

Machine learning challenge for monitoring cognitive load

Description

A machine learning challenge for detecting cognitive load, with results to be presented at the Ubintention workshop at UbiComp 2020, Cancun, 12 - 13th September 2020.

The goal of this machine learning challenge is to recognize 2 levels of cognitive load from a wearable wrist-band (Microsoft Band 2). The participants will have to develop a machine-learning pipeline that will process the sensor data, create models and output the recognized cognitive load. The best three teams will also receive prizes.

Prizes

1. 1000 EUR
2. 650 EUR
3. 350 EUR

*Note: Prizes may increase subject to additional sponsors.

Challenge Deadlines

- **15.03.2020** - Labelled training data is provided
- ~~07.04.2020~~ **(1.06.2020)** Participants registration deadline
- **23.04.2020 (already provided)** - Unlabelled test data is provided
- ~~18.05.2020~~ **19.06.2020** - Submission deadline: predictions + workshop poster
- ~~01.06.2020~~ **3.07.2020** - Review notification
- ~~15.06.2020~~ **17.07.2020** - Camera-ready poster submissions
- **12.09.2020** - Top N (max 5) teams present their work at the workshop
- **20.09.2020** - Release of the ground-truth of the test data

Registration

Each team should send a registration email to **cogload.challenge@gmail.com** as soon as possible but no later than 07.04.2020 (new deadline 1.06.2020). The email should include the following information:

- The name of the team
- The names of the participants in the team
- The organization/company (individuals are also encouraged)
- The contact person with email address

Submission of predictions on the test dataset

The participants should submit a csv file (e.g. "teamName_predictions.csv") containing the predictions for the test dataset. The structure of the file should be the same as the file **test/exampleTeam_predictions.csv**, i.e., the first column contains the predictions and the second column contains the user_id. The number of lines (rows) is equal to the number of instances. Thus, each team should only modify the first column of this file. The user_ids are provided for personalization purposes.

The participants' predictions should be submitted online by sending an email to cogload.challenge@gmail.com. Services such as Dropbox, Google Drive, can be also used.

To be part of the final ranking, participants **will be required to submit a poster** at the Ubittention workshop at UbiComp 2020. **The date for the poster submission is 18.05.2020 (new deadline 19.06.2020)**.

Two submissions are allowed per team, however, **it is required that the team presents 2 posters** for Ubittention and the posters should be substantially different (>70%). This can be the case if the same team is using 2 different approaches to tackle the problem.

Dataset description and format

The dataset contains data from 23 participants, 5 of them for testing and 18 for training. The train and test data was generated using windows of 30 seconds.

The **train data** contains the following files:

1. gsr_train.csv - Galvanic skin response sampled at 1Hz.
2. hr_train.csv - Heart Rate sampled at 1Hz.
3. rr_train.csv - RR intervals resampled at 1Hz.
4. temp_train.csv - Skin temperature sampled at 1Hz.
5. labels_train.csv - The first column contains the data labels (0 - cognitive load, 1-resting).
The second column contains the user_id.

The **test data** contains the following files:

1. gsr_test.csv - Galvanic skin response sampled at 1Hz.
2. hr_test.csv - Heart Rate sampled at 1Hz.
3. rr_test.csv - RR intervals resampled at 1Hz.
4. temp_test.csv - Skin temperature sampled at 1Hz.
5. exampleTeam_predictions.csv - first column contains example predictions and it should be modified by the participants. The second column contains the user_id.

The sensor files in the training data (*gsr*, *gr*, *rr* and *temp*) contain 632 lines x 30 columns, corresponding to 632 instances each containing 30 samples (30 seconds at the sampling rate 1 Hz). The train instances are randomly permuted.

The sensor files in the test data (*gsr*, *gr*, *rr* and *temp*) contain 193 lines x 30 columns, corresponding to 193 instances each containing 30 samples (30 seconds at the sampling rate 1 Hz). The test instances are randomly permuted.

***Note:** Additional preprocessing steps (e.g., filtering) were not used in the preparation of the dataset, besides the sampling/resampling. Thus, the data is raw, as provided by the MicrosoftBand2.

Download train data

<https://www.dropbox.com/sh/om48sxy6c9cng8/AADk44tan6HuPPQaa7gYfPf-a?dl=0>

Download test data

<https://www.dropbox.com/sh/0i6to1k0myo03yu/AACj1cltoFGSJ58F-06g55Xka?dl=0>

Evaluation

Accuracy will be used as a metric for evaluation. The final evaluation will be performed at the end of the competition.

Rules

- The organizers reserve the right to perform code inspection after the submission. In case of a denial, the participants may be disqualified.
- Approaches that are taking into account the timeline of the data, i.e., the temporal dependence among the labels are not allowed, as it may mirror the experimental scenario and it would defeat the purpose of the challenge. For clarification, exploiting temporal dependence within instances, i.e., within the 30 seconds windows is allowed. Also, using User_ids for personalization is allowed.
- If an unforeseen or unexpected event (including, but not limited to: someone cheating; a virus, bug, or catastrophic event corrupting data or the submission platform; someone discovering a flaw in the data or modalities of the challenge) that cannot be reasonably anticipated or controlled (also referred to as force majeure) and which affects the fairness and/or integrity of this contest, we the organizers the right to cancel, change or suspend this contest at any time.

Eligibility

- You do not work/collaborate with any of the challenge organizers for the development of solutions for the challenge.
- If you submit an entry, but are not qualified to enter the contest, this entry is voluntary. The organizers reserve the right to evaluate it for scientific purposes. If you are not qualified to submit a contest entry and still choose to submit one, under no circumstances will such entries qualify for sponsored prizes.

Entry

- Registration (see above): as soon as possible but not later than 01.04.2020 (new deadline **1.06.2020**).
- Challenge: Participants will submit prediction results on test data.
- Workshop poster: To be part of the final ranking, participants will be required to submit a poster explaining the approach.
- Submission: The participants' predictions should be submitted online by sending an email to shldataset.challenge@gmail.com.
- Two submissions are allowed per team, provided that the team submits 2 posters and the posters should be substantially different (>70%).

Winners announcement

The final results and ranking of the teams will be announced at the Ubintention workshop at the UbiComp 2020.

Contact

All inquiries should be directed to: cogload.challenge@gmail.com

Challenge Organizers:

- Dr. Veljko Pejovic - University of Ljubljana, Slovenia
- Dr. Matjaž Gams - Jozef Stefan Institute, Slovenia
- Dr. Jens Garbas - Fraunhofer IIS, Germany
- Dr. Mitja Luštek - Jozef Stefan Institute, Slovenia
- Dr. Hristijan Gjoreski - Ss. Cyril and Methodius, N. Macedonia
- Tine Kolenik - Jozef Stefan Institute, Slovenia
- Martin Gjoreski - Jozef Stefan Institute, Slovenia
- Dominik Seus - Fraunhofer IIS, Germany



