

Machine Learning of Multi-channel Electroencephalographic Data*

Aurora Saibene¹

Department of Informatics, Systems and Communication, University of Milano -
Bicocca, Milan, Italy
a.saibene2@campus.unimib.it

Abstract. Machine Learning techniques have been recently applied in the healthcare field and particularly for electroencephalographic signal classification, opening new possibilities for brain activities and diseases analysis through peculiar applications like the Brain Computer Interfaces.

The project proposal for the Ph.D. thesis work briefly described in the following wants to address the problems arising from these biomedical heterogeneous data, starting from the preliminary signal processing for noise removal, moving to possible data normalisation for subject and population based analysis and exploiting the outputted manipulated data to create classifiers for peculiar brain activities labelling, diseases identification, Brain Computer Interface development.

These steps will require an evaluation of the state-of-the-art, which present mostly semi-automatic or manual signal processing techniques, that will be used to create fully automated denoising modules for every type of data and integrated for scenario-dependent signal reconstruction procedures. Also, there is a narrow number of studies addressing the normalisation problem, which is to be considered for population-based analysis. Finally, the recent works on electrophysiological signal classification will be used to evaluate commonly used Machine Learning algorithms and to create best-practices for feature extraction, a benchmark for deep learning techniques application and the study of Brain Computer Interface mainly for rehabilitation purposes.

Keywords: Brain Computer Interface · Deep learning · Electroencephalogram · Machine Learning · Signal processing.

1 Introduction

In the last decades the constant technological improvement and the availability of a greater amount of data have led to an increasing interest over the application of Machine Learning (ML) techniques in the biomedical field, posing new challenges for the development of faster and more accurate classification algorithms.

* Main advisor: Francesca Gasparini, Department of Informatics, Systems and Communication, University of Milano - Bicocca, Milan, Italy.

The project proposal for the Ph.D. thesis work described in the following, wants to address the problems arising from the heterogeneity of a peculiar kind of biomedical data, i.e. the Electroencephalographic (EEG) signals.

Recently, EEG has begun to be extensively used in the medical and research fields due to its characteristics: it is non-invasive, records the cerebral bio-electric potentials through multiple sensors (electrodes) placed on the scalp [12], has a high temporal resolution [8]. The recording produces a multi-channel signal, i.e. the EEG data structure is usually in the form of a matrix, whose rows correspond to the sensors and the columns to the electric potential recorded in a specific time.

The EEG has been used, for example, in face recognition experiments [9], in the analysis of vegetative or minimally conscious states [5], for the development of Brain Computer Interfaces (BCIs) for rehabilitation purposes and to allow the control of medical devices (e.g. wheelchairs) [6].

However, the development of these kinds of applications encounters some difficulties due to the fact that the EEG signal is weak, time varying and easily affected by biological (ocular, muscular, cardiac movements) and non-physiological (direct current, electrical leakage) noises [11], which must be removed to allow a better analysis without losing useful experimental data.

Also, to make assumptions over a population, the specificity of each recording must be considered and so the difficulty of normalising the data, i.e. trying to fit a specific recording into a canonical space, arises. This necessity, to which an unique solution is yet to be found, may allow researchers to move from a classical subject-based analysis to a population-based one.

2 Problem statement

Starting from an experiment of face recognition conducted in collaboration with professor Daini¹ [9], the Ph.D. project mainly wants to refine the denoising techniques used in the aforementioned work and move to the identification of metrics for discrimination of noisy patterns in the EEG signal.

Afterwards, the obtained features will be used to train a classifier to develop a tool for semi- to completely-automated noise removal, which will be expected to run with slight variations in any kind of case scenario.

Finally, the Ph.D. work will consider the problem of normalisation to allow analysis inter-subject or between different populations (e.g. access the differences in the brain activations between a normal recogniser population and an impaired one). This last step will be introduced as an addition to the defined pipeline, lacking of state-of-the-art references and having the heterogeneity and subject-specificity of the EEG data as difficult issues to address.

¹ Department of Psychology, University of Milano - Bicocca, Milan, Italy

3 State-of-the-art and Methodology

As well depicted by *Urigüen et al.* [11], concerning the EEG signal pre-processing, the state-of-the-art presents a great number of noise removal algorithms, but there are no automatic procedures for the identification and correction of noisy patterns that could be considered completely effective and efficient, preferring more consolidated methods for semi-automatic inspection and noise suppression or even manual rejection of the noisy recording portions.

As stated in the introduction, the difficulties of developing an automated procedure arise from the specific nature of the EEG signal: it varies from patient to patient, depends on the recording hardware, may present a mixture of the actual electrophysiological data and contaminated components.

Denoising techniques are mostly scenarios-dependent, but despite that, some recent studies confirm the success of methods used in the identification of noisy components and suggests some useful tips for EEG automatic signal manipulation. For example, *Al-Qazzaz et al.* [1] present an automatic noise removal pipeline specifically suitable for working memory tasks in normal and affected by dementia populations.

Therefore, starting from the study of semi-automatic procedures, the noise removal methodology suggested for the Ph.D. work wants to obtain a modular set of procedures, that could be algorithms applicable to any kind of experiment and to scenario-dependent ones and move to the identification of noisy patterns through some metrics, which could be used for the training of peculiar classifiers. In fact, the state-of-the-art presents a good amount of ML techniques for EEG signal classification. The most used are k-Nearest Neighbour (k-NN), which assigns to a tested sample the label of the k-nearest training sample, and Support Vector Machine (SVM), that segregates the data through an hyper plan with maximal margins, as supervised classifiers and Naïve Bayes (NB), which is based on Bayes' theorem and determines the class of earlier probabilities through a maximum probability algorithm and uses a feature probability distribution from a training set [2], as a probabilistic one. The common characteristic of these methods is the necessity of having a validated training set.

To allow a classification that could be executed with both un- and supervised approaches, recently the research community involved in healthcare began to explore and develop applications based on deep learning techniques.

In this regard, the review edited by *Miotto et al.* [7] describes these applications challenges and opportunities, which arise even in the more specific domain of EEG classification.

Therefore, the Ph.D. project wants to (1) start from the supervised classifiers and create best practices for feature extraction in scenario-based experiment and for noise patterns classification, where the most used features would be extracted computing the Power Spectral Density (PSD) over the frequency bands that characterise the EEG signal (e.g. average spectral power, spectral power for each frequency and approximate entropy) and (2) create a benchmark through which evaluate the possibility of developing a deep learning classifier.

The last cited item opens new issues, like the fact that a deep learning model

requires a great amount of data, which should be clean and well-structured. This could be achieved by manipulating the raw state of the EEG signal to be more clean and interpretable through the signal processing procedure previously cited. Also, there are no robust and well-maintained deep learning procedures on EEG applications, even though there have been recent studies for the use of Convolutional Neural Networks (ConvNets). *Schirrneister et al.* [10] developed a new method for visualising learned features and showed how to design and train ConvNets to decode task-related information from raw EEG data. However, the outputs given by ConvNets are frequently difficult to interpret and this method involves a good number of hyperparameters, but ConvNets have some interesting characteristics that could represent a good compromise to choose them among the other ML algorithms: there is not the necessity of a priori features selection, they are scalable on large datasets and exploit the hierarchical structure typical of the natural signals [10].

Finally, from the necessity of evaluating brain activities and functions in non-and pathological conditions or between different states, comes the issue regarding a proper way to normalise the EEG signal, which is not only characterised by a high dimensionality, but also - as for other biomedical data - presents heterogeneity, temporal dependency, sparsity and irregularity. This problem has been discussed between numerous researchers (mainly on *ResearchGate*²), but an univocal solution has yet to be found.

The advised approaches emerging from the discussions and, only minimally, in the state-of-the-art are mainly (1) the normalisation of the power spectra for each frequency band and sensor [4], which could be difficult to apply due to the EEG nature and (2) the standardisation of the sensors voltage by using the z-score to detect differences between groups, usually applied on resting state EEG recording and thus inappropriate for task-based experiments.

The Ph.D. project wants to evaluate, as an additional step, the suggested solutions and find better methods for data normalisation, moving from a subject-based approach to a population-based one.

4 Conclusion

The project for the Ph.D. work is divided in three main steps: signal processing, classification of heterogeneous EEG data and normalisation.

Each of them could be divided in and expanded with different sub-modules: novel algorithms for denoising based on peculiar signal characteristics, normalisation procedures less commonly cited as the min-max normalisation [13], new features and classifiers that could be useful for BCI development and whose accuracy could be evaluated verifying for example classes balance, kappa metric, confusion matrix on offline data [6].

The cleared signal and noisy patterns classification will be validated by experts of the Department of Psychology at University of Milano - Bicocca, evaluating

² https://www.researchgate.net/post/Normalization_of_resting_EEG_data_for_comparisons_between_different_subjects

the accuracy, sensitivity and specificity of the obtained results [3]. Therefore, the presented paper wants to give guidelines for EEG signal processing and classification, given that the research field that the Ph.D. project wants to address is in constant evolution and improvement.

References

1. Al-Qazzaz, N.K., Hamid Bin Mohd Ali, S., Ahmad, S.A., Islam, M.S., Escudero, J.: Automatic artifact removal in EEG of normal and demented individuals using ICA–WT during working memory tasks. *Sensors* **17**(6), 1326 (2017)
2. Amin, H.U., Mumtaz, W., Subhani, A.R., Saad, M.N.M., Malik, A.S.: Classification of EEG signals based on pattern recognition approach. *Frontiers in computational neuroscience* **11**, 103 (2017)
3. Barua, S., Ahmed, M.U., Ahlström, C., Begum, S.: Automatic driver sleepiness detection using EEG, EOG and contextual information. *Expert systems with applications* **115**, 121–135 (2019)
4. Haegens, S., Cousijn, H., Wallis, G., Harrison, P.J., Nobre, A.C.: Inter-and intra-individual variability in alpha peak frequency. *Neuroimage* **92**, 46–55 (2014)
5. Lehembre, R., Bruno, M.A., Vanhaudenhuyse, A., Chatelle, C., Cologan, V., Leclercq, Y., Soddu, A., Macq, B., Laureys, S., Noirhomme, Q.: Resting-state EEG study of comatose patients: a connectivity and frequency analysis to find differences between vegetative and minimally conscious states. *Functional neurology* **27**(1), 41 (2012)
6. Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., Yger, F.: A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of neural engineering* **15**(3), 031005 (2018)
7. Miotto, R., Wang, F., Wang, S., Jiang, X., Dudley, J.T.: Deep learning for health-care: review, opportunities and challenges. *Briefings in bioinformatics* (2017)
8. Radüntz, T., Scouten, J., Hochmuth, O., Meffert, B.: EEG artifact elimination by extraction of ICA-component features using image processing algorithms. *Journal of neuroscience methods* **243**, 84–93 (2015)
9. Saibene, A., Corchs, S., Daini, R., Facchin, A., Gasparini, F.: EEG Data of Face Recognition in Case of Biological Compatible Changes: A Pilot Study on Healthy People. In: *Proceedings of the 15th International Joint Conference on e-Business and Telecommunications - Volume 1: ICETE*, pp. 414–420. INSTICC, SciTePress (2018). <https://doi.org/10.5220/0006909104140420>
10. Schirrneister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., Ball, T.: Deep learning with convolutional neural networks for EEG decoding and visualization. *Human brain mapping* **38**(11), 5391–5420 (2017)
11. Urigüen, J.A., Garcia-Zapirain, B.: EEG artifact removal state-of-the-art and guidelines. *Journal of neural engineering* **12**(3), 031001 (2015)
12. Zani, A., Mado Proverbio, A., Mangun, G., M. Fletcher, E., Brattico, E., Olcese, C., Tervaniemi, M., Ntnen, R., L. Wilding, E., Federmeier, K., Kutas, M., T. Knight, R., Scabini, D., Luu, P., Tucker, D.: *Metodi Strumentali nelle Neuroscienze Cognitive. EEG ed ERP- Instrumental Methods in Cognitive Neuroscience. EEG and ERP* (11 2013)
13. Zhang, X., Yao, L., Zhang, D., Wang, X., Sheng, Q.Z., Gu, T.: Multi-person brain activity recognition via comprehensive EEG signal analysis. *arXiv preprint arXiv:1709.09077* (2017)