\*nm\_AHE.mat’，存储到’L:AHE’文件夹中。
2. ‘AHEEpisode.m’函数即按照急性低血压的定义编写代码，判断是否发生了急性低血压。
   1. 未发生急性低血压的病例的筛选

除发生急性低血压的病例外，还需要筛选未发生的病例进行对照。由于未给出非急性低血压病例的定义，所以根据急性低血压的定义，自行设定筛选规则如下：（1）数据长度大于11小时；（2）预测窗口内大于60mmHg的平均动脉压值在70%以上；（3）全部11小时数据中0值所占比例小于5%；（4）全部11小时数据中负值所占比例小于5%

筛选过程：

1. 将’L:/Available’中，发生急性低血压的病例所对应的原始数据文件夹移动移动至’L:/AHEdir’，则’L:/Available’文件夹中剩余的数据即为可能符合未发生急性低血压规则的样本，从中进行筛选。（mvdir.m）
2. 按照非急性低血压病例的筛选规则，从’L:/Available’中筛选样本（selectnono.m）

2018年3月15日星期四 修改

selectAHE.m文件用来对AHE样本重新进行筛选，相比于之前的程序，主要做了如下修改：

1. 将“偏移与增益的处理”和“AHE样本的筛选”分开进行，因为之前已经保存了处理的结果。
2. 筛选过程中，只取第一次发生AHE的数据段：对于同一病人的一段数据记录，可能会存在多个AHE段，之前的程序是将所有的AHE段全部找出，然后最后取的是低于62的血压值的百分比最大的一段（认为是急性低血压最严重的部分），在此次修改过程中，只取了第一次发生急性低血压的数据段，因为同一病人多次发生AHE，可能是由于相同的原因，而且第一次发生后可能已经进行了相应的处理，混杂因素较多。

## 本文件夹中程序运行顺序：

1. Loaddata.m:数据去除基线、增益后按照特定顺序：hr、sbp、dbp、mbp、pulse、resp、spo2，nbp，对数据进行整理排序，存储为”\_selected.mat”形式
2. Age.m:从头文件中提取年龄和性别，缺失的用-100代替，后期从临床数据库中查找补全，临床数据库中也没有的就插值
3. SelectAHE.m:筛选急性低血压样本，并存储到path=‘D:\1yj\_AHE\'中

子程序：

***FindAHE,m***:按照急性低血压的定义，从整个数据记录段中筛选出满足AHE定义的11小时数据段；

***AHEEpisode.m***:判断长度60min的数据段是否满足AHE定义

2019年6月11日星期二

新增

testAHEfor10hour\_20190530.m:该程序主要用于测试之前筛选出的数据段T0之前的10个小时内是否有发生急性低血压的。

FindT0.m:该程序主要用于找出所有可用数据段的T0位置，该T0为第一次发生AHE的时刻。

Loaddata.m:加入了对采样频率的相应处理，对于采样频率为1的数据段，提取具体时刻，例如，15：28：17，15时28分17秒，再提取出17秒，然后频率转换时，从第17个数据记录开始算，即每个实际的1min数据求一个平均值。

Select\_avilable\_20190601.m:

1. 该程序基于findT0.m的运行结果，该运行结果输出一个记录文件，记录的是文件名称、subject\_id、数据段总长度、及第一次发生AHE对应的位置T0。
2. 在此基础上筛选出T0大于600的数据段，即保证T0之前至少有10个小时的数据。然后对于同一subject\_id筛选出多个AHE记录的情况，按照时间排序后，只保留第一次记录的AHE样本。
3. 将发生AHE的subject\_id对应的数据文件夹转移，从剩余的subject\_id对应的文件夹中筛选未发生AHE的样本。
4. 筛选未发生AHE的样本主要用之前的selectnonAHE.m，依据T0之前数据缺失比例以及T0之后大于60mmHg的MAP个数。由于筛选AHE对应的T0时，只保留了第一次检测到的，并且T0大于600的，所以在筛选出的未发生AHE的样本，在T0之前可能发生AHE，所以需要进一步判断并排除。只要在T0之前发生了AHE，不再往后寻找可用数据段，即保证任何时刻，T0之前都没有AHE的发生。
5. 筛选出的未发生AHE的数据段，同样针对同一subject\_id筛选出的多条记录，只按时间先后顺序，保留时间靠前的记录。
6. 将筛选出的可用的AHE样本放在AHE\_2019文件夹中，将筛选出的未发生AHE的样本放在nonAHE\_2019文件夹中。
7. 最后输出一个D:\01袁晶\AHEdata\total\_stat.xlsx文件，共5列，分别未文件名称，subject\_id、数据段总长度、T0位置、以及标签列，将所有AHE及非AHE的样本进行统计。