WESAD Dataset: WEarable Stress and Affect Detection

I. General information

Contact persons: Philip Schmidt and Attila Reiss, contact at: firstname.lastname@de.bosch.com.
If you publish material based on this dataset, please reference the publication [1].

I.1. Dataset structure

The dataset is organised so that each subject has a folder (SX, where X = subject ID). Each subject folder contains the following files:
  - SX_readme.txt: contains information about the subject (SX) and information about data collection and data quality (if applicable)
  - SX_quest.csv: contains all relevant information to obtain ground truth, including the protocol schedule for SX and answers to the self-report questionnaires; see details below
  - SX_respiban.txt: contains data from the RespiBAN device; see details below
  - SX_E4_Data.zip: contains data from the Empatica E4 device; see details below
  - SX.pkl: contains synchronised data and labels; see details below

I.2. Subjects

17 subjects participated in the study. However, due to sensor malfunction, data of two subjects (S1 and S12) had to be discarded. Thus, the respective folders are missing in WESAD. Information on each subject can be found in SX_readme.txt, in the respective subject's folder. Please refer to [1] for overall information on the subjects (see Section 3.1 there).

II. Data format

Raw sensor data was recorded with two devices: a chest-worn device (RespiBAN) and a wrist-worn device (Empatica E4). The study protocol labels (see Section III.1 below) are synchronised with the RespiBAN raw data (same start time). However, the RespiBAN and the Empatica E4 data need to be manually synchronised. Subjects performed a double tapping gesture with their non-dominant hand (where they wore the E4) on their chest. The resulting characteristic pattern in the acceleration signal can be used for synchronising the two devices' data. Moreover, the dataset also includes the file SX.pkl, which includes synchronised raw sensor data and labels, see details in Section II.3 below.

II.1. Data from RespiBAN

The RespiBAN Professional was used: http://www.biosignalsplux.com/en/respiban-professional.
Please refer to [1] for details on sensor placement (see Section 3.2 there). All signals were sampled at 700 Hz. Raw data is contained in SX_respiban.txt. There are 10 columns here. First column: sequential line number. Second column: ignore. Columns 3-10: raw data of the 8 sensor channels. The order of the channels is defined in the header. The entries "XYZ" refer to the 3-channel accelerometer (thus, acceleration data is provided in 3 columns).
In order to convert the raw sensor values into SI units, each channel has to transformed based on the formulas given below (signal contains the raw sensor values, vcc=3, chan_bit=2^{16}).

- ECG (mV): ((signal/chan_bit-0.5)*vcc)
  Details: http://www.biosignalsplux.com/datasheets/ECG_Sensor_Datasheet.pdf
- EDA (µS): (((signal/chan_bit)*vcc)/0.12)
  Details: http://www.biosignalsplux.com/datasheets/EDA_Sensor_Datasheet.pdf

- EMG (mV): ((signal/chan_bit-0.5)*vcc)
  Details: http://www.biosignalsplux.com/datasheets/EMG_Sensor_Datasheet.pdf
- TEMP (°C):
  vout = (signal*vcc)/(chan_bit-1.)
  rntc = ((10^4)*vout)/(vcc-vout)
  - 273.15 + 1./(1.12764514*(10^(-3)) + 2.34282709*(10^(-4))*log(rntc) +
    + 8.77303013*(10^(-8))*(log(rntc)^3))
  Details: http://www.biosignalsplux.com/datasheets/TMP_Sensor_Datasheet.pdf
- XYZ (g): (signal-Cmin)/(Cmax-Cmin)*2-1, where Cmin = 28000 and Cmax = 38000
  Details: http://www.biosignalsplux.com/datasheets/ACC_Sensor_Datasheet.pdf
- RESPIRATION (%): (signal / chan_bit - 0.5) * 100
  Details: http://www.biosignalsplux.com/datasheets/PZT_Sensor_Datasheet.pdf

II.2. Data from Empatica E4

The Empatica E4 was used: http://www.empatica.com/research/e4/. The E4 device was worn on the subjects' non-dominant wrist. Sampling rate of the different sensors was different, see below. Raw data is contained in SX_E4_Data.zip. When unzipped, the following files contain derived information and thus should be ignored in this dataset: HR.csv, IBI.csv, tags.csv. The file info.txt contains some details on the folder's content. Raw data from the E4 device is contained in the following files (in each file, first line refers to the sensor channel's global timestamp at start, second line refers to the sensor channel's sampling rate):

- ACC.csv: sampled at 32 Hz. The 3 data columns refer to the 3 accelerometer channels. Data is provided in units of 1/64g.
- BVP.csv: sampled at 64 Hz. Data from photoplethysmograph (PPG).
- EDA.csv: sampled at 4 Hz. Data is provided in µS.
- TEMP.csv: sampled at 4 Hz. Data is provided in °C.

II.3. Synchronised data

The double-tap signal pattern was used to manually synchronise the two devices' raw data. The result is provided in the files SX.pkl, one file per subject. This file is a dictionary, with the following keys:

- 'subject': SX, the subject ID
- 'signal': includes all the raw data, in two fields:
  - 'chest': RespiBAN data (all the modalities: ACC, ECG, EDA, EMG, RESP, TEMP)
  - 'wrist': Empatica E4 data (all the modalities: ACC, BVP, EDA, TEMP)
- 'label': ID of the respective study protocol condition, sampled at 700 Hz. The following IDs are provided: 0 = not defined / transient, 1 = baseline, 2 = stress, 3 = amusement, 4 = meditation, 5/6/7 = should be ignored in this dataset

III. Ground truth

All relevant information can be found in SX_quest.csv, where X = subject ID. One can use either the study protocol conditions as labels, or the answers to the self-report questionnaires.

III.1. Study protocol

The order of the different conditions is defined on the second line in SX_quest.csv. Please refer to [1] for further details on each of the conditions (see Section 3.3 there). Please ignore the elements "bRead", "fRead", and "sRead": these are not relevant for this dataset.

The time interval of each condition is defined as start and end time, see the lines 3 and 4 in SX_quest.csv. Time is given in the format [minutes.seconds]. Time is counted from the start of the RespiBAN device's start of recording.

III.2. Self-report questionnaires

Within the study protocol, after each of the five defined conditions (baseline, amusement, stress, meditation 1, meditation 2), the subjects were asked to fill in a self-report. The self-reports consist of the following questionnaires: PANAS, shortened STAI, SAM (Self-Assessment Manikins, for valence and arousal). Answers are provided on the respective lines in SX_quest.csv. The order of the questionnaires (e.g. the five PANAS-questionnaire lines) is the same as the order of the different conditions. Additionally, after the stress condition, a shortened SSSQ-questionnaire is filled in by the subjects, see the last line in SX_quest.csv. Please refer to [1] for further details on the self-report questionnaires (see Section 3.4 there).

In the following, the items of each questionnaire are listed, in the order as stored in SX_quest.csv. For each questionnaire, the answering options are given as well. Considering the PANAS-questionnaire, four items (Stressed, Frustrated, Happy, Sad) were added by us, see explanation in [1]. These items were scored by the subjects using the same scale as all other PANAS items. Considering the SSSQ-questionnaire: this consists of nine items, but the first three items (Annoyed, Angry, Irritated) are included in the PANAS-questionnaire. The item Annoyed is by default part of PANAS, while the items Angry and Irritated (in brackets below) are only asked after the stress condition, since only relevant for the SSSQ.

PANAS questionnaire items (1 = Not at all, 2 = A little bit, 3 = Somewhat, 4 = Very much, 5 = Extremely)

- Active
- Distressed
- Interested
- Inspired
- Annoyed
- Strong
- Guilty
- Scared
- Hostile
- Excited
- Proud
- Irritable
- Enthusiastic
- Ashamed
- Alert
- Nervous
- Determined
- Attentive
- Jittery
- Afraid
- Stressed
- Frustrated
- Happy

- (Angry)
- (Irritated)
- Sad

STAI questionnaire items (1 = Not at all, 2 = Somewhat, 3 = Moderately so, 4 = Very much so)

- I feel at ease
- I feel nervous
- I am jittery
- I am relaxed
- I am worried
- I feel pleasant

SAM questionnaire items (scale 1-9)

- Valence (1 = low valence, 9 = high valence)
- Arousal (1 = low arousal, 9 = high arousal)

SSSQ questionnaire items (1 = Not at all, 2 = A little bit, 3 = Somewhat, 4 = Very much,
5 = Extremely):

- I was committed to attaining my performance goals
- I wanted to succeed on the task
- I was motivated to do the task
- I reflected about myself
- I was worried about what other people think of me
- I felt concerned about the impression I was making

References

[1] Philip Schmidt, Attila Reiss, Robert Duerichen, Claus Marberger and Kristof Van Laerhoven. 2018. Introducing WESAD, a multimodal dataset for Wearable Stress and Affect Detection. In 2018 International Conference on Multimodal Interaction (ICMI '18), October 16–20, 2018, Boulder, CO, USA. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3242969.3242985