Data Screening

Nazrul Anuar Nayan

I. OBJECTIVE

The objective of this task is run MatLab scripts on all the uploaded datasets for data drop-outs inspection.

II. COMPLETED TASKS

A. Data sets

A screening for data-drop out has been carried out. Each variables (ECG, PPG and reference RESP) has been checked for the availability (no NaN) of the data. The scripts for the screening has been uploaded into each dataset folder in 'bspprojects9\OURR_results\' of 'dataset names_Screen\Codes'. The screening procedure is done prior to signal quality index (SQI) which will begin next week. The summary of the screening is as shown in Table I. This week, the scripts for extracting the data of Dialysis1,2 and 3 from their original raw data, taken from 'bspprojects7' have also been uploaded into 'dataset names_Screen\Codes' folders . The scripts to extract data of MIMIC, CapnoBase and Fantasia are not available as the extractions were done by Mr. Marco and Mr. Alistair.

1) MIMIC: As shown in the Table I, MIMIC, the data has been differentiated (please refer 'bspprojects9\OURR_results\') accordingly into; (1) the data which include the ECG,PPG and reference Resp in 'EcgPpgResp' folder, (2) data which contain only ECG and reference Resp in 'EcgResp' folder, and (3) data which has PPG and reference Resp in'PpgResp' folder. There are 33 drop-out in MIMIC, either has no value for Resp or, no data for both ECG and PPG.

2) CapnoBase: For CapnoBase all 42 data (8 min recording) have ECG, PPG and RESP information.

3) Fantasia: For Fantasia, we only have 40 ECG and RESP. All data are valid and no drop-outs. No PPG data for these Fantasia dataset.

4) Dialysis1: There are 96 data which have all ECG, PPG and RESP in each of the recorded data.

5) Dialysis2: For Dialysis2, out of 575 data, 574 data has valid ECG, PPG and RESP. Only one data has only ECG,RESP. No drop-outs for this dataset.

6) Dialysis3: After the screening, 373 data are identified to have all 3 variables (ECG,PPG and RESP). Only one drop-out, which has no ECG and RESP.

TABLE I
DATA SCREENING FOR DATA AVAILABILITY

Name	MIMICII	CapnoBase	Fantasia	Dialysis 1	Dialysis 2	Dialysis 3
Data ⇑ 'bspprojects9\OURR':A	1017	42	40	96	575	374
ECG-PPG-RESP :B	949	42	0	96	574	373
ECG-RESP :C	29	0	40	0	1	0
PPG-RESP :D	6	0	0	0	0	0
Drop-out : A-(B+C+D)	33	0	0	0	0	1
'bspprojects9\OURR_results\' ⇑	MIMIC_Screen	CapnoBase_Screen	Fantasia_Screen	Dialysis1_Screen	Dialysis2_Screen	Dialysis3_Screen
ECG Signal	✓	✓	✓	✓	✓	✓
PPG Signal	✓	✓	✗	✓	✓	✓
Reference RESP	✓	✓	✓	✓	✓	✓
Data extraction codes ⇑	✗	✗	✗	✓	✓	✓
Data screening codes ⇑	✓	✓	✓	✓	✓	✓

B. Testing on the RR estimation using 40 MIMIC data

A simple Matlab code has been generated and a testing for RR extraction using 40 good quality signal data of MIMIC has been done. This is more into a learning process of the extraction rather than a proposal to this project. The selection of the 40 data was done manually as the signal quality code is not available. The result of the comparison between ref RESP and extraction of RR using RSA and RPA are as shown in Figure 1. The result only shows the relationship between x and y axis. Using RPA, the result is $y = 1.1x + 0.58$ and using RSA method, $y = x + 1.1$. Further statistical analysis e.g. MSE and R^2 will be carried out. The pulse data and age of the subjects are also included in the graph.

C. Machine Learning Conference

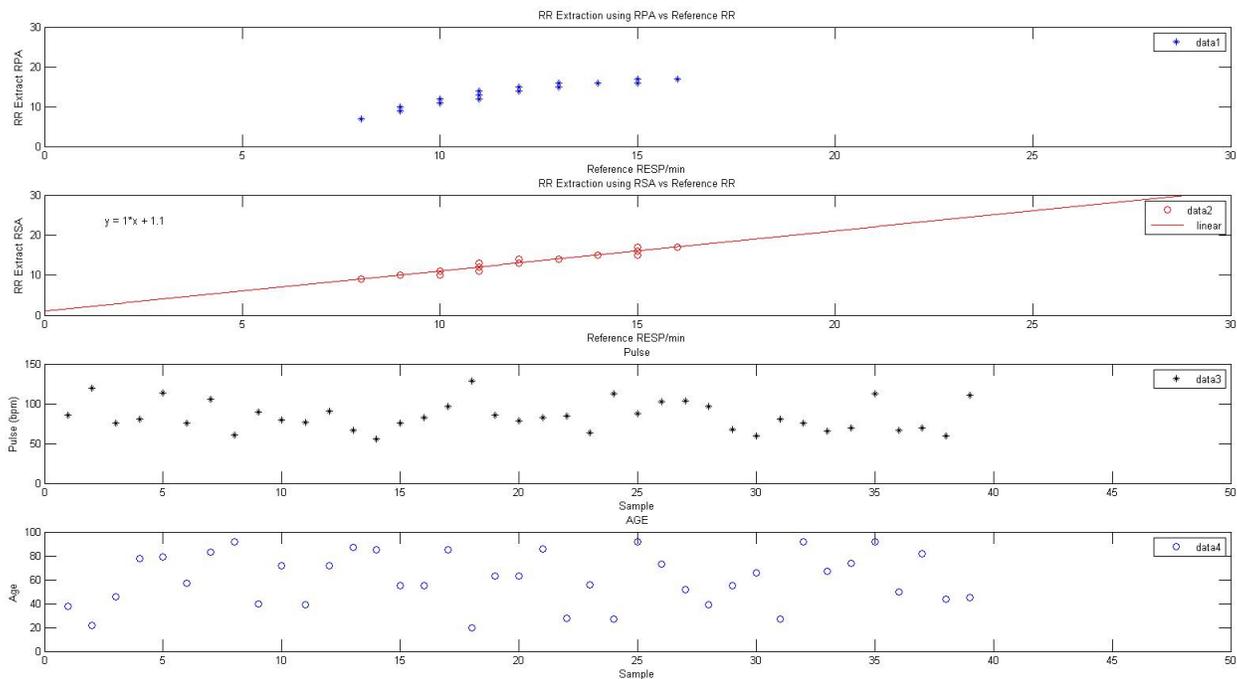The program is pending. The discussion on the final number of speakers will be done prior finalizing the program.

Fig. 1. The result of reference RESP and the extraction using a basic algorithm on 40 MIMIC good signal quality dataset

III. Previous reports

All the progress reports are stored in http://goo.gl/5yASrs
- 20 Feb 2015 - Datasets for respiratory estimation
- 06 Mar 2015 - Format of the Datasets for Respiratory Rate Estimation
- 20 Mar 2015 - Dialysis 1, 2 and 3 data sets
- 01 Apr 2015 - Completion of Dialysis 1 and 2
- 17 Apr 2015 - Data Screening