ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 327-332

Neural Network Piecewise Linear Preprocessing for Time-series Prediction

T.W.S. Chow and C.T. Leung

Department of Electronic Engineering
City University of Hong Kong
83 Tat Chee Avenue
Kowloon, Hong Kong

Abstract. This paper presents a piecewise linear preprocessing method to enhance performance of feedforward neural network. The introduction of the piecewise preprocessing method reduces the correlation coefficient of input variables and results in facilitating the network optimization and generalization. This preprocessing technique enables the correlation coefficient and test set error to be enormously reduced. The applications of time-series prediction applied to evaluate the performance of the proposed method is also presented.

1. Introduction

Artificial Neural Network (ANN) becomes a versatile tool for many physical sciences and engineering applications such as pattern recognition, robotics, nonlinear system identification and control, and time-series modeling. ANN can be viewed as circuit of highly interconnected nonlinear units with modifiable interconnection weights. It is inherently an abstract simulation of a real nervous system that contains a collection of neuron units communicating with each other via axon connections. This wide range of applications is completely due to its functional approximation property. A function defined on a compact set in $C[a,b]$ or $L^p[a,b]$, can be approximated arbitrarily well by an ANN with one hidden layer [1]. In practice, because of finite number of hidden neurons, ANN alone cannot arbitrarily approximate the target function. Data processing prior to ANN is completely indispensable. The prominent features or components can be extracted via data preprocessing that can facilitate network optimization and generalization. The preprocessing, such as cepstrum analysis for speech recognition, is commonly based on prior information. However, the prior knowledge of the empirical data from a new research area is totally inadequate. Feature extraction preprocessing method is hardly established at this stage. Multivariate Statistical Analysis is a popular choice. For examples, the factor analysis and the principal component analysis are two popular tools of statistical analysis. However, those analysis tools are used to find the linear relationship of the system. For highly nonlinear system, such analysis tools are sometimes not quite applicable. A piecewise linear preprocessing method is proposed for such situation. This preprocessing method reducing mutual information redundancy among input variables is proposed. Section 2 of this paper describes the proposed preprocessing method. Results and Discussions of the proposed method applying to two different time-series predictions are given in Section 3. The final section concludes the significance of the proposed preprocessing method.

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 327-332

2. Piecewise Linear Preprocessing Method

In the introduction, the importance of preprocessing was described. For the situation, the system characteristics has been thoroughly examined and preprocessing technique can be easily derived from its characteristics. However, in many cases, the target system has not been fully studied. Statistical information is the one of the easiest obtained and valuable information. Correlation coefficient is one of the most important statistical measurement. It measures the amount of linear association between two random variables [2] and it is also an index for measuring the extent of redundancy of the information contents [3]. Xue *et al.* concluded that redundancy of dimensionality formed by hidden neurons deters to get both optimal approximation and computation of learning iteration. Likewise, the extent of redundancy in the input variables is another dominant factor to network performance. We propose a preprocessing method that can reduce the mutual correlation coefficient of input variables. Assume that the input variables are all Gaussian distributed. A piecewise linear transformation to the input variables can diminish the magnitude of mutual correlation coefficient based on the following Oh's theorem:

If two jointly Gaussian random variables u and v with means u_o and v_o are transformed into y and z as follows:

$$y = T_1(u) = \begin{cases} au + y_o, & u < u_o \\ bu + (a-b)u_o + y_o, & u \geq u_o \end{cases}$$

$$z = T_2(v) = \begin{cases} cv + z_o, & v < v_o \\ ev + (c-e)v_o + z_o, & v \geq v_o \end{cases}$$

where a, b, c, and e are non-zero real constants, then

$$|\mathrm{Cor}[T_1(u), T_2(v)]| = |\mathrm{Cor}[u,v]| \qquad \text{if } a = b,\ c = e \text{ or } \mathrm{Cor}[u,v] = 0,$$

$$|\mathrm{Cor}[T_1(u), T_2(v)]| < |\mathrm{Cor}[u,v]| \qquad \text{otherwise,}$$

where Cor[x,y] denotes the correlation coefficient between x and y, and it is defined by

$$\mathrm{Cor}[x, y] = \frac{E[xy] - E[x]E[y]}{\sigma_x \sigma_y}.$$

The formal proof of the above theorem can be referred to [4]. From the above theorem, $|\mathrm{Cor}[T_1(u),T_2(v)]|$ is maximum, i.e. $|\mathrm{Cor}[u,v]|$, when T_1 and T_2 are linear transformation. In other words, the introduction of a piecewise linear function in the input nodes can decrease the absolute values of the pairwise correlation coefficient of input variables. This result indicates that piecewise linear transformations reduce the information redundancy amongst input variables. The transformed input variables will be more "orthogonal" so that the network optimization will be facilitated. The feedforward neural network integrating piecewise linear preprocessing is as shown in Fig. 1.

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 327-332

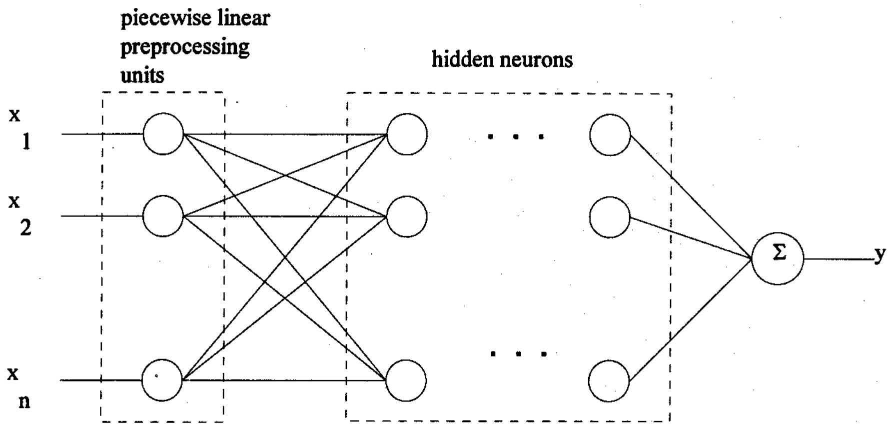

Fig. 1. The integration of piecewise linear preprocessing units prior to feedforward ANN.

3. Simulation Results and Discussions

To verify our idea, two applications of time-series prediction are considered. In the following verification, Extended backpropagation algorithm [5] was selected instead of error backpropagation algorithm [6] due to impractical long convergence time. The extended backpropagation is an adaptive gradient descent via the adaptation of learning rate and momentum factor. The learning algorithm was written in C language and the training phase was performed on a 486DX2-66 PC compatible. The sigmoid function of the hidden neurons is defined by

$$(x) = \frac{1-e^{-x}}{1+e^{-x}}.$$

To compare the performance of the feedforward network with different piecewise linear transformation, the network was applied to Henon Attractor Prediction, Mackey-Glass time-series Prediction. The equations describing the Henon attractor and the Mackey-Glass time-series [6] are shown below:

Henon Attractor: $x_{n+1} = 1 - 1.4x_n^2 + 0.3x_{n-1}$,

Mackey-Glass time-series: $\dot{x}(t) = \frac{0.2x(t-17)}{1+x(t-17)^{10}} - 0.1x(t)$.

a. *Henon Attractor Prediction*: the values of x_n and x_{n-1} is the input variables of the networks. The output variable is x_{n+1}. 250 training patterns and another 240 different test patterns were used in this study. Figure 2 illustrates the snapshot of the testing time-series.

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 327-332

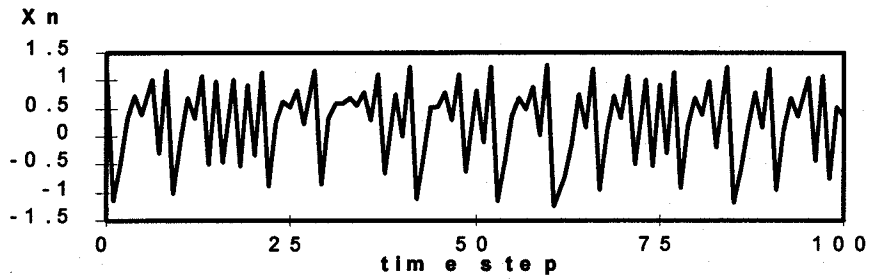

Fig. 2. The waveform generated by the Henon attractor.

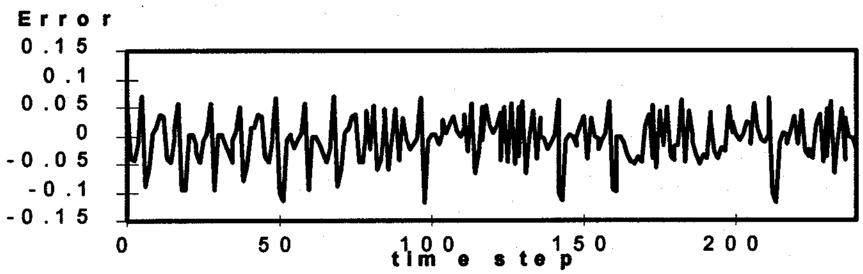

Fig. 3a. The error for each point in the testing time-series of the network without preprocessing.

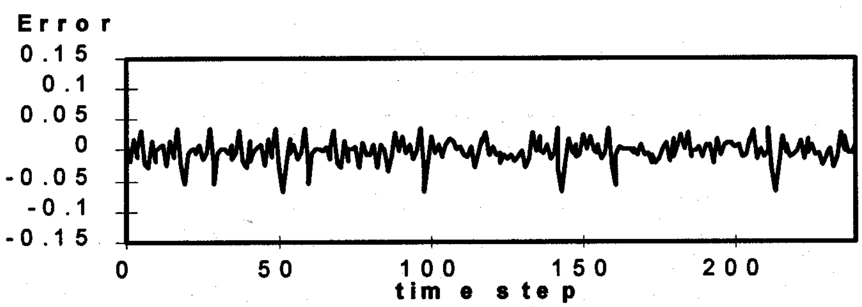

Fig. 3b. The error for each point in the testing time-series of the network with arctanh(x) preprocessing.

b. *Mackey-Glass Time-series Prediction*: the input variables now consists of 17 past values of $x(t)$,i.e., $(x(t),x(t\text{-}1),...,x(t\text{-}16))$. The output variable is $x(t+1)$. In this study, 1000 training patterns and another 900 different test patterns were used.

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 327-332

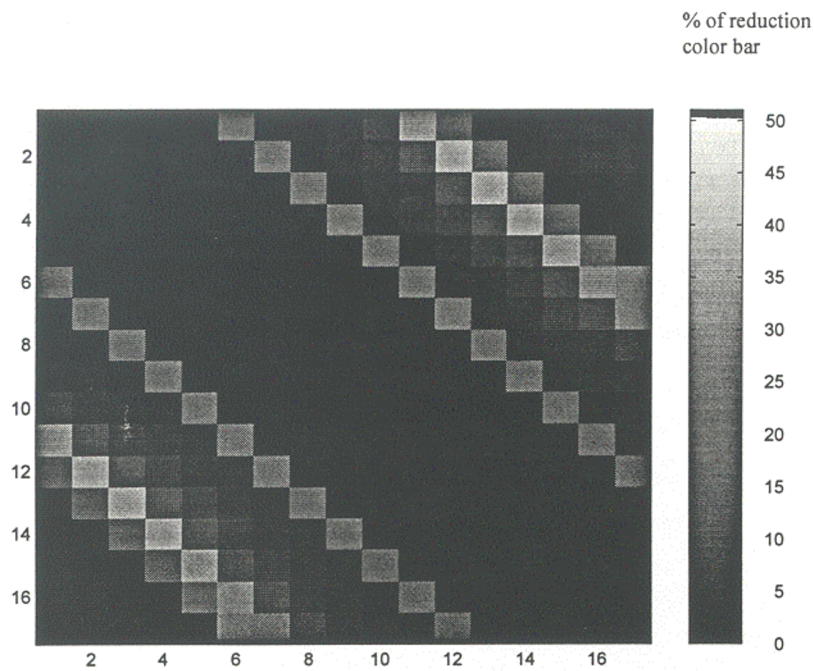

Fig. 4. The gray map summarizes the percentage reduction in correlation coefficient matrix of input variables in Mackey-Glass time-series prediction. The numbers 2, 4, 6 ... 16 on the sides of the above gray map stand for the input variables of the network $x_2, x_4, x_6, \ldots x_{16}$.

Two piecewise linear preprocessing transformations, tanh(x) and arctanh(x), were thoroughly studied for the above two applications. The number of training iterations for both time-series prediction is 30,000. Extensive tests were carried out and very encouraging results were obtained as tabulated in Tab. 1. With the introduction of the proposed preprocessing, the training error are found to be generally smaller. Comparing to the network without preprocessing, the RMSs of the test patterns for Henon and Mackey-Glass time-series prediction are decreased by a magnitudes of 0.0337 and 0.0079 respectively, and the test error and training error of the network with preprocessing is much more closer. We can notice that piecewise linear preprocessing makes feedforward network more generic. Figures 3a and 3b compare the test error of the network with piecewise preprocessing to that without preprocessing. It can be noticed that the networks using arctanh(x) preprocessing exhibit low prediction error. The error sparks are suppressed by our preprocessing technique. It is the another example that network with proposed preprocessing is more generic. Figure 4 manifests the percentage reduction in correlation coefficient matrix of input variables in Mackey-Glass time-series prediction. The arctanh(x) preprocessing enables at most 40% of reduction in correlation coefficient that corroborate the theorem in Section 2. It really make the input variables more “orthogonal”.

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 327-332

	pre-processing function	RMS error (training phase)	RMS error (test phase)	no. of training patterns	no. of test patterns	network structure
Henon	nil	0.0359	0.0431	250	240	2-25-1
	tanh(x)	0.0240	0.0284			
	arctanh(x)	0.0081	0.0094			
Mackey-Glass	nil	0.0089	0.0111	1000	900	17-36-1
	tanh(x)	0.0071	0.0086			
	arctanh(x)	0.0027	0.0032			

Tab. 1. The simulation results for the applications of Henon Attractor Prediction and Mackey-Glass time-series Prediction.

4. Conclusion

The introduction of piecewise linear preprocessing technique has a significant effect in facilitating the network optimization and generalization because the preprocessing technique is capable of reducing redundancy in information contents of input variables. Consequently, the network is apt at providing an excellent function approximation. This technique has been developed and thoroughly examined by applying to Mackey-Glass time-series and Henon time-series. Remarkable success has been achieved in reducing training and testing error.

References

1. T. Chen and H. Chen: Approximations of Continuous Functionals by Neural Networks with Application to Dynamic Systems. IEEE Trans. on Neural Networks, VOL. 4, NO. 6, pp. 910 - 918 (1993).
2. R. A. Johnson and D. W. Wichern: Applied Multivariate Statistical Analysis. Prentice-Hall, Englewood Cliffs, N.J. (1982).
3. Q. Xue, Y. Hu and W. J. Tompkins: Analyses of the hidden units of back-propagation model by singular value decomposition. Proc. IJCNN'90, Wash-DC, VOL. I, pp. I-739 - I-742 (Jan 1990).
4. S.-H. Oh and Y. Lee: Effect of Nonlinear Transformations on Correlation between Weighted Sums in Multilayer Perceptrons. IEEE Trans. on Neural Networks, VOL. 5, NO. 3, pp. 508 - 510 (1994).
5. Y. F. Yam and T. W. S. Chow: Extended Backpropagation Algorithm. IEE Electronics Lett., VOL. 29, pp. 1701 - 1702 (1993).
6. M. C. Mackey and L. Glass: Oscillation and Chaos in Physiological Control Systems. Science, VOL. 179, pp. 287 - 289 (1977).