

Data Science in Neuroscience: A Review of The EEG Analytical Workflow

Cugliari G* and Ivaldi M

Department of Medical Sciences, University of Turin, Turin, Italy

ISSN: 2577-1914



***Corresponding author:** Cugliari, Giovanni, Department of Medical Sciences, University of Turin, Turin, Italy

Submission: 📅 February 17, 2020

Published: 📅 March 17, 2020

Volume 6 - Issue 2

How to cite this article: Cugliari G, Ivaldi M. Data Science in Neuroscience: A Review of The EEG Analytical Workflow. *Res Inves Sports Med*, 6(2): RISM.000633.2020
DOI: [10.31031/RISM.2020.06.000633](https://doi.org/10.31031/RISM.2020.06.000633)

Copyright@ Cugliari G, This article is distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use and redistribution provided that the original author and source are credited.

Abstract

In the era of big data, quantitative-based approach has become a very useful tool in neuroscience studies. Neural phenomena that occur at the level of the cerebral cortex generates electric activity that can be recorded using electroencephalography (EEG). We can divide into three macro areas, considering non-pathological studies in human movement, the suite of interesting topics as i) physical stimuli and body postures, ii) visual stimuli and experience, and iii) auditory stimuli and motor imagery. In this context, data analysis represents a fundamental core of tasks with the aim to extract accurate and consistent information. To facilitate the replicable identity, scientific research is therefore interested in developing statistical procedures of biological data analysis. The aim of this review is to explain the analytical workflow applied to the EEG signal. Through theoretical and practical feedback, this work will be useful for data scientists, neuroscientists, statistics, engineers or physiologists.

Keywords: EEG; Data science; Human movement; Neuroscience; Statistical modeling; Clustering; ICA

Introduction

Electroencephalography (EEG) is a non-invasive technique that allows better performance compared to other neuroimaging techniques, such as magnetoencephalography (MEG), considering the time resolution of the recorded signal; it can be considered as the representation of the postsynaptic potentials generated at the level of the cortex of the brain by neurons synchronous activity. In detail, EEG represents the sum of several waves at different frequencies and amplitudes, generated by specific brain activity related to particular tasks (sensorial, movement-related or cognitive). Some event-related potentials (ERPs) are associated with certain states of consciousness specially induced or brain-specific pathological conditions (like epilepsy). From the comparison between the spontaneous activity (used as a baseline) and its variation during the induced activity is possible to identify in real-time the areas and the modalities of the change electrical activity. The strength of this type of analysis is the time resolution, in fact, is possible to obtain the information every millisecond of electrical activity; its major limitation is the spatial resolution, the electrodes pick up only the electrical signal that reaches the surface of the skull and just due to reconstruction signal algorithms is possible to locate the real source of signal. In this context, data science represents a very informative tool from the signal recording to the study funding elaboration. As explained in the discussion section, the analytical workflow can be divided into five macro-areas, each one with a suite of peculiar characteristics: data analysis tool, preprocessing, data preparation, component calculation, clustering optimization, and statistical modeling.

Discussion

Different study designs are required to answer different research questions. EEG experiments require preparation of participants and set-up of the equipment making sure that the methods return the desired outcomes to maximize scientific research standards such as objectivity, reliability, and validity. After data mining procedures, filtering the continuous data minimizes the introduction of artifacts; during data preparation, artifacts were evaluated with the aim to control for outliers and missing values; the third phase consists in the component calculation, independent component analysis (ICA) produces the maximally temporally independent signals that can be considered in the inferential analyses; the goal of the clustering function is to compute an N-dimensional cluster position vector for each component; Finally, the statistical modeling phase shows the potential alternatives to analyze data considering the aim of the study with the aim to validate the association estimates.

Data analysis tool

For data mining, recording, processing, and modeling event-related and continuous electroencephalography (EEG) is important the choice of the appropriate analytical tool: a MATLAB toolbox named EEGLAB provides a programming environment and interactive graphical user interface (GUI) for accessing, visualizing, measuring, manipulating, and storing electrophysiological data; it allows several modes of visualization of the single-trial and averaged data [1,2].

Preprocessing

To build EEG scalp maps, the dataset must contain information about the locations of the recording electrodes: if the data have been recorded with a reference, they can usually be re-referenced to any other reference channel. Filtering the continuous data, using high/low/bandpass, minimizes the introduction of artifacts (linear trends); this step uses the linear finite impulse response (FIR), forward and backward, to ensure that the delay phase introduced by the filter is nullified. Procedures to study the event-related EEG signals of continuously recorded EEG data can be summarized in three tasks: i) specifying the considered baseline data epochs time; ii) extracting data epochs time to events of interest; iii) removing a mean baseline value from each epoch to extract the potential effect of the indagated outcome.

Data preparation

Data scrolling is useful to reject epochs of data (channels with outliers, after kurtosis or probability tests) which contains artifact; the normalization procedure of the voltage allows better visualization of data. A signal is analyzed using power spectrum, which provides information on the signal power at each frequency. The Fourier transform decomposes the EEG time series into the power spectrum, in which the power is the square of the EEG amplitude, and the amplitude is the integral average of the EEG signal during the epoch sampled. The frequency resolution is given by the inverse of the time value of the epoch. Lastly, it possible to evaluate the event-related cross-coherence, to determine the degree of synchronization between the activations of two channels.

Component calculation

ICA components of EEG data are maximally temporally independent but spatially unconstrained and therefore they are able to find maps representing the projection of a partially synchronized region of cortex. All three algorithms available return near-equivalent components (Runica, Jader, and Fastica). Missing data treatment may be replaced using an approach based on spherical interpolation. Parametric or bootstrap statistics (non-parametric statistics) may be used to compare a given outcome in experimental study design.

Clustering optimization

After post hoc analysis (to control for multiple comparisons) means and neural network clustering methods are available. Several measures to construct the cluster can be used: ERP, power

spectrum, ITC (Intertrial coherence), component scalp maps and their equivalent dipole model locations. The pre-clustering function represents the first step in this process: the goal is to compute an N-dimensional cluster position vector for each component: then, cluster position vectors will be used to measure the distance of components from each other considering the N-dimensional cluster space. At this phase the normalization procedure is required, this involves dividing the measurement data of all principal components by the standard deviation of the first PCA (principal component analysis) component. There are several steps involved in the independent component clustering procedure: i) identifying a set of EEG datasets containing ICA weights ii) specifying group, task condition, and session for each dataset, iii) identifying the component in each dataset to cluster, iv) specifying and computing measures to use in clustering, v) performing component clustering for each investigated measures, vi) viewing the scalp maps and activity measures of the component clusters, vii) performing signal processing and statistical estimation on the clusters, viii) studying the consistency and properties, using validation procedures, of the generated component clusters.

Statistical modeling

It's possible to perform parametric and non-parametric tests (paired t-test, unpaired t-test, ANOVA) on each of the investigated measures. Matched/unmatched data samples can be used as an extension of paired/unpaired data samples when there are more than two samples. Resampling methods help to perform statistical inference without assuming a known probability distribution for the data. Instead, the bootstrap method consists of drawing random sub-samples followed by the randomization method (shuffling data samples). Recent progress in signal processing and information theory has seen the development of blind source separation methods, which attempt to find a coordinate frame onto which the data projections have minimal overlap. Second this concept, ICA is a family of linear blind source separation methods: the core mathematical concept of ICA is to minimize the mutual information among the data projections. ICA is being applied to various biomedical signal processing problems that include: i) performing a speech from noise separation, ii) decomposing functional resonance imaging data, and iii) separating brain area activities and artifacts mixed in the electro-encephalographic activity. When performing a large number of statistical inferences, it is necessary to correct for multiple comparisons (Bonferroni, Holms method, False Discovery Rate, Max method, Cluster method).

Conclusion

In this section, we explain a brief overview of potential experiments in the context of human movement. Some studies that we have conducted in recent years have in fact been dedicated to deepening peculiar and niche areas. We will show three macro-categories of studies ordered by type of stimulus and considered sub-task: i) physical stimuli and body postures; ii) visual stimuli and experience; iii) auditory stimuli and motor imagery. In detail we valuated all bands considering ERP, scalp maps and time-frequency.

Physical stimuli and body postures

We have indeed studied to evaluate the timing and characteristics of electrocortical activity during stretch reflex evocation of the quadriceps femoris; we also studied the variations during different body postures. Our findings improve the understanding of the neurophysiological dynamics following the stretch reflex after concussion of the patellar tendon, executed in different postures, considering scalp-map, power spectra, and time-frequency analysis. The use of scalp-map and power spectra analysis techniques represents a sophisticated use of advanced signal processing methodologies to analyze brain activity during movement, considering the posture-related correlation, and in specificity in sport science [3].

Visual stimuli and experience

Experience may be a very important outcome with the aim to increase the understanding of a specific context: it represents the level of confidence during the execution of multiple tasks. The evaluation of electrocortical activity during visual stimuli may reveal differences depending on human being experience level. These findings confirm the relationship between EEG activity and vision of specific physical movements and extensive knowledge on the electrocortical response to visual stimuli emphasizing the difference between experienced and inexperienced subjects [4].

Auditory stimuli and motor imagery

Changes in electrocortical activity during motor imagery are among the most interesting findings in the recent neuroscientific studies. In detail, the statistically significant differences between expert dancers and controls could indicate a difference in the attentional effort during the dance imagery task. Remarks extended the knowledge on the EEG response to auditory stimulus

during motor imagery in particular for Beta rhythm components, emphasizing specific characteristics in function of the level of familiarity to the dance motor imagery task [5].

Overall considerations

This review describes some best procedures for the experimental design, data visualization and descriptive or inferential statistical analysis applied to the neuroscience context using EEG signal [6]. The use of EEG in the study of human electrocortical activity is an extremely promising scientific branch due to recent technological advances both as regards instruments, but also for the new frontiers of computational capacity in artificial intelligence.

References

1. Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods* 134(1): 9-21.
2. Makeig S, Debener S, Onton J, Delorme A (2004) Mining event-related brain dynamics. *Trends Cogn Sci* 8(5): 204-210.
3. Ivaldi M, Pretari F, Cugliari G (2018) Electrocortical activity during stretch reflex in athletes. *Sport Sci Health* 14: 625-631.
4. Ivaldi M, Cugliari G, Fiorenti E, Rainoldi A (2018) Delta and alpha rhythms are modulated by the physical movement knowledge in acrobatic gymnastics: an EEG study in visual context. *Sport Sci Health* 14: 563-569.
5. Ivaldi M, Cugliari G, Peracchione S, Rainoldi A (2017) Familiarity affects electrocortical power spectra during dance imagery, listening to different music genres: independent component analysis of Alpha and Beta rhythms. *Sport Sci Health* 13: 535-548.
6. Cugliari G, Ivaldi M (2015) *Multivariate statistical analysis in neuroscience: advanced mathematical modeling applied to electroencephalographic signals in complex data problems*. Grin Verlag, Munich, Germany.

For possible submissions Click below:

[Submit Article](#)