ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 323-328

Time Series Prediction using Neural Networks and its Application to Artificial Human Walking

R.S. Venema, A. Ypma, J.A.G. Nijhuis and L. Spaanenburg

Department of Computing Science, Rijksuniversiteit Groningen

P.O. Box 800, 9700 AV Groningen, The Netherlands

Abstract. *One of the main issues in the research on time series is its prediction. Using a tapped–delay neural network we formulate the optimal network size from the signal correlation time. Then the biofeedback–driven neurocontrol for artificial human walking is developed with much detail on the signal preprocessing. Finally we indicate the need for a hierarchical network architecture to eliminate oscillatory effects inherent to handling the human motor control problem.*

1. Introduction

The most important aspect in time series prediction is the modelling of the series. Before we are able to predict the future values based on its history, we have to derive a model for the underlying behaviour. Statistics gives useful tools for this, but a major objection to most of these techniques (for example the well–known ARMA method [1]) is that they assume a priori that the time series was generated by a linear process. Most series one encounters in practice, however, are of a nonlinear nature; therefore much effort has been spent to achieve a more effective, that is nonlinear, model building. The most important statistical approaches to model a nonlinear time series are listed in [2], where these models are applied to predict the famous sunspot activity data.

A powerful and very promising alternative is the use of an artificial neural network (ANN). ANNs are considered well–suited in the adaptive identification of nonlinear systems [3]. A careful comparison of different neural methodologies for time series characterization and prediction is presented in [4]. A methodology includes three stages: (i) *Preprocessing and model identification*, where the architecture of the network is designed. In the preprocessing–stage statistical measures for determinism, chaos, randomness, nonlinearities and dependencies are used. Another important aspect is the calculation of characteristic features containing information about the time series, that aid the network in the learning process. (ii) *Model building*, where the training of the network is performed with a training data set. (iii). *Verification*, where the optimized net that results from the training stage is checked against the test data set, that is data that had not been used in the training process.

Most studies put emphasis on the last two stages. However, an ANN has limited ability in interpreting input–data: when offered conflicting or badly preprocessed data, it has a hard time in establishing any relations between inputs (and outputs) and so bad learning and generalization takes place. The topic of preprocessing vs. network architecture is thus of major importance in neural network design. In this paper we will therefore focus on the preprocessing–stage, in which the data is carefully studied before building an appropriate ANN.

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 323-328

As an application we present the use of ANNs in the construction of an artificial walking system, i.e. a system which enables paraplegics to walk again. We will focus on the inverse dynamics problem, trying to relate movement to its underlying muscle activation pattern. We will show that ANNs are capable of predicting future muscle activation using past activation and movement data.

2. Preprocessing and model identification

We now turn to the central theme of our research: how to capture temporal features of the data vector in order to make reliable future predictions? The following techniques are available for a carefully exercised preprocessing:
– *Data analysis*; before building a predictor one should study the process that did generate the observed realisations, i.e. is it of a deterministic (e.g. chaotic) or a stochastic (e.g. noisy) nature. The entropy measure together with the correlation– and reconstruction dimensions [4] give valuable indications whether the process is deterministic or stochastic. When it is generated by a stochastic linear process one can use statistical theory; when dealing with nonlinear data they are not adequate anymore. In this case we may proceed by modelling the nonlinear temporal function directly, e.g. using interpolation, local linear approximation [5], least–squares methods or ANNs.
– *Characteristic feature extraction*; when predicting a signal some timesteps ahead one could choose to present the network only the signal's own history so that there is an analogy with statistical predictors. Another approach is that one first determines the characteristic features of the data and collect these in the input vector of the network. By presenting the network arithmetical functions and statistical features that cover all necessary information of the raw signal, one reduces the size of the network. The smallest network that can learn the training data is usually desirable.
–*Additional data*; utilization of additional data (i.e. related to the signal to be predicted) is in most cases essential for good results. Using additional data features for further fine–tuning can conveniently be done when an ANN–based predictor is constructed. The choice of additional data, giving useful information on the to be predicted signal, can be based on cross–correlation of the data with the signal. This choice is a nontrivial task as much redundancy could be present in the inputs supplied to the network. When an ANN is confronted with too much freedom, its generalization capabilities will deteriorate. Practical utilization of the system requires a reduction of redundancy, for instance by the well–known Principal Component Analysis (PCA).

A major issue in our quest for the optimal neural predictor is to take decisions about preprocessing, network architecture and learning rule in a reproducible way. After preprocessing and data analysis, the next choice is on the network architecture and learning rule. We will mainly debate the first choice; as learning rule we choose the standard error backpropagation rule. An important decision in neural system design concerns the dimensions of the network, i.e. the number of input, hidden and output units. Many techniques for optimization of network dimensions exist, such as network pruning and growing. In [2] several criteria based on Akaike's information criterion (AIC) are introduced that determine the optimal number of ANN hidden and input units.

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 323-328

In time series prediction the most popular architectures are time−delay and real−time recurrent networks [6]. Here we will focus on a frequently used ANN in temporal processing, namely the multilayer Perceptron with on its inputs a tapped delay−line, i.e. a device that stores past samples of the inputs. We want to find the optimal size of the delay−line to learn the signal characteristics and to generalize it adequately. We state the assumption that the number of time delays in the tapped delay−line to capture the temporal properties of the input data depends on the input data autocorrelation function. More specifically, we estimate the size of the tapped delay−line by the number of reasonably autocorrelated lags: $ND \approx \min(k : k \in \mathbb{N} : r(k) \leq \epsilon)$ where ND denotes the number of data samples in the delay−line, ϵ is a suitable non−negative constant and r denotes the sample AutoCorrelation Function (ACF) of the input−signal x of length n, defined by [1]: $r(h) = \gamma(h)/\gamma(0)$,$0 \leq h \leq n-1$. Here $\gamma(h)$ is the sample autocovariance $\frac{1}{n}\sum_{j=1}^{n-h}(x(j) - \bar{x})\,(x(j+h) - \bar{x})$ with $\bar{x}$ denoting the sample mean. For $\epsilon = 0$, the upper bound is also denoted as *correlation time,* i.e. the minimum lag for which the value of the ACF is below 0.

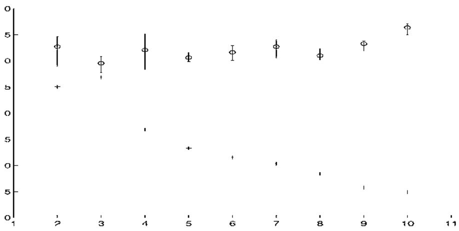

Fig.1. Train and test error for a signal with corr.3

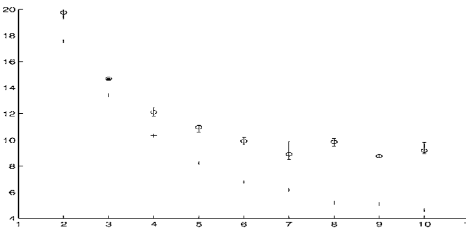

Fig.2. Train and test error for a signal with corr.7

The hypothesis about correlation content and time−delay length is investigated for signals with predefined autocorrelation (using a first−order moving average scheme). The network generalization errors show to be minimal about the length that corresponds to the correlation time of the test signal. Results for two test signals (with predefined autocorrelation 3 and 7) are shown in figure 1 and figure 2 respectively. The circle within a rectangular bar denotes the Mean Steady−state Test Error (MSTE), i.e. the 3σ −error after 5.000 epochs, averaged over 5 experiment repetitions, whereas the horizontal bars on the outside of the vertical bar indicate minimum and maximum MSTE over 5 experiments; the plus−signs denote the, uniformly decreasing, mean squared train error. At delay−lines shorter than the input signal correlation time, there is still “much to be learned”, so on average similar generalization capability is shown; at larger delay lines, the minimum level has already been reached, so that adding more inputs will increase the degrees of freedom, while no extra information is provided. This will worsen the generalisation potential. We conclude from these simulations that the correlation time is a reasonable estimator for the time−delay size in the input layer, when dealing with signals with simple linear dynamics.

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 323-328

3. Application to artificial human walking

Two basic approaches are pursued in the research on rehabilitation of paraplegia. In Functional Electrical Stimulation (FES), the injured muscles are electrically stimulated to regain functional gait [7]. Biofeedback provides patients with sensory information to support the rehabilitation process [8]. Lately, one strives to combine these approaches by utilizing specific information about balance (distribution of ground reaction forces), body configuration (joint angles) and muscle activity (EMG−signals) to control the stimulation of a paraplegic's disabled lower extremity musculature.

We will focus on the *inverse dynamics problem* to relate movement to the underlying muscle activation pattern. In other words, we are looking for *"a mapping between joint angles and ground reaction forces on the one hand and EMG patterns on the other"*. Since walking is a cyclic activity, we expect the body signals concomitant with walking to be also repetitive and therefore likely to be predictable. Specifically, the EMG−signal from the right Tibialis Anterior muscle (see figure 3) will be predicted during walking by a neural network using (i) the EMG−signal's own history, (ii) information about joint angles and ground reaction forces and (iii) statistical features from (a time−window of) the EMG−signal.

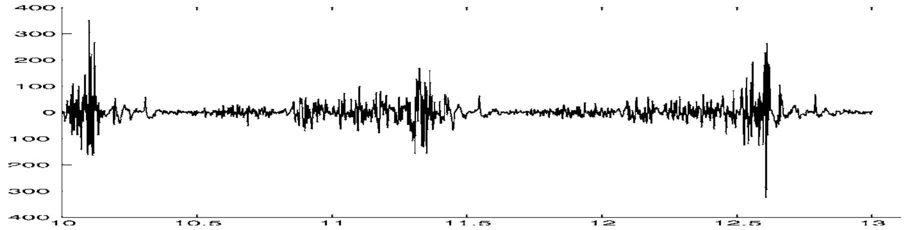

Fig.3. EMG Signal from the right Tibialis Anterior muscle

The measurement data contain 3000 samples (representing 2.5 steps). The first 2000 samples are to be used for training whereas the last 1000 samples are kept for verification. From estimates for the correlation− and reconstruction dimensions [4] it is noticed that the first increases uniformly with the second, indicative for a random signal. This was affirmed in further experiments. In [7] it is stated, that the standard deviation of the signal is indicative of the walking cycle and hence can be used as a feature for the neural predictor. Other useful parameters are: mean, maximum and minimum values of a small segment of EMG.

In [8] it is stated that the analysis of the EMG requires to take also the associated movement data into account. The selection of suitable movement signals can be done by looking at the Cross−Correlation Function (CCF) of muscle and movement signals with the EMG. Two periodic (walking) signals with strong correlation will have a CCF that is both clean (lacking local distortions) and periodic. Experiments are negative for ground reaction forces of the right lower extremity, but positive for the joint excursions in the sagittal plane (i.e. hip extension and knee flexion). This was verified by a neural emulation of PCA using the generalized Hebbian learning rule. The different input sets range from (i) joint angles from the frontal and transversal plane, i.e. only knee rotation

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 323-328

and hip abduction (this resulted in the eigenvalues 0.700, 0.517, 0.026, 0.011) till (ii) all joint angles to/from all three planes, i.e. knee flexion and hip abduction was also included (eigenvalues 1.277, 1.076, 0.447, 0.245, 0.099, 0.030, 0.019, 0.015). We have noted that large new eigenvalues emerge after addition of hip extension or knee flexion, therefore these can be expected to contain most of the information of all angles.

As neural predictor a multilayer Perceptron is chosen, having n inputs (the predictor order), 5 hidden units and 1 output with a 0.7 learning rate and a 0.5 momentum term.

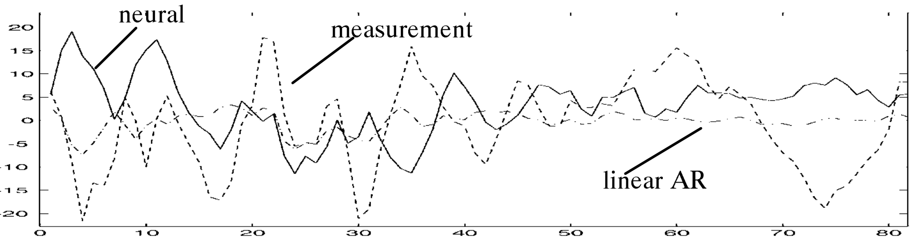

Fig.4. Recursive linear and neural predictions of order 62

After checking the correlation time and the partial autocorrelations of the data segment, the predictor order is set. From a comparison of neural to linear prediction errors (using the Durbin–Levinson algorithm [1]) we find that the latter gives the smaller rms error over all test samples (1.02 vs. 1.54). But from figure 4 (this snapshot of the first 80 predictions for predictor order 62 is typical for much predictor orders) we see, that the linear predictor quickly fades out to zero, while the neural prediction reflects the correct dynamical behaviour but is frequently out of phase (which is the cause of the large rms errors). From [7] it is known that despite the phase lag this already suffices for an estimation of exerted muscle force. To improve performance, we add as features to the network the mean value and standard deviation of the delayed samples as well as the optimal linear AR–prediction. The test errors for order 3 is now 2.93, but already 1.06 for order 62 (see figure 5 for order 137). Although adding statistical features may have a positive influence for larger order, adding movement data has a distinctly negative effect: rms errors are for order 3, 62, 137 equal to 2.56, 1.06, 5.79 respectively.

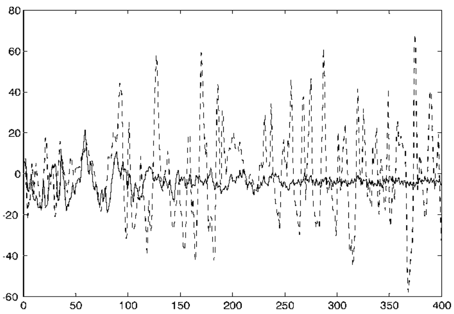

Fig.5. Neural pred.(order 137) using stat. features

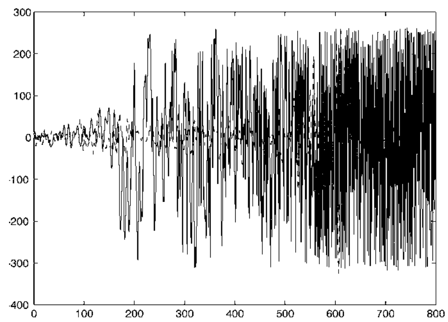

Fig.6. Neural pred.(order 137) using all features

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 323-328

Figure 6 shows a 137−order 800−lag predictor. Now, there appears erroneous behaviour for large delay−lines when using movement features. Note that the oscillations arise only after many recursive predictions, so it may well be that this phenomenon is caused by the propagation and accumulation of the one−lag prediction errors. In figure 7 the final 62−order response is depicted.

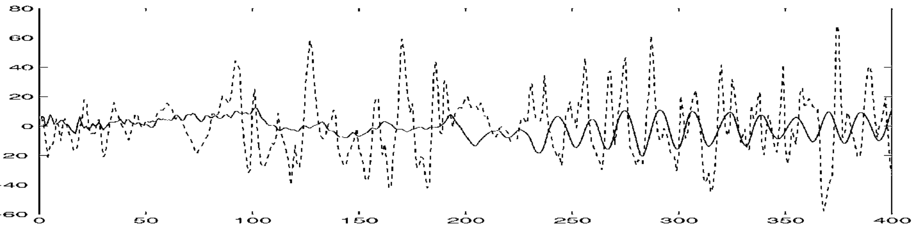

Fig.7. Neural pred.(order 62) using all features

Conclusions.

For signals with linear dynamics, it seems indeed possible to perform prediction with a delay−line only, and we can base its length on the signal's correlation time. Recursive n−lag prediction is likely to result eventually in large artefacts, ultimately leading to saturation of the neuron transfer functions and oscillations. For the application at hand, only the "amount of activity" is required and the neural predictor is fully justified. However, because of the large time constants in the human motor control system, predictions have to be made many samples ahead. This can obviously not be realized by a single ANN with long delay−lines but requires a more heterogeneous set−up.

Acknowledgement

The authors would like to thank J.A.M. Maathuis of Infotronic Medical & Industrial Engineering for introducing the inverse dynamics problem and for supplying the data.

References

[1]. P.Brockwell,R.Davis: Time series: theory and methods. *Springer−Verlag* (1987)
[2]. C. de Groot, D. Wurtz: Analysis of univariate time series with connectionist nets: a case study of two classical examples. *Neurocomputing No. 3* (1991)
[3]. K. Narendra, K. Parthasarathy: Identification and control of dynamical systems using neural networks. *IEEE Trans. Neural Networks Vol. 1, No. 1* (1990)
[4]. A. Weigend, N. Gershenfeld: Time series prediction: forecasting the future and understanding the past. *Addison Wesley* (1994)
[5]. J. Farmer, J. Sidorowich: Predicting chaotic time series. *Phys. Rev. Lett. Vol. 59, No.8* (1987)
[6]. S. Haykin: Neural networks. *MacMillan* (1994)
[7]. Hermens et al.: The clinical use of surface EMG. *Medica Physica No. 9* (1986)
[8]. G. Kleissen et al.: Simultaneous measurement of surface EMG and movements for clinical use. *Med. & Biol. Eng. & Comput. No. 27* (1989)