Short description

The goal of this challenge is to predict the age of subject from his brain activity during deep sleep and the sequence of his sleep stages over a night.

Challenge context

Brain activity is of particular interest to study sleep and its disorders. Sleep doctors have proposed to distinguish different sleep stages with respect to specific temporal and frequency patterns. They have elaborated scoring rules [1, 2] in order to assign each 30s sleep sample with a sleep stage among: Wake, N1, N2, N3 (deep sleep), REM (paradoxical sleep). One can thus assign to each night a sequence of sleep stages which is called a hypnogram. It represents the sleep trends of a subject during a night.

Sleep scorers take into account several patterns which are known to present some variability due to age, [3, 4, 5]. Deep sleep (also referred to as Slow Wave Sleep) characteristics seem to exhibit particular dependencies in age. Deep sleep time period is reduced and more fragmented with aging.

[1] Iber C, Ancoli-israel S, Chesson A, Quan SF, A. A. of S. M. (2007). The AASM manual for the scoring of sleep and associated events; rules, terminology and technical specifications (2007).

[2] Rechtschaffen, A., & Kales, A. (1968). A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects.

[3] Ohayon, M. M., Carskadon, M. a, Guilleminault, C., & Vitiello, M. V. (2004). Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. Sleep, 27(7), 1255–73.

[4] Novelli, L., Ferri, R., & Bruni, O. (2010). Sleep classification according to AASM and Rechtschaffen and Kales: Effects on sleep scoring parameters of children and adolescents. Journal of Sleep Research, 19(1 PART. 2), 238–247.

[5] Dijk, D.-J., Groeger, J. a, Stanley, N., & Deacon, S. (2010). Age-related reduction in daytime sleep propensity and nocturnal slow wave sleep. Sleep, 33(2), 211–223.

Challenge goals

We propose to deeper investigate the relation between sleep characteristics and aging through a regression age problem. More precisely, we want to predict the age of a subject given (a) 5 minutes of EEG record during deep sleep and (b) its hypnogram. The performance metric used is the mean average percentage error (MAPE) (https://en.wikipedia.org/wiki/Mean\_absolute\_percentage\_error)

Data description

3 .csv files are provided:

- a training input file

- a training output file

- a testing input file

The separator used is ";".

The training and testing input files have the following keys:

- ID: id of the sample. Data samples are recorded onto different subjects which may be a source of variability in the EEG data. ID ranges from 0 to 580

- DEVICE: device used to record the EEG data. The EEG signals come from 2 devices (0 or 1). The signals are recorded from the same channels but may exhibit a change in the gain value of the amplifier of the device.

- EEG\_0 … EEG\_74999: this corresponds to a continuous 5min period of EEG signal (derivation Fp2 - A2), sampled at 250Hz, during deep sleep.

- HYPNOGRAM: list of variable length corresponding to the hypnogram associated to the deep sleep EEG sample given. The values are 0 (for Wake), 1 (for N1), 2 (N2), 3 (Deep sleep - N3), 4 (REM), -1 (when scoring was not possible).

Please find below an overview of the first lines of the training input file:

ID DEVICE EEG\_0 EEG\_1 EEG\_2 EEG\_3 EEG\_4 \

0 0.0 18.513157 17.887674 17.442404 16.944677 16.080168

1 0.0 3.377115 3.720635 4.122524 4.086743 3.476264

2 0.0 20.753752 33.122223 41.263256 44.764065 45.441994

3 0.0 14.535362 14.571295 14.288245 13.808476 13.219709

4 1.0 0.000006 0.000006 0.000006 0.000009 0.000008

EEG\_74999 HYPNOGRAM

0 -9.003139 ['0', '0', '0', '0', '0', '0', '0', '0', '0', ...

1 4.335758 ['0', '0', '0', '0', '0', '0', '2', '0', '0', ...

2 -4.994258 ['0', '0', '0', '0', '0', '0', '0', '0', '0', ...

3 2.302796 ['0', '0', '0', '0', '0', '0', '0', '0', '0', ...

4 0.000022 ['0', '0', '0', '0', '0', '0', '0', '0', '0', ...

The training output file has the following keys:

- ID: id of the sample. ID ranges from 581 to 830

- TARGET: age of the subject correspond to this sample

Here after are displayed the first lines of the training output file

ID TARGET

0 32

1 29

2 36

3 56

4 60

The training set has 581 samples. The testing set has 249 samples.

You must provide a file which has the same format as the training output, with ID's ranging from 581 to 829