Available online at www.sciencedirect.com
ScienceDirect
Neurocomputing 70 (2007) 1360–1371
NEUROCOMPUTING
www.elsevier.com/locate/neucom

A novel approach for estimation of optimal embedding parameters of nonlinear time series by structural learning of neural network

Yusuke Manabe, Basabi Chakraborty[a,*]

[a]*Graduate School of Software and Information Science, Iwate Prefectural University, Iwate 020-0193, Japan*

Received 17 November 2004; received in revised form 7 June 2006; accepted 27 June 2006
Communicated by J. Zhang
Available online 20 October 2006

Abstract

In this work a novel approach for estimation of embedding parameters for reconstruction of underlying dynamical system from the observed nonlinear time series by a feedforward neural network with structural learning is proposed. The proposed scheme of optimal estimation of embedding parameters can be viewed as a global non-uniform embedding. It has been found that the proposed method is more efficient for estimating embedding parameters for reconstruction of the attractor in the phase space than conventional uniform embedding methods. The simulation has been done with Henon series and three other real benchmark data sets. The simulation results for short term prediction of Henon Series and the bench mark time series with the estimated embedding parameters also show that the estimated parameters with proposed technique are better than the estimated parameters with the conventional method in terms of the prediction accuracy. The proposed technique seems to be an efficient candidate for prediction of future values of noisy real world time series.

Keywords: Chaos; Nonlinear time series; Strange attractor; Embedding theorem; Embedding parameters; Structural learning; Neural network

1. Introduction

Nonlinear time series are ubiquitous and a lot of them e.g., stock market or exchange rate, meteorological data or network traffic flow, are significant to our society and life. In order to understand and predict future values of nonlinear time series, it is important to analyze the time series to extract knowledge of the underlying dynamical system. Time series are generally sequences of measurements of one or more observable variables of an underlying dynamical system, whose state changes with time as a function of its current state vector. Linear dynamical systems evolve over time to an attracting set of points that are called fixed point attractors and the time series derived from such a system have a regular appearance. There are many linear modelling algorithms for analysing those time series. However, some nonlinear dynamical systems evolve to a chaotic attractor or a strange attractor. The path of the state vector through the attractor is non-periodic, exhibits highly irregular geometrical pattern and sensitive to initial condition. The generated time series shows a complex appearance and behaviour. Many real world observed time series are of such chaotic nature, the actual variables of the underlying dynamical system that contribute to state vector are unknown. We need nonlinear methods for their analysis and modelling in order to predict their future behaviour.

One standard approach for prediction of future values of such a chaotic time series involves the reconstruction of the chaotic dynamics of the phase space from the observed (measured) values of the state vector and thereby predicting the evolution of the measured variable. The *embedding theorem* of Takens [37] guarantees that the space of time delayed vectors with sufficiently large dimension (actually there is an upper bound provided for the embedding dimension) will capture the structure of the original phase space. Unfortunately embedding theorem does not provide any values for the embedding parameters i.e. delay time τ

*Corresponding author. Tel.: +81 19 694 2580; fax: +81 19 694 2573.
E-mail addresses: g236d009@edu.soft.iwate-pu.ac.jp (Y. Manabe), basabi@soft.iwate-pu.ac.jp (B. Chakraborty).

0925-2312/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2006.06.005

and embedding dimension m, a good choice is needed for correct reconstruction of the attractor. Therefore, estimation of the optimal embedding parameters for reconstruction of the nonlinear dynamics has been studied as an important problem. Abarbanel [1] suggested some heuristics for estimation of delay time and embedding dimension.

Artificial neural networks are popular and efficient models for modelling nonlinear phenomenon where various inputs are combined to predict unknown data and they have been extensively used for nonlinear time series analysis. The most popular choice of neural architecture is the multilayer feedforward neural network with back propagation (BP) learning algorithm. Though widely used, the multilayer network with BP learning suffers from some serious drawbacks. The most severe shortcoming is inability to reveal the interpretation of hidden units and connection weights in connection to the underlying distribution of the data model. Structural learning or pruning techniques are one of the approaches that aim to minimize the problems of BP learning algorithm. A variety of structural learning techniques have been proposed in several papers [10,16,27,32,33,26,6].

Although the initial researches on structural learning [32,33] aimed to improve generalization ability of neural network, later it is shown that structural learning or pruning techniques with regularization are capable of discovering rules in classification problems [10,16] and selection of essential input components or analysis of input–output relations in time series prediction problems [27,26,6]. The use of regularizer in neural network learning has also has been used in automatic relevance detection of input units to the target concept [24,29] and the magnitude of connection weights has been used in selection of inputs [40,23]. Neural network with several tuning parameters is also a good candidate for optimal model selection from various alternatives. Neural network with hierarchical Bayesian learning has also been successfully applied for estimation of embedding dimension for generated time series data in [28].

In order to accurately reconstruct the chaotic dynamics of nonlinear time series for prediction of its future values, we propose an algorithm for optimal estimation of embedding parameters from the nonlinear time series by neural network model trained by structural learning with regularizer. An initial version of the algorithm has been compared with existing theoretically motivated heuristic approaches using Lorenz data in [25]. In this paper we present the proposed algorithm in more detail and simulation results of artificial data as well as some benchmark real data sets to show its effectiveness. In the next two sections, the problem of estimation of embedding parameters for modelling and prediction of time series followed by the related work for its implementation have been discussed. In the following section the proposed scheme for optimal model selection for estimation of embedding parameters has been described. The next section represents simulation experiments and their results followed by the final section with discussion and concluding remarks.

2. Reconstruction of chaotic attractor

2.1. Time series and dynamical system

The relation between a time series and its underlying dynamical system can be expressed by the following equations:

$$\boldsymbol{u}(t+1) = \boldsymbol{F}[\boldsymbol{u}(t)] + \boldsymbol{\xi}(t), \tag{1}$$

$$y(t) = \boldsymbol{g}[\boldsymbol{u}(t)] + \boldsymbol{\eta}(t), \tag{2}$$

where $\boldsymbol{u}(t)$ is the state vector of the dynamical system at time t, the function $\boldsymbol{F}$ represents the state change of dynamical system. $y(t)$ is the observed value at time t while $\boldsymbol{g}$ represents the observation function, $\boldsymbol{\xi}(t)$ and $\boldsymbol{\eta}(t)$ are dynamical noise and observational noise at time t, respectively. Eqs. (1) and (2) represent that we cannot directly obtain the state $\boldsymbol{u}(t)$ of original dynamical system.

2.2. Embedding theorem and delay coordinate embedding

Embedding theorem, which is developed by Takens [37] and expanded by Sauer et al. [3], guarantees that, with a single observed time series $y(t)$, we can obtain the following $\boldsymbol{f}$ which has one-to-one correspondence to the original dynamical system.

$$\begin{aligned} &\boldsymbol{v}(t+1) = \boldsymbol{f}(\boldsymbol{v}(t)), \\ &\boldsymbol{v}(t) = (y(t), y(t+\tau), \ldots, y(t+(m-1)\tau)), \end{aligned} \tag{3}$$

where $\boldsymbol{f}$ denotes reconstructed dynamical system, $\boldsymbol{v}$ denotes time delay coordinate vector, m is called *embedding dimension* and τ is called *delay time*. An example of embedding of Lorenz System with the effect of different values of m and τ on reconstruction of the original system is shown in Fig. 1 where (a) is original Lorenz attractor with x, y, z variables, (b) is a single time series of variable x, (c)–(f) are reconstructed attractor from variable x with various m and τ. From the example it is clear that for correct reconstruction of the attraction, a fine estimation of the parameters (m and τ) is needed.

2.3. Estimation of embedding parameters

There are variety of heuristic techniques for estimating the embedding parameters, m and τ, the details can be found in [1,3,14]. Here we briefly review the most representative ones. For the estimation of delay time τ, average mutual information (AMI) [8] at varying sample rates is computed and the first minimum is taken as the appropriate sample rate. The most commonly used techniques to estimate the embedding dimension m are false nearest neighbour (FNN) [15] and singular value analysis [2].

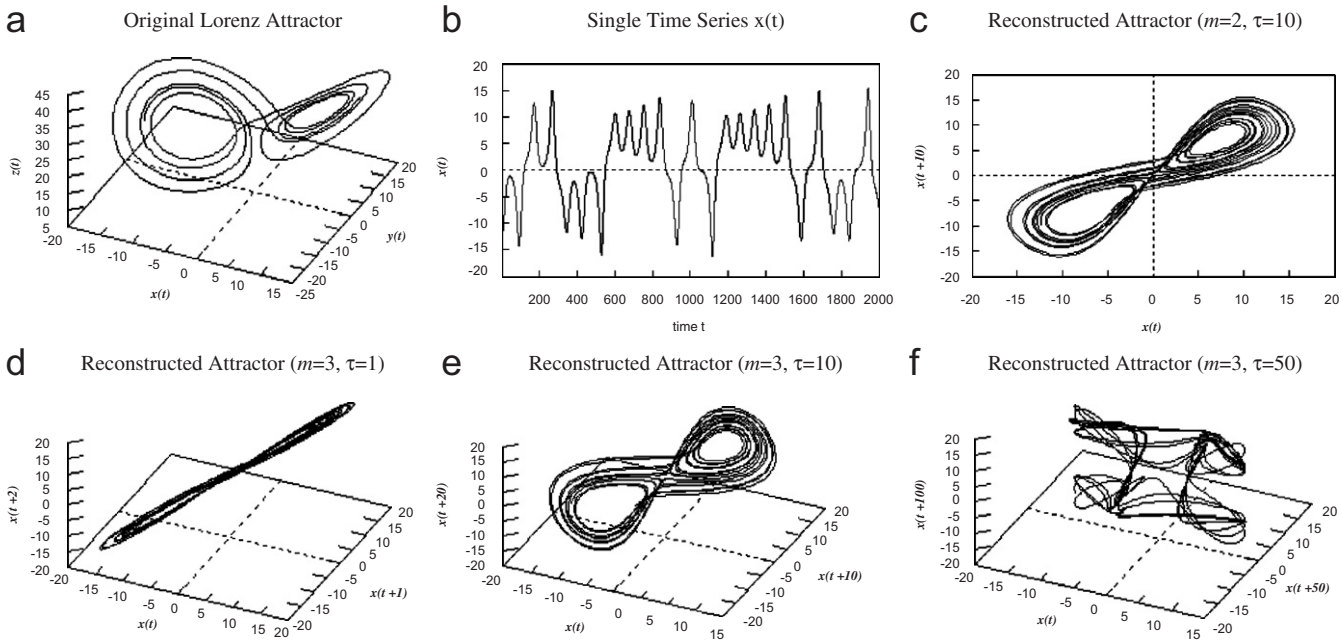


Fig. 1. An example of attractor reconstruction.

2.3.1. Estimation of delay time τ: ami

AMI $I(\tau)$ proposed by Fraser and Swinney [8], one of the popular techniques to estimate the optimal τ, is as follows:

$$I(\tau) = \sum_{y(n),y(n+\tau)} P(y(n), y(n+\tau)) \times \log_2 \frac{P(y(n), y(n+\tau))}{P(y(n))P(y(n+\tau))}, \quad (4)$$

where $P(y(n))$ or $P(y(n+\tau))$ denote probability density functions of $y(n)$ or $y(n+\tau)$, respectively, $P(y(n), y(n+\tau))$ denotes joint probability density function of $y(n)$ and $y(n+\tau)$. These probability density functions are calculated from histograms of each time series. When τ becomes large, $y(n)$ and $y(n+\tau)$ become independent, and $I(\tau)$ will tend to zero. It was the suggestion of Fraser that one use the function $I(\tau)$ as a kind of autocorrelation function to determine optimal delay time τ. In general, the maximum of function $I(\tau)$ shows coherence between $y(n)$ and $y(n+\tau)$ is very strong and a minimum shows a point where sample points are maximally decorrelated. The optimal delay time τ is the first minimum of function $I(\tau)$.

2.3.2. Estimation of embedding dimension m: fnn

The most popular method for estimation of m is FNN, proposed by Kennel [15]. FNN makes use of topological feature of dynamical system. The reconstructed attractor must be a one-to-one image of the attractor in the original phase space. The algorithm is as follows:

Delay coordinate vector in m dimension is

$$\boldsymbol{v}(t) = (y(t), y(t+\tau), y(t+2\tau), \ldots, y(t+(m-1)\tau)),$$

and nearest vector of $\boldsymbol{v}(t)$ is

$$\boldsymbol{v}^{NN}(t) = (y^{NN}(t), y^{NN}(t+\tau), \ldots, y^{NN}(t+(m-1)\tau)).$$

Then, Euclidean distance between $\boldsymbol{v}(t)$ and $\boldsymbol{v}^{NN}(t)$ in m dimension is

$$R_m(t) = \sqrt{\sum_{k=1}^{m} \{y(t+(k-1)\tau) - y^{NN}(t+(k-1)\tau)\}^2}.$$

On the other hand, Euclidean distance between $\boldsymbol{v}(t)$ and $\boldsymbol{v}^{NN}(t)$ in $m+1$ dimension is

$$R_{m+1}(t) = \sqrt{\sum_{k=1}^{m+1} \{y(t+(k-1)\tau) - y^{NN}(t+(k-1)\tau)\}^2}.$$

Then, relative distance between $\boldsymbol{v}(t)$ and $\boldsymbol{v}^{NN}(t)$ in $m+1$ dimension is

$$R_L = \sqrt{\frac{R_{m+1}(t)^2 - R_m(t)^2}{R_m(t)^2}} = \frac{|y(t+m\tau) - y^{NN}(t+m\tau)|}{R_m(t)}.$$

When this quantity R_L is larger than some heuristic threshold, we have a false neighbour. The optimal embedding dimension, which is sufficiently high for reconstructing attractor, is the dimension in case fraction of false neighbour is zero or at least sufficiently small.

3. Related work on estimation of optimal embedding parameters

The standard approach to time delay embedding for reconstruction of non-linear time series is to choose an embedding dimension m and sample the time series $y(t)$ at equal time intervals τ to obtain the function of Eq. (3). This is known as uniform embedding.

Several authors [5,17] argued that the optimal model of the dynamics in certain situations depends on embedding window ($d_w = m \times \tau$) rather than on embedding lag τ only. But in practice we need to estimate both the parameters for prediction problems. Optimal embedding parameters exist because real world observed time series is finite and noisy. If m is too large, a complicated attractor is reconstructed and it is difficult to predict and control it. If τ is too small, delay coordinate vectors are plotted as an overborne attractor in reconstructed phase space owing to strong coherence between the vectors. If τ is too large, owing to orbital instability, the vectors scatter (non-coherence) in the state space.

Many competing criteria to select the embedding parameters have been developed and all are heuristic. As selection of τ is essentially dependent on the chosen model [11], Small and Tse [36] proposed algorithm to find optimal d_w which minimises one step prediction error for long noise free time series followed by derivation of suitable τ. However, uniform embedding is not the only possibility for correct approach of embedding for every chaotic system, it is quite suitable for classical systems such as Rossler or Lorenz system but it fails for time series with multiple periodicity. Judd et al. [12] suggested non-uniform embedding as a better approach over conventional uniform embedding. The problem of estimating the optimum embedding parameters is to find the embedding window d_w and the optimal parameters $\{l_i | i = 1, \ldots, k\}$, where

$$\boldsymbol{v}(t) = \{y(t+l_1), y(t+l_2), \ldots, y(t+l_k)\} \quad (0 \leqslant l_1 < l_2 < \cdots < l_k \leqslant d_w). \tag{5}$$

They used cylindrical basis models, a class of standard radial basis model, and minimum description length as the criterion function for optimal model selection for estimation of optimum embedding parameters. But in this approach, for large d_w, the computational burden of selection of optimal set of $\{l_i | i = 1, \ldots, k\}$ from $(2^{d_w} - 1)$ candidates by exhaustive search is quite heavy. Small [34] proposed genetic algorithm for searching the best candidate when d_w is moderate ($d_w > 10$) by minimizing the model prediction error with randomly selected subset of data to minimize computational effort and showed that non-uniform embedding strategy is better for capturing the dynamics of the original systems than comparable uniform embedding. But the problem with adopting this strategy is that the selection of the optimal embedding parameters becomes computationally intractable for even moderate d_w with large N, the data length.

Thus, the problem of finding the correct embedding parameters for a particular time series depends on the model and the model selection scheme. Neural network models are one of the popular tools for modelling nonlinear phenomenon though their performances largely depend on successful training. Multilayer feedforward models and basis function networks are extensively used for modelling nonlinear time series [18,20,30]. For optimal model selection of highly parameterized neural network models, information theoretic criterion introduced by Akaike [4] and minimum description length or predictive minimum description length criterion have been used in several works [7,35,19]. Those methods usually require a maximum likelihood estimator which is difficult to derive analytically for neural networks. The key point is to find out an appropriate structure of the neural network.

In the present work, we propose an algorithm for estimating optimal embedding parameters by using feedforward neural network model with structural learning and cross validation for successful training and model selection. Our approach can be viewed as a global non-uniform embedding because of utilizing neural network model. On the other hand, the approach by Judd et al. [11,12] utilizing cylindrical basis models, can be viewed as local non-uniform embedding. Our proposed algorithm is introduced in the next section.

4. Proposed algorithm for refined estimation of embedding parameters

4.1. Refinement scheme by neural network learning

In Section 2, we described conventional estimation methods of embedding parameters. The estimation of parameters for non-uniform embedding strategy, i.e. $\{l_i | i = 1, \ldots, k\}$ and k, resembles the problem of selecting the best set of inputs to model some unknown quantity. Here considering this analogy we propose a feed forward multilayer neural network model with structural learning for estimation of optimal refined embedding parameters for nonlinear time series. The proposed algorithm consists of four steps which are described below.

4.1.1. Step 1: Estimation of provisional embedding vector

In the first step embedding parameters m and τ are estimated by conventional heuristic algorithms AMI and FNN, respectively. The value of m and τ are then used to construct provisional embedding vector as follows:

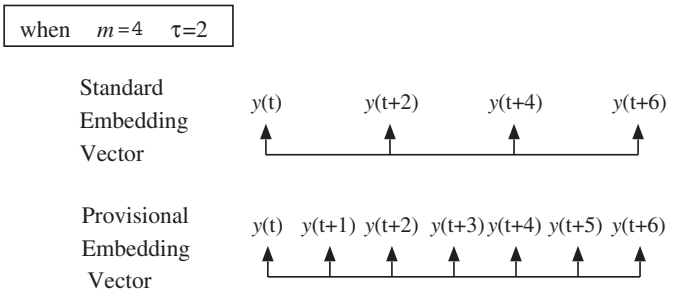
$$\boldsymbol{v}_p(t) = \{y(t), y(t+1), y(t+2), \ldots, y(t+(m-1)\tau)\}$$

instead of standard embedding vector expressed in Eq. (4). The constructed provisional vector should contain the standard estimated vector by conventional methods. Fig. 2 represents the embedding for the case when $m = 4$ and $\tau = 2$.

4.1.2. Step 2: Learning the predictive mapping by neural network with pruning

In Step 2 neural network is used to learn the predictive mapping $f : \boldsymbol{v}_p(t) \in R^{(m-1)\tau+1} \to y(t+(m-1)\tau)+\rho) \in R^1$ as follows:

$$y(t+(m-1)\tau)+\rho) = f_{NN}(y(t), y(t+1), y(t+2), \ldots, y(t+(m-1)\tau)). \tag{6}$$



Provisional embedding vector contains general embedding vector.

Fig. 2. Standard embedding vector and proposed provisional embedding vector.

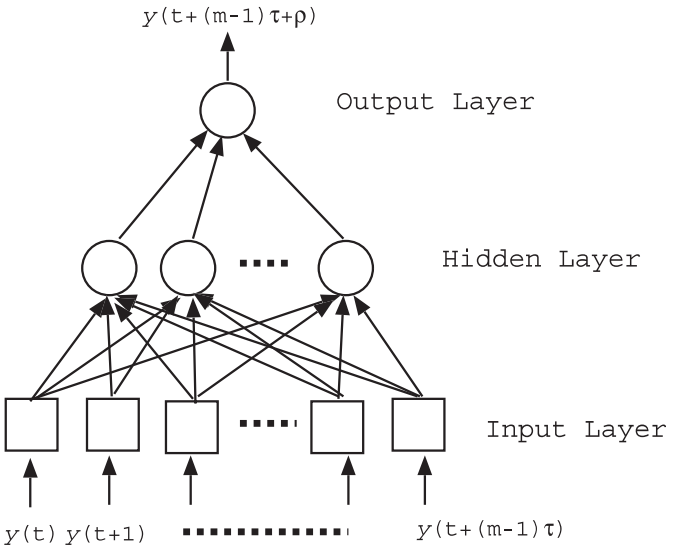


Fig. 3. Neural network structure for learning predictive mapping.

Feedforward three layer neural network for learning the mapping shown in Fig. 3 with $((m-1)\tau+1)$ neurons in the input layer and 1 neuron in the output layer. The initial number of neurons in the hidden layer is adequate and it is optimized by pruning algorithm during learning. The difference between reconstructed dynamics in Eq. (3) and predictive mapping in Eq. (6) is to be noted here. Although the reconstructed dynamics is of the form of multi-input multi-output, the mapping consists of several nonlinear filters and the nonlinear predictive mapping [31]. The filters does the simple shifting operation without estimation. The learning of nonlinear predictive mapping is the most important task for us. We tried multi-input multi-output neural network with pruning algorithm for learning and found that after learning the emerging network structure reflects both the shifting operations and the nonlinear mapping. The connections reflecting the shifting operations are meaningless for our purpose and only add to the computational cost. So we used multi-input single output neural network for our simulation experiment.

S-fold cross validation is used for validation of the trained neural network. The objective of structural learning is to optimize the network topology and here we have used structural learning with Laplacian regularizer proposed by Ishikawa [10] with hidden unit minimization proposed in [33]. The learning phase comprizes of the following two criteria functions which are to be applied in order:

$$J_{\mathrm{I}} = \frac{1}{2}\sum_{k}(o_k - t_k)^2 + \lambda \sum_{ij} |w_{ij}|, \tag{7}$$

$$J_{\mathrm{II}} = \frac{1}{2}\sum_{k}(o_k - t_k)^2 + \lambda \sum_{|w_{ij}|<\theta_\lambda} |w_{ij}|, \tag{8}$$

where λ denotes relative weights of the Laplace regularizer term, w_{ij} is a connection weight from ith unit to jth unit and θ_λ is a threshold parameter. The learning by J_{I} achieves structural optimization by data fitting and detecting redundant connection weights while the learning by J_{II} contributes to the minimization of training error for fine tuning of the training. Finally the hidden layer units whose output variance become near zero are deleted by adding their output values as the bias to the next higher layer units as follows:

$$b_k^{\mathrm{new}} = b_k^{\mathrm{old}} + w_{jk}\bar{h_j} \quad (\text{if } \sigma_{h_j}^2 \simeq 0), \tag{9}$$

where $\sigma_{h_j}^2$ denotes variance of jth hidden unit, b_k denotes bias of the kth unit in the next higher layer and $\bar{h_j}$ denotes average output of jth hidden unit. Now to achieve the optimal structure of the network fine tuning of λ is necessary. We trained neural network several times by varying λ from higher to lower values and selected optimal network model which is explained in the next step.

4.1.3. Step 3: Selection of optimal neural network model

The optimal neural network model has been chosen for the value of λ for which relative normalized score (RNS) $RNS(\lambda)$ is minimum. Now $RNS(\lambda)$ is calculated as follows:

$$\begin{aligned} RNS(\lambda) &= K(\lambda)' + V(\lambda)', \\ K(\lambda)' &= (K(\lambda) - K_{\min})/(K_{\max} - K_{\min}), \\ V(\lambda)' &= (V(\lambda) - V_{\min})/(V_{\max} - V_{\min}), \end{aligned} \tag{10}$$

where $K(\lambda)$ represents the average number of network parameters (the summation of connection weights and bias parameters greater than θ_λ) and $V(\lambda)$ denotes average validation error in S-fold cross validation. $K_{\max}(K_{\min})$ and $V_{\max}(V_{\min})$ represent maximum (minimum) values of K and V, respectively. As here model selection is based not only on the cross validation error but also on the number of parameters of the network, the selected model reflects the data distribution. Use of this score helps to achieve optimal parameter value that can balance between training and pruning. For optimal model selection, AIC (Akaike information criterion), MDL (minimum description length) and BIC (Bayes information criterion) are often used. The justification of using a new criterion RNS is explained in a later section.

4.1.4. Step 4: Estimation of embedding parameters from neural network model

In the last step, the embedding parameters are estimated from the selected neural network model. The input units with the total of connection weights $|W_i|$ greater than the average value are selected as the important input units. The $|W_i|$ is calculated as follows:

$$|W_i| = \sum_{k=1}^{S} \sum_{j=1}^{H} |w_{ijk}|, \tag{11}$$

where S represents the number of partition of the data set for cross validation and H represents the number of hidden units of the trained network. The set of optimal input units serve as the embedding vector. The number and the interval of the important input units represent embedding dimension and delay, respectively. $|W_i|$ can provide the priority with each component of embedding vector according to weight values connected to each input unit. Thus, one can select various embedding vectors from a set of embedding vectors according to the priority.

4.2. Related criterion on optimal model selection

One of the popular criterion for optimal model selection is to utilize information criterion. Information criterion (IC) is formed as follows:

$$\text{IC} = \text{log likelihood for data fitting} + \text{number of model parameters.}$$

Generally the model with the smallest information criterion represents optimal model, which should have good data fitting as well as small degree of parameters. In case of linear models the large number of parameters is associated with good data fitting. But in case of nonlinear model like neural network, the model with extremely small parameters can often achieve good data fitting. In such a case, the model with the small first term is selected as the optimal model without considering the second term and information criterion does not work well. It has also been found that AIC and BIC both work well for linear models [41] and in practical application AIC tends to overfit data [21]. Moreover, when the number of parameters are large AIC is not so effective [13]. In our problem, first term is extremely stronger than the second term and several IC, i.e. AIC, BIC, MDL, did not work well. Therefore, we suggested RNS to find out optimal model which is also straightforward and easy to compute.

As an alternative approach, MacKay proposed Bayesian estimation framework for learning of neural network and estimation of hyperparameter [22]. Hyper parameter denotes the relative weights for regularization term, i.e. λ in our work. Moreover, Matsumoto et al. [28] proposed hierarchical Bayes approach in line with MacKay's approach. Matsumoto et al. successfully applied the neural network with hierarchical Bayes learning to estimate embedding dimension for some artificial data, i.e. Henon system and Rossler system. However, they considered uniform embedding vector and only estimated embedding dimension.

5. Simulation experiments

In order to evaluate our proposed refinement scheme an artificial time series data generated from Henon System with added noise has been used for simulation. Furthermore, we have used for simulation three real data sets named as NH3 Laser data, Sunspot data and Measles data used as benchmark data in chaotic time series prediction problems.

5.1. Description of the data sets

5.1.1. Henon time series

Henon system is represented by the following equation:

$$x(t+1) = 1 - ax(t)^2 + z(t),$$
$$z(t+1) = bx(t). \tag{12}$$

The time series is generated with the parameters $a = 1.4$, $b = 0.3$, $x(0) = 0.1$, $z(0) = 0.2$. In this simulation, we use $x(t)$ with $t = 101$–300 as observed time series and $x(t)$ with $t = 301$–350 as test time series. $x(t)$ is added 30 dB gaussian noise $\eta(t)$.[1] The generated time series $y(t)$ is given by

$$y(t) = x(t) + \eta(t). \tag{13}$$

5.1.2. Real data sets

NH3 Laser data were recorded from a NH3 far-infrared-laser in a chaotic state. The details can be found in [38]. Total number of the data is 1000, first 500 has been used for analysis and modelling and next 500 has been used for prediction.

Sunspot data were recorded by Wolf from 1700 to 1988. This data is downloaded from Time Series Data Library [39]. Total number of the data is 289, first 150 was used for analysis and modelling and next 139 has been used for prediction.

Measles data is monthly reported number of cases of measles in New York city from 1928 to 1972. This data is downloaded from Time Series Data Library, too. The total number of the data in the set is 534 of which 300 have been used for analysis and 234 data have been used for prediction.

5.2. Simulation results by conventional method

The simulation experiments have been done by TISEAN software package programs [9]. The results of conventional methods are shown in Figs. 4 and 5. From these graphs, the delay time τ and embedding dimension m are $\tau = 6$, $m = 3$ (Henon), $\tau = 2$, $m = 4$ (NH3 Laser), $\tau = 5$, $m = 4$

[1] Noise level (dB) is calculated by $10\log_{10}(\sigma_s^2/\sigma_n^2)$ where σ_s^2 denotes variance of signal time series and σ_n^2 denotes variance of noise time series.

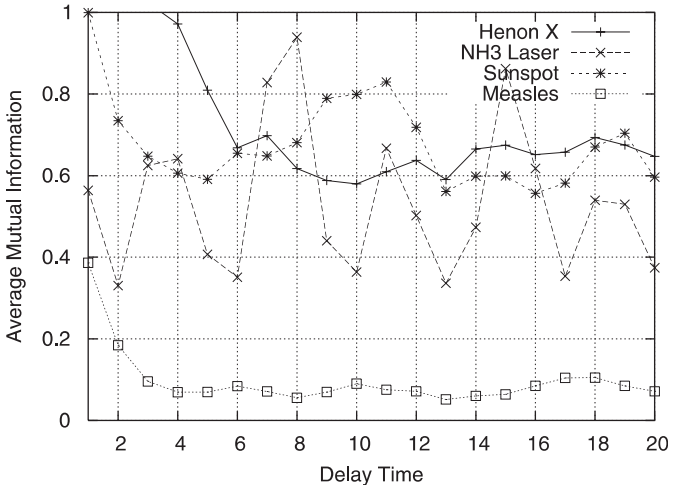


Fig. 4. Average mutual information.

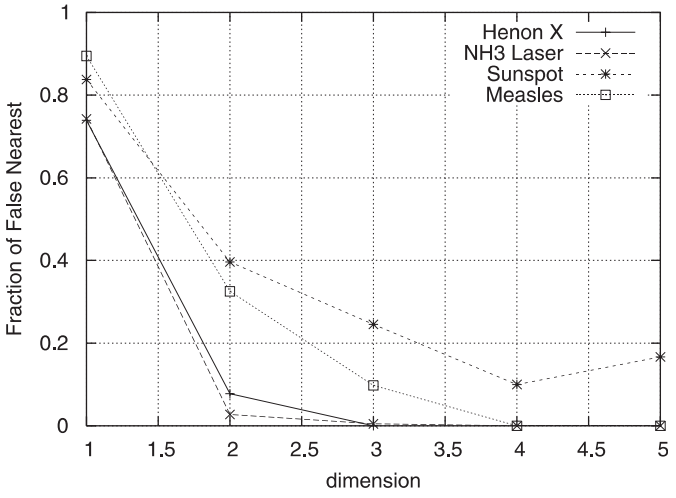


Fig. 5. False nearest neighbour.

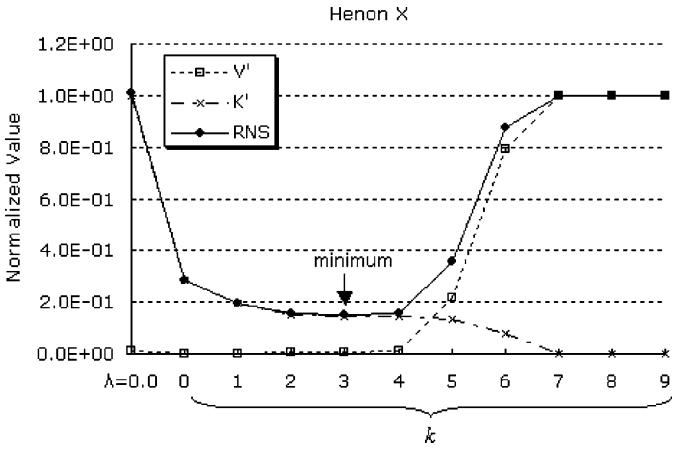


Fig. 6. Relative normalized score for Henon system's data.

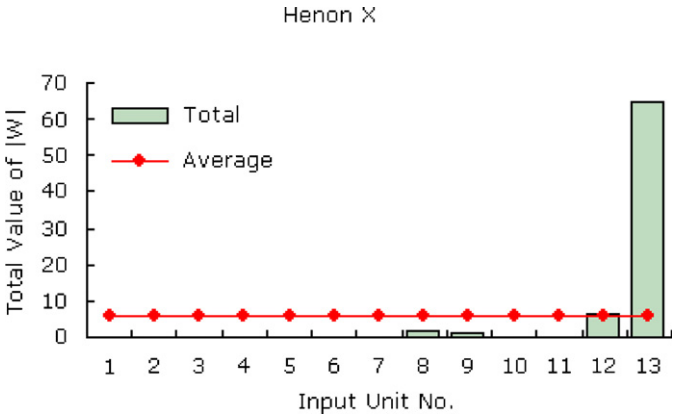


Fig. 7. Total absolute value of weights for Henon system's data.

(Sunspot) and $\tau = 4$, $m = 4$ (Measles), respectively. Henon system is two dimensional because of two variables x and y. However, in case of conventional methods, embedding dimension is estimated high because of observation noise.

5.3. Simulation results by proposed method

The simulation results with the proposed refinement scheme have been presented in the following sections.

5.3.1. Henon system

From the result of conventional methods, provisional embedding vector is set as follows:

$$\boldsymbol{v}_p(t) = \{y(t), y(t+1), \ldots, y(t+12)\}.$$

The neural network parameters are taken as follows: no. of input units = 13, no. of hidden units = 6, no. of output units = 1. The parameter settings for learning are learning rate $\eta = 0.1$, momentum rate $\alpha = 0.2$, threshold for selective regularizer $\theta_\lambda = 0.1$. BP learning with learning epochs $\{J_{\mathrm{I}}, J_{\mathrm{II}}\} = \{30\,000, 20\,000\}$ has been used. Regularization parameter is varied to 11 patterns, i.e. $\lambda = 0.0$, $10^{-5} \times 1.7^k$ $(k = 0, 1, \ldots, 9)$. No. of partition S for cross validation learning is taken as 5. Henon data for training is scaled from 0.1 to 0.9.

Fig. 6 represents RNS. This shows that RNS is minimum when $k = 3$. Fig. 7 shows cumulative absolute value of weights connected with input units when RNS is minimum. From this figure, we found that the priority of essential input units are #13, #12, #8 .... Especially, input unit #13 is quite prominent component. Thus, one can select several set of embedding vectors from 1 to 3 dimension as follows:

$$\boldsymbol{v}(t) = \{y(t)\},$$
$$\boldsymbol{v}(t) = \{y(t), y(t+1)\},$$
$$\boldsymbol{v}(t) = \{y(t), y(t+4), y(t+5)\}.$$

The selection of embedding parameters are evaluated by short term prediction performance.

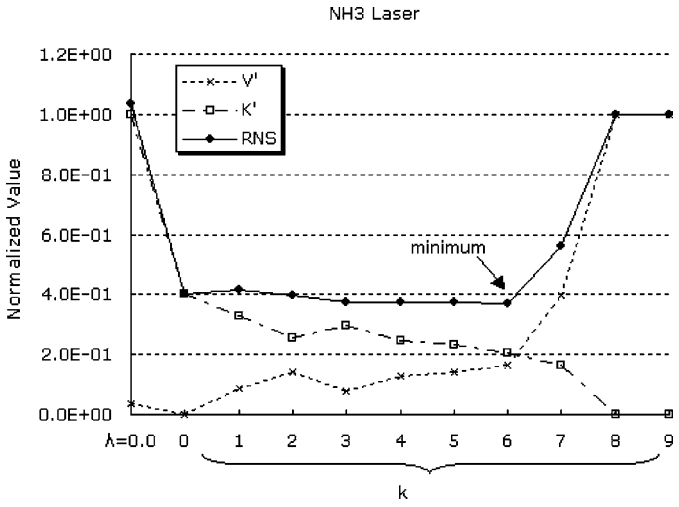


Fig. 8. Relative normalized score for NH3 Laser data.

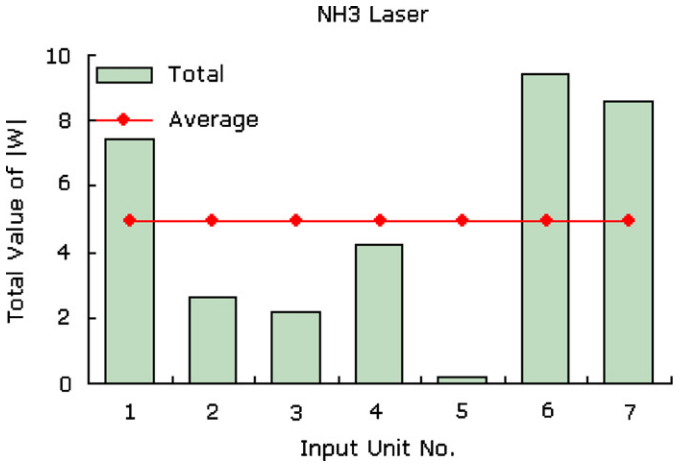


Fig. 9. Total absolute value of weights for NH3 Laser data.

5.3.2. NH3 laser data

From the estimated parameters of the conventional method the provisional embedding vector as follows:

$$\boldsymbol{v}_p(t) = \{y(t), y(t+1), y(t+2), \ldots, y(t+6)\}.$$

The neural network parameters are taken as follows: no. of input units = 7, no. of hidden units = 4, no. of output units = 1. All learning parameters are set as same as the simulation experiment for Henon map. RNS has been calculated as before and found to be minimum for $k = 6$ as is evident from Fig. 8, Cumulative absolute value of weights connected with input unit when RNS is minimum shown in Fig. 9. The priority of essential inputs come out to be input unit #6, #7, #1, #4 .... We can select several set of embedding vectors from 2 to 4 dimension as follows:

$$\boldsymbol{v}(t) = \{y(t), y(t+1)\},$$

$$\boldsymbol{v}(t) = \{y(t), y(t+5), y(t+6)\},$$

$$\boldsymbol{v}(t) = \{y(t), y(t+3), y(t+5), y(t+6)\}.$$

5.3.3. Sunspot data

For this data the provisional embedding vector is as following (based on estimation from conventional method):

$$\boldsymbol{v}_p(t) = \{y(t), y(t+1), y(t+2), \ldots, y(t+15)\}.$$

The neural network parameters are taken as follows: no. of input units = 16, no. of hidden units = 8, no. of output units = 1. All learning parameters are set as same as the simulation experiment for Henon map. RNS has been calculated as before and found to be minimum for $k = 4$ as shown in Fig. 10. Cumulative absolute value of weights connected with input unit when RNS is minimum shown in Fig. 11. The priority of essential inputs come out to be input unit #16, #15, #9, #13 .... One can select embedding vectors from 2 to 4 dimension as follows:

$$\boldsymbol{v}(t) = \{y(t), y(t+1)\},$$

$$\boldsymbol{v}(t) = \{y(t), y(t+6), y(t+7)\},$$

$$\boldsymbol{v}(t) = \{y(t), y(t+4), y(t+6), y(t+7)\}.$$

5.3.4. Measles data

The provisional embedding vector calculated from the estimated embedding parameter by conventional method is

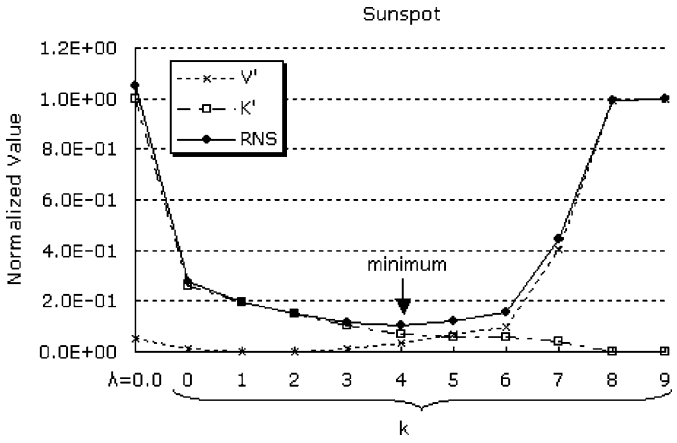


Fig. 10. Relative normalized score for Sunspot data.

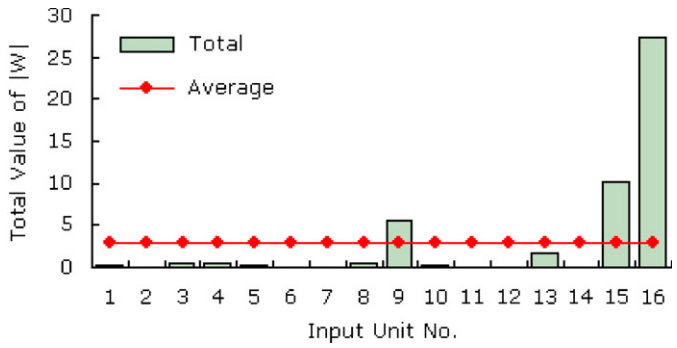


Fig. 11. Total absolute value of weights for Sunspot data.

the following:

$$\boldsymbol{v}_p(t) = \{y(t), y(t+1), y(t+2), \ldots, y(t+12)\}.$$

The neural network parameters are taken as follows: no. of input units = 13, no. of hidden units = 6, no. of output units = 1. All learning parameters are set as same as the simulation experiment for Henon map. RNS has been calculated as before and found to be minimum for $k = 2$ as shown Fig. 12. Cumulative absolute value of weights connected with input unit when RNS is minimum shown in Fig. 13. The priority of essential inputs come out to be input unit #13, #12, #1, #11 .... One can select embedding vectors from 2 to 4 dimension as follows:

$$\boldsymbol{v}(t) = \{y(t), y(t+1)\},$$

$$\boldsymbol{v}(t) = \{y(t), y(t+11), y(t+12)\},$$

$$\boldsymbol{v}(t) = \{y(t), y(t+10), y(t+11), y(t+12)\}.$$

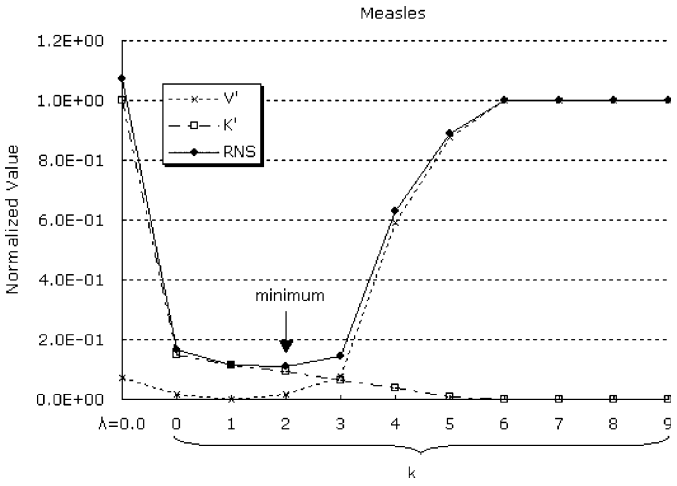


Fig. 12. Relative normalized score for Measles data.

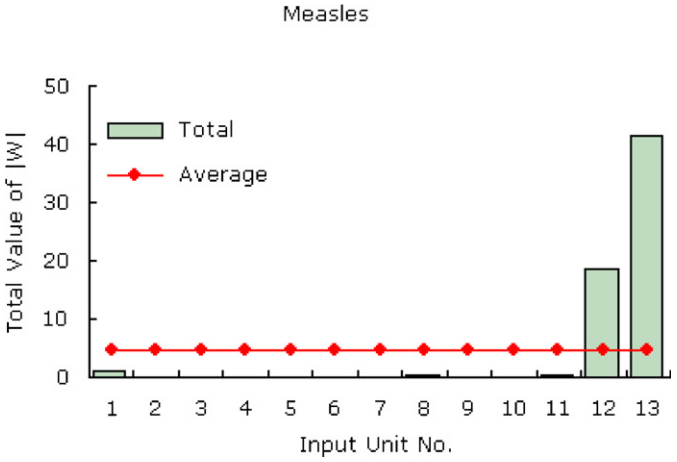


Fig. 13. Total absolute value of weights for Measles data.

5.4. Validation of estimated parameters by short term prediction

In order to evaluate the features of reconstructed attractor with the estimated parameters we have done short term prediction of all the time series used for analysis. Local linear predictor has been used for predicting future values of the time series. Local linear predictor for ρ-step previous is shown as follows:

$$\hat{\boldsymbol{v}}(t+\rho) = \sum_{i=1}^{K} \boldsymbol{v}(t_i + \rho), \tag{14}$$

where $\hat{\boldsymbol{v}}(t+\rho)$ denotes the predicted value of $\boldsymbol{v}(t)$, K denotes the number of nearest neighbour vector of $\boldsymbol{v}(t)$ and $\boldsymbol{v}(t_i+\rho)$ denotes ρ-step previous vector of the nearest neighbour vector of $\boldsymbol{v}(t)$. The prediction accuracy of this method strongly depends on the topological feature of reconstructed attractor. This shows that the prediction accuracy reflects whether reconstructed attractor is better or not. One can quantitatively evaluate the reconstructed attractor, that is embedding parameters, by the prediction accuracy. That is why we selected the method as predictor. Correlation coefficient defined as below is used as evaluation index for comparison of prediction results

$$R = \frac{\sum_{t=1}^{P}(z(t)-\bar{z})(\hat{z}(t)-\bar{\hat{z}})}{\sqrt{\sum_{t=1}^{P}(z(t)-\bar{z})^2}\sqrt{\sum_{t=1}^{P}(\hat{z}(t)-\bar{\hat{z}})^2}}, \tag{15}$$

where $z(t)$ denotes real time series, $\hat{z}(t)$ denotes predicted time series, $\bar{z}$ and $\bar{\hat{z}}$ denote the average of real time series and predicted time series, respectively. P denotes the number of prediction data. One step prediction results for different number of nearest neighbour vectors for three real data sets have been shown in Figs. 14–17, respectively. In all the cases, embedding vectors estimated by our method provide high prediction accuracy. It is found that the proposed estimation scheme seems estimate the parameters better than conventional techniques and the

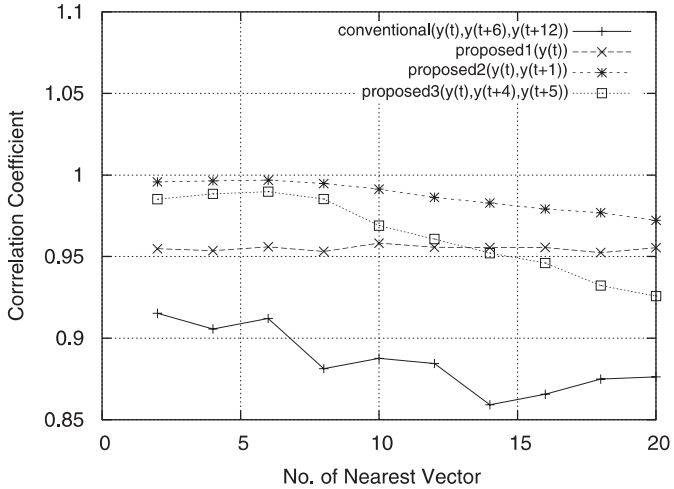


Fig. 14. One step prediction accuracy for Henon time series.

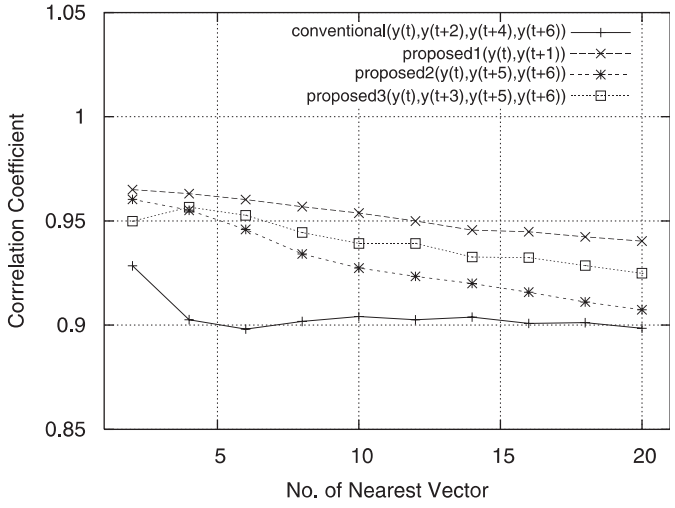


Fig. 15. One step prediction accuracy for NH3 Laser data.

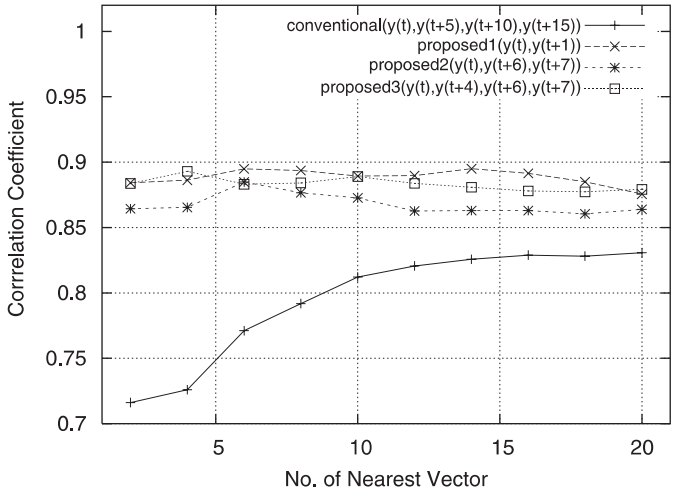


Fig. 16. One step prediction accuracy for Sunspot data.

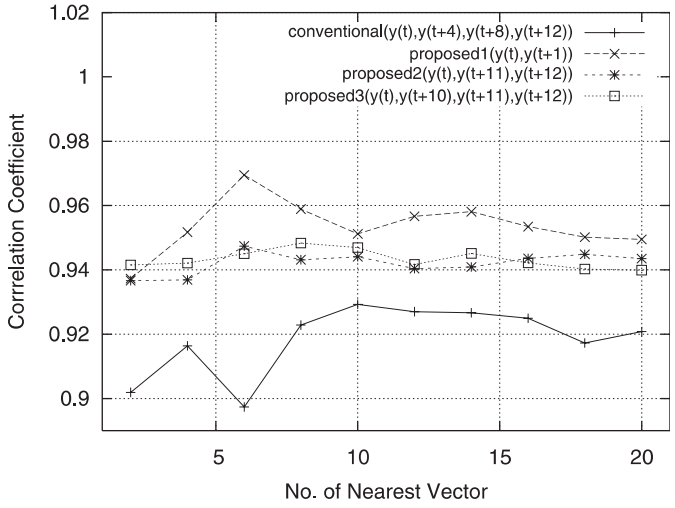


Fig. 17. One step prediction accuracy for Measles data.

average of $|W_i|$ gives an indication of essential components.

5.5. Possibility of long term prediction

Here, we examined about long term prediction accuracy of non-uniform embedding vectors by using three real time series. Embedding vectors estimated by proposed method are composed of the embedding parameters greater than the average of $|W_i|$.

The results of long term prediction for different prediction steps are shown in Figs. 18–20, respectively. Local linear predictor used here considers five nearest vectors. The result shows that for 3 step or 4 step prediction, the prediction accuracy with the refined estimated parameters by proposed scheme is better than the conventional methods. In the proposed algorithm the neural network used for refinement of embedding parameters has been trained for one step prediction. For good

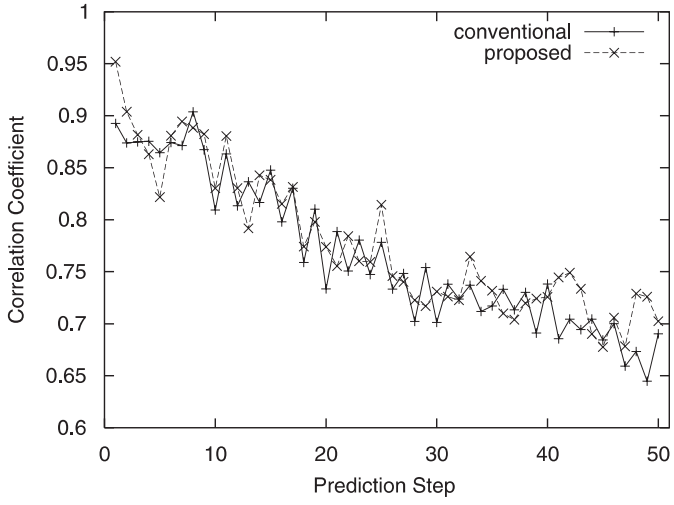


Fig. 18. Long term predictability for NH3 Laser data.

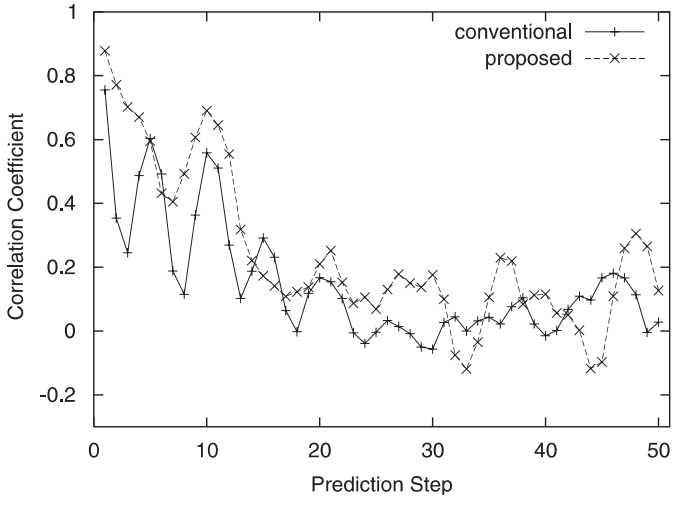


Fig. 19. Long term predictability for Sunspot data.

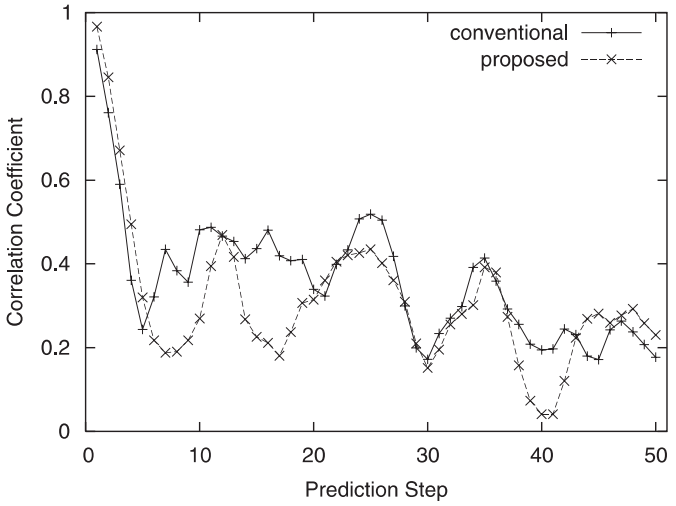


Fig. 20. Long term predictability for Measles data.

prediction accuracy for long term prediction, the optimal parameter estimation by the proposed technique may need learning of neural network with training data for long term prediction.

6. Conclusion

Analysis and prediction of nonlinear time series is crucial as we encounter them frequently in our daily life. Proper estimation of optimal embedding parameters for the reconstruction of the underlying dynamical system from the observed chaotic time series is greatly needed for the purpose of its understanding and correct prediction of future values. Conventional methods for uniform embedding are heuristic. Non-uniform embedding seems to capture the dynamics of real time series better than uniform embedding. Estimation of optimal embedding parameters for non-uniform embedding is computationally heavy. In this work an efficient scheme for estimation of optimal embedding parameters for global non-uniform embedding has been proposed. The proposed technique uses neural network model with structural learning.

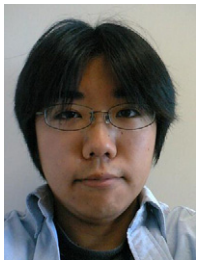
The simulation results with Henon series shows that the estimated embedding parameters are good enough for reconstruction of the phase space attractor as is evident from comparison with the original attractor. It is found from the simulation results with the noisy real data sets and also from Henon series that the estimated embedding parameters can better represent the dynamics of the generated time series than the estimated parameters from the conventional method because the prediction accuracy for one step prediction with the refined estimated parameters are better than the estimated embedding parameters by conventional method. So the proposed technique can be used for efficient short term prediction of future values of noisy real world time series. Long term prediction is difficult and we are trying to explore now to estimate the optimal embedding parameters which can be useful for long term prediction. The prediction accuracy is also dependent on the efficiency of the predictor algorithm. In our experiment using simple local linear predictor also shows improvement of prediction accuracy with the estimated embedding parameters with the proposed technique. With the promising results of our proposed scheme of estimation of refined estimation parameters for global non-uniform embedding currently we are investigating to address the following problems:

- effect of uniform and non-uniform embedding on long term prediction accuracy;
- application of non-uniform embedding vector to noise reduction;
- and finding a better algorithm for optimal regularizer parameter for optimal model selection.

References

[1] H.D.I. Abarbanel, Analysis of Observed Chaotic Data, Springer, New York, 1996.
[2] A.M. Albano, J. Muench, C. Schwartz, A.I. Mees, P.E. Rapp, Singular-value decomposition and the Grassberger–Procaccia algorithm, Phys. Rev. A 38 (1988) 3017–3026.
[3] K. Alligood, T. Sauer, J.A. Yorke, Chaos: An Introduction to Dynamical Systems, Springer, New York, 1997.
[4] H. Akaike, A new look at the statistical model identification, IEEE Trans. Autom. Control 19 (1974) 716–723.
[5] C.J. Cellucci, A.M. Albano, P.E. Rapp, Comparative study of embedding methods, Phys. Rev. E 67 (2003) 066210-1-13.
[6] B. Chakraborty, Y. Manabe, Structural Learning of Neural Network for Continuous Valued Output: Effect of Penalty Term to Hidden Units, Lecture Notes in Computer Science, vol. 3316, Springer, Berlin, 2004, pp. 599–605.
[7] D.B. Fogel, An information criterion for optimal neural network selection, IEEE Trans. Neural Networks 2 (1991) 490–497.
[8] A.M. Fraser, H.L. Swinney, Independent coordinates for strange attractors from mutual information, Phys. Rev. A 33 (1986) 1134–1140.
[9] R. Hegger, H. Kantz, T. Schreiber, Practical implementation of nonlinear time series methods: the TISEAN package, CHAOS 9, 1999, pp. 413–435.
[10] M. Ishikawa, Structural learning with forgetting, Neural Networks 9 (3) (1996) 509–521.
[11] K. Judd, A. Mees, Embedding as a modelling problem, Physica D 120 (1998) 273–286.
[12] K. Judd, M. Small, A.I. Mees, Achieving good nonlinear models: keep it simple, vary the embedding, and get the dynamics right, in: A.I. Mees (Ed.), Nonlinear Dynamics and Statistics, Birkhauser, Boston, 2001, pp. 65–80.
[13] R. Kamimura, T. Takagi, S. Nakanishi, Improving generalization performance by information minimization, IEICE Trans. Inform. Syst. E78-D (2) (1995) 163–173.
[14] H. Kantz, T. Schreiber, Nonlinear Time Series Analysis, Cambridge University Press, Cambridge, 1997.
[15] M.B. Kennel, Determining embedding dimension for phase-space reconstruction using a geometrical construction, Phys. Rev. A 45 (1992) 3403–3411.
[16] S. Kikuchi, M. Nakanishi, Recurrent neural network with short-term memory and fast structural learning method, Syst. Comput. Japan 34 (6) (2003) 69–79.
[17] S. Kim, R. Eykholt, J.D. Salas, Delay time window and plateau onset of the correlation for small data sets, Phys. Rev. E 58 (1998) 5676–5682.

[18] R. Lapedes, R. Farber, Nonlinear signal processing using neural networks: prediction and system modelling, Los Alamos Nat. lab., Los Alamos, NM, Technical Report, LA-UR87-2662, 1987.
[19] M. Lehtokangas, et al., Predictive minimum description length criterion for the time series modelling with neural networks, Neural Comput. 8 (1996) 583–593.
[20] H. Leung, T. Lo, S. Wang, Prediction of noisy chaotic time series using an optimal radial basis function neural network, IEEE Trans. Neural Networks 12 (5) (2001) 1163–1172.
[21] Z. Liang, R.J. Jasaczak, R.E. Coleman, Parameter estimation of finite mixtures using the EM algorithm and information criteria with application to medical image processing, IEEE Trans. Nucl. Sci. 39 (4) (1992) 1126–1133.
[22] D.J.C. MacKay, Bayesian methods for adaptive models, Ph.D. Dissertation, Calif. Inst. Tech., Pasadena, 1991.
[23] D.J.C. MacKay, Bayesian interpolation, Neural Comput. 4 (1992) 415–447.
[24] D.J.C. MacKay, Bayesian non-linear modeling for the energy prediction competition, ASHRAE Trans. 100 (Part 2) (1994) 1053–1062.
[25] Y. Manabe, B. Chakraborty, Estimating embedding parameters using structural learning of neural network, IEEE International Workshop on NSIP 2005, Sapporo, Japan, May, 2005.
[26] Y. Manabe, B. Chakraborty, H. Fujita, Structural learning of multilayer feed forward neural networks for continuous valued functions, in: Proceedings of IEEE-MWSCAS 2004, July, 2004, pp. III77–III80.
[27] T. Matsui, T. Iizaka, Y. Fukuyama, Peak load forecasting using analyzable structured neural network, IEEE Power Eng. Soc. Winter Meeting (2001) 405–410.
[28] T. Matsumoto, Y. Nakajima, M. Saito, J. Sugi, H. Hamagishi, Reconstructions and predictions of nonlinear dynamical systems: a hierarchical Bayesian approach, IEEE Trans. Signal Process. 49 (9) (2001) 2138–2155.
[29] R.M. Neal, Bayesian Learning for Neural Networks, Springer, New York, 1996.
[30] J.C. Principe, J.M. Kuo, Dynamic modelling of chaotic time series with neural networks, in: Proceedings of Neural Information Processing Systems NIPS, vol. 7, 1995, pp. 311–318.
[31] J.C. Principe, L. Wang, M.A. Motter, Local dynamic modelling with self-organizing maps and applications to nonlinear system identification and control, Proc. IEEE 86 (11) (1998).
[32] R. Reed, Pruning algorithms—a survey, IEEE Trans. Neural Networks 4 (5) (1993) 740–747.
[33] J. Sietsma, R.J.F. Dow, Creating artificial neural networks that generalize, Neural Networks 4 (1991) 67–79.
[34] M. Small, Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance, Nonlinear Science Series A, vol. 52, World Scientific, Singapore, 2005.
[35] M. Small, C.K. Tse, Minimum description length neural networks for time series prediction, Phys. Rev. E (2002) 66:066701.
[36] M. Small, C.K. Tse, Optimal embedding parameters: a modelling paradigm, Physica D 194 (2004) 283–296.
[37] F. Takens, Detecting strange attractors in turbulence, in: D.A. Rand, L.S. Young (Eds.), Dynamical Systems and Turbulence, Springer, New York, 1981, pp. 366–381.
[38] The Santa Fe Time Series Competition Data ⟨URL:http://www-psych.stanford.edu/~andreas/Time-Series/SantaFe.html⟩.
[39] Time Series Data Library ⟨URL:http://www-personal.buseco.monash.edu.au/~hyndman/TSDL/⟩.
[40] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer, Berlin, 1995.
[41] Y. Zhao, M. Small, Minimum description length criterion for modelling of chaotic attractors with multi-layer perceptron networks, IEEE Trans. Circuits Syst. I 52 (2005).



Yusuke Manabe received his Master's degree in Software and Information Science from Graduate School of Iwate Prefectural University, Japan. Currently he is pursuing his doctoral studies in the same school. His main research interests are in Dynamic Biometrics, Gesture Recognition and Action description of CG Character.

Basabi Chakraborty received M. Tech and Ph.D degrees in RadioPhysics and Electronics from Calcutta University, India. She received another Ph.D. in Information Science from Tohoku University, Japan. Currently she is a faculty in the Department of Software and Information Science, Iwate Prefectural University, Japan. Her main research interests are in Pattern Recognition, Image Processing, Biometrics and Soft Computing Techniques.