ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6.
Available from http://www.i6doc.com/en/.

Graph Based Neural Networks for Automatic Classification of Multiple Sclerosis Clinical Courses

Francesco Calimeri[1], Aldo Marzullo[1], Claudio Stamile[2], Giorgio Terracina[1] *

1 - Department of Mathematics and Computer Science
University of Calabria - Italy

2 - CREATIS; CNRS UMR5220; INSERM U1044;
Université de Lyon, Université Lyon 1, INSA-Lyon, Villeurbanne - France

Abstract. Automatic classification of biomedical imaging became an important field of research within the scientific community, in the latest years. Indeed, advances in image acquisition and processing techniques, along with the success of novel deep learning methods and architectures, represented a considerable support in providing better biomarkers for the characterization of several diseases, and brain diseases in particular. In this work we propose a novel neural network approach that is applied to graphs generated from MRI data in order to make predictions about the clinical status of a patient. Results show high performances in classification tasks and open interesting perspectives in the field.

1 Introduction

Analyzing and processing Magnetic Resonance Imaging (MRI) data is a very important task in biomedical imaging: pathological mechanisms in different brain diseases still remain unknown, and hence the investigation of their temporal progression by means of noninvasive techniques, such as neuroimaging, is crucial for a better understanding, and then prediction, of disease evolutions, as well as for a proper management of therapeutic treatments [13, 15]. In the last decade, new approaches have been proposed to study and quantify damages in the brain white matter (WM) network; the most successful, so far, is based on the analysis of WM network by means of tools and techniques from the graph theory [10]. WM networks consist of nodes, corresponding to segmented cortical regions, and links, reconstructed by tractography from White Matter (WM) fibers-tracts.

Complex network analysis allows to describe highly-structured data simply via geometric representations [12]; similar models have already been used for studying social behaviors, and recently opened new perspectives in neuroscience for the study of functional and structural brain connectivity using graph-derived metrics. This new approach paved the way to the characterization of either cognitive impairments or pathological alterations caused by different brain diseases, including Multiple Sclerosis (MS).

*The work is partially funded by an European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 690974 and by the Italian Ministry of University and Research under project "Dottorato innovativo a caratterizzazione industriale" PON R&I FSE-FESR 2014-2020.

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

In the last decade, neural networks (NN) have been proven capable of extracting highly meaningful statistical patterns in large-scale and high-dimensional datasets, showing great power and flexibility by learning to represent data as a nested hierarchy of concepts. In this exciting scenario, peculiar type of architectures achieved outstanding results in many different applications [7]. In particular, in the past few years, a significant number of research works have been devoted to encode the graph structure directly using a neural network model [5, 8, 11]. These kind of techniques showed promising results in capturing temporal information and properly managing structured data.

In this work we propose a method for the classification of Multiple Sclerosis (MS) patients into four clinical profiles ([Clinical Isolated Syndrome (CIS), Relapsing Remitting (RR), Secondary Progressive (SP), Primary Progressive (PP)]) by using structural connectivity information related to WM networks. The main goal is to take advantage from structural brain connectivity and Neural Networks for a better characterization and classification of four MS clinical profiles. To the best of our knowledge this is the first attempt to classify MS patients by considering together the four classes.

The rest of the paper is structured as follows. In Section 2 we provide a detailed description of our approach; we then present experimental classification results in Section 3 and discuss them in Section 4. Eventually, we draw our conclusions in Section 5.

2 Neural Networks for the Classification of MS Patients

The approach herein proposed exploits structural representations of MR images which are extracted and used to train a NN to classify MS patients. In the following, we provide a detailed description of the method.

2.1 Brain Structural Connectivity Graph

The connectivity graph of each subject is obtained by applying the method originally proposed in Kocevar *et al.* [6] and Stamile *et al.* [14]. For completeness, we briefly recall the mentioned approach.

In a first phase, FreeSurfer [1] is used to obtain cortical and sub-cortical parcellations from T1-weighted MR images. Then, using the resulting segmentation, the graph nodes are defined, and a label corresponding on the five tissue-type (cortical grey matter (GM), sub-cortical GM, WM, cerebrospinal fluid (CSF) and abnormal appearing tissue) was assigned to each voxel. Each image is then pre-processed using correction of Eddy-current distortions [3] and skull stripping. MRtrix spherical deconvolution algorithm [16] is applied to compute main diffusion directions per each voxel. Spherical harmonic (order of $L_{max} = 4$) is used to estimate both orientation distribution function (ODF) and response function. Probabilistic streamline tractography algorithm [16] is applied to generate fibertracks in voxels labeled as WM voxels. The combination of GM segmentation and WM tractography produces, for each subject, a symmetrical connectivity

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6.
Available from http://www.i6doc.com/en/.

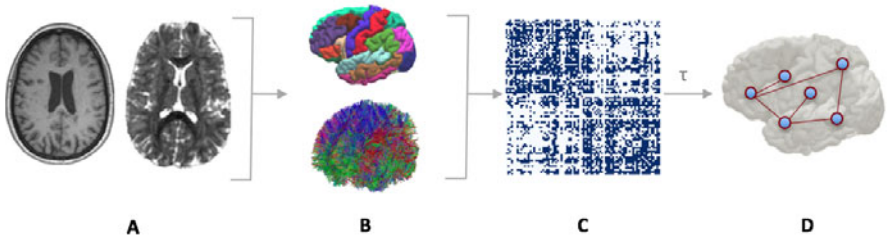

Fig. 1: Illustration of different graph creation steps: (A) T1 and diffusion weighted MR images are used to generate cortical parcellation and fiber tractography (B), which are combined to generate connectivity matrix (C). Finally, a threshold τ is applied to the connectivity matrices to generate adjacency matrices (D)

matrix $A \in \mathbb{N}_{+}^{q \times q}$. A representation of graph construction workflow is reported in Figure 1.

Let $\Psi : \mathbb{N}_1^2 \rightarrow \mathbb{N}$ indicating number of fibers connecting two nodes. Then, each element of A is defined as $a_{i,j} = \Psi(i,j)$. In particular, the connectivity matrix A represents the weighted undirected graph $G = (V, E, \omega)$ where V ($|V| = q$) is the set containing the segmented GM brain regions, E is the graph edges set defined as:

$$E = \{\{i,j\} \mid \Psi(i,j) > 0 \; \forall \; 1 \leq i,j \leq q\}$$

Finally, the graph binarization function $\Upsilon : G \rightarrow G^{'}$ is applied to the weighted undirected graph $G = (V, E, \omega)$ to generate an unweighted undirected graph $G^{'} = (V^{'}, E^{'})$. More in detail, the graph binarization function $\Upsilon : G \rightarrow G^{'}$ performs the following mapping:

$$V^{'} = V \quad E^{'} = L(1, \ldots, T), \; T = \frac{(q^2 - q)\tau}{2}$$

where L is the list of graph edges (E) sorted in ascending order of weight. In other words, the function Υ creates, starting from a weighted graph, an unweighted graph containing only the T strongest connections of G with respect to a given threshold $\tau \in \mathbb{R}_{[0,1]}$.

2.2 Graph-Based Neural Networks

In this work, we take advantages from one of the architectures originally proposed by Manessi *et al.* [8] in order to deal with graph-focused applications. In this context, indeed, the goal is to learn a function f which maps each graph to integers values. This task can usually be accomplished by using a supervised classification approach on the graph structures [8]. The model we used for the classification task is composed by stacking together two layers, namely the *Vertex Sequential Fully Connected* (vs-FC) layer and the *Graph Sequential Fully*

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

Connected (gs-FC) layer: it takes as input a graph and the corresponding vertex features, and performs a supervised classification of the graph in four MS clinical profiles. In order to provide a description of each unit, we will denote by $[\mathbf{Y}]_{i,j}$ the i-th row, j-th column element of the matrix $\mathbf{Y}$, and by $\mathbf{Y}^{'}$ the transpose of $\mathbf{Y}$; also, we denote by *softmax* and $ReLU$ the soft-maximum and the rectified linear unit functions (see [2]). $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is an undirected graph, where $\mathcal{V}$ is the set of vertex and $\mathcal{E}$ is the set of edges and, for each vertex $v^k \in \mathcal{V}$, $\mathbf{x}^k \in \mathbb{R}^d$ is the associated *feature vector*.

Vertex Sequential Fully Connected. Consider $\mathbf{Z} \in \mathbb{R}^{L \times N}$. The *Vertex Sequential Fully Connected layer* (vs-FC layer) with c output nodes is given by the function vs-FC$_c$ parametrized by the weight matrix $\mathbf{W} \in \mathbb{R}^{N \times c}$ and the bias matrix $\mathbf{B} \in \mathbb{R}^{L \times c}$:

$$\text{gs-FC}_c : \mathbf{Z} \to softmax(\mathbf{WZ} + \mathbf{B})$$

with $(softmax(\mathbf{WZ} + \mathbf{B})) \in \mathbb{R}^{L \times c}$. The vs-FC layer is actually a Fully Connected layer [2] with softmax activation. The vs-FC layer outputs L c-class probability vectors.

Graph Sequential Fully Connected. Consider $\mathbf{Z} \in \mathbb{R}^{L \times N}$. The *Graph Sequential Fully Connected layer* (gs-FC layer) with k output nodes is given by the function gs-FC$_k$ parametrized by the weight matrices $\mathbf{W}_1 \in \mathbb{R}^{N \times k}$, $\mathbf{W}_2 \in \mathbb{R}^{1 \times L}$ and the bias matrices $\mathbf{B}_1 \in \mathbb{R}^{L \times k}$and $\mathbf{B}_2 \in \mathbb{R}^{1 \times k}$:

$$\text{vs-FC}_k : \mathbf{Z} \to softmax(\mathbf{W}_2 ReLU(\mathbf{W}_1 \mathbf{Z} + \mathbf{B}_1) + \mathbf{B}_2)$$

with $(softmax(\mathbf{W}_2 ReLU(\mathbf{W}_1 \mathbf{Z}_i + \mathbf{B}_1) + \mathbf{B}_2)) \in \mathbb{R}^{1 \times k}$. The gs-FC layer acts as two FC layers with softmax-ReLU activation. This layer outputs one k-class probability vector.

It is worth noting that the architectures proposed in Manessi *et al.* are here extended in order to deal with several time steps; in this work, only one time step was used for the classification task. Furthermore, we perform a featureless approach, so that the input of the model is the adjacency matrix $\mathbf{A} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$.

3 Results

We carried out our experiment on a dataset consisting of 90 MS patients, distributed into the four classes as 12 CIS, 30 RR, 28 SP, 20 PP, respectively. Multiple examination at different time points were collected per each subject. However, we considered each of the time steps as a single data point, such that the dataset is actually composed by 578 samples (63 CIS, 199 RR, 190 SP, 126 PP). $T1$ and diffusion tensor imaging (DTI) were used to obtain structural connectivity matrices for each subject. 80% and 20% of the dataset was used as training and test set respectively. The NN has 84×84 input nodes, followed by a vs-FC and a gs-FC layer of 50 nodes each. Dropout (factor 0.6) between

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

	Precision ($\pm stdev$)	Recall ($\pm stdev$)	F-Measure ($\pm stdev$)
CIS	0.64 ($\pm$0.17)	0.82 ($\pm$0.19)	0.71 ($\pm$0.16)
PP	0.71 ($\pm$0.16)	0.79 ($\pm$0.12)	0.74 ($\pm$0.11)
RR	0.82 ($\pm$0.09)	0.80 ($\pm$0.10)	0.80 ($\pm$0.07)
SP	0.87 ($\pm$0.12)	0.81 ($\pm$0.10)	0.84 ($\pm$0.09)
Tot	**0.82** ($\pm$0.07)	**0.79** ($\pm$0.07)	**0.80** ($\pm$0.07)

Table 1: Average Precision, Recall and F-Measure ($\pm$ standard deviation) achieved during cross validation (10 folds). Results are computed per class (CIS-PP-RR-SP) and with respect to all the classes (Tot)

vs-FC and gs-FC was used and $l2$ regularization with a factor of 0.01 to reduce overfitting. The parameter τ was set to 0.15 after a grid search over different threshold values (from $\tau = 0.05$ to $\tau = 1$ with 0.05 steps). Adamax [4] optimizer was used for training with learning rate 0.002 over 100 epochs. Because of the limited size of the dataset due to the high costs required to perform screenings, cross validation with 10 folds was used to provide a more robust evaluation of the model. Quality of the classification was evaluated by means of Precision, Recall and F-Measure [9].

4 Discussion

Results of our experiments are reported in Table 1: our model shows high levels of accuracy on the selected task. Indeed, the experiments show that the herein proposed model allows to classify MS patients with an average F-Measure of 80% on the test set, which is a promising result based on the 70.6% reached by [6] working on three classes only, even if not directly comparable. Furthermore, these results were achieved after about 100 epochs with a limited amount of data, showing that NNs combined with structural representation of brain networks offers new opportunities to identify potential biomarkers for the characterization of global as well as local effects of pathological mechanisms on brain networks.

Nevertheless, it is worth noting that it can be observed how the level of accuracy significantly decreases when classifying CIS patients, where the overall F-Measure is also lower with respect to the other classes. It is reasonable to think that the main reason relies in the small amount of cases in the dataset used for the training; indeed, it might not be sufficiently representative of this class (which is the minority one among the four). Hence, we can conclude that the misclassification might be solved by training the network on a larger dataset.

5 Conclusion

In this work we proposed a graph-based method for the classification of multiple sclerosis patients according to their clinical forms. We make use of proper techniques from graph theory to describe brain network topology, and then employ

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

Graph-based NNs to classify each subject. Experimental results show the capability of NNs of capturing information useful for the classification of MS patients directly from graph structures, achieving promising results. As future work, we aim at improving results by using the whole trail of longitudinal data, collected from each patient, as input for the model, also trying to take advantage from different novel models proposed in literature.

References

[1] B. Fischl, A. Van der Kouwe, C. Destrieux, E. Halgren, F. Segonne, D.H. Salat, E. Busa, L.J. Seidman, J. Goldstein, D. Kennedy, V. Caviness, N. Makris, B. Rosen, A.M. Dale, "Automatically parcellating the human cerebral cortex. Cerebral Cortex," vol. 14, pp. 11-22, 2004.

[2] I. Goodfellow, Y. Bengio, A. Courville, "Deep Learning," MIT Press, 2016.

[3] M. Jenkinson, C.F. Beckmann, T.E.J. Behrens, M.W. Woolrich, S.M. Smith, "FSL. Neuroimage," vol. 62(2), pp. 782-790, 2012.

[4] D. Kingma, J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2016.

[5] T.N. Kipf, M. Welling, "Semi-supervised classification with graph convolutional networks," arXiv preprint arXiv:1609.02907, 2016.

[6] G. Kocevar, C. Stamile, S. Hannoun, F. Cotton, S. Vukusic, F. Durand-Dubief, D. Sappey-Marinier, "Graph theory-based brain connectivity for automatic classification of multiple sclerosis clinical courses," Frontiers in neuroscience, vol. 10, 2016.

[7] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86(11), pp. 2278-2324, 1998.

[8] F. Manessi, A. Rozza, M. Manzo, "Dynamic Graph Convolutional Networks," arXiv preprint arXiv:1704.06199, 2017.

[9] D. Powers, "Evaluation: from precision, recall and f-factor to roc, informedness, markedness & correlation (Technical report)," Adelaide (Australia), 2007.

[10] M. Rubinov, O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," Neuroimage, vol. 52, pp. 1059-1069, 2010.

[11] F. Scarselli, M. Gori, A.C. Tsoi, M. Hagenbuchner, G. Monfardini, "The graph neural network model. IEEE Transactions on Neural Networks," vol. 20(1), pp. 61-80, 2009.

[12] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, P. Vandergheynst: "The Emerging Field of Signal Processing on Graphs," IEEE signal processing magazine. vol. 5, pp. 83-98, 2013.

[13] C. Stamile, G. Kocevar, F. Cotton, F. Durand-Dubief, S. Hannoun, C. Frindel, D. Sappey-Marinier, , "A longitudinal model for variations detection in white matter fiber-bundles," in Proc. Int. Conf. Syst., Signals Image Process., pp. 57-60, 2015.

[14] C. Stamile, G. Kocevar, S. Hannoun, F. Durand-Dubief, D. Sappey-Marinier, "A Graph Based Classification Method for Multiple Sclerosis Clinical Forms Using Support Vector Machine," Medical Learning Meets Medical Imaging, Springer, pp. 57-64, 2015.

[15] C. Stamile, G. Kocevar, F. Cotton, F. Maes, D. Sappey-Marinier, S. Van Huffel, "Multiparametric Nonnegative Matrix Factorization for Longitudinal Variations Detection in White-Matter Fiber Bundles," IEEE journal of biomedical and health informatics, 2016, *in press.*

[16] J. Tournier, F. Calamante, A. Connelly, "MRtrix: Diffusion tractography in crossing fiber regions," IJIST. vol. 22(1), pp. 53-66, 2012.