

# Fault Prediction of Nonlinear System Using Time Series Novelty Estimation

Shengchao Su, Wei Zhang  
Industrial Engineering Training Centre  
Shanghai University of Engineering Science  
Shanghai 201620, China  
[jnssc@sues.edu.cn](mailto:jnssc@sues.edu.cn)



**ABSTRACT:** Many complex nonlinear systems have the characteristics such as difficult modeling, dangerous testing, and high cost to experiment with fault. Aim at these points, a new kind of novelty estimation method based on an on-line Least Square Support Vector Regression (LS-SVR) algorithm is presented for fault prediction of nonlinear system in this paper. Firstly, a set of characteristic token values, which represent normal state of time series at the beginning, are selected as the training data to establish the on-line LS-SVR model. Then, the output values of the on-line model are analyzed to estimate novel pattern. Once the novel pattern is affirmed, the fault of system is forecasted immediately. The proposed method needs neither the math model of system nor the fault training data and primary knowledge, and it is believed with good real-time capability. The results of simulation on Continuous Stirred Tank Reactor (CSTR) show the efficiency.

## Categories and Subject Descriptors

G.1.2 [Approximation]: Nonlinear Approximation; D.4.5 [Reliability]: Fault-tolerance

**General Terms:** Nonlinear System, Time series

**Keywords:** Novelty Estimation, Nonlinear System, Time Series, Fault Prediction

**Received:** 1 February 2013, Revised 26 March 2013, Accepted 2 April 2013

## 1. Introduction

Fault detection and diagnosis has been a new research

direction since 1970s, and becomes one of the hotspots in information and control theory presently. With the further requirements of system reliability and security, people expect that they can get the failure information before the fault damage the system. As a result, the fault prediction was put forward.

Fault prediction is the special area that combines fault detection with event forecast [1]. Now little progress has been made in fault prediction and there are only two main categories in this field. One of the common approaches is to use model-based methods [2-4] for prediction and detection. This requires modeling of the system, filtering the measure data, and estimation of the future state. The basic idea is detect the fault based on the prediction output of the model, and make a decision about the future state of system. As a premise, the system model must be known and accurate for these methods to be highly effective. Unfortunately, the nonlinearity and uncertainty widely exist in the system model. For some complex nonlinear system, the model even cannot be obtained. Those can easily degrade the estimation output and cause either missed detections or false alarms. As another kind of common approaches, qualitative approaches [5-9] such as neural networks, wavelet technique, and rough set that do not require system models, has elicited considerable research interests in the last 10 years. These methods can avoid the dependence on the model of system, so that they have robustness, adaptation, self-study and disposal of complex nonlinearity which enable the methods more widely applied than the model-based methods. However, it is worth to mention that these methods need fault training data or prior knowledge. For example, the neural networks need many data of all kinds of faults and

its performance is limited by the number and distributed situation of the chosen sample. For many complex engineering systems, whose acquisition of fault data generally is at the cost of equipment damage and loss of economy, it's unpractical to fully obtain the fault training data or prior knowledge. Even though the training data of the fault can be gained, the obtained training data will gradually be out of use because of time-variance and uncertainty of the actual system. So the knowledge-based method probably brings about bigger false-positive and false-negative error, which result in the degradation of prediction performance even without any use.

Time series data mining (TSDM) method [10-11] proposed in recent years combines time series analysis and data mining. The method predicts when events will happen through identifying temporal pattern in time series, which can predict the defined events directly by the past and current state without estimating the future state. So it is a new way for nonlinear time series prediction. Su et al. [12] applied one-class Support Vector Machine (SVM) to TSDM, and defined fault as the event in nonlinear time series to realize the fault prediction of CSTR only with data of normal condition. But the training of SVM must solve the problem of convex quadratic programming. Although the obtained solution is an optimal one, the more the sample data is, the slower computing speed will be and the bigger spaces it takes because the complexity of the algorithm mainly relies on the number of the sample data.

In this paper, Least Square Support Vector Regression (LS-SVR) is adopted to replace SVM, and an online LS-SVR algorithm is proposed. Based on this algorithm, we devise a real-time novelty estimation method without needs of any novel training data or the prior knowledge. The online estimation method is then extended to be a real-time fault prediction approach of nonlinear time series. As a result, the experiment of online fault prediction for CSTR is fulfilled under the circumstances of only the normal operating data of system. The remainder of this paper is organized as follows. Section 2 provides the problem formulation of fault prediction for nonlinear time series. Section 3 presents an online LS-SVR algorithm. Based on the online algorithm, an online novelty estimation method of nonlinear time series is constructed in section 4. Furthermore, the novelty estimation method is extended to be an online fault prediction approach of nonlinear system. In section 5, a simulation example is provided to demonstrate the detailed implementation procedures and the effectiveness of the proposed approach. Finally, conclusions are given in section 6.

## 2. Problem Formulation

Consider a fault prediction problem as follows.

For a non-linear system whose mathematics model is unknown, suppose that the system work in normal state at the beginning and the time series sample data  $x_t, t =$

1, 2, ...,  $N$  since the system starts to run is given. Now we're required to judge whether the fault will happen in the future. In terms of TSDM, the question above is namely the novel pattern identification problem of the time series as follows:

Let the time series data of system output  $x_t \in R^L, t = 1, 2, \dots, N$ , and  $X_t$  is a  $m \cdot L$  dimensions column vector obtained at time  $t$ , where  $X_t = \{x_{t-(m-1)}^T, \dots, x_{t-2}^T, x_{t-1}^T, x_t^T\}^T$ .  $X_t$  presents the state at the current time. It is assumed that system is normal until the time  $N$  and state variable  $X_t (t = m, m + 1, \dots, N)$  are known, we want to judge whether the state variable  $X_t (t = N + 1)$  belongs to the novel pattern or not.

## 3. Online LS-SVR Algorithm

For a given training data sets  $(x_i, y_i), i = 1, \dots, l, x_i \in R^n, y_i \in R$ , a non-linear mapping  $\phi(\bullet)$  is used to map the data sets from an input space to a high dimension feature space, which converts the non-linear question in the input space into linear one in the high dimension space. In this high dimension space the linear function is so constructed as follows:

$$y(x) = w^T \phi(x) + b \quad (1)$$

Due to the structure risk minimization fundamental, consider the function complexity and fitting error synthetically. The problem of regression can be represented as one of restriction optimization:

$$\begin{cases} \min_{w, b, e} J(w, e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^l e_i^2 \\ s.t. : y_i = w^T \phi(x_i) + b + e_i, i = 1, 2, \dots, l \end{cases} \quad (2)$$

Where,  $\gamma$  is regularization parameter while  $e_i$  is error variable.

Generally, it's extremely difficult to compute the equation (2) directly since  $w$  may be of infinite dimensions. Therefore, this optimization problem is converted into its antithesis spaces. Define Lagrange function:

$$L(w, b, e, a) = J(w, e) - \sum_{i=1}^l a_i \{w^T \phi(x_i) + b + e_i - y_i\} \quad (3)$$

$a_i$  is the Lagrange coefficient. Based on Karush-Kuhn-Tucker (KKT) optimization condition, a linear equation group below can be gained by expurgating  $w, e$ .

$$\begin{bmatrix} 0 & e^T \\ e & Q + I/\gamma \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (4)$$

Where,

$y = [y_1, \dots, y_l]^T, e = [1, \dots, 1]^T, a = [a_1, \dots, a_l]^T, Q_{ij} = \phi(x_j) \cdot \phi(x_j), i, j = 1, \dots, l$ . The Mercer condition is applied in variable  $Q$ , consequently  $Q_{ij} = \phi(x_j) \cdot \phi(x_j) = K(x_i, x_j), i, j = 1, \dots, l$ .

$K(x_i, x_j)$  is kernel function. Usually Gaussian radial basis

function (RBF)  $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2)$  is selected. Ultimately the LS-SVR model is obtained as follows:

$$y(x) = \sum_{i=1}^l a_i K(x, x_i) + b \quad (5)$$

For the online learning of LS-SVR, suppose the sample  $(x_i, y_i), i = k, \dots, k+l-1, x_i \in R^n, y_i \in R$  at the moment  $k+l$ . The learning sample sets can be represented as  $\{X_S(k), Y_S(k)\}$ , Where,  $X_S(k) = [x_k, x_{k+1}, \dots, x_{k+l-1}]^T, Y_S(k) = [y_k, y_{k+1}, \dots, y_{k+l-1}]^T, x_k \in R^n, y_k \in R$ . So the kernel function matrix  $Q$ , the to-be-computed Lagrange coefficient  $a$  and constant warp  $b$  are all functions of  $k$ . That is to say at the moment  $k$ , they can be denoted separately as  $Q_{ij}(k) = K(x_{i+k-1}, x_{j+k-1}), i, j = 1, \dots, l, a(k) = [a_k, a_{k+1}, \dots, a_{k+l-1}]^T, b(k) = b_k$ , so formulation (5) is transformed to be equation (6).

$$y(x) = \sum_{i=k}^{k+l-1} a_i K(x, x_i) + b(k) \quad (6)$$

Let  $U(k) = Q(k) + I / \gamma$ , where  $I$  is an unit matrix, then equation (4) is come into equation (7).

$$\begin{bmatrix} 0 & eI^T \\ eI & U(k) \end{bmatrix} \begin{bmatrix} b(k) \\ a(k) \end{bmatrix} = \begin{bmatrix} 0 \\ y(k) \end{bmatrix} \quad (7)$$

Let  $P(k) = U(k)^{-1}, h(k) = K(x_k, x_k) + 1 / \gamma, H(k) = [K(x_{k+1}, x_k) + 1 / \gamma, \dots, K(x_{k+l-1}, x_k)]^T$ , then

$$P(k) = [Q(k) + I / \gamma]^{-1} = \begin{bmatrix} h(k) & H(k)^T \\ H(k) & D(k) \end{bmatrix}^{-1} \quad (8)$$

$$\begin{bmatrix} 0 & 0 \\ 0 & D(k)^{-1} \end{bmatrix} + S_k(k) S_k(k)^T c_k(k)$$

Where,  $c_h(k) = 1 / (h(k) - H(k)^T D(k)^{-1} H(k))$ ,

$$D(k) = \begin{bmatrix} K(x_{k+1}, x_{k+1}) + 1 / \gamma & \dots & K(x_{k+l-1}, x_{k+1}) + 1 / \gamma \\ \vdots & \ddots & \vdots \\ K(x_{k+1}, x_{k+l-1}) + 1 / \gamma & \dots & K(x_{k+l-1}, x_{k+l-1}) + 1 / \gamma \end{bmatrix}$$

$$s_h(k) = [-1, H(k)^T D(k)^{-1}]^T$$

Put equation (8) into equation (7), then

$$a(k) = P(k) \left[ Y(k) - \frac{eIeI^T P(k) Y_S(k)}{eI^T P(k) eI} \right] P(k) - Y_S(k) eI b(k) \quad (9)$$

$$b(k) = \frac{eI^T P(k) Y_S(k)}{eI^T P(k) eI} \quad (10)$$

At the moment  $k+l+1$ , a new sample  $(x_{k+1}, y_{k+1})$  is added and the oldest one  $(x_k, y_k)$  is forgotten, and the kernel matrix

$$Q_{ij}(k+1) = K(x_{i+k}, x_{j+k}), i, j = 1, \dots, l$$

$$P(k+1) = U(k+1)^{-1} = [Q(k+1) + I / \gamma]^{-1}$$

To sum up, online learning LS-SVR is a optimization process along with time rolling, whose algorithm summarized as below.

Step 1: Initialization,  $k = 1$ ;

Step 2: Select new data, while throw away the oldest ones;

Step 3: Compute kernel matrix  $Q(k)$  and  $P(k)$ ;

Step 4: Compute  $b(k), a(k)$  and predict  $y(x_{k+l})$ ;

Step 5:  $k \leftarrow k+1$ , return to step 2.

## 4. Online Fault Prediction Using Time Series Novelty Estimation

### 4.1 LS-SVR Based Online Novelty Estimation

In this paper, we firstly set a group of characteristic token values as the normal regression ones in which system runs at the beginning and then establish an online LS-SVR model. Then the novel pattern is estimated by comparing the output value of LS-SVR.

Let the initial sample time sequence of the system be  $X(t) = \{x_t \in R, t = 1, 2, \dots, N\}$ . The novel patterns are considered as the state vectors which are represented by time series data in the past and current time. So the phase space reconstruction of the initial sample sequences  $X(t)$  is required. The primary time sequences are then embedded into state space  $Q, Q \subseteq R$ , as a result the  $m$  time sequences data at current time  $t$  correspond to state vector  $X_t = \{x_{t-(m-1)}, \dots, x_{t-1}, x_t\}^T$  in the reconstruction phase space. Accordingly, the primary time sequences are converted into the vector sets  $T_m(l), T_m(l) = \{X_t, t = m, m+1, \dots, N\}, l = N - m + 1$ . When the system is running, new sample data are gained continuously and vector sets are  $T_m(l)$  updated together.

Suppose a group of initial regression values, which represent the normal system characteristics,  $Y = \{y_i\} = \{\beta \xi_i\}, i = 1, \dots, l$ . Where,  $\{\xi_i\}$  are a group of random numbers whose average is 0 and variance is 1. And  $\beta$  is a parameter to regulate the range of normal regression value. Usually  $y_i$  approaches to 0. Regarding  $(X_t, y_t)$  as training sample  $(x_i, y_i), i = k, \dots, k+l-1, t = i+m-1, x_i \in R^{m \times l}, y_i \in R, k \in Z^+$ , the LS-SVR online estimation model can be established.

Analyzing the regression value of LS-SVR output, we consider  $X_{k+l}$  as temporal state currently while there are  $n$  continuous regression values whose norm  $\|y(x_{k+l})\| > \theta\beta$ , where  $\theta$  is a adjustment parameter. When  $\theta$  increases, the temporal state false-positive error of system will decrease, and its false-negative error will rise. On the contrary, when  $\theta$  decreases the former increases while the latter rises. The performance of the method can be adjusted through modulating parameter  $\theta$  according to actual system.

### 4.2 Online Fault Prediction

For a general system, there always exists a time delay  $\Delta T$  from discovering the fault temporal state to the time when the failure happens. If  $\Delta T$  is much bigger than the sample time of a group of samples, there will be enough time to predict the fault. Therefore, to discover the novel

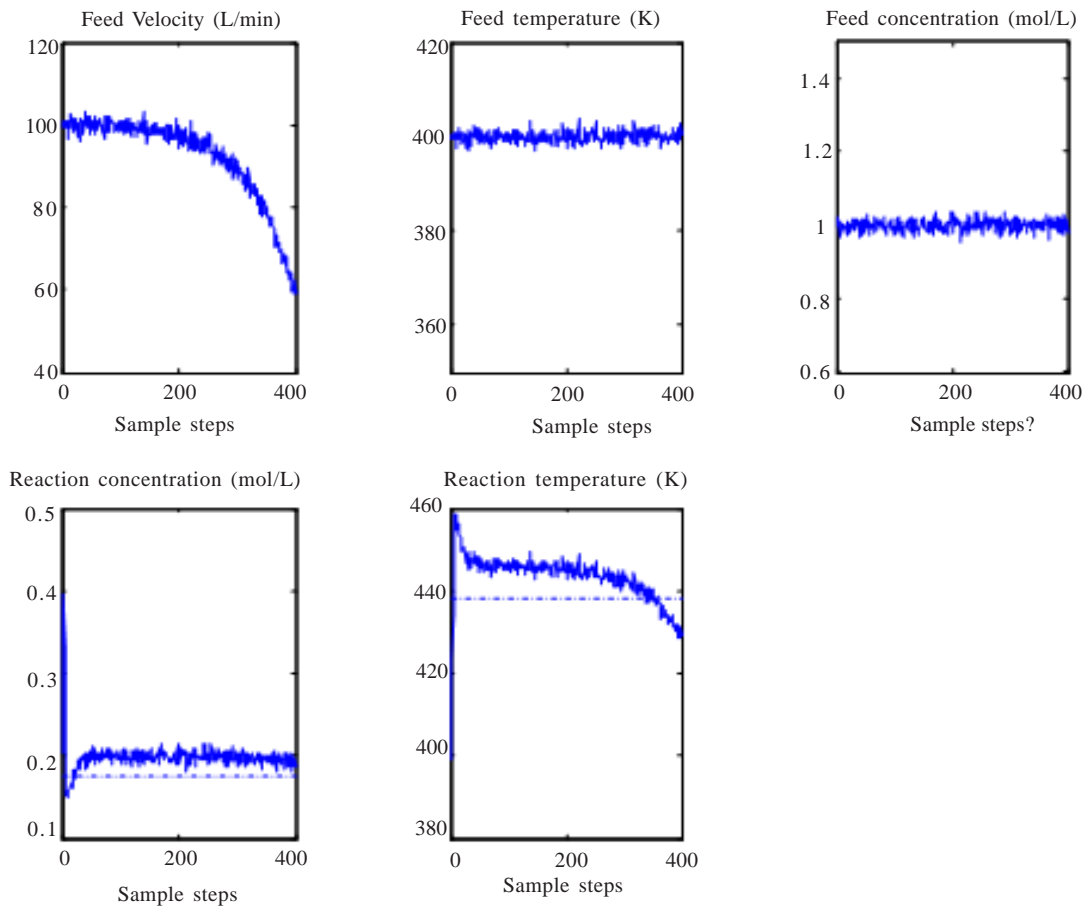


Figure 1. The operating curve when the CSTR feed velocity fault happens with noise

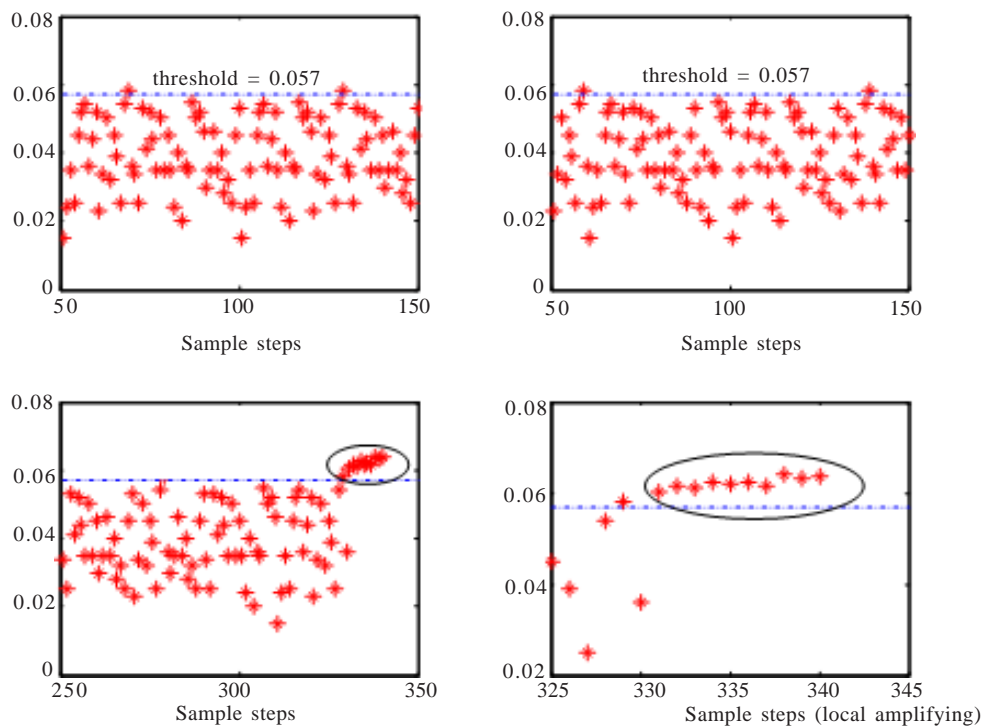


Figure 2. Online estimation results of  $T$

state in time is just the same as predicting the system failures.

### 5. Application Example

The Continuous Stirred Tank Reactor (CSTR) [3] is

considered as a research object in the paper. Suppose that the initial reaction concentration  $C_A$  of the reactor tank is  $0.2 \text{ mol/L}$ , the reaction temperature  $T$  is  $400\text{K}$ , the flow velocity of feed  $q$  is  $100\text{L/min}$ . It's known that the CSTR work in normal state at the beginning, and the



sample cycle of system is 0.2 min. In order to guarantee the normal operation of the reaction process, we monitor the reaction concentration  $C_A$  and the reaction temperature online. When they fluctuate on the range of threshold, the system can be considered normal, otherwise they exceed the system is believed to be broken-down. Let the concentration threshold be  $0.88 * C_A$ ,  $C_A = 0.2$  and the temporal threshold be  $0.98 * T_d$ ,  $T_d = 446K$ . From the step  $S = 1000$  set a system fault, namely the flow velocity of feed  $q$  declines following the exponent curve  $\dot{q}(s) = q(s) + 1 - \exp\left(\frac{s-1000}{80}\right)$ . The noise mentioned in literature [3] is then added to the system model, so the CSTR operating curve can be gained which is depicted in Figure 1.

In the simulation experiment, we select the reaction temperature as the evidence which earliest reports whether the system fault happens. Experiment parameters are set as follows: the initial sample number  $N = 55$ , the imbedded dimension number  $m = 8$ , the threshold upper limit of temporal state regression values  $\theta\beta = 0.057$ ,  $n = 10$ , parameter  $\gamma = 10$  and the kernel function  $K(x, x_i) = \exp\left(-\frac{\|x - x_i\|}{15}\right)$ .

Figure 2 shows the fault prediction results of our method, where  $y$ : the state regression value. In this chart, the ellipse part also represents the novel regression value and the fault state is judged based on them. There're already 10 continuous regression values which are bigger than the setting threshold when the sample step  $S = 341$ , so we can consequently report that system has faults at this time. The prediction time is 9 sample cycles ahead of fault happening time ( $S = 350$  [3]).

Considering the randomness of noise, we do the simulation 80 times to test the efficiency of proposed method. Of these simulations the primary 40 times are used to verify the false-positive error when the system is normal, while the last 40 times to test the advance time of prediction, the average deviation of time prediction and false-negative error when the system is abnormal. In view of the damage to actual system when fault happens, the threshold parameters are adjusted according to the principle "prefer to false alarm rather than missing alarm". The testing results are showed in Table 1.

The average advance time of prediction (min)	The average deviation of time prediction (min)	False-positive error (%)	False-negative error (%)
2.05	0.40	7.2	0

Table 1. The Prediction Performance Using the Proposed Method

From the results above, the efficiency of the proposed method is not affected by noise; meanwhile the accuracy of prediction is less influenced with noise. So it can predict the fault of CSTR feed in real-time with satisfactory prediction performance.

## 6. Conclusions

In this paper, an online fault prediction method has been proposed based on time series novelty estimation. The method can predict the fault of nonlinear system accurately without any fault data or prior knowledge. And it can study and predict with fewer amounts of calculation while system is running. So, it is reliable with good real-time capability and can be served as a new approach to predict faults in other complex nonlinear systems.

The purpose of the paper is to supply an effective approach for fault prediction of some actual systems with complex nonlinearity. How to utilize the existing information to estimate the approximate time when fault will happen is still the problems that need study in the future.

## 7. Acknowledgements

This work is supported in part by the Innovation Program of Shanghai Municipal Education Commission under Grant 12YZ156, and the innovation program of Shanghai college student (CS1124008).

## References

- [1] Zhang, Z. D., HU, S. S. (2008). A new method for fault prediction of model-unknown nonlinear system. *Journal of the Franklin Institute*, 345 (2) 136–153.
- [2] Yang, S. K., Liu, T. S. (1999). State estimation for predictive maintenance using Kalman filter. *Reliability Engineering and System Safety*, 66 (1) 29–39.
- [3] Chen, M. Z., Zhou, D. H. (2001). An Adaptive Fault Prediction Method Based on Strong Tracking Filter. *Journal of Shanghai Maritime University*, 22 (3) 35–40.
- [4] Tang, G. Z., Zhang, G. M., Gong, J. M. et al. (2008). Turbo-Generator Vibration Fault Prediction Using Gray Prediction Model. *In: Proceedings of the 7<sup>th</sup> World Congress on Intelligent Control and Automation*, p. 8542–8545. *IEEE Conference Publishing Services*.
- [5] Tse, P. W., Atherton, D. P. (1999). Prediction of machine deterioration using vibration based fault trends and recurrent neural networks. *Journal of Vibration and Acoustics*, 121 (7) 355–362.
- [6] Ho, S. L., Xie, M., Goh, T. N. (2002). A Comparative Study of Neural Network and Box-Jenkins ARIMA Modeling in Time Series Prediction. *Computer & Industrial Engineering*, 42 (2–4) 371-375.
- [7] Xu, K., Xie, M., Tang, L. C., Ho, S. L. (2003). Application of Neural Networks in forecasting Engine Systems Reliability. *Applied Soft Computing*, 2 (4) 255–268.
- [8] Policker, S., Geva, A. B. (2000). A New Algorithm for Time Series Prediction by Temporal Fuzzy Clustering. *In: Proceedings of 15<sup>th</sup> International Conference on Pattern Recognition*, p. 732–735.

[9] Pena, J. M., Letourneau, S., Famili, F. (1999). Application of rough sets algorithms to prediction aircraft component failure. *In: Lecture Notes in Computer Science*, 1642, p. 473–484. Springer.

[10] Povinelli, R. J., Xin, F. (2003). A New Temporal Pattern Identification Method for Characterization and Prediction of Complex Time Series Events. *IEEE Transactions on Knowledge and Data Engineering*, 15 (2) 339–352.

[11] Povinelli, R. J. (1999). *Time Series Data Mining: Identifying Temporal Patterns for Characterization and Prediction of Time Series Events*. USA: Marquette Univ..

[12] Su, S.C., Zhang, Z. D., Zhu, D. Q. (2008). Fault prediction for nonlinear time series based on temporal pattern mining. *Systems Engineering and Electronics*, 30 (10) 2023–2027.