The Sixteenth International Symposium on Artificial Life and Robotics 2011 (AROB 16th '11),
B-Con Plaza, Beppu,Oita, Japan, January 27-29, 2011

An Efficient Identification Scheme for Nonlinear Polynomial NARX Model

Yu CHENG, Miao YU, Lan WANG and Jinglu HU

Graduate School of Information, Production and Systems, Waseda University
Hibikino 2-7, Wakamatsu, Kitakyushu, Fukuoka, 808-0135, Japan
(Tel/Fax : 81-93-692-5271)
(chengyu0930@fuji.waseda.jp)

Abstract: Nonlinear polynomial NARX model identification often faces the problem of huge pool of candidate terms, which makes the evolutionary optimization based identification algorithm work with low efficiency. This paper proposes an efficient identification scheme with pre-processing to reduce the searching space effectively. Both the input selection and term selection are implemented to truncate the candidate pool with the help of correlation based orthogonal forward selection (COFS) algorithm and simplified orthogonal least square (OLS) algorithm, respectively. Then multi-objective evolutionary algorithm (MOEA) is used to identify the polynomial model in a relative small searching space.
Keywords: nonlinear polynomial model identification, input selection, term selection, efficient

I. INTRODUCTION

Recently, nonlinear polynomial NARX (Nonlinear AutoRegressive with eXogenous inputs) model has attracted much attention because it has shown great potential in the ability of approximating nonlinear input-output relationship. As a kind of effective approach, evolutionary optimization algorithms, such as genetic algorithm (GA) and multi-objective evolutionary algorithm (MOEA) have been commonly used for identification of nonlinear polynomial NARX model. However, it is still considered as a difficult task because the size of candidate terms increase drastically with maximum time delay of input-output data and nonlinearity of polynomial model [1].

So far, some related research has been devoted to complex nonlinear system identification. Ref. [2] claims that a hierarchical encoding technique is introduced to be effective for identifying polynomial models with relatively high-nonlinearity, however, it is just an improvement of GA, and as told by the authors, the process is still time-consuming and easily traps into a local optimum. Moreover, a two-step scheme for polynomial NARX model identification has been proposed in our previous research [3]. It combines heuristic optimization approach with pre-screening process, in which the simplified orthogonal least square (OLS) based term selection method is used to formulate a relative small searching space. However, the selection of input variables is not considered, which has significant influence on the pruning of searching space.

In this paper, an efficient identification scheme is proposed for nonlinear polynomial NARX model with both input selection and term selection methods [4], which can be seem as pre-processing for evolutionary optimization based searching processing. Firstly, correlation based orthogonal forward search (COFS) algorithm is applied, which makes the orthogonal input variable with maximum correlation coefficient of output select one by one. The final library consisting of all the necessary variables could be determined according to the threshold by the user. Although it is considered somehow not very accurate, it could exclude most of the redundant inputs thus reduce the original candidate pool effectively. Then, term selection will be implemented by using correlation analysis and the simplified OLS algorithm, hence the searching space could be limited within small size. At last, MOEA is used to identify the polynomial model in the reduced space. Simulations are intent to show the effectiveness of the proposed method.

This paper is organized as follows: Section 2 briefly describes the problem to be solved. Section 3 discusses the identification scheme in detail. Section 4 provides numerical simulations to demonstrate effectiveness of the new method, and Section 5 presents the conclusions.

II. STATEMENT OF THE PROBLEM

Consider a single-input-single-output (SISO) nonlinear time-invariant system whose input-output dynamics is described as

$$y(t) = f(x(t)) + e(t) \tag{1}$$
$$x(t) = [x_1(t), x_2(t)\cdots, x_n(t)]^T$$

where $x(t)$ is an n-dimensional input vector, **y** and **e** denote an output and white noise vector respectively, and $f(\bullet)$ is a nonlinear function.

In the case of nonlinear NARX model, $x(t) = [y(t-1), y(t-2), ..., y(t-n_y), u(t-1), u(t-2), ..., u(t-n_u)]$, where *y* and *u* are the system input and output. n_u and n_y are unknown maximum delays of input and output.

However, many of the input variables are often redundant and only a subset of them is significant. It is pointed that there are at least two problems induced [5]. First, the model complexity will increase drastically with the number of variables. Second, including irrelevant variables leads to the over-fitting problem, and as a consequence the model may tend to be oversensitive to training data and is likely to exhibit poor generalization properties.

III. IMPLEMENTATION OF EFFICIENT IDENTIFICATION SCHEME

It is found that evolutionary optimization based approaches are very efficient when the size of searching space is not too large. Based on this fact, input selection and term selection method are used as pre-processing for the efficient identification scheme. Then MOEA is applied to determine a set of significant terms to be included in the polynomial model with the help of independent validation data. Readers interested in the details of MOEA based identification may refer to our previous research [3]. In the following, input selection and term selection for pre-processing will be discussed.

1. COFS Based Input Selection

To select significant inputs with big contribution to the output vector from the whole input space, correlation coefficient could be used to evaluate the relationship between each input variable and the output vector and denoted as

$$C(x_i, y) = \left| \frac{\sum_{t=1}^{N} (x_i(t) - \overline{x_i})(y(t) - \overline{y})}{\sqrt{\sum_{t=1}^{N} (x_i(t) - \overline{x_i})^2 \sum_{t=1}^{N} (y(t) - \overline{y})^2}} \right|. \tag{2}$$

x_i and *y* represent input and output variable, *N* is the length of measurement. The bigger the correlation coefficient is, the more important the input variable is considered.

In order to exclude the influence from other input variables, orthogonal forward search algorithm is used and significant input variables will be selected one by one. At the first step, let

$$l_1 = arg\,\max_{1 \le i \le n} C(\mathbf{x}_i, \mathbf{y}) \tag{3}$$

where l_1 is the first important input variable selected from the whole library, and the associated orthogonal variable can be chosen as $\mathbf{q}_1 = \mathbf{x}_{l_1}$.

From the second step, every remained input variable is orthogonalized with all the selected inputs, and each correlation coefficient of output will be calculated. Assumed there are already *m*-1 input variables have been selected, the *m*-th significant input is selected from remaining pool, and orthogonalized with $\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_{m-1}$ as below

$$\mathbf{q}_j = \mathbf{x}_j - \frac{\mathbf{x}_j^T \mathbf{q}_1}{\mathbf{q}_1^T \mathbf{q}_1} \mathbf{q}_1 - \cdots - \frac{\mathbf{x}_j^T \mathbf{q}_{m-1}}{\mathbf{q}_{m-1}^T \mathbf{q}_{m-1}} \mathbf{q}_{m-1}. \tag{4}$$

Then the orthogonalized input variable which has maximum correlation with the output is chosen as following

$$l_j = arg\,\max_{i \in D} C(\mathbf{q}_i, \mathbf{y}) \tag{5}$$

where *D* is the subset contains all the remained input variables. As the index to reflect the importance of an input variable, ERR is used to represent the contribution of each input to the output vector. The input variables have very little contribution to the output will be ignored. Based on this fact, a threshold is defined which should make sure all the necessary inputs will be included.

2. Term Selection for Identification

Candidate terms formed by all the selected inputs are still too large if system nonlinearity is high. Therefore, term selection is needed to reduce the candidate pool. In this paper, two importance indices are introduced to evaluate the contribution of each term.

A. Importance Index 1

Importance Index 1 $\mathcal{I}_i^{(1)}$ is used to evaluate the correlation of each monomial term to the system output. Let ρ_i denotes the correlation coefficient of monomial term $y_i(t)$ and system output $y(t)$, calculated by

$$\rho_i = \frac{\sum_{t=1}^{N} (y_i(t) - \overline{y_i})(y(t) - \overline{y})}{\sqrt{\sum_{t=1}^{N} (y_i(t) - \overline{y_i})^2 \sum_{t=1}^{N} (y(t) - \overline{y})^2}} \tag{6}$$

Therefore, based on the principle of simplicity, Importance Index 1 should be given by

$$\mathcal{I}_i^{(1)} = \frac{|\rho_i|}{e^{O_i}} \tag{7}$$

where O_i is the order of the *i*-th term.

B. *Importance Index 2*

Instead of recursive manner of the original OLS method, a simplified OLS algorithm is introduced, in which all the terms are orthogonalized in one time, and the contribution to the output of each orthogonal one is calculated and denoted as ERR_i [3]. The principal of simplicity is also applied here, and Importance index 2 is given by

$$\mathcal{I}_i^{(2)} = \frac{ERR_i}{e^{O_i}} \qquad (8)$$

Although the two importance indices based term selection scheme is not very accurate, it could be used to prune candidate pool efficiently with all the necessary terms included. Therefore, evolutionary optimization for identification can work efficiently with a small searching space.

IV. NUMERICAL SIMULATIONS

To show efficiency of the proposed identification scheme, two experiments are simulated in this section, which are tested with the assumption that the time-delays are unknown thus big values are given.

1. Example Data Sets

In both two examples, 1000 input-output data sets are sampled for training from each model when the systems are excited using random input sequences with amplitude between -1.0 and +1.0.

2. Systems under Study

Example 1: The system is governed by a polynomial model, described by

$$\begin{aligned} y(t) &= -0.5y(t-2) + 0.7u(t-1)y(t-1) \\ &+ 0.6u^2(t-2) + 0.2y^3(t-1) \\ &- 0.7u^2(t-2)y(t-2) + e(t). \end{aligned}$$

Example 2: The system is a nonlinear rational model studied by Narendra in 1990

$$y(t) = f[y(t-1), y(t-2), y(t-3), u(t-1), u(t-2)] + e(t)$$

where

$$f[x_1, x_2, x_3, x_4, x_5] = \frac{x_1 x_2 x_3 x_5 (x_3 - 1) + x_4}{1 + x_2^2 + x_3^2}$$

Here, $\mathbf{e} \in (0, 0.1)$ is a white Gaussian noise.

3. Parameter Setting

It is assumed that the time delay for Example 1 and 2 are unknown thus initialized by 10 for both inputs and outputs (totally 20 input variables contained), which are considered big enough. However, when it is assumed the maximum order of each case is five, it is found the candidate terms pool is too large for evolutionary algorithms to search directly. Therefore, COFS based input selection and simplified OLS based term selection are implemented, then NSGA-II [6] is applied to extract all the possible polynomial terms. The details of NSGA-II for system identification are from our previous research [3].

4. Identification Results

In both cases, the thresholds for COFS algorithm are set as 0.9 and 0.98 after normalization in Example 1 and 2, respectively, then all the input variables selected satisfied with the threshold condition are listed in Tab. 1.

Table 1. Input selection for Example 1 and 2

No.	Inputs selected in Example 1	Inputs selected in Example 2
1	y(t-2)	u(t-1)
2	y(t-1)	y(t-1)
3	y(t-3)	u(t-2)
4	u(t-1)	y(t-3)
5	u(t-2)	y(t-2)

To make comparison with some other state-of-art input selection methods, subset from Delta Test method [7], linear OLS method [4], and FOS-MOD algorithm [8] are also generated. The minimal subsets include all the real input variables for each method are given in Tab. 2 and Tab. 3.

Table 2. Results comparison for Example 1

Method	Minimal input subsets	size
COFS	y(t-1),y(t-2),y(t-3),u(t-1),u(t-2)	5
Delta Test	y(t-1),y(t-2),y(t-3),y(t-4),u(t-1), u(t-2),u(t-3)	7
Linear OLS	y(t-1),y(t-2),y(t-3),y(t-4),y(t-5), y(t-7),u(t-1),u(t-2)	8
FOS-MOD	y(t-1),y(t-2),y(t-5),y(t-7),y(t-9), y(t-10),u(t-1),u(t-2),u(t-4),u(t-5),u(t-7),u(t-9),u(t-10)	13

Table 3. Results comparison for Example 2

Method	Minimal input subsets	size
COFS	y(t-1),y(t-2),y(t-3),u(t-1),u(t-2)	5
Delta Test	y(t-1),y(t-2),y(t-3),u(t-1),u(t-2)	5
Linear OLS	with all the inputs selected	20
FOS-MOD	with all the inputs selected	20

We can know from the table that COFS method could get the minimal subset to contain all the true input variables with the smallest size. However, linear OLS and FOS-MOD worked not well. In Example 2, with the further insight of the rank list, the inputs $y(t\text{-}1)$, $y(t\text{-}3)$, $u(t\text{-}1)$, $u(t\text{-}2)$ ranked top by Linear OLS method, however, the $y(t\text{-}2)$ is ranked at the end of the list, therefore, this input variable would be easy to lost in this input selection procedure. What's more, it is believable that the method FOS-MOD could not deal with this case appropriately because all the important inputs are scattered in the rank list.

Furthermore, the results of term selection are shown in Tab. 4 and Tab. 5.

Table 4. Term selection for Example 1

True model term	Initial rank	Selected rank
y(t-2)	2	1
u^2(t-2)	39	3
u(t-1)*y*(t-1)	13	5
u^2(t-2)*y*(t-2)	**103**	10
y^3(t-1)	45	**13**

Table 5. Term selection for Example 2

True model term	Initial rank	Selected rank
u(t-1)	5	1
y^2(t-3)*u*(t-1)	87	10
u(t-1)*u*(t-3)*y*(t-2)	81	15
u(t-3)y^2(t-1)*y*(t-3)	277	261
y(t-1)y^2(t-3)*u*(t-2)*u*(t-3)	**483**	**308**

It can be found that all the true model terms become more significant after the term selection process. The ranking value of each term is improved, and the minimal candidate pool is also reduced from 103 to only 13 to contain all the necessary terms in Example 1, and the similar situation could be found in Example 2, in which, the candidate pool is pruned from 483 to 308. In fact, 300 and 500 terms are selected as the searching space, which is considered big enough to include all the necessary terms.

In the phase of evolutionary optimization based system identification, NSGA-II is applied to identify the model structure. The identified polynomial model for Example 1 is:

$$\begin{aligned}\hat{y}(\hat{t}) &= -0.4989y(t-2)+0.6486y(t-1)u(t-1)\\ &+0.6922u^2(t-2)+0.1914\hat{y}^3(t-1)\\ &-0.6545\hat{y}(t-2)u(t-2)^2.\end{aligned}$$

From the final model it is found that although the irrelevant input variable $y(t$-3) is selected by COFS, it is eliminated and only true inputs are included in the final results.

In Example 2, the identified model could be expressed as:

$$\begin{aligned}y(t) &= 0.9151u(t-1)-0.2998y^2(t-3)u(t-1)\\ &-0.4218y(t-2)u(t-1)u(t-3)\\ &-0.3855y^2(t-1)y(t-3)u(t-3)\\ &+0.3563y(t-1)y^2(t-3)u(t-2)u(t-3).\end{aligned}$$

To test the obtained polynomial model, a 800 input-output data is sampled as test data, and the input data is described as

$$u(t)=\begin{cases}\sin(2\pi t/250) & \text{if } t\le 500\\ 0.8\sin(2\pi t/250)+0.2\sin(2\pi t/25) & \text{otherwise.}\end{cases}$$

It's found that the simulation result is as small as 0.1603.

V. CONCLUSIONS

In this paper, our contribution is to introduce an efficient scheme for nonlinear polynomial NARX model identification. In order to make evolutionary optimization based identification worked efficiently, input selection and term selection are implemented to choose the minimal subset with all the necessary terms included. As two examples shown, the proposed COFS algorithm outperforms other input selection methods, and the pre-processed identification could work efficiently with a relative small searching space.

REFERENCES

[1] L. Ljung, System Identification: Theory for the User, Second Edition, Prentice-Hall PTR, Upper Saddle River, N.J., 1999.
[2] Z.J. Yang, T. Fujimoto, and K. Kumamaru, A genetic algorithm approach to identification of nonlinear polynomial models, in Proc. of the 12th IFAC Symposium on Identification, Santa Barbara, 2000.
[3] Y. Cheng, L. Wang and J. Hu, "A Two-step Scheme for Polynomial NARX Model Identification Based on MOEA with Pre-screening Process", IEEJ Trans. on Electrical and Electonic Engineering, Vol.6, No.3, 2011.
[4] H. L. WEI, S. A. BILLINGS and J. LIU, "Term and Variable Selection or Non-linear System Identification ," INT. J. Control, vol. 77, No. 1, pp. 86–110, Jan. 2004.
[5] K. Z. Mao and S. A. Billings, "Variable Selection in Non-linear Systems Modelling," Mechanical Systems and Signal Processing, vol. 12, no. 2, pp. 351–366, 1999.
[6] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and T. Meyarivan, "A Fast and Elitist Multi-objective Genetic Algorithm: NSGA-II," IEEE Transaction on Evolutionary Computation, vol. 6, no. 2, pp. 182–197, 2002.
[7] Emil Eirola, Elia Liitiainen, Amaury Lendasse, Francesco Corona and Michel Verleysen, "Using the Delta Test for Variable Selection ," Proc. of European Symposium on Artificial Neural Networks, Bruges, Belgium, April, 2008, pp. 25–30.
[8] H. L. WEI, S. A. BILLINGS "Feature Subset Selection and Ranking for Data Dimensionality Reduction ," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 1, pp. 162–166, Jan. 2007.