# A Comparative Analysis of Correlation and Correntropy in Graph-Based Brain Computer Interfaces

Luisa Fernanda Suárez Uribe, Carlos Alberto Stefano Filho, Vanessa Brischi Olivatto, Vinícius Alves de Oliveira, Diogo Coutinho Soriano, Levy Boccato, Gabriela Castellano and Romis Attux

*Abstract*— **This work presents a comparative analysis of correlation and correntropy in the context of graph-based brain-computer interfaces using motor imagery. These two statistical entities are used in the construction of the graphs, from which features are extracted. The results indicate that correntropy has a more consistent performance over the different graph measures, hence deserving to be considered as a relevant option by researchers of the field.**

*Keywords*— **Brain-computer interfaces, information-theoretic learning, correntropy, graph measures, motor imagery.**

## I. Introduction

Brain-computer interfaces (BCIs) aim to control an external device by directly employing the user brain signals [1]. For a BCI to appropriately function, an accurate classification scheme that is capable of identifying a useful brain response and associating it with specific commands is mandatory. In the case of recorded signals from an electroencephalography (EEG) device, an important strategy to evoke such responses is left / right hand motor imagery (MI). This stimulation results in event-related desynchronizations (ERDs) amongst motor neurons, generating a decrease in the power spectral density (PSD) of the mu frequency band (7 - 13 Hz) on the contralateral hemisphere to which the hand MI is performed [2]. Therefore, extracting features from the signal spectral domain and using them to distinguish between MI tasks has become a common approach [3], [4].

Although this method is relatively well-established in the MI-BCI community, no optimum manner of proceeding through the stages of a BCI has yet been found. In fact, there are still some limitations, such as the great inter-subject variability and inconsistencies regarding the responsive MI frequency bands and response scalp locations for each subject [5].

Luisa Uribe, Vanessa Olivatto, Vinícius de Oliveira, Levy Boccato and Romis Attux, School of Electrical and Computer Engineering, University of Campinas (UNICAMP), Campinas - SP, Brazil, E-mails: lsuarez@dca.fee.unicamp.br, vanessa.olivatto@gmail.com, viniciusalvesoli@gmail.com, lboccato@dca.fee.unicamp.br, attux@dca.fee.unicamp.br. Vanessa Olivatto is also with Instituto Eldorado, Campinas - SP - Brazil. Carlos Stefano Filho and Gabriela Castellano, Physics Intitute, University of Campinas (UNICAMP), Campinas - SP, Brazil, E-mails: carlos.stefanofilho@gmail.com, gabriela@ifi.unicamp.br. Diogo Soriano, CECS, UFABC, São Bernardo - SP, E-mail: diogo.soriano@gmail.com. Luisa Uribe, Romis Attux, Carlos Stefano Filho, Diogo Soriano and Gabriela Castellano are also with Brazilian Institute of Neuroscience and Neurotechnology (BRAINN). This work was supported by CAPES, CNPq, FINEP and FAPESP (grant 2013/07559-3).

The use of graph theory in neuroscience has led to new insights regarding the functional relationship between distinct brain areas. It has been highlighted that a further understanding of the functional connectivity during MI tasks might improve the existing BCI technology [5].

In this work, we compare the performance of graph measures to classify signals from right and left hands MI, extracted from graphs built according to two different approaches: correntropy - an entity from the field of information-theoretic learning [7] - and correlation. We chose to work with three graph measures based on centrality (degree centrality, betweenness centrality and eigenvector centrality), as they provide complementary information regarding a node's importance within the network. Linear discriminant analysis was adopted in the classification stage.

This work is organized as follows. In section II, we describe the materials and methods used, including a brief exposition of correntropy; in section III, we present the obtained results, while section IV brings our conclusions and final remarks.

## II. Materials and Methods

### A. Data acquisition

Data from six healthy subjects were acquired using a 64-Ag/AgCl electrode array positioned according to the 10-10 positioning system in the BrainCap. A set of two BrainAmp amplifiers (BrainProducts, Germany) were used. All subjects signed an informed consent term, previously approved by the Ethics Committee of UNICAMP (n. 791/2010).

For the EEG recordings, volunteers were sitting comfortably in a chair in front of a computer screen, where a chronometer was displayed, so they could keep track of the time associated with each block. The acquisition protocol consisted of blocks of 10 seconds each, alternating between periods of task (left or right hand movement imagination) and rest (Figure 1). This procedure was carried out for 170 s, constituting one run. Two runs were performed for each subject. To ensure that all subjects properly understood the protocol, two other runs took place before these, but with actual hands movement instead of MI.

For both movement and MI acquisitions, volunteers were instructed to perform the same type of motor action, which consisted on the opening and closing of their hands (left or right, depending on the task block being executed). Moreover, we asked subjects to attempt to perform this movement at a 1

Hz frequency (which should account for 10 motor actions by the end of the task block)

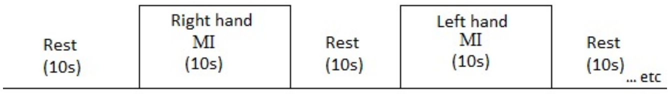



Fig. 1: Experimental protocol scheme.

### B. Data pre-processing

The collected data were downsampled to 256 Hz, since the original 5 kHz sampling rate proved to be unnecessarily high for our study. Next, data were filtered in the mu frequency band (7 - 13 Hz). Both of these operations were performed in EEGLab, a MATLAB suite.

For artifact removal, we designed a common average removal (CAR) filter. The basic idea is to calculate the mean signal across all electrodes and to subtract this value from each individual channel signal [6].

### C. Graphs and adjacency matrices calculation

Consider an undirected, connected, weighted graph $G = \{V, E, W\}$, where $V$ is a finite set of vertices with $|V| = N$, $E$ is a set of edges, and $W$ is the adjacency matrix. A signal $f : V \to \mathcal{R}^N$ can be defined on the vertices of the graph represented as a vector $f \in \mathcal{R}^N$, where the $n^{th}$ component, $f_n$, represents the signal value at the $n^{th}$ vertex of $V$.

There are several possible forms to define a similarity function $s(f_i, f_j)$ between all pairs of data points $f_i$ and $f_j$ to model the local neighborhood relationships between the vertices. The similarity graphs establish that two vertices are connected if the similarity $s(f_i, f_j)$ between the corresponding data points is positive or larger than a certain threshold $T$. The corresponding edge is therefore weighted by $w_{i,j} = s(f_i, f_j)$, and the weighted adjacency matrix of $G$ is simply defined as $W = (w_{i,j})_{i,j=\{1,\cdots,n\}}$. If $w_{i,j} = 0$, the vertices $v_i$ and $v_j$ are not connected since either $s(f_i, f_j) = 0$ or $s(f_i, f_j) < T$. If $G$ is an undirected graph, then $w_{i,j} = w_{j,i}$ and $W$ is a symmetric matrix with a complete set of real eigenvalues and an orthogonal eigenvector basis.

In our study, we modeled interactions amongst electrodes by a graph, with each node being associated to a single electrode. The edges represent interactions between these nodes, which were estimated by two methods - Pearson correlation and correntropy -, and the corresponding classification performances of the measures obtained from these graphs were compared. The purpose of this comparison is basically to verify whether the higher-order information brought by correntropy can play an effective role in improving the rate of correct classification. Notice that each element of the graph adjacency matrix corresponds to a measure of similarity amongst the time series of the associated electrode pair. Let us now introduce the measure of correntropy.

### D. Correntropy

Correntropy is an emblematic entity in the context of information-theoretic learning, and can be interpreted as a generalized correlation function that is capable of measuring the similarity between random variables / stationary random processes [7], [8]. The concept of correntropy may be extended to cross-correntropy when it expresses a degree of similarity between two arbitrary scalar random variables X and Y. The cross-correntropy is defined as

$$V(X,Y) = E_{XY}\left[\kappa(X,Y)\right] = \int\int \kappa(x,y)\, p_{X,Y}(x,y)dxdy, \tag{1}$$

where $\kappa(\cdot,\cdot)$ denotes a positive-definite kernel function. Equation (1) can be reduced to the standard cross-correlation if $\kappa(x, y)$ is replaced by $xy$. The most widely used definite kernel function is the translation-invariant Gaussian kernel, defined as

$$\kappa(X,Y) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(\frac{-(X-Y)^2}{2\sigma^2}\right), \tag{2}$$

where $\sigma$, known as kernel size, acts as an adjustable and sensitive parameter that is responsible for controlling the "observation window" in which similarity is obtained. In practical applications, the statistical mean required for computing the correntropy is replaced by a sample mean over a finite number of data, resulting in the following estimator:

$$\hat{V}_{\sigma,N}(X,Y) = \frac{1}{N}\sum_{i=1}^{N} G_\sigma(x_i - y_i), \tag{3}$$

where $G_\sigma(alpha) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-|\alpha|^2}{2\sigma^2}\right)$.

By representing the Gaussian kernel according to the Taylor series expansion, the correntropy function can be rewritten as:

$$V_\sigma(X,Y) = \frac{1}{\sqrt{2\pi}\sigma}\sum_{n=0}^{\infty}\frac{(-1)^n}{2^n\sigma^{2n}n!}E\left[(X-Y)^{2n}\right], \tag{4}$$

which reveals that all even-order moments of the random variable $(X - Y)$ are being implicitly explored in correntropy. It is possible to verify that the information provided by the conventional cross-correlation is part of the new function, more specifically in the term corresponding to $n = 1$, which is proportional to:

$$E_X\left[X^2\right] + E_Y\left[Y^2\right] - 2R_{XY}(X,Y). \tag{5}$$

This result shows that the generalized cross-correntropy keeps the conventional cross-correlation embedded, but, at the same time, carries information about higher-order moments of the random variables. This extra information may express a better notion of similarity between two random variables in many different applications.

In the context of our application, considering $X \in \mathcal{R}^N$, from Eq. (3), the correntropy can be expressed in terms of the following matrix $V$, which will be the graph adjacency matrix for the next steps of processing. This will be compared to a conventional correlation matrix.

$$V = \begin{pmatrix} E[k(x_1x_1)] & E[k(x_1x_2)] & ... & E[k(x_1x_N)] \\ E[k(x_2x_1)] & E[k(x_2x_2)] & ... & E[k(x_2x_N)] \\ . & & & \\ . & & & \\ . & & & \\ E[k(x_Nx_1)] & E[k(x_Nx_2)] & ... & E[k(x_Nx_N)] \end{pmatrix} \quad (6)$$

### *E. Graph centrality measures*

Graph centrality measures are metrics that define the importance of a node under certain criteria. There are a variety of centrality measures that can be defined. In this study, we focused our analysis on three types of measures, which give complementary information: degree centrality (DC), betweenness centrality (BC) and eigenvector centrality (EC).

The DC for a node '$i$' is defined as the sum of all the weights related to that node and all other nodes '$j$' ($w_{ij}$), that is:

$$DC(i) = \sum_j w_{ij}. \quad (7)$$

The BC is a measure that considers, for a node '$i$', all the possible shortest paths between any other two node pairs ($l_{jk}$) that pass through node '$i$' ($l_{jk}(i)$). In other words, it is a measure of how much that node serves as a "bridge" between any other two nodes (under a shortest path consideration).

$$BC(i) = \frac{2}{N(N-1)} \sum_{j \neq i \neq l} \frac{l_{jk}(i)}{l_{jk}}, \quad (8)$$

where $N$ is the total number of nodes in the graph.

The EC of node '$i$' is calculated by the i-th position of the eigenvector associated with the largest eigenvalue of the adjacency matrix.

It is important to note that a large node centrality value for one measure does not guarantee that the same node will display, also, a large value for another centrality measure. As an example, please refer to Fig. 2.

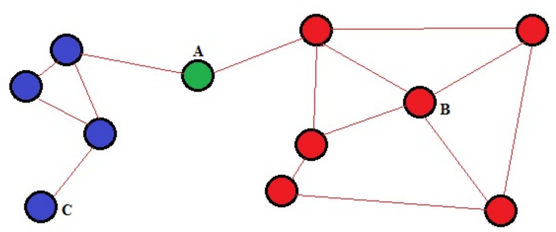



Fig. 2: Centrality values illustration. Node A has low DC, but possesses a fundamental role regarding communication between the red and blue nodes (high BC).

### *F. Classification*

From the analysis of graph centrality measures, vectors of 64 features were obtained, each feature related to an electrode. From these vectors, smaller, 32-feature vectors were created, consisting of the difference of the graph measures between equivalent electrodes from the left and right hemispheres. Next, a selection algorithm was used to find the best combination of features.

The approach used for feature selection was that of wrappers [9]. The employed wrapper was based on forward selection. The algorithm is based on a "bottom–up" approach, which means that it starts with an empty feature set. In each step, one feature is added until 32 features are reached.

Classification was based on a Linear Discriminant Analysis (LDA) classifier [9]. The total data set of one session was divided into two sets of 70% for training and 30% for test. This division was done ten times, i.e., the elements that were part of the two sets were different in each case. This procedure is similar to k-fold crossvalidation, with the difference that our second set is for test and not for validation.

## III. Results and Discussion

Figure 3 displays a bar graph of the classification error according to the $\sigma$ value of the kernel function in Equation (2). Each centrality measure is shown in a different color: EC in blue, DC in green and BC in red. It can be seen that for BC, there is a $\sigma$ value above which no result is shown. This is due to the fact that, for these values, BC calculation did not converge, probably due to sparse adjacency matrices. Still, for every $\sigma$ value that BC existed, it had the worst classification performance.

The lowest classification errors were obtained either for EC or DC, depending on the $\sigma$ value. EC performed best with a $\sigma$ of 5.00, whereas DC yielded the lowest classification error for a $\sigma$ of 1.70.

Considering an average performance of both EC and DC, there seems to exist a general decrease in classification error up to a $\sigma$ of 1.70. An increase can be observed above this point, mainly represented by the worse performance of DC. The $\sigma$ value is associated with the spread of the Gaussian function. Therefore, the results of Figure 3 suggest that the best classification scenario is not, necessarily, the one where this spread has the lowest or the highest value. In fact, for our case, an optimum $\sigma$ may be chosen as 1.70.

Figure 4 shows the classification error when the Pearson correlation is used for the adjacency matrices calculation, instead of the correntropy. In contrast with Figure 3, BC yielded the lowest classification error in this case. For EC and DC, classification accuracy was considerably lower in this case.

The contrast between results from Figures 3 and 4 indicates that the two distinct approaches for constructing the graphs can result in different networks, thus considerably changing the centrality measures, at least regarding the classification problem.

This is also confirmed by focusing on the results of classification error for each subject in Table I. Good results are obtained when using the combinations of correntropy + EC and correntropy + DC. The correlation has a comparable result for some subjects with correlation + BC. The values of the classification error for the correntropy in Table 1 were the best values obtained among all different $\sigma$ values for each subject.

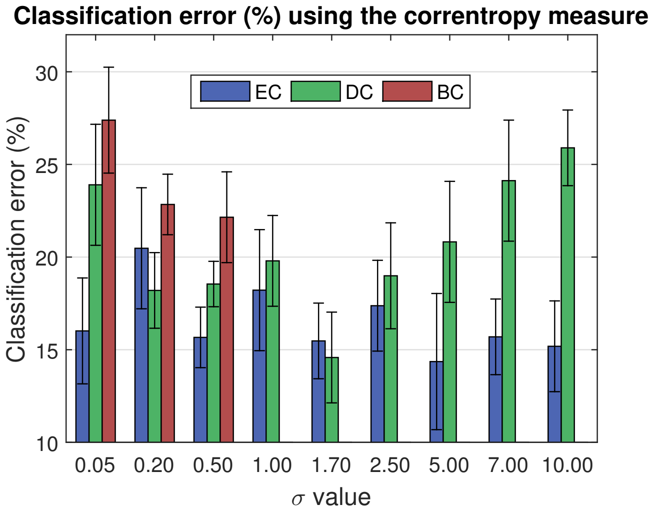



Fig. 3: Mean classification error for all subjects using the correntropy measure. Distinct $\sigma$ values were tested, and are shown in the horizontal axis. The three centrality measures are displayed as different colors: EC (blue), DC (green) and BC (red). Error bars constitute the standard error across all subjects for a given condition.

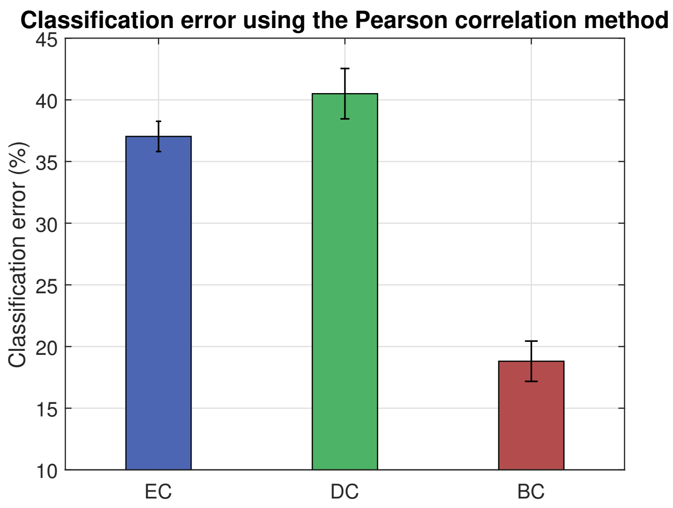



Fig. 4: Mean classification between all subjects using the Pearson correlation method. All three centrality measures are shown: EC (blue), DC (green) and BC (red). Error bars represent the standard error across all subjects for a given centrality measure.

A statistical test was also performed to make a comparison of the results obtained for the six cases that were tested. Although there was a not high number of subjects to make these comparison, after statistical analysis was assumed gausinity of the data. The comparison was made using the ANOVA test for multiple comparisons of pairs of columns from the results of Table I. Table II presents the results where the difference was significant, these values shows this difference. The value of significance used for this test was p $\leq$ 0.05. In the first column below of each presented technique its average value is presented also. The results of this test show that, above all, the techniques that used correntropia were the ones that presented the most significant difference in almost all the cases compared.

TABLE I: Mean classification error for each subject for all the cases tested. S indicates the different subjects tested.

| S | EC +Corren | EC +Correl | BC +Corren | BC +Correl | DC +Corren | DC +Correl |
|---|---|---|---|---|---|---|
| 1 | 0,135 | 0,283 | 0,167 | 0,190 | 0,076 | 0,391 |
| 2 | 0,120 | 0,349 | 0,208 | 0,186 | 0,094 | 0,422 |
| 3 | 0,054 | 0,415 | 0,207 | 0,093 | 0,092 | 0,418 |
| 4 | 0,077 | 0,363 | 0,205 | 0,221 | 0,161 | 0,381 |
| 5 | 0,107 | 0,377 | 0,174 | 0,193 | 0,094 | 0,419 |
| 6 | 0,103 | 0,410 | 0,094 | 0,081 | 0,183 | 0,263 |

TABLE II: Significant mean difference for the multiple comparison ANOVA test between the six methods used.

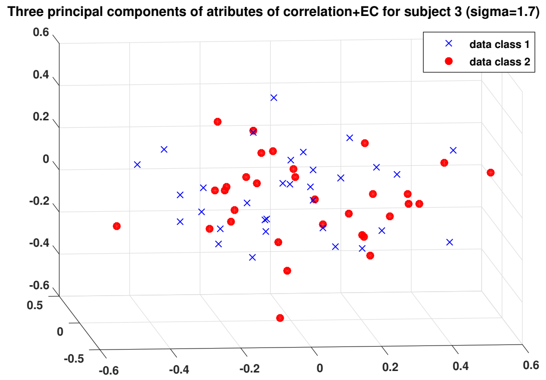
| | EC+ Corren | BC+ Corren | DC+ Corren | EC+ Correl | BC+ Correl | DC+ Correl |
|---|---|---|---|---|---|---|
| **EC+ Corren m=0,072** | | -0,103 | | -0,293 | -0,087 | -0,301 |
| **BC+ Corren m=0,176** | | | | -0,190 | | -0,206 |
| **DC+ Corren m=0,097** | | | | -0,269 | | -0,285 |
| **EC+ Correl m=0,366** | | | | | 0,205 | |
| **BC+ Correl m=0,166** | | | | | | -0,222 |
| **DC +Correl m=0,382** | | | | | | |

A visual comparison of the results was done using Principal Component Analysis (PCA), extracting the three most significant components of the data that were used for classification. Figure 5 shows these attributes in a three-dimensional space for Subject 3, that had one of the the best separation results. The beneficial effect of the higher-order information brought by correntropy is quite clear.

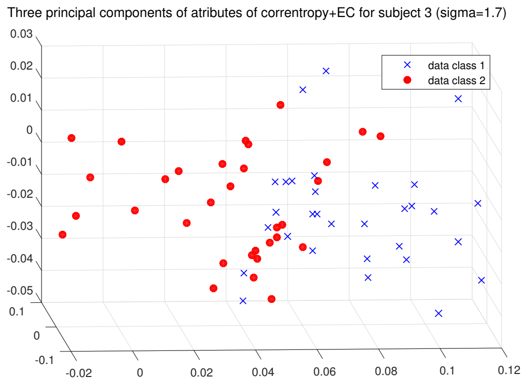
## IV. Conclusions

This work focused on the use of graph analysis applied to EEG signals in the context of BCI, directed towards the construction of activity maps that allow to relate the temporal behavior of different electrode series to distinct two types of MI task (left / right motor imagery). In the literature, responses have already been seen using techniques such as correlation to carry out this task, but in this work we included the correntropy measure. This allowed us to find similarities between channels using, in theory, more information than when the simple correlation is used.

The results showed that similar classification performances are obtained using correntropy combined to different graph metrics to generate the classification attributes. The only case in which the correlation achieved a comparable performance,




(a) Atributes of correlation+EC




(b) Atributes of correntropy+EC

Fig. 5: Using PCA for showing the attributes of the two classes for subject 3.

with a relatively low classification error, was when BC was used; therefore, we believe correntropy must be considered by the community as a most relevant tool in this problem.

Additionally, we also analyzed the influence of the kernel size $\sigma$ in the performance. The obtained results indicate that this parameter should be adapted according to the subject under observation, since the temporal brain response of each subject can vary considerably.

This initial study can be the basis for a deeper analysis of the use of information-theoretic learning in the context of brain-computer interfaces, at least when a graph-based approach is used for feature generation. Possible future steps could involve investigating how to process the signals in the spectral domain, using the definition of the *correntropy spectral density (CSD)*, which can be evaluated from the autocorrentropy function of a random process (stationary). A cross spectral density could also be defined by selecting two random stationary processes. This would imply that in addition to the temporal information considered so far to link different channels (electrodes), also a frequency related information could be used. Adding this information related to the frequency domain could also be valuable for analyzing MI-BCI, presenting an interesting point to be evaluated in upcoming works.

## Acknowledgements

The authors thank CAPES, CNPq, FINEP and FAPESP (grant 2013/07559-3) for the financial support.

## References

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller and T. M. Vaughan, *Brain-computer interfaces for communication and control.* Clinical neurophysiology, 113(6): 767–791, 2002.
[2] G. Pfurtscheller and C. Neuper, *Motor Imagery and Direct Brain–Computer Communication.* Proceedings of the IEEE, 89(7), 2001.
[3] Sivakami A., Devi S. S. *ANALYSIS OF EEG FOR MOTOR IMAGERY BASED CLASSIFICATION OF HAND ACTIVITIES.* International Journal of Biomedical Engineering and Science, 2(3), 2015.
[4] Cheng M., Jia W., Gao X., Gao S., Yang F. *Mu rhythm-based cursor control: an offline analysis.* Clinical Neurophysiology, 115: 745 - 751, 2004.
[5] Hamedi M., Salleh SnH., Noor AM. *Electroencephalographic Motor Imagery Brain Connectivity Analysis for BCI: A Review.* Neural Comput., 28(6): 999 - 1041, 2016.
[6] McFarland D.J., McCane L.M., David S.V., Wolpaw J.R. *Spatial filter selection for EEG-based communication.* Electroencephalography and Clinical Neurophysiology, 103(3): 386–94, 1997.
[7] Santamaria I., Pokharel P.P., and Principe J.C. *Generalized correlation function: definition, properties, and application to blind equalization.* IEEE Transactions on Signal Processing, 54(6): 2187–2197, 2006.
[8] Principe J.C. *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives. Edition 1.* Springer Publishing Company, Incorporated, 2010.
[9] Theodoridis S., Koutroumbas K. *Pattern Recognition*, fourth Edition. Academic Press, 2008.