Remaining useful life estimation based on the hybrid Support Vector Regression with the Nonlinear Autoregressive with Exogenous Input (SVR-NARX)

Estimativa da vida útil remanescente com base no método híbrido Support Vector Regression with the Nonlinear Autoregressive with Exogenous Input (SVR-NARX)

DOI: 10.54021/seesv5n1-168

Recebimento dos originais: 24/05/2024
Aceitação para publicação: 14/06/2024

Euldji Riadh
PhD Student in Electronic
Institution: Laboratory of Applied Automation and Industrial Diagnostics (LAADI), Faculty of Science and Technology, University of Djelfa
Address: BP 3117, Djelfa, Algeria
E-mail: r.euldji@univ-djelfa.dz

Boumahdi Mouloud
PhD in Mechanical Engineering
Institution: Department of Mechanical Engineering, Faculty of Technology, University of Yahia Fares
Address: 26000, Medea, Algeria
E-mail: boumahdi_m@yahoo.fr

Bachene Mourad
PhD in Mechanical Engineering
Institution: Department of Mechanical Engineering, Faculty of Technology, University of Yahia Fares
Address: 26000, Medea, Algeria
E-mail: bachene_medea@yahoo.fr

Euldji Rafik
PhD in Automatic
Institution: Laboratory of Mechanics, Physics, Mathematical modeling (LMP2M), University of Medea
Address: 26000, Medea, Algeria
E-mail: rafikeuldji@gmail.com

Euldji Imane
PhD in Chemical Engineering
Institution: Department of Process and Environmental Engineering Biomaterials and Transport Phenomena Laboratory (LBMPT), Faculty of Technology, University of Yahia Fares
Address: 26000, Medea, Algeria
E-mail: euldji.imane@univ-medea.dz
**ABSTRACT**

The rotating machines took an important role in the industries and manufacturing technology, the continually using of these tools leads to its breakdown, which manages to several loess, including high economies loss. This paper aims to avoid the unexpected failure of those tools by estimating the Remaining Useful Life (RUL) of the ball bearing, for this sake a couple of methods namely Decision Tree (DT), and the hybrid Support Vector Regression (SVR) with the Nonlinear Autoregressive with Exogenous Input (NARX) named as SVR-NARX which is applied to determine the RUL, first Time Domain Features (TDF) are extracted from the raw vibration signal and then this TDF are selected using the DT method, after that the Discrete Wavelet Transform (DWT) is applied on the selected features to separate the high and low frequencies from the selected features, the extracted frequencies components (EFC) are used as input which are used to train and test the SVR-NARX, the obtained model is then used to determine the RUL, The online PRONOSTIA database is applied for the training and testing the SVR-NARX, the SVR-NARX is compared to its primitives the SVR and NARX trained and tested using the EFC and the original selected feature, the overall of the applied strategy indicate that the SVR-NARX trained by the EFC gave high results in terms of Root Mean Squared Error (RMSE=0.0090, 0.0085) and Factor of determination ($R^2 = 0.999, 0.997$) for both training and testing respectively, the applied strategy gave high result which should be further considered for other machine related tasks.

**Keywords:** condition monitoring, decision tree, support vector regression, nonlinear autoregressive with exogenous input.

**RESUMO**

As máquinas rotativas desempenharam um papel importante nas indústrias e na tecnologia de manufatura, e o uso contínuo dessas ferramentas leva à sua quebra, resultando em várias perdas, incluindo grandes prejuízos econômicos. Este artigo visa evitar a falha inesperada dessas ferramentas estimando a Vida Útil Remanescente (RUL) do roloamento de esferas. Para isso, são aplicados dois métodos, a saber: Árvore de Decisão (DT) e a Regressão de Vetor de Suporte híbrida (SVR) com o modelo Autorregressivo Não Linear com Entrada Exógena (NARX), denominado SVR-NARX, para determinar a RUL. Primeiro, são extraídas as Características do Domínio do Tempo (TDF) do sinal de vibração bruto e, em seguida, essas TDF são selecionadas usando o método DT. Depois disso, a Transformada Wavelet Discreta (DWT) é aplicada nas características selecionadas para separar as frequências altas e baixas das características selecionadas. Os componentes de frequências extraídos (EFC) são usados como entrada para treinar e testar o SVR-NARX. O modelo obtido é então utilizado para determinar a RUL. O banco de dados online PRONOSTIA é aplicado para o treinamento e teste do SVR-NARX. O SVR-NARX é comparado aos seus modelos primitivos, o SVR e o NARX, treinados e testados usando os EFC e a característica original selecionada. A estratégia aplicada no geral indica que o SVR-NARX treinado pelos EFC apresentou altos resultados em termos de Erro Quadrático Médio (RMSE=0.0090, 0.0085) e Fator de Determinação ($R^2 = 0.999, 0.997$) tanto para o treinamento quanto para o teste, respectivamente. A estratégia aplicada apresentou resultados elevados que devem ser considerados para outras tarefas relacionadas a máquinas.
**1 INTRODUCTION**

Ball bearings are a key component that ensures the rotation phase in all rotating machines, and their deterioration can affect the equipment and further the whole production line. Ball bearing condition monitoring belongs to the condition-based maintenance (CBM), which allows for avoiding any unexpected breakdown (Goyal; Pabla, 2015). Condition monitoring (CM) of ball bearings includes acoustic emission, temperature analysis, oil debris, and among all of them is vibration-based condition monitoring analysis (Malla; Panigrahi, 2019).

The vibration-based condition monitoring is performed by utilizing sensors that are attached to the bearings, such as accelerometers, to collect the data. Another essential task in CM is determining the amount of time that is left for the bearing to operate before failure, which is known as RUL. The estimation of RUL before failure offers many advantages, such as avoiding the machine from complete damage and saving high maintenance money costs. The RUL is determined using data driven techniques, which gave high results with accurate RUL prediction. Saidi et al. (Saidi; Ben Ali; Fnaiech, 2015), and Benkedjouh et al. (Benkedjouh et al., 2013). applied the SVR to determine the RUL of bearings in a mechanical application. SVR is a regression algorithm that enables the use of continuous values, unlike SVM classification. In the study presented by Chaochao Chen et al. (Chen; Vachtsevanos; Orchard, 2012), the remaining useful life of a UH-60 helicopter planetary gear plate was estimated based on the ANFIS approach and the particle filter, the ANFIS approach was used to model evolution trends of the fault, while the Particle filter is applied to integrate the ANFIS approach for multi-step ahead prediction, and the RUL is obtained from the probability density function (PDF), the results indicates that using both models gave more accurate result than using each model alone. Rabiei et al. (Rabiei; Droguett; Modarres, 2016). Also utilized the SVR technique to establish a correlation between input data variables, aiming to predict damages and estimate the onset of cracks in a metallic alloy prone to fatigue cracking. Benkedjouh et Rechak (Benkedjouh; Rechak, 2017). applied the vibration condition monitoring for tool wear degradation assessment and it RUL estimation,
based on the Empirical Mode Decomposition (EMD) and the Improved Extreme Learning Machine (IELM), the relation between the cutting forces data and tool wear in terms of correlation gave good results. Fumeo et al. (Fumeo; Oneto; Anguita, 2015) introduce an online SVR approach for estimating the RUL of bearings by optimizing the trade-off between precision and computational efficiency. Benkedjouh et al. (Benkedjouh et al., 2015) Developed an approach based on SVR to predict wear progression in a high-speed milling machine and predict its RUL. A hybrid model was applied by Nieto et al. (García Nieto et al., 2015) by integrating SVR and PSO (Particle Swarm Optimization) for predicting the RUL of aircraft engines. They utilized PSO in the process of estimating the best hyper-parameters for the SVR approach. The purpose of this contribution is to build a novel hybrid approach based on data-driven techniques for monitoring the condition and determining the RUL of the ball bearing.

Motivate by the previous works, vibration signals are first collected from the external race of a ball bearing, TDF are extracted from the vibration signal and then selected using the DT algorithm based on determined thresholds which comes in second, in the third step the DWT is used, The DWT is used to extract the high (details) and low (approximation) frequencies of the applied signal, the extracted frequencies components are used as input (Jin; Kim, 2015) to train the SVR-NARX. The SVR-NARX approach is used to model the prediction values of the evolution degradation of the ball bearing, the development of the bearing degradation is used to determine the RUL, the performance of the SVR-NARX is judged based on performance measurement. Highlighting the main contributions in this work: (1) using the DT for effective feature selection, (2) using the DWT to extract wavelet feature from the selected feature by the DT approach, (3) using the hybrid SVR-NARX model to predicted and estimate the RUL, The rest of this contribution is structured as follows: the second section presents the applied methodology and a brief discretion of the DT, SVM and NARX model, third section presents the results and discussion, and at the last section a conclusion is presented.
2 METHODOLOGY

The applied methodology is presented in the flowchart of Figure 1 which is about to extract TDF and then to selected the most appropriate ounces using the DT, the DWT approach is then applied on the selected features from the training set, the wavelet features are used to train the SVR, the obtained approach is then tested using DWT features extracted from the test data (based on the DT model), the results are then smoothed using the RLOWESS function and then RUL is estimated.

Figure 1. Applied Methodology

2.1 DECISION TREE

The DT is a data-mining technique used applied of the time to build classification models and to give decisions (Euldji; Boumahdi; Bachene, 2021), for solving various problems in a Boolean form like yes or no answers based on input data divided into attributes (features) and classes, the visualization of the DT and its simple structure makes it easy to read and understand, it can handle both categorical and numerical data, and its deals also with multiple output systems, The dataset includes features and classes which are required to build the DT. criteria entropy (CE) and the Gain Ratio (GR), both helps for selecting the most
appropriate feature for high classification accuracy, the CE is a measure related to uncertainty associated with a random variable (Bhosale; Ade; Deshmukh, 2014), the selected feature comes at top of the tree which has the max GR value (Euldji; Boumahdi; Bachene, 2021).

These criteria are defined as follows:

\[
Info(T) = - \sum_{i=1}^{k} \frac{|C_j|}{|T|} \log_2 \left( \frac{C_j}{|T|} \right) 
\]  

\[
Info(X_i, T) = - \sum_{i=1}^{n} \frac{|T_i|}{|T|} \log_2 \left( \frac{T_i}{|T|} \right) = - \sum_{i=1}^{n} \left( \frac{|T_i|}{|T|} \cdot \sum_{j=1}^{k} \frac{|C_j|}{|T_i|} \log_2 \left( \frac{C_j}{T_i} \right) \right) 
\]  

Gain of entropy, or informational gain, due to attribute \(X_i\), defined as:

\[
Gain(X_i, T) = Info(T) - Info(X_i, T) 
\]  

Gain ratio, or Gain of normalized entropy, defined as:

\[
GR(X_i, T) = \frac{Gain(X_i, T)}{Split\ info(X_i, T)} 
\]  

Where:

\[
Split\ info(X_i, T) = - \sum_{i=1}^{n} \frac{|T_i|}{|T|} \log_2 \left( \frac{T_i}{|T|} \right) 
\]  

Where \(X = X_1, X_2, . . . , X_i, . . . , X_n\) are the attributes set and \(n\) is the number of attributes, \(C = \{C_1, C_2, . . . , C_k\}\) the classes set and \(k\) is the number of classes, \(|C_j|, j = 1, 2, . . . , k\) is the number of examples which belongs to the class \(C_j\), \(T\) and \(|T|\) the set of training examples and the total number of examples respectively.
2.2 SUPPORT VECTOR REGRESSION

Support Vector Machine (SVM) is a powerful data mining technique used in various supervised learning applications such as classification, clustering and regression problems (Raja et al., 2021). For regression tasks, the SVR approach is based on mapping the input raw data $x$ into a high dimensional attributes space $F$ using a kernel function and obtaining the optimum hyper-plane that separates each of them (Li et al., 2017; Xu et al., 2020) using a nonlinear projecting function $\phi (x)$, and then to conduct a linear regression in $F$ space (Pourbasheer et al., 2011; Vapnik, 1995; Xu et al., 2020). The linear regression function is given as follows (Cheng; Cai, 2017; Li et al., 2017; Nait Amar; Zeraibi, 2020; Tatar et al., 2016):

$$f(x) = \omega \cdot \phi(x) + b$$  (6)

Where $\phi (x)$ is the kernel function, $\omega$ and $b$ refers to the weight vector and the bias term, respectively.

That can be achieved by minimizing the cost function (Farhat, [s.d.]; Nait Amar; Zeraibi, 2020):

$$\text{cost function} = \frac{1}{2} \omega^2 + C \sum_{i=1}^{k} (\xi_i^- + \xi_i^+)$$  (7)

$$ \begin{cases} y_i - (w \cdot \phi(x_i) + b) \leq \varepsilon + \xi_i^+ \\ (w \cdot \phi(x_i) + b) - y_i \leq \varepsilon + \xi_i^- \\ \xi_i^+,\xi_i^- \geq 0, \quad i = 1,2,3,...,n \end{cases}$$

A sample form of the support vector machine method is shown in Figure 2.

2.3 NONLINEAR AUTOREGRESSIVE MODEL WITH EXOGENOUS INPUTS

The NARX is the non-linear form of the Autoregressive with Exogenous Input (ARX) technique, ARX means that the dynamic characteristics of the process are fully characterized. Advantageously, model parameters can be estimated recursively, that relates the exogenous input and output that is destroyed by noise.
If we suppose that the exogenous input process is generated independently of the noise process with standard deviation, we can write the ARX approach as (Wang; Wu; Chen, 2001):

\[
y(k) = \delta_1 y(k - 1) + \delta_2 y(k - 2) + \ldots + \delta_r y(k - r) + \omega_0 x(k - b) \\
+ \omega_1 x(k - b - 1) + \omega_2 x(k - b - 1) + \ldots + \omega_s x(k - b - s) \\
+ N(k)
\]

Where \( b \geq 0 \).

Let \( B \) denote the back shift operator, which is defined by \( Bx(k) = x(k - 1) \). Hence \( B^k x(k) = x(k - b) \).

The above difference equation can be written as:

\[
y(k) = \delta_1 By(k) + \delta_2 B^2 y(k) + \ldots + \delta_r B^r y(k) + \omega_0 x(k - b) + \omega_1 Bx(k - b) \\
+ \ldots + \omega_s B^s x(k - b) + N(k)
\]

\[
= (\delta_1 B + \delta_2 B^2 + \ldots + \delta_r B^r) y(k) + (\omega_0 + \omega_1 B + \ldots + \omega_s B^s) B^b x(k - b) + N(k)
\]

If we define two operators by:

\[
\omega_s(B) = \omega_0 + \omega_1 B + \omega_s B^s
\]
\[ \delta_r(B) = \delta_1 B + \cdots + \delta_r B^T \] (11)

The ARX approach can be rewritten as:

\[ y(K) = \delta_r(B)y(k) + \omega_s(B)B^b x(k) + N(k) \] (12)

The approach order \( r, s \) and \( b \) are identified based on different criteria, and the unknown parameters \( \delta_1, \delta_2, \ldots, \delta_r, \omega_0 + \omega_1 + \omega_s, \sigma_N^2 \), which are predicted from the observation data (Yang; Makis, 2010).

3 RESULT AND DISCUSSION
3.1 EXPERIMENTAL DATASETS

To prove the success of the applied methodology, an online database named PRONOSTIA is used. The PRONOSTIA database is dedicated to testing and verifying the fault detection, diagnosis, and prognosis methods of ball bearings. Only vibration measurements are used in this contribution, collected from the NSK 6804DD bearing with a capability of operating at speeds of 1800, 1650, and 1500 rpm and a maximum load of 5000N. The data are collected by conducting a speed-up bearing degradation test rig. Accelerometers adhered to the external race of the bearing to monitor both horizontal and vertical vibration signals. More details about this platform are available in (Nectoux et al., 2012). The database includes two parts training and testing set at different operation condition and load. Table 1 includes the PRONOSTIA database train and test data (Nectoux et al., 2012).

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Operation Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conditions 1</td>
</tr>
<tr>
<td><strong>Load (N)</strong></td>
<td></td>
</tr>
<tr>
<td>4000</td>
<td></td>
</tr>
<tr>
<td><strong>Speed(rpm)</strong></td>
<td></td>
</tr>
<tr>
<td>1800</td>
<td></td>
</tr>
<tr>
<td><strong>Training set</strong></td>
<td></td>
</tr>
<tr>
<td>Bearing1-1</td>
<td></td>
</tr>
<tr>
<td>Bearing1-2</td>
<td></td>
</tr>
<tr>
<td><strong>Testing set</strong></td>
<td></td>
</tr>
<tr>
<td>Bearing1-3</td>
<td></td>
</tr>
<tr>
<td>Bearing1-4</td>
<td></td>
</tr>
<tr>
<td>Bearing1-5</td>
<td></td>
</tr>
<tr>
<td>Bearing1-6</td>
<td></td>
</tr>
<tr>
<td>Bearing1-7</td>
<td></td>
</tr>
</tbody>
</table>

Source: (Nectoux et al., 2012)
3.2 FEATURE EXTRACTION

TDF are extracted from the PRONOSTIA database. Only vibration data of bearing by condition 02 are applied in this application. According to Table 1, two bearing sets named Bearing2-1 and Bearing2-2 are for training purpose, TDF are extracted from both sets and their evolution other the time is presented in Figure 3 and Figure 4 while the mathematical formula of the TDF are present in Table 2.

![Figure 3. Feature evolution other the time for bearing2-1](image)

3.3 FEATURE SELECTION

In the feature selection phase, a Dt algorithm, namely J48, is implemented using WEKA data mining tools. For the classification of the data, the Weka merged C4.5 (J48) is used, and the theoretical background of C4.5 is referred (Hormann, 1964). is applied to both training sets. These sets are treated separately, each set contains three classes: normal, abnormal, and dangerous, with are the output, with 6 characteristics as inputs (see Table 2) (Euldji; Boumahdi; Bachene, 2021) (Euldji et al., 2023)
Both inputs and output are necessary to construct the tree. These classes are defined by two thresholds: the alarm threshold and the danger threshold. The alarm threshold is the first measurement multiplied by two, which indicates the beginning of degradation. The danger threshold is the first measurement multiplied by six, which defines the danger zone.

Table 3 shows the results of the classification based on the J48 algorithm.
Figure 5. Decision tree for Bearing2-1

From Table 3: it is obvious that the selected feature by the decision tree is the RMS from both bearings, bearing2-2 is presenting higher performance and small mean absolute error, but returning to the leaves the only tree that present all thresholds together (Normal, Abnormal and Danger) is Figure 5,
which indicate that the RMS from bearing2-1 is chosen to be the candidate feature for the proposed model.

3.4 PROPOSED STRATEGY

After the selection of the RMS from Bearing2-1, to be used to train the model, the DWT is applied on the RMS to extract the high (details) and low (approximation) frequencies of a selected feature. This is very helpful, because we can use the extracted frequency components as new inputs to the proposed approach, which is named as DWRMS. More information about the extraction of these components can be found in (Jin; Kim, 2015).

The result of using the proposed SVR-NARX is highlighted in Table 4 based on the RMSE and the $R^2$, which is compared against the SVR and NARX using both DWRMS and RMS.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Train</th>
<th></th>
<th>Test</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>$R^2$</td>
<td>RMSE</td>
<td>$R^2$</td>
</tr>
<tr>
<td>SVR-RMS</td>
<td>0.0692</td>
<td>0.927</td>
<td>0.0422</td>
<td>0.935</td>
</tr>
<tr>
<td>SVR-DWRMS</td>
<td>0.0189</td>
<td>0.954</td>
<td>0.0170</td>
<td>0.962</td>
</tr>
<tr>
<td>NARX-RMS</td>
<td>0.0662</td>
<td>0.937</td>
<td>0.0370</td>
<td>0.925</td>
</tr>
<tr>
<td>NARX-DWRMS</td>
<td>0.0182</td>
<td>0.966</td>
<td>0.0160</td>
<td>0.954</td>
</tr>
<tr>
<td>SVR-NARX-RMS</td>
<td>0.0598</td>
<td>0.958</td>
<td>0.0339</td>
<td>0.976</td>
</tr>
<tr>
<td>SVR-NARX-DWRMS</td>
<td>0.0090</td>
<td>0.999</td>
<td>0.0085</td>
<td>0.997</td>
</tr>
</tbody>
</table>

Source: Author

From Table 4 it is clear that using the SVR-NARX trained and tested using the DWRMS gave higher performance according the RMSE and $R^2$ value, the result of using the DWRMS on the SVR and NARX separate gave also good result compared to the obtained result using only the RMS, the overall of obtained result is very promising for the RUL estimation, Figure 7 and Figure 8 show the output measured by the proposed model for both trained test, respectively.
Figure 7. training phase for the proposed model

Figure 8. Testing phase for the proposed model

Figure 9 show the result of the Factor of determination ($R^2$) for the proposed model.

3.5 RUL ESTIMATION

After the tested is done the Smoothing techniques based on the RLOWESS method (Aftab; Moghadam, 2022) is applied to remove the noise, from the predicted test data for accurate RUL prediction.
The RUL is defined by the relation:

\[
RUL(t) = t_f - t_c
\]  

(13)

Where, and are the final and current time respectively, the predicted RUL by the proposed model at its total failure.

Figure 10 denoted the result of the actual and predicated Smoothed RUL.

From Figure 10 at the begging the predicted RUL is above the real RUL and then the predicted RUL at about 2000 seconds which is equivalent to 55 minutes
is under the actual RUL which indicates that the proposed model can predict the RUL before the real failure during the whole operation of the machine which very promising for avoiding unexpected failure.

4 CONCLUSION

In this work, the failure of the ball bearing is estimated to avoid the unexpected breakdown based on the Decision Tree and the hybrid Support Vector Regression with the Nonlinear Autoregressive with Exogenous Input (SVR-NARX) modeled by vibration signals from the online database PRONOSTIA, only vibration data from condition 02 are used in this work, six TDF are extracted from each training base and then these features are selected using the j48 DT method implemented in the WEKA software, to selected the appropriate features to train the model, after that the DWT is applied on the selected features to separate the high (details) and low (approximation) frequencies, the separate frequencies components are used as new features, the SVR-NARX is trained and tested using the wavelet features and evaluated based on the RMSE and the $R^2$, the predicted values from the test data are smoothed using the RLOWESS function to eliminate the noise and unwanted outliers, the smoothed values are then used to compute the RUL, the main contributions in this paper are: (1) using the DT for effective feature selection, (2) using the DWT to extract wavelet feature from the selected feature by the DT approach, (3) using the hybrid SVR-NARX model to predicted and estimate the RUL, the proposed strategy can benefit society and academia through developing of new technologies by creating new diagnostic tools based on hybrid models, reduction in maintenance costs, Enhanced Safety by decreasing the unexpected breakdowns, the applied strategy have one limitation regarding the extrapolation of the future values beyond the range of testing data which is not considered in this study, and should be announced as a perspective for future work, the overall of the applied strategy can gave high performance regarding RUL prediction which should be applied for other rotatory machines, determining the RUL of battery and other domains such as the prediction of the corona virus.
ACKNOWLEDGMENTS

Thanks to General Directorate of Scientific Research and Technological Development (DGRSDT), MESRS, Algeria; Laboratory of Applied Automation and Industrial Diagnostics LAADI, Faculty of Science and Technology University Ziane Achour of Djelfa, Algeria, and to Dr Abdessalem Benterki, Laboratory of LMP2M, Department of Mathematics, University of Médéa, Algeria.
REFERENCES


