EEG-driven RNN Classification for Prognosis of Neurodegeneration in At-Risk Patients

Conference Paper · October 2016

DOI: 10.1007/978-3-319-44778-0_36

5 authors, including:

Giulio Ruffini
Starlab Barcelona SL / Neuroelectrics Corporation

Ibanez-Soria David
Starlab Barcelona SL

Marta Castellano
Starlab Barcelona SL

Stephen Dunne
Starlab Barcelona SL

EEG-driven RNN classification for prognosis of neurodegeneration in at-risk patients

Giulio Ruffini*, David Ibañez, Marta Castellano, Stephen Dunne, and Aureli Soria-Frisch

Starlab Barcelona / Neuroelectrics,
Avda. Tibidabo 47 bis, 08035 Barcelona, Spain
{giulio.ruffini,david.ibanez,stephen.dunne,aureli.soria-Frisch}@starlab-int.com
http://starlab-int.com,http://neuroelectrics.com

Abstract. REM Behavior Disorder (RBD) is a serious risk factor for neurodegenerative diseases such as Parkinson's disease (PD). We describe here a recurrent neural network (RNN) for classification of EEG data collected from RBD patients and healthy controls (HC) forming a balanced cohort of 118 subjects in which 50% of the RBD patients eventually developed either PD or Lewy Body Dementia (LBD). In earlier work [1, 2], we implemented support vector machine classifiers (SVMs) using EEG mean spectral features to predict the course of disease in the dual HC vs. PD problem with an accuracy of 85%. Although largely successful, this approach did not attempt to exploit the non-linear dynamic characteristics of EEG signals, which are believed to contain useful information. Here we describe an Echo State Network (ESN) classifier capable of processing the dynamic features of EEG power at different spectral bands. The inputs to the classifier are the time series of 1 second-averaged EEG power at several selected frequencies and channels. The performance of the ESN reaches 85% test-set accuracy in the HC vs. PD problem using the same subset of channels and bands we selected in our prior work on this problem using SVMs.

Keywords: Echo State Networks, RNNs, EEG, Parkinson's Disease, Reservoir computing

1 Introduction

The human brain can be modeled as a highly dimensional complex dynamical system in which electrochemical communication and computation play a central role. Electroencephalographic (EEG) and magnetoencephalographic (MEG) signals contain rich information associated with these processes. To a large extent, progress in the analysis of such signals has been driven by the study of classical temporal and spectral features in electrode space, which has proven useful to

* This work partly supported by the Michael J. Fox Foundation within the project "Discovery of EEG biomarkers for Parkinson Disease and Lewy Body Dementia using advanced Machine Learning techniques".

study the human brain in health and disease. For example, the "slowing down" of EEG is known to characterize neurodegenerative diseases [3, 4]. However, brain activity measurements exhibit non-linear dynamics and non-stationarity across temporal scales that cannot be addressed well by classical, linear approaches. The complexity of these signals calls for the use of novel tools capable of exploiting such features and representing rich spatio-temporal hierarchical structures. Interestingly, deep learning techniques in particular and neural networks in general are bio-inspired by the brain —the same biological system generating the electric signals we aim to decode. They should be well suited for the task.

Here we explore a particular class of recurrent neural networks (RNNs) called Echo State Networks (ESNs) that combine the power of RNNs for classification of temporal patterns and ease of training. RNNs and, in particular, ESNs implement non-linear dynamics with memory and seem ideally poised for the classification of complex time series data. The main concept in ESNs and related types of so-called "reservoir computation" systems is to have data inputs drive a semi-randomly connected, large, fixed recurrent neural network (the "reservoir") where each node/neuron in the reservoir is activated in a non-linear fashion— see Figure 2 . The interior nodes with random weights constitute what is called the "dynamic reservoir" (DR) of the ESN. The motivation for keeping interior connection weights random but fixed (not to be learned) is, on the one hand, to allow for high dimensional feature mapping of the inputs (in a sense much like a kernel method) while, on the other, to avoid the complex problem of training recurrent neural network architectures (such as the vanishing of training error gradients [5]). An important feature of ESNs is that only the output weights (and various hyperparameters) are trained [7, 8]. Although we explore ESN architectures here, other relevant RNN options include long-short term memory networks (LSTMs) [6].

2 The dataset

The data in this study consisted of resting-state EEG collected from awake patients using 14 scalp electrodes [4]. The recording protocol consisted of conditions with periods of "eyes open" of variable duration (~ 2 minutes) followed by periods of "eyes closed" in which patients were not asked to perform any particular task. EEG signals were digitized with 16 bit resolution at a sampling rate of 256 S/s. The amplification device implemented hardware band pass filtering between 0.3 and 100 Hz and notch filtering at 60 Hz to minimize the influence of power line noise. All recordings were referenced to linked ears. The dataset includes a total of 59 patients diagnosed of REM (random eye movement sleep) Behavioral Disorder (RBD) and 53 healthy controls without sleep complaints in which RBD was excluded. EEG data was collected in every patient at baseline, i.e., when they were still RBD. After 1-10 years of clinical follow-up, 14 patients developed Parkinson disease (PD), 14 Lewy body dementia (LBD) and the remaining 31 remained idiopathic. The data was collected by the Hopital du Sacre-Coeur, Montrèal [4]. Our classification efforts here focus on the HC *vs.*

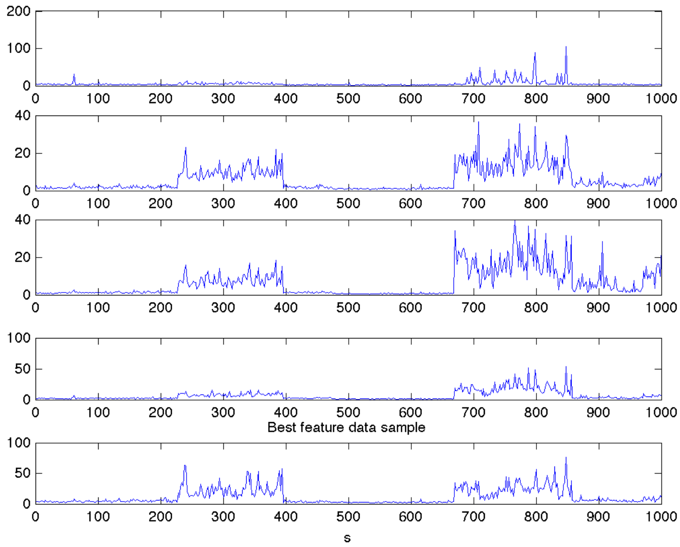

Fig. 1. Example of ESN input data. Five streams of EEG power time series at different bands and scalp channels. From top to bottom: Delta-T4 power, Theta-F8 power, Theta-T4 power, Theta-F7 and Theta-F4 power (horizontal units are samples corresponding to 1 second window power averages sliding every 0.5 seconds).

PD dual problem involving the available 14 PD converters and 14 HCs randomly selected for each classification train/test cycle. Each data snipped per subject contains information of power in 10 bands, 14 elecrtrodes and about 200 samples

EEG feature time series to feed the ESN (typically five channel-band signal streams—see Figure 1) were extracted after manual quality control and artifact correction of the data. Only eyes-closed sequences were considered for further analysis. Here we computed essentially a spectrogram per subject to extract temporal series of power for each electrode and band. In particular, we used a set of features selected in prior work for average power SVM classification [1] which include the combination of delta and theta band power from frontal and temporal channels. While we explored the use of 4 second and 1 second spectrogram windowing, the latter provided superior performance. As we discuss below, we hypothesize that this improvement reflects a better capture of relevant signal dynamics which can be used by the ESN. Our present study aimed to explore whether there is useful dynamic information on the EEG power data time series from subjects. We did not carry out an exhaustive test of the performance of classifiers using multiple feature and channel combinations, which is left for future work.

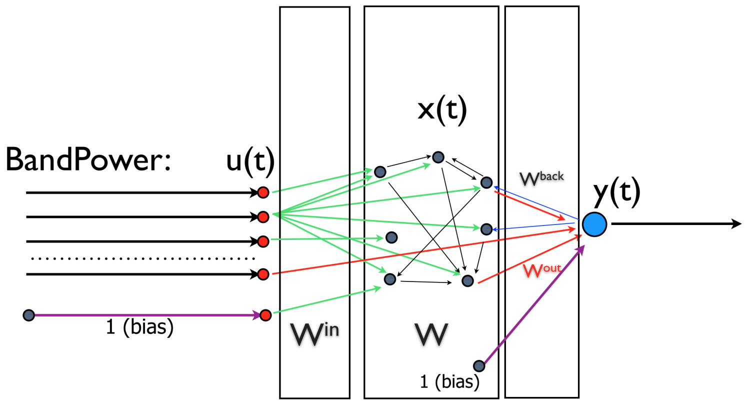

Fig. 2. ESN model displaying input, internal and output nodes and bias terms.

3 Echo State Network description

Following [7] and [8] we have designed ESNs driven by multi-channel temporally-varying power data as inputs, and providing as desired output the class label (after mapping it to a square like signal taking values in $\{-0.5, 0.5\}$, see Figure 3). The ESN node dynamics are captured by the state variable $x \in \mathbb{R}^D$, and driven by inputs $u \in \mathbb{R}^{N_{in}}$. Evolution is described iteratively by

$$x(n+1) = \alpha_{leak}\, f(W^{in}[u(n+1); 1] + Wx(n) + W^{back}y(n) + noise) \\ +(1-\alpha_{leak})\, x(n) \quad (1)$$

and

$$y(n) = f(W^{out}[u(n); x(n); 1])\,, \quad (2)$$

with $f(x) = \tanh x$ (point by point hyperbolic tangent) [7, 8]— see Figure 2. The internal weight matrices are initialized semi-randomly using a sparsity criterion and an important parameter—called α_W here and usually known as the *spectral radius*—that determines the damping of the system. We explored ranges of $\alpha_W =$ 0.5 to 2. We note that while it is normally stated that values less than one are required for a stable ESN, greater values applicable in some cases when the ESN inputs are non-zero [9]. W^{out} is computed using regularized least squares to ensure a match of output with the target signal in the training phase.

Overall, the parameters in our implementation of the ESN model are:

- D: DR dimension / the number of internal nodes. In our problem, the best performances were obtained with reservoirs with 3000 nodes (the maximum we tested).
- Sparsity threshold S: enforces sparsity in the W matrix when it is first created using a random uniform distribution in the range ± 1. Values with absolute value below the threshold are set to 0.

- α_W: spectral radius. Used to map the thresholded random $W \to \alpha_W\, W/S(W)$, where $S(A)$ denotes the spectral radius of a matrix A. Defines the "gain" of the DR.
- α_{in}: sets the scale of random input connections W_{in} to the range of $\pm\frac{1}{2}\alpha_{in}$.
- α_{back}: sets the scale of random connections of the feedback teacher signal W_{back} to the range of $\pm\frac{1}{2}\alpha_{back}$.
- α_{leak}: leaking rate of the neurons, useful for time smoothing of the dynamics.
- λ: Tickhonov parameter for regularization of the inverse problem of the output weights W_{out} (in the sense of modifying least squares cost to $F(w) = \frac{1}{\sigma}||Aw - y||^2 + \lambda||w||^2$).
- Input Channels and bands: The selection of N_{input} input time series.
- Initial conditions of the nodes of the ESN: $x(t_0)$.
- Training noise level: *noise*

Different parameter configurations have been tested, with emphasis on the relevance of spectral radius, DR size, the role of feedback teacher signal W_{back} and the initial conditions of the nodes of the ESN $x(t_0)$. In particular, three scenarios are discussed here after the ESN has been trained to assess the stability of the network following the analysis in [7]:

- Teaching signal (output for feedback) 'on' for the whole duration of ESN run (see Figure 3 first row). All other parameters of the ESN are kept constant, including the initial conditions of the nodes of the ESN.
- Teaching signal 'on' for a short interval, then turned off and provided by the ESN output (see Figure 3 second row). All other parameters of the ESN are kept constant, including the initial conditions of the nodes of the ESN.
- ESN provides its own feedback all the time (see Figure 3 third row). All other parameters of the ESN are kept constant, except the initial conditions of the nodes of the ESN.

4 Classification performance assessment

In order to map out the classification performance of the ESN for different parameter sets, we implemented a set of algorithms in Matlab (run on a MacBook Pro laptop) as described by the following pseudocode:

```
    FOR each parameter set:
             REPEAT M times (runs):
                   A- Create (random, balanced) training and test sets
                   B- Find a good DR (W_out matrix) with respect to the
                      training set (see below)
                   C- Evaluate its performance on training and test set
             END
           Compute mean performances over the M runs, save
    END
Provide a map of the saved mean performances in parameter space
```

For each run, 10% of the BandPower data of the PD and HC groups (subject-wise) was left out as a test set (step A above). Cycles of 50 iterations of *leave-10%-out* train/test were carried out to obtain an estimation of the classification performance for each set of parameters. Typically, classification accuracy in the training set was near 95-100% across all parameter configurations.

To select a good ESN DR (i.e, the W_{out} matrices) using the *training data* (step B above), the following steps were employed:

1. The DR network is initialized with random weights for all weight matrices (except output connections W_{out}).
2. In the parameter configuration where the feedback teacher signal W_{back} is on, the network is "forced" with the desired *teacher* signal (with values of $W_{back} = \pm\, 0.5$), and the associate state dynamics saved. Otherwise, the state dynamics of the network with no feedback are saved.
3. The best output connections W_{out} are found in the optimized, Tikhonov regularized least squares sense by comparing the output dynamics with the desired teacher signal. This provides the following quality metrics: the (L2) mean squared error (MSE) of output to teacher, and an accuracy metric. The accuracy per subject is measured by finding which teacher signal the output is closest to in the L2 sense.
4. Once the W_{out} are computed, a second run is then carried out to extract the training MSE and accuracy.
5. The above process is repeated N times or until good training set accuracy is provided, to find a good W_{out} matrix set based on the prior step. The best W_{out} and its corresponding parameter set are saved, and the ESN classifier is thus fully defined.

Finally, the performance of the ESN classification is estimated the test set with the selected W_{out} and its corresponding parameter set on the *testing data* (step C above).

5 Discussion

A first observation is that —as proposed in [8]—our best results were obtained with large dynamic reservoirs (D=3000) with least-squares regularization, with accuracies reaching an 85% average on the test set. In addition, spectral radii larger than one were effective as well—with good results with $\alpha_W \sim 2$—and feedback did not seem to play an important role in the problem.

These early results are promising, as is the fact that the match of ESN outputs to target was excellent on training set and often on test sets (see Figure 3 bottom row for an example of 100% classification performance on the test set). However, from this alone it does not follow that the ESN was actually exploiting dynamical features in the data streams (which is what we wished to demonstrate). One simple way to test if this is happening is to reshuffle the data time-wise (within each subject) to see if classification accuracy is affected. We indeed found that classification performance degraded if the data was reshuffled in time (85/55%

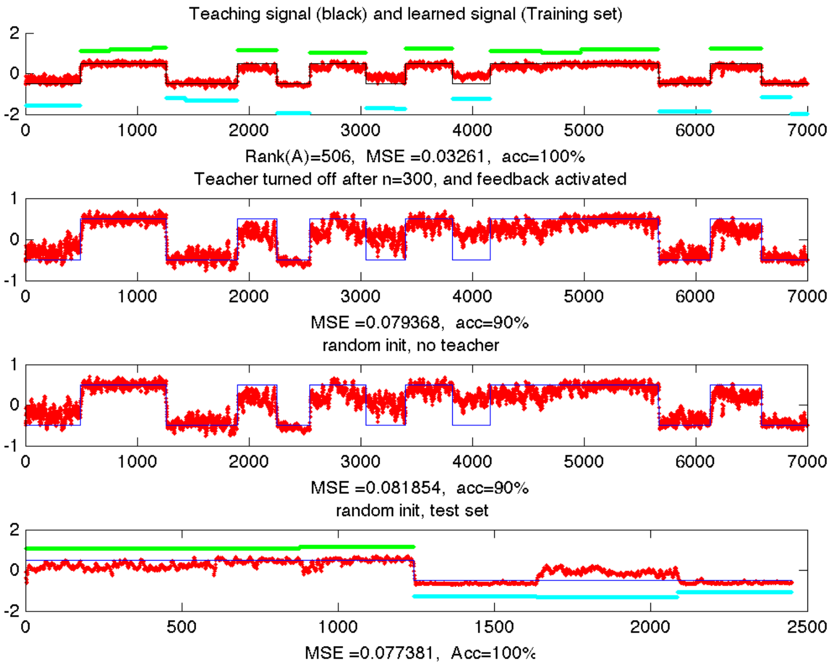

Fig. 3. Sample pot of the match of ESN output to desired target with accuracy for the trained signal (first three plots in sequence: ESN with teacher forcing, ESN with teacher forcing discontinued after N=300, and ESN with no teacher forcing at any time) and for the test set signal in the last row. The cyan and green lines denote the rescaled subject ID (cyan is HC, green PD). The horizontal axis denotes sample.

train/test accuracy). While suggestive, however, this may simply highlight the fact that ESNs require some "smooth" input dynamics to emulate data streams at all.

A better test to check for the used amount of information in the dynamics as opposed to mean amplitude is to normalize each input stream—independently per channel and per subject—to unit standard deviation. This is rather extreme—it would make any spectral power based classification impossible. We found that this did indeed cause a degradation of performance both for training and test sets, but not as much as temporal reshuffling (with train/test accuracy up to 95/65%). Thus, although mean power amplitude information is used by the network, we conclude that the ESN is also using dynamic information in the inputs. Along these lines, it is especially interesting to note that we also saw an improvement in performance using 1 second vs. 4 second sliding windowed data (about 10%), with accuracy consistently in the range 80-85%. This result is rather interesting in itself and suggests we could get additional classification performance by fusing ESN classifiers with SVM-spectral ones.

Finally, we note that in this particular study we did not carry out an exhaustive search in the feature space of channels and bands, but relied on the best performing features found in prior SVM classification studies which were blind to dynamically encoded information. More tests exploring the feature space and different feature combinations should therefore be carried out even if the search is restricted to bandpower signals (not the only choice with EEG data). Future work should also explore the role of dynamic reservoir architecture, which with large dimensions ought to be studied as a complex network [10].

Acknowledgments. This work has been partially funded by The Michael J. Fox Foundation for Parkinsons Research under Rapid Response Innovation Awards 2013. The data was collected by Jacques Montplaisir's team at the Center for Advanced Research in Sleep Medicine affiliated to the University of Montrèal, Hopital du Sacre-Coeur, Montrèal and provided within the scope of the aforementioned project.

References

1. A. Soria-Frisch, J. Marin, D. Ibañez, S. Dunne, C. Grau, G. Ruffini, J. Rodrigues-Brazète, R. Postuma, J.-F. Gagnon, J. Montplaisir, A. Pascual- Leone (2014) Machine Learning for a Parkinsons prognosis and diagnosis system based on EEG. Proc. International Pharmaco-EEG Society Meeting IPEG 2014, Leipzig, Germany
2. A. Soria-Frisch, J. Marin, D. Ibañez, S. Dunne, C. Grau, G. Ruffini, J. Rodrigues-Brazète, R. Postuma, J.-F. Gagnon, J. Montplaisir, A. Pascual-Leone (in prep) Evaluation of EEG biomarkers for Parkinson Disease and other Lewy Body Diseases based on the use of Machine Learning techniques
3. L. Fantini, M., Gagnon, J. F., Petit, D., Rompr, S., Dcary, A., Carrier, J., & Montplaisir, J. (2003). Slowing of electroencephalogram in rapid eye movement sleep behavior disorder. Annals of neurology, 53(6), 774-780.
4. Rodrigues Brazète J, Gagnon JF, Postuma RB, Bertrand JA, Petit D, Montplaisir J, Electroencephalogram slowing predicts neurodegeneration in rapid eye movement sleep behavior disorder, Neurobiol Aging. 2016 Jan;37:74-81. doi: 10.1016/j.neurobiolaging.2015.10.007.
5. Y. Bengio, P. Simard and P. Frasconi, IEEE Trans. Neur. Networks, vol 5, no 2, March 1994
6. S. Hochreiter and J. Schmidhuber, Long Short-Term Memory, Neural Computation, 9(8):17351780, 1997.
7. H. Jaeger, The "echo state" approach to analysing and training recurrent neural networks, GMD Report 148, German National Research Center for Information Technology, 2001.Fourth revision, July 2013, available online at http://minds.jacobs-university.de/sites/default/files/uploads/papers/EchoStatesTechRep.pdf
8. M. Lukosevicius, A practical guide to applying echo state networks, to appear in Neural Networks: tricks and trade, reloaded, G. Montavon, G.B. Orr, and K.-R. Muller, eds, Springer 2012
9. Izzet B. Yildiz, Herbert Jaeger, Stefan J. Kiebel, Re-visiting the echo state property, Neural Networks, Volume 35, November 2012, Pages 1-9, ISSN 0893-6080, http://dx.doi.org/10.1016/j.neunet.2012.07.005.
10. R. Albert, A.-L. Barabasi, Statistical mechanics of complex networksks, Rev. Mod. Phys. 74 (47), 2002.