A DIFFERENTIAL ENTROPY BASED METHOD FOR DETERMINING THE OPTIMAL EMBEDDING PARAMETERS OF A SIGNAL

Temujin Gautama, Danilo P. Mandic† and Marc M. Van Hulle‡*

*,‡Laboratorium voor Neuro- en Psychofysiologie
K.U.Leuven, Campus Gasthuisberg, Herestraat 49
B-3000 Leuven, BELGIUM
Email: {temu,marc}@neuro.kuleuven.ac.be

† Department of Electrical and Electronic Engineering
Imperial College of Science, Technology and Medicine
Exhibition Road, SW7 2BT, London, UK
Email: d.mandic@ic.ac.uk

ABSTRACT

A novel method for determining the set of parameters for a phase space representation of a time series is proposed. Based upon the differential entropy, both the optimal embedding dimension m, and time lag τ, are simultaneously determined. The choice of these parameters is closely related to the length of the optimal tap input delay line of an adaptive filter or time-delay neural network. The method employs a single criterion – the "entropy ratio" between the phase space representation of a signal and an ensemble of its surrogates – and is first systematically tested on synthetic time series for which the optimal embedding parameters are known, after which it is verified on a number of benchmark real-world time series. The proposed entropy ratio method is shown to consistently outperform some well-established methods.

1. INTRODUCTION

For processing of signals with structure, as is the case with most biomedical signals, an established method for visualising an attractor of the underlying nonlinear dynamical system is by means of time delay embedding [1]. This way, for a given time lag τ, a time series $\{x_k\}$ is represented in the so-called 'phase space' by a set of delay vectors (DVs) $\mathbf{x}(k) = [x_{k-\tau}, \ldots, x_{k-m\tau}]$ of a given embedding dimension m. In digital signal processing, these two parameters are crucial for determining the optimal tap input length of an adaptive filter or a time-delay neural network. For instance, if the temporal span of $(m \cdot \tau)$ is too small, the signal variation within the delay vector is mostly governed by noise and either m or τ should be increased. However, there is no established criterion for choosing which of the two parameters to modify. In practice, it is common to have a fixed time lag τ (sampling rate) and to adjust the embedding dimension m (length of a filter) accordingly. Notice that if τ were too small to cover the minimal time span needed to capture the dynamics of a signal, the tap input length m (and thus the number of filter parameters) would become rather large, resulting in an increased complexity of training. In turn, if τ is greater than optimal, the nature of the resulting model becomes too discrete, resulting in a failure of the filter to capture the underlying signal dynamics. Recall that, although in principle, the phase space representation of a time series is independent of the value of τ, this is only the case for an infinite amount of data, hence the need for an optimisation method to jointly determine m and τ.

Several methods exist for determining the optimal embedding parameters, whereby the optimal time lag τ_{opt} and embedding dimension m_{opt} are optimised separately. This way, the time lag is first determined as that for which the mutual information between time samples separated by τ, that is x_k and $x_{k+\tau}$, is minimal [2]. Using τ_{opt}, the optimal embedding dimension is found next, as that for which the number of false nearest neighbours (a measure of the consistency of the Euclidean distance between neighbouring DVs in $\mathcal{R}^m$ when the embedding dimension is increased from m to $m+1$), is small [3]. We refer to the combination of the time delayed mutual information and false nearest neighbour methods as the TDMI/FNN method. An interpretation of the first step is that the axes of the 2D phase space signal representation are being chosen as independent as possible, not necessarily a good criterion if the embedding dimension exceeds two [1]. The second step verifies the preservation of the topological structure of a signal in $\mathcal{R}^m$, the presence of which suggests strong dependence between the dimensions in the phase space. Clearly, there is an incongruence between the two stages.

To this cause, we propose a unified and unambiguous optimisation procedure for simultaneously determining both the time lag τ, and the embedding dimension m. The method is based on estimates of the differential entropy ratio of the

*supported by a scholarship from the Flemish Regional Ministry of Education (GOA 2000/11).

†supported by QLG3-CT-2000-30161 and a visiting fellowship from the K.U.Leuven (F/01/079) while at the Laboratorium voor Neuro- & Psychofysiologie, K.U.Leuven, Belgium.

‡supported by research grants received from the Fund for Scientific Research (G.0185.96N), the National Lottery (Belgium) (9.0185.96), the Flemish Regional Ministry of Education (Belgium) (GOA 95/99-06; 2000/11), the Flemish Ministry for Science and Technology (VIS/98/012), and the European Commission, 5th framework programme (QLG3-CT-2000-30161 and IST-2001-32114).

phase space representation of a sampled time signal and an ensemble of its surrogates. For rigour, the proposed method is compensated for the dimensionality and temporal correlatedness.

2. THE ENTROPY RATIO (ER) METHOD

To measure the 'amount of disorder', based upon the probability density function (pdf) $p(\mathbf{x})$ of data, the differential entropy is used: $H(\mathbf{x}) = -\int_{-\infty}^{+\infty} p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x}$. Particularly convenient is the Kozachenko-Leonenko (K-L) estimate of the differential entropy [4]

$$H(\mathbf{x}) = \sum_{j=1}^{N} \ln (N\rho_j) + \ln 2 + C_E \tag{1}$$

owing to its flexibility with respect to the dimensionality of the data set. In Eq. 1, N is the number of samples in the data set, ρ_j is the Euclidean distance of the j-th delay vector to its nearest neighbour, and $C_E (\approx 0.5772)$ is the Euler constant. For a given embedding dimension, m, and time lag, τ, let $H(x, m, \tau)$ denote the differential entropies estimated for time delay embedded versions of a time series, x, which shall be used as an inverse measure of the *structure* in the phase space.

2.1. Determining the Optimal Embedding Parameters

The set of optimal parameters, $\{m_{\text{opt}}, \tau_{\text{opt}}\}$, yields a phase space representation which best reflects the dynamics of the underlying signal production system. Therefore, it is expected that this representation has a minimal differential entropy (minimal disorder), and that a deviation from $\{m_{\text{opt}}, \tau_{\text{opt}}\}$ results in an increase. Thus, we optimise the differential entropy (Eq. 1) for m and τ, the minimum of $H(x, m, \tau)$ yielding the optimal set of embedding parameters $\{m_{\text{opt}}, \tau_{\text{opt}}\}$.

Notice that the K-L entropy estimate (Eq. 1) is not robust with respect to dimensionality. This is compensated for by standardising $H(x, m, \tau)$ with respect to an ensemble of so-called 'surrogates' of signal x (for an overview, see [5]). In the simplest case, N_{s} surrogates $x_{\text{s},i} \quad i = 1, \dots, N_{\text{s}}$ of a signal x are generated by performing a random permutation of the time samples. This way, the signal distribution is unaffected and the serial correlations are randomised (yielding a whitened signal with a signal distribution identical to that of the original x). The K-L estimates for the time delay embedded versions of the original time series $H(x, m, \tau)$, and its surrogates $H(x_{\text{s},i}, m, \tau)$ are computed using Eq. 1 for increasing m and τ (index i refers to the i-th surrogate). To determine the optimal embedding parameters, the ratio

$$I(m, \tau) = \frac{H(x, m, \tau)}{\langle H(x_{\text{s},i}, m, \tau)\rangle_i}, \tag{2}$$

needs to be minimised, where $\langle \cdot \rangle_i$ denotes the average over i. To penalise for higher embedding dimensions, the minimum description length (MDL) method is superimposed, yielding the "entropy ratio" (ER):

$$R_{\text{ent}}(m, \tau) = I(m, \tau) + \frac{m \ln N}{N}, \tag{3}$$

where N is the number of delay vectors, which is kept constant for all values of m and τ under consideration. This way, a difference in K-L estimate (Eq. 1) cannot be attributed to the number of time samples or DVs.

For time series exhibiting strong serial correlations, however, applying the method directly yields embedding parameters which have a preference for $\tau_{\text{opt}} = 1$. Indeed, for $\tau = 1$, the presence of time correlations implies a higher degree of structure, thus, a lower amount of disorder. To that cause, surrogate data are best generated using the iterative Amplitude Adjusted Fourier Transform (iAAFT) method [5], which retains within the surrogates both the signal distributions *and* also approximately the autocorrelation structure of the original signal. This way, the serial correlations are present in both the original and the surrogate time series, as desired. The minimum of the plot of the entropy ratio yields the optimal set of embedding parameters.

3. SIMULATIONS

To validate the proposed entropy ratio criterion, it is first tested on a variant of the Hénon map, which is implemented using different delays. Next, to highlight the robustness of the ER method, the effect of resampling a time series at a higher rate is examined. Finally, the proposed method is tested on benchmark time series. In all simulations, $N_{\text{s}} = 5$ surrogates were generated using the iAAFT method, and the entropy ratios were evaluated for $m = 2, \dots, 15$ and $\tau = 1, \dots, 10$. Increasing the number of surrogates did not affect the results.

The obtained parameters are compared to those obtained by combining two established methods for estimating τ_{opt} and m_{opt} separately. The time delayed mutual information (TDMI) method [2] estimates the optimal time lag as that for which the time delayed mutual information function, $A(\tau)$, shows the first local minimum. This function can be approximated by [1]:

$$A(\tau) = -\sum_{ij} p_{ij}(\tau) \ln \frac{p_{ij}(\tau)}{p_i p_j}, \tag{4}$$

where p_i is the probability to find a signal value in the i-th interval, and $p_{ij}(\tau)$ is the joint probability of finding a signal value in the i-th interval, and a value at time τ later in the j-th interval. The pdfs are estimated using a binning approach (200 bins). The obtained time lag is consequently used for

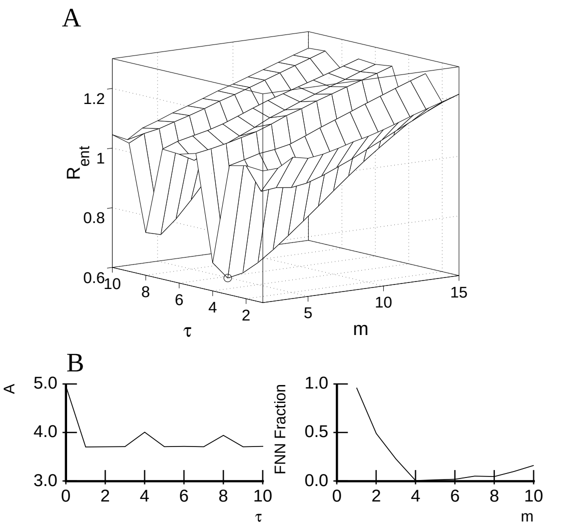

Fig. 1. Analyses for a realisation of the Hénon Map, with $d = 4$, using the ER (A) and the TDMI/FNN method (B). In panel A, the minimum of the ER-plot is indicated by an open circle.

estimating the optimal embedding dimension using the False Nearest Neighbour method (FNN; [3]), which checks for the consistency of the distance to the nearest neighbour for increasing values of m.

Hénon Map

The variants of the Hénon map considered were realisations of 500 samples obtained from

$$x_k \;=\; 1 - a\, x_{k-d}^2 + b\, y_{k-d};\;\; y_k = x_{k-d}, \qquad (5)$$

where $a = 1.4$, $b = 0.3$, and d is a time delay. It is desired that $m_{\rm opt}$ is invariant to the change of d, and that $\tau_{\rm opt} = d$.

Figure 1 shows the entropy ratio $R_{ent}(m, \tau)$ for $d = 4$. The minimum of the plot, indicated by an open circle, yields $m_{\rm opt} = 3$ and $\tau_{\rm opt} = 4$. The method is tested further for $d = 1, \ldots, 8$, and in each analysis the minimum was found (correctly) for $m_{\rm opt} = 3$ and $\tau_{\rm opt} = d$ (results not shown). For comparison, the results obtained using the TDMI/FNN method are shown in Fig. 1B. Using the TDMI method, the first local is consistently found at $\tau_{\rm opt} = 1$. As a consequence, the FNN method compensates for this underestimate of $\tau_{\rm opt}$ and yields an *over*estimate of $m_{\rm opt}$, which in nearly all simulations is $m_{\rm opt} = d$.

3.1. Effect of Long Time Correlation

Consider once more a 500 samples realisation of Eq. 5 with $d = 2$. To investigate time correlations, the time series is resampled using linear interpolation, whereby s denotes the interpolation ratio. It is desired that the embedding dimension remains constant with respect to the resampling factor, and that the time lag is proportional to s.

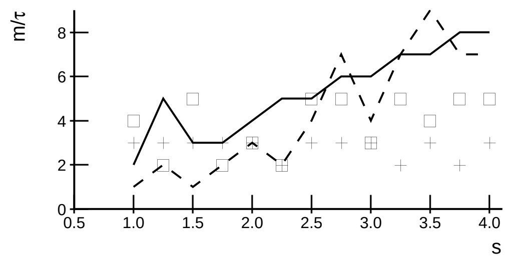

Fig. 2. Estimated embedding dimension, $m_{\rm opt}$, and time lag, $\tau_{\rm opt}$, as a function of the resampling factor, s, for the Hénon series with $d = 2$. The results for the ER are shown as crosses and the solid curve, and those for the TDMI/FNN method as squares and the dashed curve, for embedding dimensions and time lags, respectively.

	ER		TDMI/FNN	
time series	$m_{\rm opt}$	$\tau_{\rm opt}$	$m_{\rm opt}$	$\tau_{\rm opt}$
laser	5	7	2	2
EEG	5	9	7	11
ECG	5	2	6	10
HRV	4	1	5	10

Table 1. Optimal embedding parameters, $m_{\rm opt}$ and $\tau_{\rm opt}$, obtained using the entropy ratio, and the TDMI/FNN method.

The results of this experiment are shown in Fig. 2. The ER method yields robust estimates of the embedding dimension (crosses), namely either two or three, and the estimated time lags (solid curve) increase fairly linearly with the resampling factor s. The TDMI/FNN method does not yield a robust estimate ($2 \leq m_{\rm opt} \leq 5$, squares), or $\tau_{\rm opt}$ (dashed curve), albeit it shows an increasing trend in the time lag estimates.

3.2. Real-World Examples

The proposed method is next illustrated on four real-world examples, namely the chaotic laser series from the Santa Fe Competition, and three physiological signals, from an electroencephalogram (EEG) recording, from an electrocardiogram (ECG) recording and, from the latter, the obtained heart rate variability (HRV) time series. The results are shown in Fig. 3. In all the plots, a clear minimum is observed, indicated by an open circle. The resulting embedding parameters are compared to those obtained using the TDMI/FNN method in Table 1. Except for the embedding parameters for the EEG signal, the results are markedly different.

4. DISCUSSION

The choice of the embedding dimension, m, and the time lag, τ, is important both for signal nonlinearity analysis and for determining the optimal length of tap input delay line of

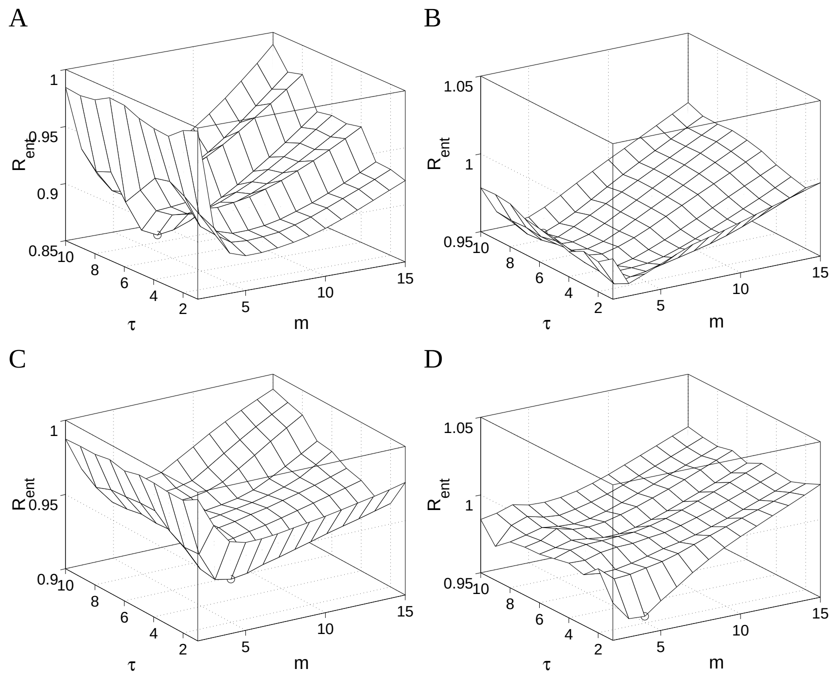

Fig. 3. Plots of the ER for the real-world examples: laser series (A), EEG (B), ECG (C) and HRV (D). The minima of the ER-plots are indicated by open circles.

an adaptive filter or time-delay neural network. To this cause, we have introduced a novel measure, the "entropy ratio" (ER), for determining an optimal set of embedding parameters. The ER method rests upon the amount of structure which is present in the time delay embedded (phase space) representation of the time series, compared to its randomised version. The random surrogate have been introduced to correct the ER for effects of dimensionality and autocorrelation structure, yielding a robust (inverse) measure for the amount of structure present in the phase space representation. The proposed method has been systematically tested on synthetic examples, in which the embedding dimension remained constant, and the time lag was varied, showing excellent accuracy. The ER has consistently outperformed the traditional method, a combination of the Time Delayed Mutual Information (TDMI; [2]) and the False Nearest Neighbours (FNN; [3]) methods. The ER has further been illustrated on four real-world examples, and in most cases the obtained embedding parameters were markedly different from those obtained using the TDMI/FNN method.

The main advantage of the proposed method is that a single measure is simultaneously used for optimising both the embedding dimension and the time lag. This way, the incongruence in methods which combine variations of the time delayed mutual information and the false nearest neighbour methods is avoided. The entropy ratio criterion requires a time series to display a clear structure in phase space, a fairly common phenomenon in physiological signals. For signals with no clear structure, the method will not yield a clear minimum, and a different approach needs to be adopted, possibly one that does not rely on a phase space representation. The entropy ratio has been shown to be robust to the dimensionality and serial correlations of a signal.

References

[1] R. Hegger, H. Kantz, and T. Schreiber, "Practical implementation of nonlinear time series methods: The TISEAN package," *Chaos*, vol. 9, pp. 413–435, 1999.

[2] A.M. Fraser and H.L. Swinney, "Independent coordinates for strange attractors from mutual information," *Phys. Rev. A*, vol. 33, no. 2, pp. 1134–1140, 1986.

[3] M.B. Kennel, R. Brown, and H.D.I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev. A*, vol. 45, pp. 3403–3411, 1992.

[4] J. Beirlant, E.J. Dudewicz, L. Györfi, and E.C. van der Meulen, "Nonparametric entropy estimation: An overview," *International J. Mathematical and Statistical Sciences*, vol. 6, pp. 17–39, 1997.

[5] T. Schreiber and A. Schmitz, "Surrogate time series," *Physica D*, vol. 142, no. 3-4, pp. 346–382, 2000.