DATA SCIENCE AWARDS SPAIN 2018

Project:

Detection of Alzheimer's Disease by means of Magnetoencephalography

Author:

Antonio Díaz Parra

Email: antodipar@gmail.com

Linkedin: https://www.linkedin.com/in/antodipar

Organizado por synergic partners A Telefónica COMPANY

Telefónica

Google Cloud

August 2018

Contents

1. Introduction

The brain is composed of massively connected elements arranged into modules that form hierarchical networks. The different modules do not operate in isolation; on the contrary, interactions at multiple levels occur giving rise to the characteristic fluctuations of brain activity. Theory shows that connecting networks yield an infrastructure in which small and local perturbations can be amplified in cascade, resulting in frequent catastrophic failures and wide-range alterations (Reis et al., 2014). Experimental evidence from humans and rodents reveals a well-defined connectivity design, characterized by the presence of strategically connected core nodes that critically contribute to resilience and maintain stability in interacting brain networks (Gallos et al., 2012; van den Heuvel and Sporns, 2013, 2011). These findings predict that modifying activity in this set of core nodes could drastically alter global patterns of brain activity; which, in turn, raises the possibility that certain brain pathologies might be a consequence of cascading maladaptive processes that alter normal connectivity (Aerts et al., 2016; Fornito et al., 2015; Stam, 2014). Interestingly, this network perspective of the brain is greatly contributing to our understanding of the mechanisms behind different neurological and psychiatric disorders, such as Alzheimer's disease (Deco and Kringelbach, 2014), and the great expectations raised in the field of network neuroscience motivated the adopted approach in this project.

2. Methodology and planning

The aim of this work is to generate a predictive model that enables distinguishing a group of healthy controls from a group of Alzheimer's disease patients. In particular, patients in a mild cognitive impairment (MCI) condition. To this end, I propose the approach presented in Fig. 1, which is described in depth in the rest of this report. Basically, the proposed approach is based on the idea that the brain can be modeled as a complex network of nodes (also called vertices) that interact with each other through links (also called edges). A network can be represented by a connectivity matrix with rows and columns encoding nodes, and matrix entries quantifying the connectivity between nodes. From this matrix, it is possible to extract a multitude of metrics that quantify the topology of the associated network. These topological measures can be then used as features for machine learning approaches.

Given the nature of the provided dataset, the recording sensors and the mean synchronization between every pair of sensors will act as the nodes and links of the brain networks, respectively. Sections 2 explains the process of brain network reconstruction from the original dataset and how analytical methods adopted from network science can be applied with the aim of generating a new set of features. Section 3 describes and justifies the technology employed in this project. Finally, section 4 explains and justifies how one can make use of network-based features for feature selection, classification and model evaluation.

3. Data description and processing

The original dataset, which is provided by the Laboratory of Cognitive and Computational Neuroscience, consists of 132 subjects, 54 of whom are healthy controls and 78 are MCI patients. All subjects underwent magnetoencephalography scanning with 102 sensors, and the synchronization between every pair of sensors was computed, hence obtaining 5,151 synchronizations. Five descriptive statistics were computed for each synchronization, namely, the mean, standard deviation, median, mean absolute deviation and covariance. Therefore, each subject (or instance) is originally characterized by a total of 25,755 (5,151 synchronizations × 5 measures) features.

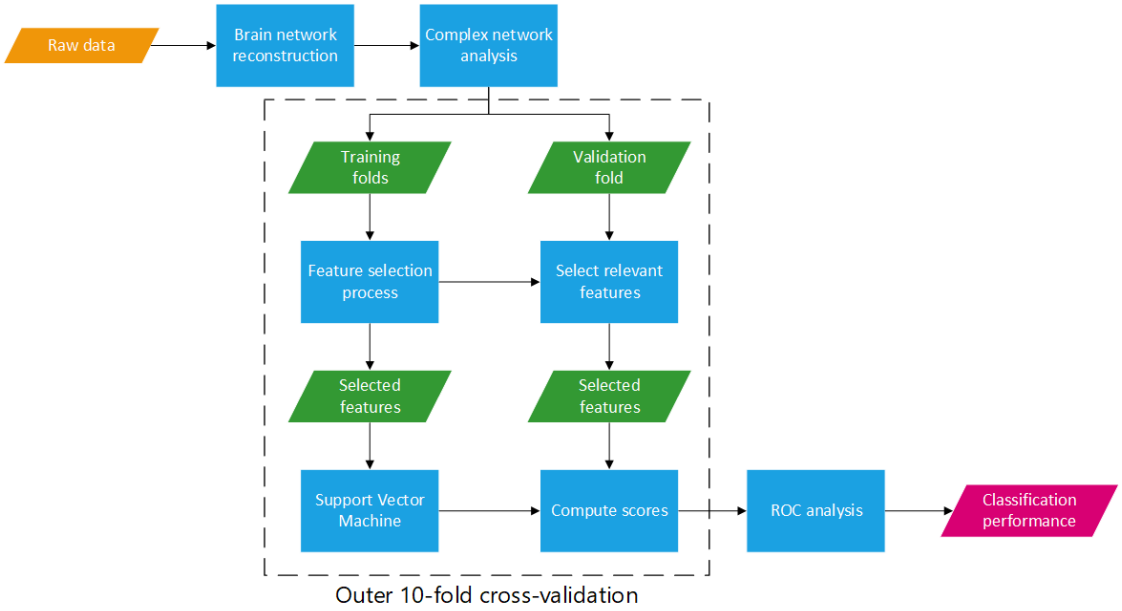

Figure 1. Basic schematic of the proposed approach. From the original dataset, brain networks are reconstructed and subsequently analyzed using complex network analysis. The resulting network measures are used as features for MCI diagnosis.

Out of the original 25,755 features, I retained the 5,151 features associated with the mean value of each synchronization, and subsequently reconstructed a subject-specific brain network. Centrality is a measure that quantifies the relative importance of a node within the overall architecture of a network (van den Heuvel and Sporns, 2013). **It has been suggested that the pathophysiological processes in Alzheimer's disease largely disturb brain nodes with a high centrality or hubs (Stam, 2014; Tijms et al., 2013).** Thus, four centrality measures were extracted from each individual node (or sensor), namely, the strength, closeness, betweenness and eigenvector. In addition, the clustering coefficient of each node was also computed. Therefore, a total of 510 (102 sensors × 5 network measures) features were generated for each sample.

Sensor metrics were scaled within each sample to have zero mean and unit variance. For example, for the strength of sensor i, k_i, feature scaling was performed using the mean, $\overline{k}$, and standard deviation, σ_k, of the strength distribution of the network:

$$k_i^z = \frac{k_i - \overline{k}}{\sigma_k}$$

This process can be also seen as a dimensionality technique reduction. Indeed, whereas the original dataset is composed of 25,755 features per subject, the process of network reconstruction and analysis generates a new dataset consisting of 510 features to describe each of the 132 subjects. **Given the connected nature of the brain, one would expect a great proportion of the original features being highly correlated. Thus, the proposed approach exploits the fact that the effects of Alzheimer's disease are not confined to a single connection (synchronization) and node (sensor). Rather, changes on brain networks caused by pathology are likely to encompass multiple connections and nodes (Zalesky et al., 2010).**

4. Justification of the design and implementation

I used a personal computer with the following specifications:

- Processor – Intel Core i7-4720HQ @ 2.60 GHz
- RAM – 16 GB
- Operating System – Ubuntu 14.04 LTS

With respect to the software, my code is based on Python (version 2.7.12) since it is widely used for data science application and there are a variety of well-established libraries available for data modeling. In particular, the following libraries were used for the implementation of the algorithms:

- Numpy (version 1.14.5)
- Scipy (version 1.1.0)
- Pandas (version 0.23.1)
- Sklearn (version 0.19.1)
- Networkx (version 2.1)

Note that nor big data technologies neither cloud services were used in this project. Given the limited sample size, a personal computer as employed in this work, together with the Python Programming Language, can cope with the technological requirements demanded in the proposed approach.

5. Justification of the analytical approach

After extracting the network-based features (section 2), I applied different machine learning tools to generate relevant brain patterns (or signatures) and a predictive model that allow classifying between controls and MCI patients.

For model evaluation, I split the dataset into 10 folds by using a 10-fold cross-validation scheme (outer 10-fold cross-validation), **as this technique has shown to have a reasonable bias-variance trade-off in the estimation of the test error (James et al., 2013)**. I further ensured that classes were roughly balanced within each fold (note that there are 78 MCI patients and 54 controls). I then implemented a feature selection process using merely the training folds (Fig. 1) as follows:

- The training dataset was further partitioned into 10 folds by using an inner 10-fold cross-validation procedure (Fig. 2).
- In each iteration of this scheme, I applied a filter method to quantify the relative importance of each feature using the training folds. **To avoid making any assumptions as to the distribution of each feature, I applied the Wilcoxon rank-sum test (or Mann-Whitney *U* test)**. In particular, the $1-P$-value resulting from the comparison of every single feature between controls and MCI patients was used as measure to determine which features have a relationship with the outcome (i.e., the class label). This process generated a ranking encoding the relative importance of each feature. **I opted for a univariate method since these methods are in general suitable for a better understanding of the relationship between the features and the class to be predicted; which, in turn, can lead to better decisions in a clinical context.**
- To identify the minimum number of features providing good predictions, lineal support vector machines (SVMs) with $C = 1 \times 10^{-2}$ were trained by adding features progressively (Fig. 2). That is, a first linear SVM was computed using the most relevant feature (according to the foregoing ranking). Next, a second linear SVM was obtained using the two most important features, sequentially adding features up to a total of 102

(20 % of the total number of features). Each of the 102 lineal SVMs was evaluated on the validation fold and the area under the curve (AUC) of the receiver operating characteristic (ROC) curve was recorded as a function of the number of features, producing a performance profile. **A small value of C was selected to further control for overfitting. Furthermore, The SVM classifier has been successfully applied in the context of diverse brain pathologies (Fagerholm et al., 2015; Klöppel et al., 2008).**

- Performance profiles and feature importance scores (quantified by 1−*P*-values) generated across the 10 iterations of the inner 10-fold cross-validation were averaged. Finally, I selected the optimal features providing de max AUC. **I made use of the AUC value to evaluate classification performance due to the unbalance of the class labels.**

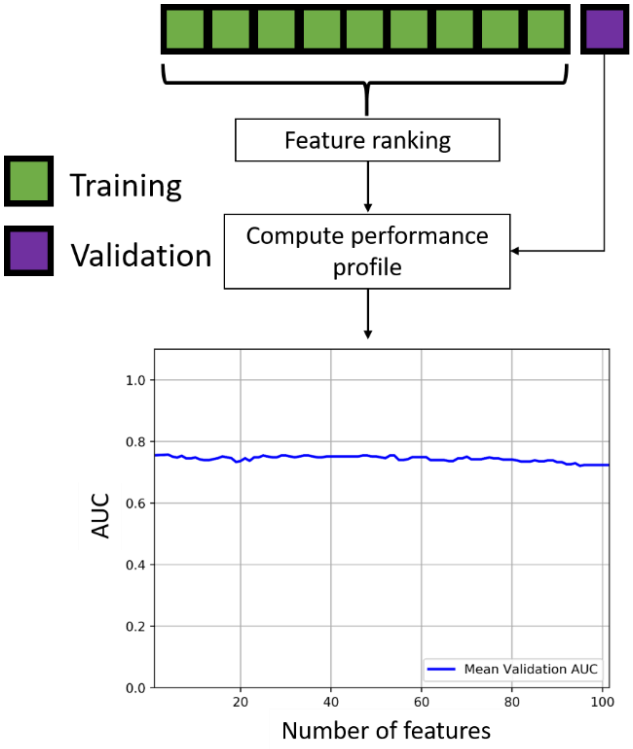

Figure 2. Inner 10-fold cross-validation for feature selection. After partitioning the input dataset into 10 folds, data from nine folds (in green) are used for feature ranking and model fitting. The remaining fold (in purple) is used for model validation, hence generating a performance profile: the classification performance (in term of AUC) as a function of the number of features, which are sorted according to their relative importance.

The just described feature selection process gives rise to the best subset of features in predicting the class label. This subset of features was used to fit a linear SVM with $C = 1 \times 10^{-2}$ in the outer 10-fold cross validation loop (Fig. 1). Table I presents the max AUC obtained in the inner 10-fold cross validation (feature selection process) and the number of features required to achieve such performance. The same subset of features was selected from the remaining validation fold and the trained linear SVM was tested on it. **Importantly, the validation fold is used for neither feature selection nor training, hence reducing the risk of overfitting.** Next, the scores provided by the SVM classifier were recorded for further analysis. Thus, after the application of the outer 10-fold cross-validation procedure, I performed a ROC analysis from the scores obtained across the 10 iterations (Fig. 3).

Table I. For each iteration of the outer 10-fold cross-validation, this table reports the max AUC obtained in the inner 10-fold cross-validation process as well as the optimal number of features for diagnosis.

Outer iteration	Max AUC (feature selection process)	Number of features
1	0.76	4
2	0.76	8
3	0.75	1
4	0.78	72
5	0.76	9
6	0.77	1
7	0.75	3
8	0.73	5
9	0.70	1
10	0.71	29
Mean	0.75	13.3

Table II reports the statistical significance of the AUC value, together with the sensitivity, specificity, and accuracy at the threshold for classification providing the best trade-off between sensitivity and specificity. I applied the one-tailed binomial test to evaluate the significance of sensitivity, specificity and accuracy indices (Pereira et al., 2009) and the non-parametric Wilcoxon rank sum test to evaluate the significance of the AUC value (Hanley and McNeil, 1982). The significant values of AUC and accuracy (P-value < 0.05) indicates that the classification outcome cannot be obtained by a random classifier and, therefore, that brain networks contain relevant information for MCI diagnosis.

It is important to acknowledge that the resulting classification performance is rather moderate (Table II). Therefore, its application in a real clinical scenario should be done with caution. Even so, there are some useful insights that can be extracted from this analysis. First, the similar values of AUC obtained with the training folds (mean AUC of 0.75, see Table I) and the validation fold (AUC of 0.71, see Table II) indicates that this approach makes a suitable control for overfitting and these results might be reproducible across samples. On the other hand, *which are the most discriminative features?* In other words, *which sensors are the most important for MCI diagnosis?* Note that the five network measures (strength, closeness, betweenness, eigenvector and clustering coefficient) were extracted from each of the 102 sensors of each subject. Table II indicates that only around 13 features are required for diagnosis. By counting the number of times that a particular features is selected across the iterations of the outer 10-fold cross-validation, Table III shows the 13 most discriminative features for distinguishing controls from MCI patients. These relevant features are associated with the recording sensors 5, 6, 10, 13, 38, 54, 25 and 44.

Finally, it would be possible to fit a final SVM model based on the features reported in Table III using the 132 subjects included in this study. This predictive model could be then applied on a new incoming sample to perform the corresponding diagnosis with a threshold of – 0.03 (optimal threshold identified in this study, see Table II). Additionally, it would be even more informative if one looked at the corresponding score and compared this value with the optimal threshold, rather than relying on a binary decision.

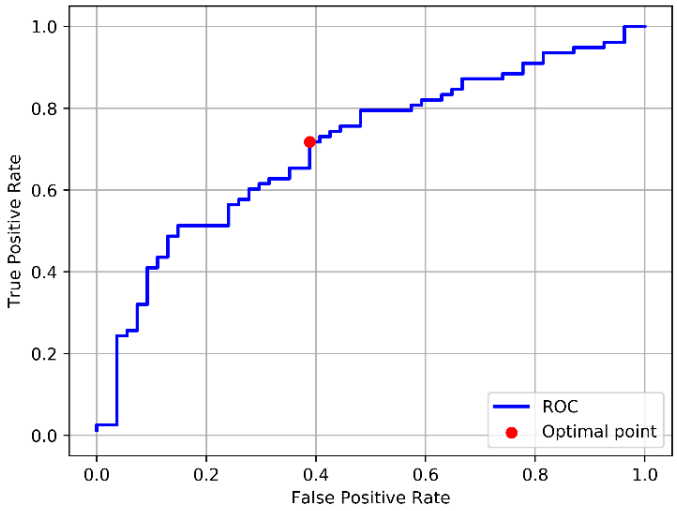

Figure 3. ROC analysis to evaluate classification performance. Sensitivity, specificity and accuracy indices were computed at the optimal point (red circle). Such point was identified as the closest point (in terms of Euclidean distance) on the ROC curve to the point defined by a true positive rate of 1 and false positive rate of 0.

Table II. Model performance at the optimal threshold for classification. The optimal threshold is represented by a score of –0.03 (note that the SVM classifier does not naturally provide class probabilities).

Measure	Value	*P*-value
AUC	0.71	2.86×10^{-5}
Sensitivity	71.79 %	7.47×10^{-5}
Specificity	61.11 %	6.68×10^{-2}
Accuracy	67.42 %	3.85×10^{-5}

Table III. Most important features in predicting MCI condition. Each feature consists of a sensor and the specific network measures associated with that sensor.

Position in the ranking	Sensor	Network measures
1	5	Betweenness
2	5	Closeness
3	6	Betweenness
4	10	Betweenness
5	5	Clustering
6	13	Betweenness
7	5	Strength
8	38	Strength
9	38	Eigenvector
10	5	Eigenvector
11	54	Strength
12	25	Eigenvector
13	44	Closeness

6. References

Aerts, H., Fias, W., Caeyenberghs, K., Marinazzo, D., 2016. Brain networks under attack: robustness properties and the impact of lesions. Brain 139, 3063–3083.

Deco, G., Kringelbach, M.L., 2014. Great expectations: using whole-brain computational connectomics for understanding neuropsychiatric disorders. Neuron 84, 892–905.

Fagerholm, E.D., Hellyer, P.J., Scott, G., Leech, R., Sharp, D.J., 2015. Disconnection of network hubs and cognitive impairment after traumatic brain injury. Brain 138, 1696–1709.

Fornito, A., Zalesky, A., Breakspear, M., 2015. The connectomics of brain disorders. Nat. Rev. Neurosci. 16, 159–172.

Gallos, L.K., Makse, H.A., Sigman, M., 2012. A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks. Proc. Natl. Acad. Sci. 109, 2825–2830.

Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, 29–36.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An introduction to statistical learning with applications in R, 1st ed. Springer, New York.

Klöppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack Clifford R., J., Ashburner, J., Frackowiak, R.S.J., 2008. Automatic classification of MR scans in Alzheimer's disease. Brain 131, 681–689.

Pereira, F., Mitchell, T., Botvinick, M., 2009. Machine learning classifiers and fMRI: a tutorial overview. Neuroimage 45, S199–S209.

Reis, S.D.S., Hu, Y., Babino, A., Andrade Jr, J.S., Canals, S., Sigman, M., Makse, H.A., 2014. Avoiding catastrophic failure in correlated networks of networks. Nat. Phys. 10, 762–767.

Stam, C.J., 2014. Modern network science of neurological disorders. Nat. Rev. Neurosci. 15, 683–695.

Tijms, B.M., Wink, A.M., de Haan, W., van der Flier, W.M., Stam, C.J., Scheltens, P., Barkhof, F., 2013. Alzheimer's disease: connecting findings from graph theoretical studies of brain networks. Neurobiol. Aging 34, 2023–2036.

van den Heuvel, M.P., Sporns, O., 2013. Network hubs in the human brain. Trends Cogn. Sci. 17, 683–696.

van den Heuvel, M.P., Sporns, O., 2011. Rich-club organization of the human connectome. J. Neurosci. 31, 15775–15786.

Zalesky, A., Fornito, A., Bullmore, E.T., 2010. Network-based statistic: identifying differences in brain networks. Neuroimage 53, 1197–1207.