Towards Robust Neuroadaptive HCI: Exploring Modern Machine Learning Methods to Estimate Mental Workload From EEG Signals

Aurélien Appriou
Inria, LaBRI (CNRS/INP/Univ. Bordeaux), RIKEN BSI
Talence, France / Wakoshi, Japan
aurelien.appriou@inria.fr

Andrzej Cichocki
SKOLTECH, RIKEN BSI
Moscow, Russia / Wakoshi Japan
a.cichocki@riken.jp

Fabien Lotte
Inria, LaBRI (CNRS/INP/Univ. Bordeaux), RIKEN BSI
Talence, France / Wakoshi, Japan
fabien.lotte@inria.fr

CHI'18 Extended Abstracts, April 21–26, 2018, Montreal, QC, Canada
ACM 978-1-4503-5621-3/18/04.
https://doi.org/10.1145/3170427.3188617

Abstract

Estimating mental workload from brain signals such as Electroencephalography (EEG) has proven very promising in multiple Human-Computer Interaction (HCI) applications, e.g., to design games or educational applications with adaptive difficulty, or to assess how cognitively difficult to use an interface can be. However, current EEG-based workload estimation may not be robust enough for some practical applications. Indeed, the currently obtained workload classification accuracies are relatively low, making the resulting estimations not fully trustable. This paper thus studies promising modern machine learning algorithms, including Riemannian geometry-based methods and Deep Learning, to estimate workload from EEG signals. We study them with both user-specific and user-independent calibration, to go towards calibration-free systems. Our results suggested that a shallow Convolutional Neural Network obtained the best performance in both conditions, outperforming state-of-the-art methods on the used data sets. This suggests that Deep Learning can bring new possibilities in HCI.

Author Keywords

Mental Workload, Neuroadaptive technology, neuroergonomics, EEG, Machine Learning, Deep Learning, Brain-Computer Interfaces

ACM Classification Keywords

Computing methodologies [Machine learning]: Machine learning algorithms; Human-centered computing [Human computer interaction (HCI)]: Empirical studies in HCI

Introduction

Brain-Computer Interfaces (BCIs) enable their users to interact with computers by using brain activity only, usually measured with Electroencephalography (EEG) [4]. EEG consists in placing electrodes on the scalp to record electrical signals produced by the brain. BCIs are notably promising for medical and assistive technologies applications [4]. For example, it can enable people with severe motor impairments to send commands to a wheelchair through brain activity only, e.g., by imagining left or right hand movements to make the wheelchair turn left or right [4]. Such BCIs are called active BCIs since users are actively sending commands to the system, here a wheelchair, by performing mental imagery tasks [17].

There is another type of BCI that proved particularly promising for Human-Computer Interaction (HCI) research and application: passive BCIs [17]. Such BCIs are not used to directly control an application, but to monitor in real-time users' mental states, in order to assess or to adapt an interface or application. Passive BCIs that are able to estimate mental workload, i.e., the amount of cognitive resources currently engaged by users in a task, appeared as particularly useful. They can notably be used to assess how cognitively difficult the manipulation of a given input device is [6]. For instance, such Passive BCIs were used to study mental workload during navigation tasks with different input devices [6], during visualization tasks [11] or during plane piloting [7]. Workload estimation was also used to design applications that dynamically adapt to the users' states, for instance to create video games with adaptive difficulty [5], to provide an optimal sequence of teaching exercises adapted to the cognitive capabilities of each learner [16], or to enable users to visualize and reflect on their own mental workload levels [8]. Despite these many promising HCI applications, in practice it is difficult to estimate mental workload reliably from brain signals, in particular from EEG signals, over time, contexts and users [10]. For instance in [10], discriminating low from high workload in oscillatory EEG activity was possible with a classification accuracy of only about 69%. There is thus a need for more robust algorithms to classify EEG signals, i.e., with higher classification accuracy, in order to obtain trustable EEG-based mental workload estimators. Therefore, we propose here to study and compare various recent machine learning algorithms for EEG-based workload classification.

We studied algorithms that proved efficient either in recent active BCI classification competitions [1, 15], or in other independent studies or fields [13]. Note that such algorithms have only been explored and compared for EEG classification of motor tasks, but not yet for workload estimation. As baseline, we use two standard methods for workload classification: 1) Common Spatial Pattern (CSP) spatial filters with a Linear Discriminant Analysis (LDA) classifier [3] and 2) the Filter Bank CSP (FBCSP) [1], which is an improved variant of the CSP. Then, we studied Riemannian approaches. Such methods represent EEG signals as covariance matrices and classify them according to their (Riemannian) distances to prototypes of covariance matrices for each class. Such methods have recently won 6 international brain signals competitions [15]. Finally, we used a Convolutional Neural Network (CNN), a Deep Learning algorithm which recently obtained promising results for many machine learning problems [9]. CNN strength is to optimize simultaneously the spatial filters, the temporal filters and the classifier, which can lead to possibly better solutions.

In this paper, we first present the workload EEG data set used for evaluation, before introducing each machine learning algorithm details. We perform two distinct evaluation studies: the first study is a user-specific study, meaning that algorithms are trained on data specific to each user, and are then evaluated on other data recorded from this same user. This is the standard way current BCI are designed, given the large between user-variability [3]. The second study is a user-independent one, where algorithms are trained on all data recorded from all users except that of the target one, on which algorithms are tested. This is much more challenging, but if successful, would enable workload monitoring without requiring any calibration for new users.

Methods

Data Set

The data set used comes from [10], and was designed to perform realistic workload classification, with various users affective states. Indeed, this experiment involved two contextual changes: cognitive workload variations and social stress variations. Signals from 28 EEG electrodes were recorded from 22 users [10]. To involve cognitive workload variations, we used N-back tasks. With this task, users saw a sequence of letters on screen, the letters being displayed one by one, every 2 seconds. For each letter the user had to indicate with a mouse click whether the displayed letter was the same one as the letter displayed N letters before. Users alternated between easy blocks with the 0-back task (the user had to identify whether the current letter was the letter 'X') and difficult block with the 2-back task (the user had to identify whether the current letter was the same letter as the one displayed 2 letters before). Each segment of 2 seconds was used as a trial, and was labeled with its workload level (low or high). Note that 2 sec is a very short time for workload classification in EEG, and that classifying longer trials would be easier, but at the expense of a lower reactivity [10]. Here, it enables us to compare algorithms in challenging conditions. 720 trials were available for each class and user.

As introduced previously, we performed both a user-specific calibration study and a user-independent one. For the user-specific calibration, the first half of the trials of each user was used as the training set for that user, and the second half as the testing set. Note the two halves were recorded in two different contextual conditions: half the users started in a relaxed state for the first half, and were then exposed to social stress for the second half using the Trier Social Stress test (see [10] for details). This order was reversed for the remaining users. This makes classification much more challenging, since the test set EEG patterns are not exactly the same as those from the training set. However, this is also more realistic for practical use. For the user-independent calibration, the training set was composed of all trials (both conditions of social stress) of all users except the current user used for testing (i.e., around 21*1440 = 30240 trials). The testing set of each user was the same testing set as with user-specific calibration, i.e., the second half of the trials (720 trials) from this user. This made comparing performances between both calibration types possible.

Machine Learning Algorithms explored

Common Spatial Patterns

The common spatial patterns (CSP) algorithm optimizes spatial filters, i.e., a linear combination of the original EEG signals, in order to improve the EEG signal-to-noise ratio. It is done such that the variance of spatially filtered signals is maximized for one class and minimized for the other class. However, this algorithm requires EEG signals to be band-pass filtered in a carefully chosen narrow frequency band, adapted to the specificity of the data analyzed. The Alpha rhythm (8-12Hz) is known to vary according to workload variations [10]. Thus, we applied CSP after band-pass filtering in 8-12 Hz. We selected 3 pairs of CSP spatial filters, as recommended in [3], to obtain 6 band-power features used to train a Linear Discriminant Analysis (LDA) classifier. This method was used as the baseline.

Filter Bank Common Spatial Patterns
The Filter Bank Common Spatial Patterns (FBCSP) algorithm proved its efficiency when winning the Fifth International BCI competition on all EEG data sets [1]. This algorithm optimizes spatial and spectral filters at the same time. To do so, the FBCSP first filters EEG signals into multiple frequency bands using a filter bank. Here we used 9 band-pass filters in 4Hz-wide bands (in 4-8 Hz, 8-12 Hz, . . . , 36-40 Hz) as in [1]. For each band-passed signals, CSP is used to optimize 2 spatial filter pairs. From the resulting 36 features (9 bands $\times$ 4 CSP filters per band), the 4 most relevant ones were selected using mutual information feature selection [12], and then used to train an LDA.

Riemannian Geometry
Riemannian approaches recently proved very promising, as they won 6 international brain signals classification competitions, see, e.g., [15, 2]. Riemannian methods represent EEG signals as covariance matrices and manipulate them with an appropriate geometry: the Riemannian geometry. Here, we represented each trial by the EEG signals covariance matrix after band-pass filtering in the alpha band (8-12 Hz), as for CSP. To design the classifier, we simply computed the average covariance matrix for each class (low vs high workload). They were to be used as class prototypes. To apply this classifier, we computed the Riemannian distance between the test trial covariance matrix, and each of these two class prototypes. The estimated class label was defined as that of the closest class prototype. In order to make the covariance matrices more discriminant, we also used Geodesic filtering, as in [15, 2].

Convolutional Neural Networks
Deep Learning are artificial neural networks with multiple layers of artificial neurones, which makes them able to approximate efficiently any function, and thus to learn virtually any task [9]. Deep Learning with convolutional neural networks (CNN) has already improved many fields, such as computer vision [9]. A very recent study presented a new type of CNN dedicated to motor task classification in EEG: the Shallow ConvNet [13]. The two first layers of this CNN perform a temporal and a spatial convolution respectively, while the following layer is a mean pooling one. Finally, a logarithmic activation function is used followed by a final fully connected layer. Thus, this CNN performs signals operations similar to that of the FBCSP. In contrast to FBCSP, all these filters are optimized simultaneously, which made it outperform the FBCSP on motor EEG signals [13]. The Shallow ConvNet uses minimally preprocessed EEG signals, so we filtered them in 4-40 Hz.
We compare these methods using classification accuracy as performance measure, i.e., the percentage of trials correctly classified in the test set.

Results

User-specific calibration study
The performances obtained by each algorithm on each user with user-specific calibration is reported on Figure 1. Paired t-tests used to compare each algorithm (FBCSP, Riemmanian geometry and CNN) to the baseline (CSP) $(mean\ accuracy\ =\ 67.0\%\pm 8.1)$ revealed that only the CNN $(mean\ =\ 72.7\%\pm 9.1)$ obtained significantly better classification performances than the CSP method $[t(1,22)=-2.189, p=0.034]$. However, the CNN did not obtain significantly better performances than the FBCSP

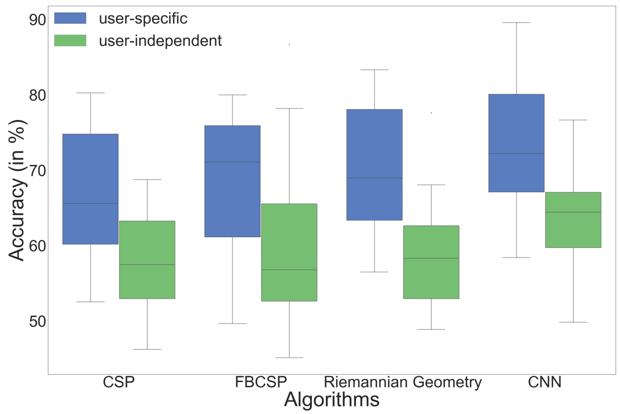

Figure 1: Classification performances of each algorithm for both user-specific and user-independent studies.

$(mean = 68.5\% \pm 8.9)$ nor the Riemannian Geometry method $(mean = 69.9\% \pm 8.1)$.

User-independent calibration study

The results of the user-independent calibration study are reported on Figure 1. We used paired t-tests to compare algorithms. As in the user-specific study, only the CNN $(mean = 63.7\% \pm 7.7)$ obtained significantly better performances than the baseline CSP method $(mean = 58.0\% \pm 6.7)$ $[t(1, 22) = -2.623, p = 0.012]$, but the FBCSP method $(mean = 60.1\% \pm 10.6)$ and the Riemannian Geometry method $(58.4\% \pm 7.1)$ did not (respectively $[t(1, 22) = -0.761, p = 0.451]$ and $[t(1, 22) = -0.179, p = 0.859]$). However, this time the CNN performed significantly better than the Riemannian Geometry method $[t(1, 22) = -2.377, p = 0.022]$, but not than the FBCSP method $[t(1, 22) = -1.313, p = 0.197]$.

Discussion and conclusion

We presented a comparison of 4 modern machine learning algorithms in order to compare EEG-based workload level classification performances, with both user-specific and user-independent calibration. Our results suggested that the CNN can obtains better performances than CSP methods in classifying two workload levels (low vs high), for both user-specific and user-independent studies. CNN could thus be used to perform more trustable neuroergonomics and neuroadaptive HCI. Obtaining reasonable performances in a user-independent calibration from only 2 sec of EEG data and only 21 users for calibration, makes the CNN particularly promising to design calibration-free neuroadaptive technologies in the future. It would also be interesting to study whether this CNN can be used to estimate robustly other cognitive states such as curiosity, attention, etc.

Our evaluations were also performed with a major difficulty for the algorithms: as mentioned, the training and testing sets differed regarding the user's affective state. For one data set the user was under social stress but not for the other. Such context changes were useful to validate the algorithms robustness to change likely to occur during real-life use. While we can easily imagine that all algorithms would have obtained better accuracies without this social stress change, the performance obtained would have been less realistic. Deep Learning in general, and CNN in particular have been a substantial improvement in artificial intelligence, leading to performance leaps on many problems [9]. Here we showed that CNN may also be useful for HCI, to perform mental state classification from EEG. In the future, other deep learning algorithms may prove promising for EEG classification as well. For example, Recurrent Neural Networks (RNN) have obtained promising performances in Natural Language Processing and videos classification

[14], and may thus prove useful for EEG-based HCI as well. Finally, we will use this CNN or other deep networks online to perform real-time adaptive training and neuroergonomics.

Acknowledgements:
This work received support from the European Research Council (grant ERC-2016-STG-714567) and the Japanese Society for the Promotion of Science (JSPS).

REFERENCES

1. KK Ang, ZY Chin, CC Wang, CT Guan, and HH Zhang. 2012. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Frontiers in Neuroscience* 6 (2012), 1–9.
2. A Barachant, S Bonnet, M Congedo, and C Jutten. 2012. Multiclass brain–computer interface classification by Riemannian geometry. *IEEE Transactions on Biomedical Engineering* 59, 4 (2012), 920–928.
3. B Blankertz, R Tomioka, S Lemm, M Kawanabe, and KR Müller. 2008. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Sig Proc Mag* (2008).
4. M Clerc, L Bougrain, and F Lotte. 2016. *Brain-Computer Interfaces 2: Technology and Applications*.
5. S Fairclough. 2008. BCI and physiological computing for computer games: Differences, similarities & intuitive control. In *Proc ACM CHI*.
6. J Frey, M Daniel, M Hachet, J Castet, and F Lotte. 2016. Framework for electroencephalography-based evaluation of user experience. In *Proc ACM CHI*.
7. T Gateau, G Durantin, F Lancelot, S Scannella, and F Dehais. 2015. Real-time state estimation in a flight simulator using fNIRS. *PloS one* 10, 3 (2015).
8. R Gervais, J Frey, A Gay, F Lotte, and M Hachet. 2016. TOBE: Tangible Out-of-Body Experience. In *Proc TEI'16*. 227–235.
9. Y Lecun, Y Bengio, and G Hinton. 2015. Deep learning. *Nature* 521, 7553 (2015), 436–444.
10. C Mühl, C Jeunet, and F Lotte. 2014. EEG-based workload estimation across affective contexts. *Frontiers in Neuroscience* 8 (2014), 1–15.
11. EMM Peck, BF Yuksel, A Ottley, RJK Jacob, and R Chang. 2013. Using fNIRS brain sensing to evaluate information visualization interfaces. In *Proc ACM CHI*.
12. BC Ross. 2014. Mutual information between discrete and continuous data sets. *PLoS ONE* 9, 2 (2014).
13. RT Schirrmeister, JT Springenberg, LDJ Fiederer, M Glasstetter, K Eggensperger, M Tangermann, F Hutter, W Burgard, and T Ball. 2017. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping* (2017).
14. Y Yang, D Krompass, and V Tresp. 2017. Tensor-Train Recurrent Neural Networks for Video Classification. (2017). `http://arxiv.org/abs/1707.01786`
15. F Yger, M Berar, and F Lotte. 2016. Riemannian approaches in Brain-Computer Interfaces: a review. *IEEE Trans Neur Sys Rehab* 25, 10 (2016).
16. BF Yuksel, KB Oleson, L Harrison, EM Peck, D Afergan, R Chang, and RJK Jacob. 2016. Learn piano with BACh: An adaptive learning interface that adjusts task difficulty based on brain state. In *Proc ACM CHI*.
17. TO Zander and C Kothe. 2011. Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. *J Neur Eng* 8 (2011).