Expert Systems with Applications 37 (2010) 3310–3317

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A hybrid of nonlinear autoregressive model with exogenous input and autoregressive moving average model for long-term machine state forecasting

Hong Thom Pham, Van Tung Tran, Bo-Suk Yang *

School of Mechanical Engineering, Pukyong National University, San 100, Yongdang-dong, Nam-gu, Busan 608-739, South Korea

ARTICLE INFO

Keywords:
Autoregressive moving average (ARMA)
Nonlinear autoregressive with exogenous input (NARX)
Long-term prediction
Machine state forecasting

ABSTRACT

This paper presents an improvement of hybrid of nonlinear autoregressive with exogenous input (NARX) model and autoregressive moving average (ARMA) model for long-term machine state forecasting based on vibration data. In this study, vibration data is considered as a combination of two components which are deterministic data and error. The deterministic component may describe the degradation index of machine, whilst the error component can depict the appearance of uncertain parts. An improved hybrid forecasting model, namely NARX–ARMA model, is carried out to obtain the forecasting results in which NARX network model which is suitable for nonlinear issue is used to forecast the deterministic component and ARMA model are used to predict the error component due to appropriate capability in linear prediction. The final forecasting results are the sum of the results obtained from these single models. The performance of the NARX–ARMA model is then evaluated by using the data of low methane compressor acquired from condition monitoring routine. In order to corroborate the advances of the proposed method, a comparative study of the forecasting results obtained from NARX–ARMA model and traditional models is also carried out. The comparative results show that NARX–ARMA model is outstanding and could be used as a potential tool to machine state forecasting.

1. Introduction

Machine state forecasting gradually plays an important role in modern industry due to its ability to foretell the states of machine in the future. This provides the necessary information for system operators to implement the essential actions in order to avoid the catastrophic failures, which lead to a costly maintenance or even human casualties. Moreover, foretelling the states of machine enables maintenance action to be scheduled more effectively, avoids unplanned breakdown, assists maintainers in estimating the remaining useful life, provides alarms before a fault reaches the critical levels to prevent machinery performance degradation and malfunction (Liu, Wang, & Golnaraghi, 2009), etc. Consequently, machine state forecasting has been considerably attracted the attention of researchers in the recent time.

In order to predict the future states of machine, the forecasting model uses the available observations that are generated from measured data by using appropriate signal processing techniques. The measured data could be vibration, acoustic, oil analysis, temperature, pressure and moisture, etc. Among of them, vibration data is commonly used because of the easy-to-measure signals and analysis. Several forecasting models have been successfully proposed in literature in which model-based techniques and data-driven based techniques were commonly utilized. Model-based techniques are applicable to where the accurate mathematical models can be constructed based on the physical fundamentals of a system, whilst data-driven based techniques utilize and require large amount of historical failure data to build a forecasting models that learn the system behavior. Obviously, data-driven based techniques are inaccurate in comparison with model-based techniques in prediction capability. However, data-driven based techniques, which are frequently based on artificial intelligence, can flexibly generate the forecasting models regardless of the complexity of system. Therefore, these techniques that some of those have been proposed in Liu et al. (2009), Tran, Yang, Oh, and Tan (2008), Vachtsevanos and Wang (2001), Wang (2007), Wang, Golnaraghi, and Ismail (2004) are the first selection of researchers' investigations.

An alternative approach to ameliorate the predicting capability in time series forecasting is the combination of model-based and data-driven based techniques. According to Zhang (2003), the reasons for hybridizing these models are: (i) in practice, it is difficult to determine whether a time series under study is generated from a linear or nonlinear underlying process or whether one particular method in more effective than the other in out-of-sample forecasting; (ii) data obtained from real-work is purely linear or nonlinear that neither model-based techniques nor data-driven based

* Corresponding author. Tel.: +82 51 629 6152; fax: +82 51 629 6150.
E-mail address: bsyang@pknu.ac.kr (B.-S. Yang).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.10.020

techniques can be adequate in modeling and forecasting. Model-based techniques can adequately capture the linear component of time series while data-driven based techniques are highly flexible in modeling the nonlinear components. Accordingly, numerous hybrid models have been depicted to provide the investors with more precise prediction. For instance, Zhang (2003) combined autoregressive integrated moving average (ARIMA) model and neural network model to forecast three well-known time series sets that were sunspot data, Canadian lynx data and the British pound/US dollar exchange rate data. Inde and Trafalis (2006) proposed a hybrid model including parametric techniques (e.g. ARIMA, vector autoregressive) and nonparametric techniques (e.g. support vector regression, artificial neuron networks) for forecasting the exchange market. A hybrid of ARIMA and support vector machines was successfully presented by Pai and Lin (2005) for predicting stock prices problems. Other outstanding hybrid approaches could be found in Rojas et al. (2008), Tseng, Yu, and Tzeng (2002), Valenzuela et al. (2008). Most of these hybrid models were implemented as a following process: first, the model-based technique was used to predict the linear relation, then the data-driven based technique was utilized to forecast the residuals between actual values and predicted results obtained from previous step. The final results were the sum of results gained each model. Furthermore, these hybrid approaches merely regarded as short-term prediction methodology.

In this study, an improved hybrid forecasting model is proposed for long-term prediction the operating states of machine. The prediction strategy used here is recursive which is one of the strategies mentioned in Sorjamaa, Hao, Reyhani, Ji, and Lendasse (2007). This forecasting model involving nonlinear autoregressive with exogenous input (NARX) (Leontaritis & Billings, 1985) and autoregressive moving average (ARMA) (Box & Jenkins, 1970) is novel in the following aspects: (1) vibration data indicating the state of machine is divided into deterministic component and error component that is the residual between the actual data and deterministic component. NARX and ARMA are simultaneously employed to forecast the former and the latter, respectively. The final forecasting results are the sum of results obtained from single model; (2) long-term forecasting, which is still a difficult and challenging task in time series prediction domain, is applied.

Additionally, the number of observations used as the input for forecasting model, so-called embedding dimension, is the problem often encountered in time series forecasting techniques. Embedding dimension could be estimated by using either Cao's method (1997) or false nearest neighbor method (FNN) (Kennel, Brown, & Abarbanel, 1992). However, FNN method depends on the chosen parameters wherein different values lead to different results. Furthermore, FNN method also depends on the number of available observations and is sensitive to additional noise. Cao's method overcomes the shortcomings of the FNN approach and therefore, it is chosen in this study.

2. Background knowledge

2.1. Nonlinear autoregressive model with exogenous inputs (NARX)

The NARX model is an important class of discrete-time nonlinear systems that can be mathematically represented as follows:

$$y(t+1) = f[y(t), y(t-1), \ldots, y(t-n_y+1); u(t), u(t-1), \ldots, u(t-n_u+1); \mathbf{W}] = f[\mathbf{y}(t); \mathbf{u}(t); \mathbf{W}] \quad (1)$$

where $u(t) \in$ and $y(t) \in$, respectively represent the input and output of the model at time t, $n_u \geqslant 1$ and $n_y \geqslant 1$ ($n_y \geqslant n_u$) are the input-memory and output-memory orders, $\mathbf{W}$ is a weights matrix, f is the nonlinear function which should be approximated by using multilayer perceptron. The structure of an NARX network is depicted in Fig. 1.

Basically, NARX network is trained under one out of two models:

Parallel (P) mode: the output is fed back to the input of the feedforward neural network as part of the standard NARX architecture:

$$\hat{y}(t+1) = \hat{f}[\mathbf{y}_p(t); \mathbf{u}(t); \mathbf{W}] = \hat{f}[\hat{y}(t), \hat{y}(t-1), \ldots, \hat{y}(t-n_y+1); u(t), u(t-1), \ldots, u(t-n_u+1); \mathbf{W}] \quad (2)$$

Series-parallel (SP) mode: the output's regressor is formed only by actual values of the system's output:

$$\hat{y}(t+1) = \hat{f}[\mathbf{y}_{sp}(t); \mathbf{u}(t); \mathbf{W}] = \hat{f}[y(t), y(t-1), \ldots, y(t-n_y+1); u(t), u(t-1), \ldots, u(t-n_u+1); \mathbf{W}] \quad (3)$$

As mentioned above, NARX network inputs include the regressors of inputs and outputs of system while a time series is one or more measured output channels with no measured input. Hence, the forecasting abilities of the NARX network may be limited when applying for time series data without regressor of inputs. In this kind of application, the tapped-delay line over the input signal is eliminated, thus the NARX is reduced to the plain focused time-delay neural network architecture (Lin, Horne, Tino, & Giles, 1997):

$$\hat{y}(t+1) = f[y(t), y(t-1), \ldots, y(t-n_y+1); \mathbf{W}] \quad (4)$$

According to Menezes and Barreto (2006), a simple strategy based on Takens' embedding theorem was proposed for solving this problem. This strategy allows the computational abilities of the original NARX network to be fully exploited in nonlinear time series prediction tasks and is described as following processes:

Firstly, the input signal regressor, denoted by $u(t)$, is defined by the delay embedding coordinates:

$$u(t) = [y(t), y(t-\tau), \ldots, y(t-(d_E-1)\tau)] \quad (5)$$

where $d_E = n_u$ is embedding dimension and τ is embedding delay.

Secondly, since the NARX network can be trained in two different modes, the output signal regressor $y(t)$ can be written as follows:

$$y_{sp}(t) = [y(t), y(t-1), \ldots, y(t-n_y+1)] \quad (6)$$

$$y_p(t) = [\hat{y}(t), \hat{y}(t-1), \ldots, \hat{y}(t-n_y+1)] \quad (7)$$

where the output regressor $y(t)$ for the SP mode in Eq. (6) contains n_y past values of the actual time series, while the output regressor

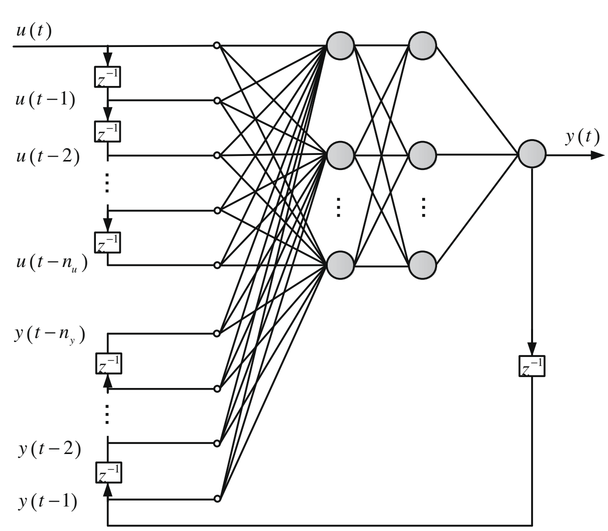

Fig. 1. The structure of NARX with n_u inputs and n_y output delays.