An Echo State Neural Network for Foetal ECG Extraction Optimised by Random Search

Joachim Behar[1]
joachim.behar@eng.ox.ac.uk

Alistair E. W. Johnson[1]
alistair.johnson@eng.ox.ac.uk

Julien Oster
julien.oster@eng.ox.ac.uk

Gari Clifford
gari@robots.ox.ac.uk

Department of Engineering Science, University of Oxford

Abstract

We present a novel application of an echo state neural network (ESN) to non-invasive foetal electrocardiogram (FECG) extraction. Extraction of the FECG is performed on abdominal recordings of pregnant women via maternal ECG cancellation. The FECG can then be used for foetal health monitoring by extracting clinically interpretable features. We show that optimising an ESN by random search gives almost equivalent performance to an exhaustive grid search with 85.6% vs. 87.9% accuracy on the test database. This is particularly useful as, while powerful, ESNs have many hyper-parameters which are not easily optimised using expert knowledge.

1 Introduction

Monitoring of foetal health during labour is of utmost importance. Many technologies exist for this task, including Doppler ultrasound and foetal scalp electrocardiography (ECG). However, it is not fully known whether Doppler ultrasound is safe for the foetus (Barnett and Maulik; 2001), and foetal heart rate (FHR) acquired in this way is less accurate than through scalp ECG (SECG). SECG involves a single electrode screwed into the scalp of the foetus and cannot capture the three dimensional electrical field emanating from the heart. Furthermore, SECG can only be recorded during labour when the cervix is dilated, preventing its use for monitoring throughout pregnancy. Abdominal ECG (AECG) involves placing electrodes on the abdomen of the mother to detect the foetal ECG (FECG). This measurement is non-invasive and could be used for monitoring during pregnancy. However, high contamination both by standard noise sources and by the maternal ECG (MECG) has limited its use. Thus there is a strong impetus for novel methods for foetal ECG extraction from AECG. Previously applied techniques include forms of blind source separation, template subtraction, Kalman filtering and adaptive filtering (see (Sameni and G.D. Clifford; 2010; Behar, Oster and G.D. Clifford; 2013) for an overview). An Echo State Neural network (ESN) (Jaeger; 2001) is a class of neural networks capable of non-linear modelling of dynamical systems and can be used to remove the MECG contribution to the AECG mixture.

2 Methods

2.1 Echo State Neural Network

The ESN is a subset of the more general class of recurrent neural networks (RNN). The primary distinction between an ESN and other classes of RNN is that the weights of the reservoir are generated randomly and only the weights of a readout layer are being trained. The aim of the ESN is to

[1]These authors contributed equally to this work.

learn a mapping in which the input function $\underline{\mathbf{u}}(n)$ (the maternal chest ECG which is FECG free) best matches a target signal $\underline{\mathbf{y}}(n)$ (the AECG). Briefly, one can consider an architecture of K inputs forming the input signal $\underline{\mathbf{u}}(n) = [u_1(n), ..., u_K(n)]$, M internal units forming the reservoir state vector $\underline{\mathbf{x}}(n) = [x_1(n), ..., x_M(n)]$, and L output units forming the output signal $\underline{\hat{\eta}}(n) = [\hat{\eta}_1(n), ..., \hat{\eta}_L(n)]$. Update of the ESN internal units is performed as shown in Eq. 1. The leaky integrator neuron model was used and considering a purely input-driven dynamical pattern recognition task, the system was simplified by setting $\mathbf{W}_b = \mathbf{0}$. In this equation α is the leakage rate.

$$\underline{\mathbf{x}}(n+1) = (1-\alpha)\underline{\mathbf{x}}(n) + f(W\underline{\mathbf{x}}(n) + \mathbf{W}_i\underline{\mathbf{u}}(n+1) + \mathbf{W}_b\underline{\hat{\eta}}(n)) \tag{1}$$

Let $\underline{\mathbf{z}}(n) = [\underline{\mathbf{x}}(n)\ \underline{\mathbf{u}}(n)]$ represent the extended system state. The output can then be computed as shown in Eq. 2, where $g(\cdot)$ represents the hyperbolic tangent.

$$\underline{\hat{\eta}}(n+1) = g(\underline{\mathbf{w}}_o(n)\underline{\mathbf{z}}(n)). \tag{2}$$

The learning of the ESN proceeds as follows; $\mathbf{W}$ (MxM) was generated sparsely with ψxMxM non-zero entries, where ψ is the sparsity parameter. $\mathbf{W}_i$ (KxM) was generated from a uniform distribution between $[-1\ 1]$. Both $\mathbf{W}$ and $\mathbf{W}_i$ are kept constant after initialisation. Eq. 2 was used to calculate the predicted signal, $\underline{\hat{\eta}}(n)$, DB_1first 30 seconds of the input chest ECG and weights $\underline{\mathbf{w}}_o$ were learnt by linear regression and kept constant (the ESN is by consequence used as a non-adaptive filter). The resultant ESN was then applied to the remaining signal.

2.2 Data and Evaluation

Two databases were considered. The first database (DB_1) is the Physionet non-invasive FECG (NI-FECG) database (PNIFECGDB), which contains 55 multichannel AECG recordings from a single subject (Goldberger et al.; 2000). Of this, one minute from 14 records with visible foetal QRS (FQRS) complexes were extracted for across three abdominal channels, (2148 FQRS per channel which were manually labelled). The second database (DB_2) is a subset of records from a private commercial database. Eleven five minute recordings from 8 pregnant women were used, and FQRS were located automatically from an associated foetal scalp ECG (6444 reference FQRS per channel using the scalp ECG). Both databases were sampled at 1 kHz with 16-bit resolution. For computational reasons the data were downsampled to 250Hz, with appropriate application of an anti aliasing filter, before the FECG extraction technique was applied. Because only one chest channel was available in DB_2 only one of the two available chest channels was used for DB_1. Each abdominal channel was considered individually. Thus $\underline{\mathbf{u}}(n) = u(n)$ and $\underline{\hat{\eta}}(n) = \hat{\eta}(n)$. As the aim is a clinical interpretable sign of foetal health, the evaluation of algorithm performance was performed by assessing how accurate FHR could be estimated from the MECG free residual signal. Estimated R-peak locations were compared with the reference annotations utilizing the F_1 accuracy measure, $F_1 = 2 * PPV * Se \div (PPV + Se)$, where PPV is the positive predictive value and Se is the sensitivity in accurately detecting foetal R-peak locations.

2.3 Hyper-parameter Search

A major issue with the ESN is the lack of optimisation methods for determining the numerous hyper-parameters associated with the model as well as the knowledge of which hyper-parameters are relevant. Furthermore, in the context of this work there are an additional two parameters corresponding to filter cut-offs for initial preprocessing of the signals; f_b and f_h, the baseline wander and high frequency filters cut-offs respectively. We compare the efficacy of a grid-search as performed by Behar, Johnson, Clifford and Oster (2013) for NI-FECG extraction using an ESN versus random search as recently proposed by Bergstra and Bengio (2012). The methodology of the grid search versus the random search is shown in Tab. 1. For parameters not specified as a range or distribution, no parametric search was performed as it was thought to be unlikely to improve model performance.

Parameters were selected as those which maximised the F_1 measure on DB_1. These parameters were then evaluated on DB_2. The parameter search proceeded as follows. For the grid search, the filters cut-off frequency ranges were fixed to reasonable a priori estimates. The leakage and spectral radius were then searched (given a high number of reservoir neurons), and well performing parameters

Table 1: Parameters search range and optimal parameters found for the preprocessing step and the ESN. GS: grid search. RS: random search. $\mathcal{U}$: uniform distribution.

Parameter	Search: GS (step size)	Search: RS	Optimal parameters: GS	Optimal parameters: RS
Low pass filter cut-off, f_b	[1, 49] (3)	$\mathcal{U} \sim [1, 50]$	20	26
High pass filter cut-off, f_h	[50, 120], (5)	$\mathcal{U} \sim [50, 120]$	95	87
Leakage, α	[0, 1] (0.1)	$\mathcal{U} \sim [0, 1]$	0.4	0.974
Spectral radius of W, ρ	[0, 1] (0.1)	$\mathcal{U} \sim [0, 1]$	0.4	0.821
Units in the reservoir, M	[10, 250] (20)	$\mathcal{U} \sim [10, 250]$	90	135
Scaling of W_{in}, γ	1	$\mathcal{U} \sim [0, 1]$	1	0.622
Seed value, s^1	–	$\mathcal{U} \sim [0, 10000]$	–	1588

Table 2: Performance comparison of the optimal parameters obtained by grid search (GS), random search (RS, best across 32 iterations $\pm$ 1 standard error), and template subtraction (TS).

Method — Statistics	DB_1 ESN-GS	DB_1 ESN-RS	DB_1 TS	DB_2 ESN-GS	DB_2 ESN-RS	DB_2 TS
Se	97.1	97.3 ± 0.29	90.3	87.6	87.6 ± 0.73	86.4
PPV	97.3	97.5 ± 0.28	90.0	86.5	85.5 ± 0.53	85.2
F_1	97.2	97.4 ± 0.27	90.1	87.9	86.5 ± 0.62	85.8

were selected. This was followed by a search for the best filter cut-off parameters. The leakage and spectral radius were then searched again, and finally the number of reservoir neurons was searched. This iterative approach avoids the prohibitively expensive exhaustive search that would be entailed if all parameters were searched in conjunction. The random search was much simpler and was performed by sampling from the specified distributions in Tab. 1 simultaneously. Both methods were also compared to a baseline template subtraction algorithm (Martens et al.; 2007). Automatic relevance determination (ARD) plots, as described in (Bergstra and Bengio; 2012), were also produced using Gaussian Processes (GP) regression in order to analyse the relevance of the parameters.

3 Results

The grid search for the parameters involved 798 repetitions, and the random search was repeated 512 times. A comparison of the optimal parameters as determined by the grid search versus the random search (for 32 iterations) is shown in Tab. 2 along with performance measures. Since 8 repetitions of a 32 iteration random search were performed, the median $\pm$ one standard error is reported. Optimising the ESN by random search gave almost equivalent performance to an exhaustive grid search with $F_1 = 85.6\%$ vs $F_1 = 87.9\%$ accuracy. The efficiency of the search, representing how quickly the random search found optimal parameters, is shown in Fig. 1. The ARD plots are shown in Fig. 2. Note that the seed value s is used as a negative control.

4 Discussion and Conclusion

The random search found extremely well performing parameters in a fraction of the searches. Less than 8% of the iterations were required to reach almost equivalent performances on both databases as compared to grid search. Furthermore, the grid search was not complete, as the hyper-parameter dimensionality was too high to allow for a fully exhaustive grid search. However was noticeable that the performance of the ESN started decreasing on the test database DB_2 after 32 iterations. This phenomena can be explained by looking at the ARD plots where the relative importance of the hyper-parameters and prefiltering parameters was different between DB_1 and DB_2 i.e. the relative importance of each parameter varies from one data set to the next. Recall that the data in the training

[1] Seed value for a Mersenne twister pseudorandom number generator.

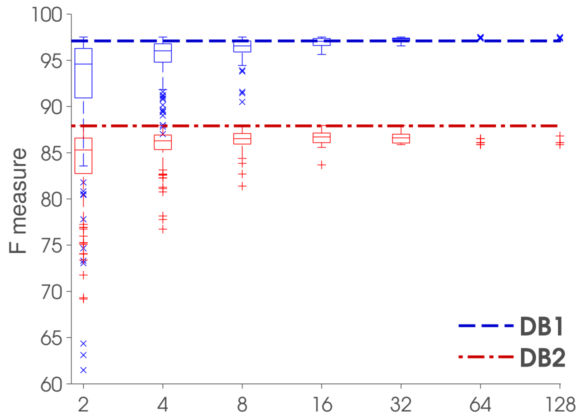

Figure 1: Performance of the ESN with respect to number of random search iterations. As each search is i.i.d., the overall 512 searches can be subsampled to replicate as many as $\frac{512}{N}$ searches each of size N. Blue 'x's are training set performances (DB_1). Red '+'s are test set performances (DB_2).

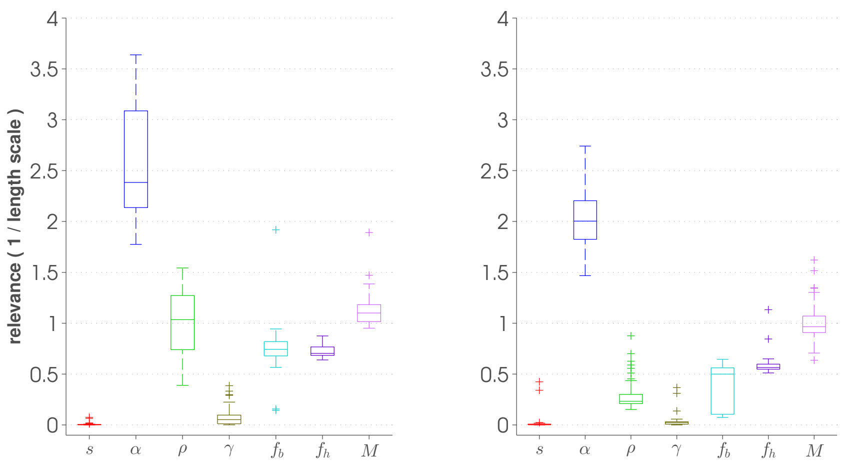

Figure 2: Automatic relevance determination (ARD) applied to the ESN hyper-parameters and the prefiltering parameters. Left: ARD performed on DB_1. Right: ARD performed on DB_2.

set (DB_1) and test set (DB_2) were recorded with different hardware, following a separate protocol and at different stage of pregnancy for different subjects.

It is unlikely that random search on its own would find 'the best' parameters of the ESN and pre-filtering step. A more sensible approach would be the use of random search followed by sequential optimisation. Nevertheless, with a very limited number of iterations, random search provides important insight on which parameters are influential in model performance. For example, we observe that input scaling is irrelevant (Fig. 2), and subsequent hyperparameter searches would likely be improved by focusing on the other parameters. Furthermore, random search provides a good estimate of the range of interest for each parameter.

In conclusion, we have shown that, in the context of NI-FECG extraction, random search was an effective method for learning the ESN parameters. As compared to grid search, random search took a fraction of the number of iteration and resulted in performances which were extremely competitive and allowed to understand which of the preprocessing and ESN hyper-parameters were relevant using ARD. Furthermore, the ESN proved to be an effective method for the removal of the MECG from the AECG for the purposes of FECG extraction, outperforming the commonly applied template subtraction. Open source code for producing the figures in this paper is made available to allow comparison and reproducibility on the public domain data http://physionet.org/physiotools/.

Acknowledgments

JB is supported by the UK Engineering and Physical Sciences Research Council, the Balliol French Anderson Scholarship Fund and MindChild Medical Inc. North Andover, MA. AJ acknowledges the support of the RCUK Digital Economy Programme grant number EP/G036861/1 (Oxford Centre for Doctoral Training in Healthcare Innovation). JO is supported by Wellcome Trust Centre Grant No. 098461/Z/12/Z (Sleep, Circadian Rhythms & Neuroscience Institute).

References

Barnett, S. and Maulik, D. (2001). Guidelines and recommendations for safe use of doppler ultrasound in perinatal applications, *Journal of Maternal-Fetal and Neonatal Medicine* **10**(2): 75–84.

Behar, J., Johnson, A., Clifford, G. D. and Oster, J. (2013). A Comparison of Single Channel Foetal ECG Extraction Methods, *Submitted to Ann. Biomed. Eng.* .

Behar, J., Oster, J. and G.D. Clifford (2013). Non Invasive FECG extraction from a set of abdominal channels, *Comp. in Card. 2013, Zaragoza, Spain, 22th-25th September, 2013.* .

Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization, *J. Mach. Learn. Res.* **13**: 281–305.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K. and Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals, *Circ.* **101**(23): 215–20.

Jaeger, H. (2001). The “echo state” approach to analysing and training recurrent neural networks., *Tecnical report GMD Report 148, German National Research Center for Information Technology* **148**.

Martens, S., Rabotti, C., Mischi, M. and Sluijter, R. (2007). A robust fetal ECG detection method for abdominal recordings, *Physiol. Meas.* **28**: 373–88.

Sameni, A. and G.D. Clifford (2010). A Review of Fetal ECG Signal Processing; Issues and Promising Directions, *The open pacing, electrophy. & therapy J.* **3**: 4–20.