

Adaptive Support Vector Regression for Long-Term Prediction under Nonstationary Environments: Prognostics of Components in Nuclear Power Plants

Jie Liu¹, Enrico Zio^{1,2}

¹Chair on Systems Science and the Energetic Challenge, EDF Foundation, CentraleSupélec, Paris, France

²Energy Department, Politecnico di Milano, Milano, Italy

Abstract

Prognostics and Health Management (PHM) aims at diagnosing the state and predicting the future evolution of Systems, Structures and Components (SSCs). In real industrial applications, SSCs are normally operated in nonstationary environments and evolving conditions. In our previous work, an adaptive online learning approach is proposed for the short-term (one-day ahead) prediction by Support Vector Regression (SVR), giving satisfactory performance under nonstationary environments. In this paper, this work is extended to long-term (one/two/three/four-weeks ahead) prediction. An application is presented, concerning of the leakage of the first seal of a Reactor Coolant Pump (RCP) of a Nuclear Power Plant (NPP). The results show the adaptability of the proposed approach and offer some reflections on the long-term prediction problem by data-driven approaches.

Key words: long-term prediction, adaptive learning, support vector regression, nuclear power plant, condition-based maintenance, predictive maintenance

1. Introduction

In the last decades, the importance of the maintenance function and management has increased dramatically. The extensive mechanization and automation of industrial equipment has reduced the number of production personnel and expanded the capital invested in the equipment. Different maintenance strategies have been developed, such as corrective maintenance, preventive maintenance, condition-based maintenance. With the development of Prognostics and Health Management (PHM), predictive maintenance is becoming more and more popular in both the research and the industry. The most important objective of PHM is to predict the future evolution of the (health) state of the Systems, Structures and Components (SSCs) of interest. The prediction can be short-term or long-term according to the prediction horizon of the specific application. Various benefits of the maintenance management have been reported in the literature, including reduction of business interruption due to failure of SSCs, reduction of scheduled maintenance costs, reduction of secondary damages, reduction of maintenance-induced failures, etc.

Nuclear Power Plants (NPPs) are typical representatives of modern complex industry systems. There are approximately 437 reactors currently in service, 73 new reactors being constructed, and as many as 481 reactors planned or proposed for construction [1]. Although the worldwide interest in nuclear power has been somewhat tempered by the incident at Fukushima Dai-Chi in Japan, as well as by recent changes in oil and gas production that may change the worldwide electricity production program, nuclear power is still considered a key element in meeting the worldwide goals for sustainable energy and green house emission [2]. In existing NPPs, the most crucial tasks to be accomplished are improving safety, maintaining availability and reducing operation and maintenance costs. Moreover, extensions in power up rates and in the average component life duration increase the need for techniques for diagnosing and predicting the NPP health, because the occurrence of component degradation and failure becomes more and more likely as load is increased or changed, and as age advances [3].

When the PHM paradigm is integrated in NPPs, the short-term prediction means multiple hours-ahead prediction which provides information for emergency action, even shutdown of NPP, while the long-term prediction considers weeks/months-ahead prediction, which assists in maintenance planning, including repair and replacement [4].

Some works have already appeared in the relevant literature concerning prognostics of NPP components. In [5] and [6], a fuzzy similarity analysis is introduced to find a combination of the reference patterns, weighed by their similarity to the observed failure pattern, to determine the future evolution of the observed failure pattern and to derive the corresponding Remaining Useful Life (RUL). In [7] and [8], Support Vector Regression (SVR) is used to predict the collapse moment for wall-thinned pipe bends and elbows. In [9], a systematic approach is introduced for the prediction of pump performance characteristics, for situations in which the experimental data are not available. In [10], the authors present a framework for the control of the steam generator water level in the secondary circuit of a NPP, based on an extension of the standard linear model predictive control algorithm to linear parameter varying systems. A back propagation Artificial Neural Network (ANN) is proposed for the prediction of thermal power in NPPs in [11]. In [12], a probabilistic neural network is applied for classifying accidents into groups of initiating events and a fuzzy neural network is used to identify the major severe accident scenarios after the initiating events. Back propagation networks are used in [13] to estimate one or more process variables, by establishing the nonlinear relationship among a set of plant variables. In [14], an ANN is used to estimate the value of the undetected next-time-step signal of the steam generator water level in a NPP. In [15], data from plant operation experience are used in combination with in-service inspections and degradation management programs to ensure that the degradation mechanisms do not adversely impact plant safety.

In these previous works, the prognostic problem has been tackled under stationary conditions. Real NPP are operated under nonstationary environments and evolving conditions, i.e. pattern drifts, changing climate, up and down rate of power according to electricity demands, ageing etc. The models trained for stationary

conditions may no longer perform well in evolved situations. Prognostics needs, then to be provided with the capability of adaptation to new and changed situations.

SVR is a supervised learning technique for regression. Developed by Vladimir Vapnik and co-workers at AT&T Bell Laboratories in 1995, SVR is based on the principle of Structural Risk Minimization (SRM) [16]. An important feature of SVR is that the solution is based only on those data points which are at the margin. These points are called Support Vectors (SVs). The linear SVR can be extended to handle nonlinear problems when the data is first transformed into a high dimensional feature space, i.e. Reproducing Kernel Hilbert Space (RKHS), using a set of nonlinear kernel functions. In our previous work in [4], an adaptive online learning framework, named Online-SVR-FID has been proposed for SVR for short-term prediction. Two types of pattern drifts are defined in [4] and different strategies are proposed to tackle these pattern drifts. In this paper, our previous work is extended to the long-term (weeks-ahead) prediction for NPP components, as the shutdown of a NPP can cost up to \$ 1.25 million per day [17] and long-term prediction can provide more time for the operators to take proper reactions to prevent the shutdown of the NPP.

The case study in this paper concerns the leak flow of the first seal in the Reactor Coolant Pump (RCP) of a NPP, with the data provided by Electricite de France (EDF). RCP pumps the coolant into the reactor to cool down the nuclear materials and to transport the heat to the steam generators for final transformation to electricity.

Results for weeks-ahead prediction show that the proposed method can give satisfactory results with good prediction adaptability and accuracy.

The rest of the paper is structured as follows. Section 2 presents SVR and the previously proposed approach for prediction [4]. The case study is described in Section 3 with data preprocessing. Section 4 introduces experiment results. And some conclusions and future work are drawn in Section 5.

2. SVR and Online-SVR-FID

In this section, details about the previous work are presented following a brief introduction of SVR.

2.1 Support Vector Regression

Suppose $T = \{(\mathbf{x}_i, y_i): i = 1, 2, \dots, T\}$ is the training dataset. SVR finds a function $f(\mathbf{x})$ that has at most ε deviation from the actually obtained targets y_i for all the training data points, and at the same time is as flat as possible. In other words, we do not care about the errors as long as they are less than ε , but will not accept any deviation larger than this. However, this may not always be the case, meaning that we also want to allow some error larger than ε . Analogously to the soft margin loss function [18], which is adapted to SVR in [19], one can introduce slack variables ξ_i, ξ_i^* in the constraints of the optimization problem of SVR.

The associated optimization problem for calculating the best function $f(\mathbf{x})$ is

$$\begin{aligned}
& \text{minimize } \frac{1}{2} \|\boldsymbol{\omega}\|^2 + C \sum_{i=1}^T (\xi_i + \xi_i^*) \\
& \text{subject to } \begin{cases} y_i - f(\mathbf{x}) \leq \varepsilon + \xi_i \\ f(\mathbf{x}) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \\ f(\mathbf{x}) = \langle \boldsymbol{\omega}, \mathbf{x} \rangle + b \end{cases} .
\end{aligned} \tag{1}$$

The constant C determines the trade-off between the flatness of $f(\mathbf{x})$ and the amount up to which deviations larger than ε are tolerated. This corresponds to dealing with an ε -insensitive loss function which is shown in Figure 1 [20]. Only the data points outside the shaded region contribute to the cost. It turns out that in most cases the optimization problem in Equation (1) can be solved more easily in its dual formulation, briefly reported below.

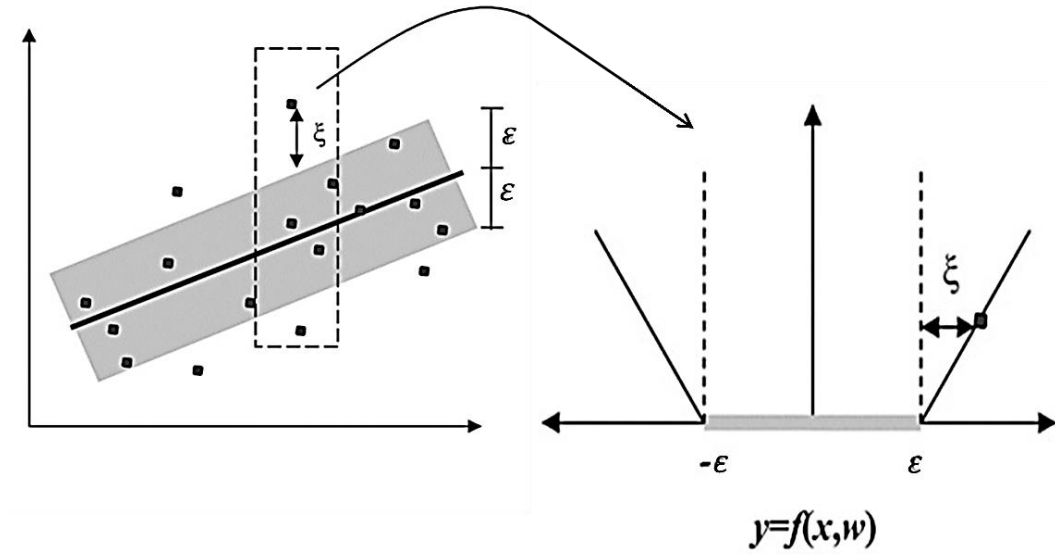


Figure 1 The soft margin loss setting for SVR [20].

The dual optimization problem of Equation (1) is

$$\begin{aligned}
L = & \frac{1}{2} \|\boldsymbol{\omega}\|^2 + C \sum_{i=1}^T (\xi_i + \xi_i^*) - \sum_{i=1}^T (\eta_i \xi_i + \eta_i^* \xi_i^*) - \sum_{i=1}^T \alpha_i (\varepsilon + \xi_i - y_i + \langle \boldsymbol{\omega}, \mathbf{x}_i \rangle + b) - \\
& \sum_{i=1}^T \alpha_i^* (\varepsilon + \xi_i^* + y_i - \langle \boldsymbol{\omega}, \mathbf{x}_i \rangle - b).
\end{aligned} \tag{2}$$

Here L is the Lagrangian and $\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$ are positive Lagrange multipliers. By substituting the partial derivatives of L with respect to the primal variables, i.e. $\boldsymbol{\omega}, b, \xi_i, \xi_i^*$, into Equation (2), the dual optimization problem becomes

$$\begin{aligned}
& \text{maximize } \frac{1}{2} \sum_{i,j=1}^T (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \langle \mathbf{x}_i, \mathbf{x}_j \rangle - \varepsilon \sum_{i=1}^T (\alpha_i + \alpha_i^*) + \sum_{i=1}^T y_i (\alpha_i - \alpha_i^*) \\
& \text{subject to } \sum_{i=1}^T (\alpha_i - \alpha_i^*) = 0 \text{ and } \alpha_i, \alpha_i^* \in [0, C].
\end{aligned} \tag{3}$$

The partial derivative of L with respect to the primal variable $\boldsymbol{\omega}$ shows that $f(\mathbf{x})$ can be rewritten as

$$f(\mathbf{x}) = \sum_{i=1}^T (\alpha_i - \alpha_i^*) \langle \mathbf{x}_i, \mathbf{x} \rangle + b. \quad (4)$$

Equation (4) is the so-called support vector expansion, i.e. ω can be described as a linear combination of the training data points, with $\langle \mathbf{x}_i, \mathbf{x} \rangle = k(\mathbf{x}_i, \mathbf{x})$ in the nonlinear case and $k(\mathbf{x}_i, \mathbf{x})$ the kernel function that maps the original data into a high dimensional space, i.e. RKHS where the relation between the data inputs and outputs becomes linear. In such sense, the complexity of a function's representation of SVR is independent of the dimensionality of the input space, and depends only on the number of SVs whose multipliers $\alpha_i - \alpha_i^*$ in Equation (4) are not zero.

2.2 Online-SVR-FID

SVR is a supervised learning approach, i.e. the data points for training and test follow an identical and independent distribution. In the case that the component is operated in nonstationary environment, the distribution generating the data points may change with time. In such case, we need to provide the model with adaptive learning ability. Online-SVR-FID has been proposed in our previous work for adaptive online learning of a single SVR model based on Feature Vector Selection (FVS) [21].

In [21], the authors propose the FVS method to select a subset of the training data points (i.e. Feature Vectors (FVs)), which can represent the dimension of the whole dataset in RKHS. The other data points can all be expressed as a linear combination of the selected FVs in RKHS.

Suppose that the FVs selected from the training dataset are $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L\}$ and the corresponding mapping is $S = \{\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \dots, \boldsymbol{\varphi}_L\}$. For a new data point $(\mathbf{x}_{new}, y_{new})$, Local Fitness $J_{S,new}$ as calculated in Equation (5) is proposed to characterize if the new data point can be well represented by the FVs in S.

$$J_{S,new} = \frac{K_{S,new}^t K_{S,S}^{-1} K_{S,new}}{k_{new,new}} \quad (5)$$

$K_{S,S} = (k_{i,j}), i, j = 1, 2, \dots, L$ is the kernel matrix of S and $K_{S,new} = (k_{i,new}), i = 1, 2, \dots, L$ is the vector of the inner product between $\boldsymbol{\varphi}_{new}$.

$$\mathbf{a}_{new} = K_{S,new}^t K_{S,S}^{-1} \quad (6)$$

The vector \mathbf{a}_{new} represent the closeness of the FVs in S to the new data point. If $1 - J_{S,new}$ is zero, the new data point is not a new FV; otherwise, it is a new FV.

Base on the work of Baudat and Anouar in [21], the Online-SVR-FID has been proposed in our previous work.

In this paper, each data point represents certain relation between the input vector and the output, and is defined as a pattern. Two types of pattern drifts, i.e. new pattern and changed pattern are considered. A new FV is defined as a new pattern and a changed pattern is the data point that shares the same input vector as previous data points but gives different outputs.

The pseudo-code of Online-SVR-FID is shown in Figure 2.

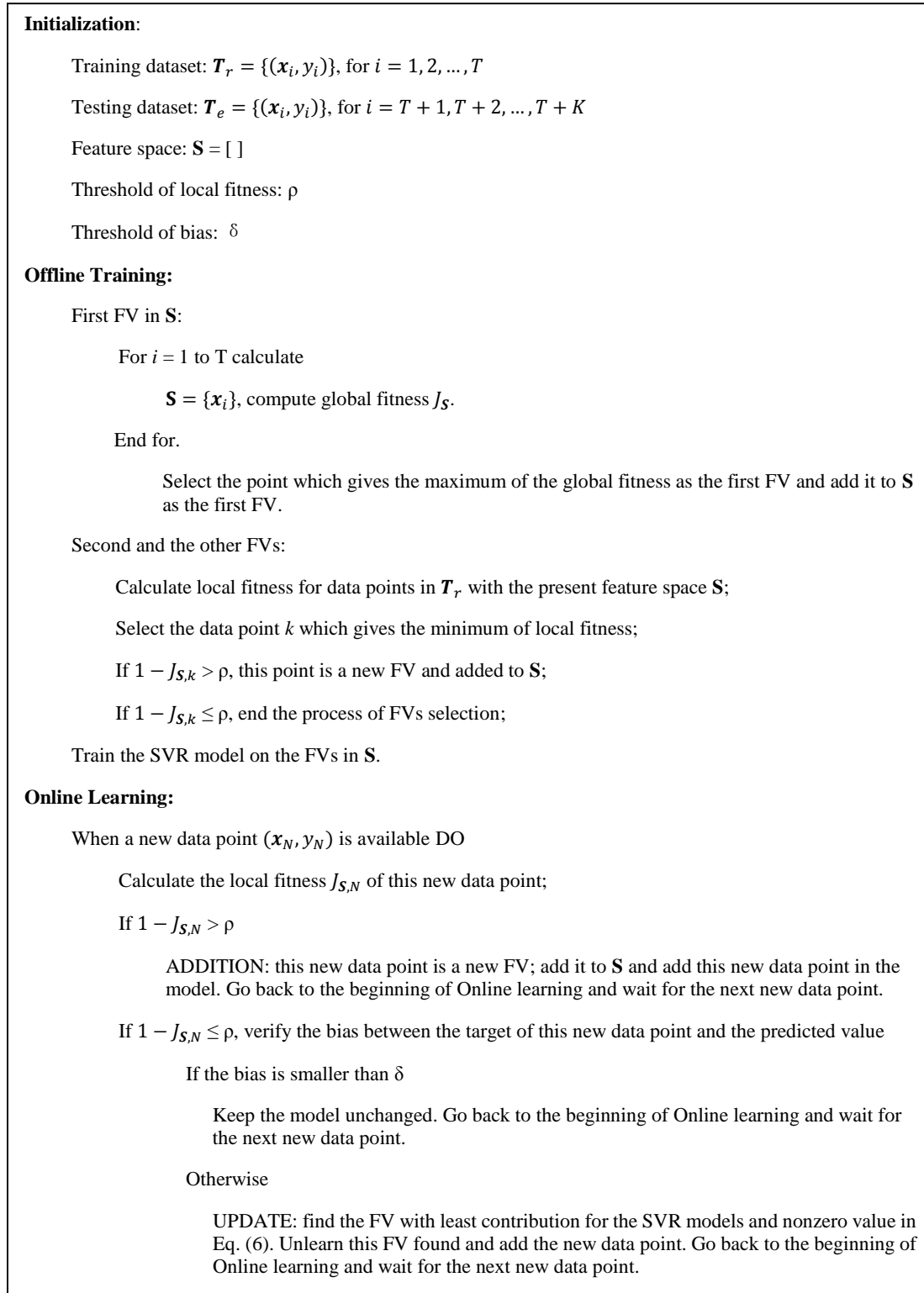


Figure 2 Pseudo-code of Online-SVR-FID.

When a new data point is judged as a changed pattern, it is used to replace one data point in the present model. The process is as follows:

1. A vector $\mathbf{m} = (m_1, m_2, \dots, m_l)$ is used to record the contribution of each FV to the SVR model. Each value in \mathbf{m} corresponds to a FV in the model.
2. \mathbf{m} is set to be a zero vector before Offline Training.
3. When the model M is trained during Offline Training with the selected FVs from the training dataset, m_i is increased by 1 if the corresponding FV is a SV. Otherwise, i.e. for a FV with zero multiplier, its contribution m_i is zero.
4. Each time the model is added with one new data point, a new m_{l+1} is added to \mathbf{m} to record the contribution of the new FV in the model. After the model is updated with ADDITION, the contribution m_i of each FV in the model is updated with the contribution update rules: if the data point is a SV in the new updated model, its new contribution is calculated as $m_i^{new} \leftarrow \tau * m_i + 1$, with τ a positive constant smaller than 1, i.e. the contribution of a FV in the new model is more weighted than that in the old models; otherwise it is kept unchanged.
5. When a change is detected with respect to the old patterns, the first step is to calculate the values \mathbf{a}_N for the new data point with Equation (5). Then, among all the FVs in the model with non-zero values in \mathbf{a}_N , the one with least contribution, say m_l , is deleted from the model and m_l is reset to zero. If there are several FVs with the same contribution and the least contribution, the FV to be replaced is selected as the oldest one among them.
6. The new data point is added to the model and it inherits the contribution m_l , which is zero for now. The vector \mathbf{m} and the feature space \mathbf{S} are updated, and also the contribution of the FV is updated according to the rules in step 4 above.

3. Real case study

Pumps play a major role in the safe operation of NPPs. Their operating characteristics play a significant role in determining the thermal and hydraulic behavior of nuclear reactors in following transients. The Reactor Coolant Pump (RCP) is a critical component of a NPP, since it guarantees enough cold water in the core of the plant to protect the nuclear materials and to deliver the heat released from the nuclear fission to the steam generator.

The RCP is composed of three main parts: the pump part, the sealing part and the motor part. The sealing system is composed of three seals, which are sequentially named the first seal, the second seal and the third seal. The sealing system prevents the boron water from leaking outside the primary circuit of a NPP. The leaked boron water may endanger the personnel working in the NPP and the equipment inside the nuclear

island. Thus, the reliable control of the leak flow is very important. If the leak flow exceeds a predefined threshold, the NPP should be shut down to decrease the chance of severe disasters.

In the reference case study, a time series data (of size 2280) of the leakage from the first seal is provided by Electricite de France (EDF). The data is measured every four hours. After the outlier deletion, missing value reconstruction and normalization, the time series data is shown in Figure 3.

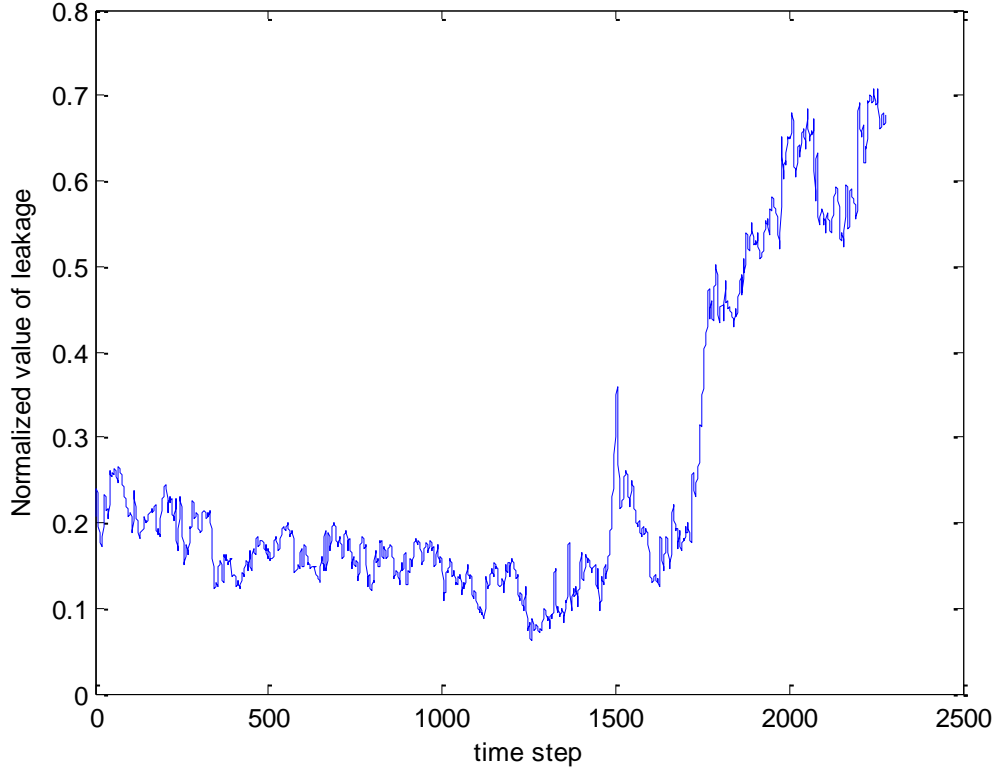


Figure 3 Time series data of the leakage from the first seal.

We can see that at the beginning (before 1200 time step), the leakage from the first seal is stable, i.e. in normal condition and starting from 1200 time step, the leakage starts increasing, i.e. abnormal condition. During 1600 and 1700 time steps, the abnormal situation is brought back to normal by the operator, but it continues to increase from time step 1700.

For this time series data, pattern drifts occur, i.e. change normal to abnormal conditions, and a model trained on the normal part can no longer perform well on the abnormal data. Thus, Online-SVR-FID is used in the experiment to gradually adapt to the pattern drifts. One/two/three/four week-ahead predictions are separately carried out and the prediction results are compared considering the prediction accuracy.

In the experiment, we, firstly, need to construct the training and test datasets. Suppose the time series data is $l(t), t = 1, \dots, 2280$, the reconstructed input-output data (x_i, y_i) is structured as $(l(i + 1), \dots, l(i + H)), l(i + H + M)), i = 0, \dots, 2280 - H - M$, with H determined by partial autocorrelation analysis, i.e. the number of historical values that are most related to the output, and M the prediction horizon, which is 42 for

one-week ahead prediction, 84 for two-week ahead prediction, 126 for three-week ahead prediction and 168 for four-week ahead prediction, as the data is measured every four hours.

For the experiment, the data points corresponding to 1000-1200 time steps in Figure 3, i.e. in normal condition, form the training dataset and the data points in abnormal part corresponding to time steps 1700 and 2200 serve as the test dataset. The two thresholds ρ and δ are fixed separately as 10^{-9} and 0.05 by trial and error.

4. Experiment results

Figures 4-7 show separately the experiments results for one-week, two-week, three-week and four-week ahead predictions, with the solid line the true values and the dotted line the predicted values.

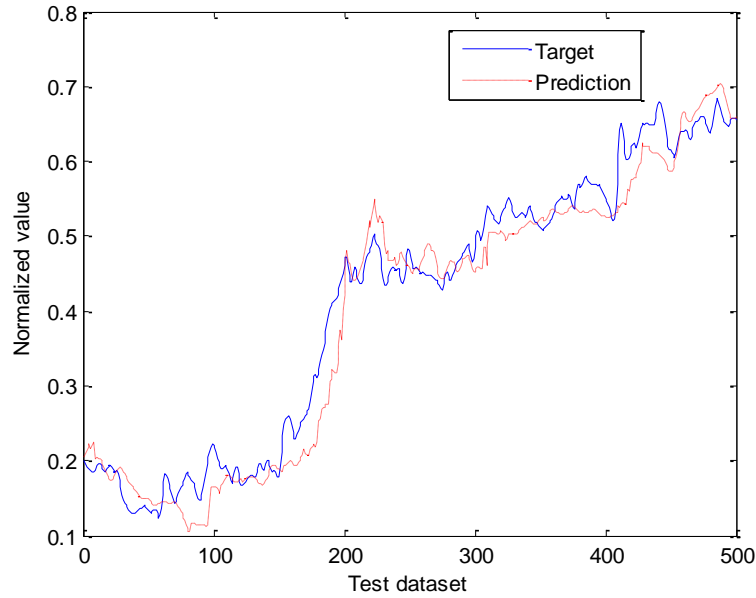


Figure 4 Experiment results for one-week ahead prediction.

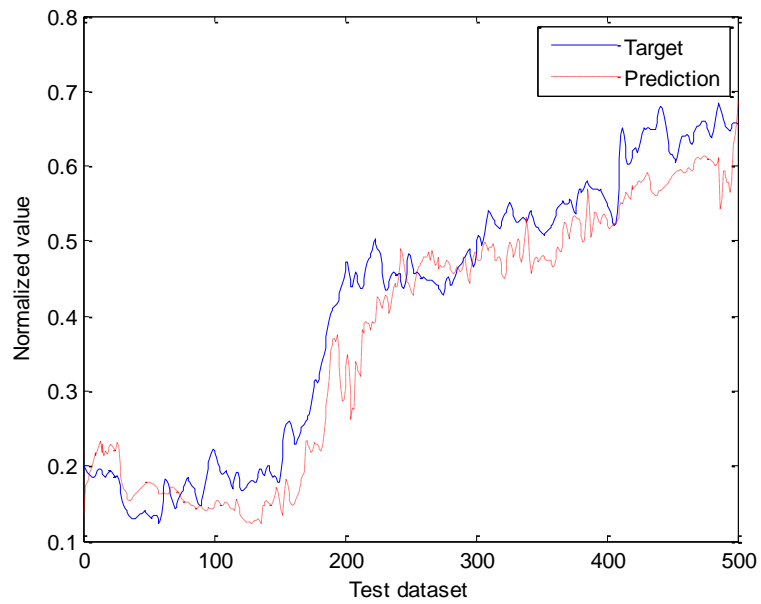


Figure 5 Experiment results for two-week ahead prediction.

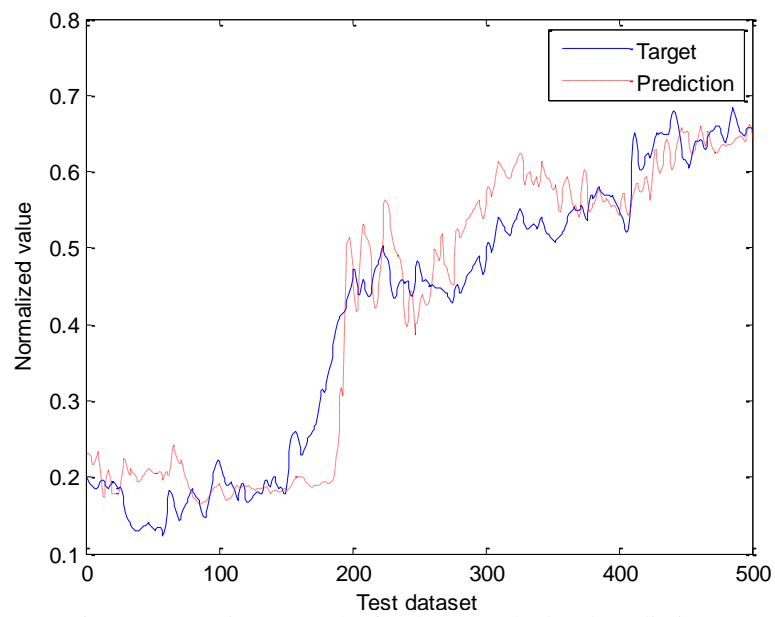


Figure 6 Experiment results for three-week ahead prediction.

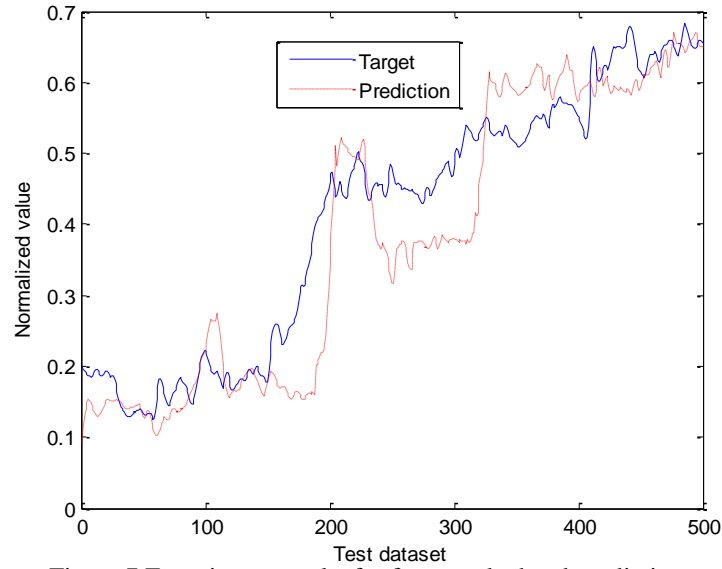


Figure 7 Experiment results for four-week ahead prediction.

From the Figures, we can see that in all cases Online-SVR-FID can adaptively and efficiently follow the change of the leakage with a reasonable delay around the 150th data point, because the model can react to the changes only with enough data. The delay is longer for a longer prediction horizon, i.e. the longer prediction horizon makes the model less flexible, because of the fact that the predictions of the data points before the pattern drifts detection cannot be changed.

We can also find that the prediction results on the stable parts (1-150 and 350-500 test data points) are similar, and the difference of the results for different prediction horizons lies on the parts where the developing trends of the leakage change, i.e. around the (test) data points at 150 and 250 time steps.

Table 1 Comparisons of the experiment results.

	One-week	Two-week	Three-week	Four-week
MSE	0.0014	0.0031	0.0028	0.0053
MARE	0.0902	1.362	0.1339	0.1512

Comparisons of the prediction results are reported in Table 1, considering Mean Squared Error (MSE) and Mean Absolute Relative Error (MARE). The comparisons show that the prediction results are better for the shorter prediction horizon, as expected, because the model can capture the change of leakage and adaptively update the model faster. Another reason for the worse results with long prediction horizon is that the correlation between the historical values in the input vectors and the outputs are weaker than the case with short prediction horizon.

5. Conclusions

In modern industry, maintenance is more and more costly and important. In NPP, the maintenance cost can take up to two thirds of the total operation cost. The failure of NPP can cause enormous economic and environmental losses. PHM framework provides important information on the future state of the SSCs of NPP. Proper maintenance can be planned to prevent or postpone failure.

In this paper, the previous work of Online-SVR-FID is extended to the long-term (weeks-ahead) prediction of the leakage in the first seal of RCP in a NPP. Experimental results show that Online-SVR-FID can efficiently capture the change in the time series data, while we should also notice that the prediction results are worse for longer prediction horizon. One reason is that the model reacts with an obviously longer delay, and the other reason is that the correlation between the inputs and output is weaker.

In the future work, more variables should be considered in the model as inputs to form a strong relation between the inputs and output for long-term prediction. As the short-term prediction is more flexible and the long-term prediction is more profitable, a natural question may rise for finding the proper strategies for the combination of short-term and long-term predictions.

References

- [1] N. E. Insider, "Nuclear new construction market map 2014," 2014.
- [2] J. Coble, P. Ramuhalli, L. Bond, J. W. Hines, and B. Upadhyaya, "A review of prognostics and health management applications in nuclear power plants," *International Journal of prognostics and health management*, submitted, 2015.
- [3] J. Liu, V. Vitelli, R. Seraoui, and E. Zio, "Nuclear power plant components condition monitoring by probabilistic support vector machine," *Annals of Nuclear Energy*, vol. 56, pp. 23-33, 2013.
- [4] J. Liu, "Failure Prognostics by Support Vector Regression of Time Series Data under Stationary / Nonstationary Environmental and Operational Conditions," *Dissertation of Ph. D., Ecole Centrale Paris*, 2015.
- [5] E. Zio, and F. Di Maio, "A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system," *Reliability Engineering & System Safety*, vol. 95, no. 1, pp. 49-57, 2010.
- [6] E. Zio, F. Di Maio, and M. Stasi, "A data-driven approach for predicting failure scenarios in nuclear systems," *Annals of Nuclear Energy*, vol. 37, no. 4, pp. 482-491, 2010.
- [7] M. G. Na, J. W. Kim, and I. J. Hwang, "Collapse moment estimation by support vector machines for wall-thinned pipe bends and elbows," *Nuclear engineering and design*, vol. 237, no. 5, pp. 451-459, 2007.

- [8] H. Y. Yang, M. G. Na, and J. W. Kim, "Fuzzy support vector regression model for the calculation of the collapse moment for wall-thinned pipes," *Nuclear Engineering and Technology*, vol. 40, no. 7, pp. 607-614, 2008.
- [9] Y. Lahssuny, and W. Jędral, "Universal correlations for predicting complete pump performance characteristics," *Journal of Power Technologies*, vol. 90, 2004.
- [10] M. A. Kon, and L. Plaskota, "Information complexity of neural networks," *Neural Networks*, vol. 13, no. 3, pp. 365-375, 2000.
- [11] M. S. Roh, S. W. Cheon, and S. H. Chang, "Power prediction in nuclear power plants using a back-propagation learning neural network," *Nuclear technology*, vol. 94, no. 2, pp. 270-278, 1991.
- [12] M. G. Na, S. H. Shin, S. M. Lee, D. W. Jung, S. P. Kim, J. H. Jeong, and B. C. Lee, "Prediction of major transient scenarios for severe accidents of nuclear power plants," *IEEE Transactions on Nuclear Science*, vol. 51, no. 2, pp. 313-321, 2004.
- [13] B. R. Upadhyaya, and E. Eryurek, "Application of neural networks for sensor validation and plant monitoring," *Nuclear Technology*, vol. 97, no. 2, pp. 170-176, 1992.
- [14] W. J. Kim, H. C. Soon, and B. H. Lee, "Application of neural networks to signal prediction in nuclear power plant," *IEEE Transactions on Nuclear Science*, vol. 40, no. 5, pp. 1337-1341, 1993.
- [15] F. A. Simonen, and S. R. Gosselin, "Life prediction and monitoring of nuclear power plant components for service-related degradation," *Journal of pressure vessel technology*, vol. 123, no. 1, pp. 58-64, 2001.
- [16] V. Vapnik, *The nature of statistical learning theory*, Springer, 2000.
- [17] NEI, "Nuclear energy's economic benefits – current and future (white paper)," Nuclear Energy Institute, Washington, D.C., 2011.
- [18] K. P. Bennett, and O. L. Mangasarian, "Robust linear programming discrimination of two linearly inseparable sets," *Optimization methods and software*, vol. 1, no. 1, pp. 23-34, 1992.
- [19] C. Cortes, and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273-297, 1995.
- [20] C. J. Lin, and R. C. Weng, "Simple probabilistic predictions for support vector regression," National Taiwan University, Taipei, 2004.
- [21] G. Baudat, and F. Anouar, "Feature vector selection and projection using kernels," *Neurocomputing*, vol. 55, no. 1-2, pp. 21-38, 2003.