Probabilistic Support Vector Regression for Short-Term Prediction of Power Plants Equipment

Jie Liu, Redouane Seraoui, Valeria Vitelli, Enrico Zio

Probabilistic Support Vector Regression for Short-Term Prediction of Power Plants Equipment

Jie Liu[a], Redouane Seraoui[b], Valeria Vitelli[a], Enrico Zio* [a c]

[a]Chair on Systems Science and the Energetic challenge, European Foundation for New Energy - Électricité de France, École Centrale Paris, Grande Voie des Vignes, F-92295, Chatenay-Malabry, France, and Supelec (École Supérieure d'Électricité), Plateau Moulon, 3 Rue Joliot Curie, F-91190, Gif-sur-Yvette, France
[b]EDF R&D, Simulation and information TEchnologies for Power generation System (STEPS) Department, 6 quai Waiter, F-78401, Chatou, France
[c]Energy Department, Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133, Milano, Italy
enrico.zio@ecp.fr

A short-term forecasting approach is proposed for the purposes of condition monitoring. The proposed approach builds on the Probabilistic Support Vector Regression (PSVR) method. The tuning of the PSVR hyerparameters, the model identification and the uncertainty analysis are conducted via novel and innovative strategies. A case study is shown, regarding the prediction of a drifting process parameter of a Nuclear Power Plant (NPP) component.

1. Introduction

In complex engineering systems, methods are sought to anticipate, diagnose and control abnormal events in a timely manner, to prevent high economic losses in case of unexpected faults (Venkatasubramanian, 2005). Condition monitoring of components and systems is usually performed at regular intervals, in order to effectively identify faults possibly affecting the system state. The prediction of the failure development is also of interest for the purposes of maintenance, and for informing decisions on the actions to take to recover the system (Zio, 2012).
Both model-based and data-driven approaches can be adopted for forecasting the evolution of a failure (Zio, 2012). Considering the difficulty of building effective physical models, and given that most components and systems are monitored throughout their life cycle, data-driven approaches are often the most suited in the context of realistic applications.
Several research works have concerned data-driven approaches for condition monitoring of engineering systems. Artificial Neural Network (ANN) (Zio and Gola, 2012), Support Vector Machine (SVM) (Liu et al., 2013), Genetic Algorithm (GA) (Baradi et al., 2011) and Auto-Associative Kernel Regression (AAKR) (Yang et al., 2006) are among the most studied and applied.
In particular, SVMs are learning machines implementing the Structural Risk Minimization (SRM) inductive principle to obtain good generalization performance on a limited number of learning patterns (Vapnik et al., 1996). It has been shown that SVMs can solve nonlinear problems after mapping the input data into a high dimensional Reproducing Kernel Hilbert Space (RKHS) (Vapnik et al., 1996). Despite the accuracy and effectiveness of SVMs, difficulties can arise in the tuning of the parameters for the specific problem at hand. Moreover, SVMs provide classification and regression results only as point estimates, while a quantification of the uncertainty associated to the prediction is crucial in most applications.
To overcome these limitations, the Bayesian probabilistic paradigm has been introduced within the SVM framework, providing approaches that can estimate the parameter and feature spaces simultaneously and also obtain a predictive distribution. These methods are called Probabilistic Support Vector Regression (PSVR). A modified Gaussian-based PSVR approach is proposed in this paper for short-term prediction with uncertainty quantification. This approach is developed and used to provide the predicted value of a parameter of interest, describing the component condition, along with the associated Prediction Intervals (PIs). A real case study is shown, regarding the prediction of a drifting process parameter of a Nuclear Power Plant (NPP) component.
The paper is structured as follows. Section 2 resumes the PSVR method for prognostics. Section 3 presents the case study with the necessary pre-processing steps, used to prepare the data and identify the prediction model. In Section 4, the results of the application of PSVR for short-term forecasting of a drifting parameter in a NPP component are presented. Some conclusions are drawn in Section 5.

2. Probabilistic Support Vector Regression (PSVR)

Different Gaussian-based versions of the PSVR method can be defined, depending on the choice of the loss function. A large number of loss functions have been proposed in the literature, e.g. Quadratic Loss Function, Laplacian Loss Function, Huber's Loss Function (Chu et al., 2002). In the proposed modified PSVR approach, the authors have chosen the Ɛ-Insensitive Loss Function, which enables a sparse set of support vectors to be obtained (Smola and Schölkopf, 2004).

2.1 PSVR with Ɛ-Insensitive Loss Function

Let us assume that the input data is a n-dimensional set of vectors $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ independently drawn in $\mathbf{R}^p$ and that we also have an independent sample of target values $\mathbf{Y} = \{y_1, y_2, \ldots, y_n\}$, where $y_i \in \mathbf{R}$, $i = 1, 2, \ldots, n$. In regression methods, the final aim is to find a function $a(\mathbf{x})$: $\mathbf{R}^p \rightarrow \mathbf{R}$ describing the relation between the input data and the target.

We hereafter briefly recall the PSVR approach to the estimation of $a(\mathbf{x})$; further mathematical details on the derivation of the method can be found in Gao et al. (2001).

We make the following assumptions:

1. The training data $\mathbf{\Gamma} = \{\mathbf{X}, \mathbf{Y}\}$ is composed by independent samples, which given $a(\mathbf{x})$, are drawn from the same probability distribution.

2. The *a priori* probability distribution of the unknown $a(\mathbf{x})$ is $P[\mathbf{a}(\mathbf{X})] \propto \exp(-\frac{1}{2}\|\hat{P}a\|^2)$, where $\|\hat{P}a\|^2$ is a positive semi-definite operator and $\mathbf{a(X)} = (a(\mathbf{x}_1), a(\mathbf{x}_2), \ldots, a(\mathbf{x}_n))^T$.

3. The Ɛ-insensitive loss function is chosen as loss function of the PSVR method (see Gao et al., 2001).

$$L(\mathbf{x}) = \begin{cases} |\mathbf{x}| - \varepsilon, & |\mathbf{x}| < \varepsilon \\ 0, & |\mathbf{x}| \geq \varepsilon \end{cases} \quad (1)$$

4. The covariance function is $K(\mathbf{X}, \mathbf{X})$, where the element K_{ij} of the i-th row and j-th column is $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{|\mathbf{x}_i - \mathbf{x}_j|^2}{2\gamma^2}\right)$, with $\mathbf{x}_i$, $\mathbf{x}_j$ are the input data points in $\mathbf{X}$, and $i,j = 1, 2, \ldots, n$.

The a posteriori probability of $\mathbf{a(X)}$ can be written as

$$P[\mathbf{a}(\mathbf{X})|\mathbf{\Gamma}] = \frac{[G(C,\ \varepsilon)]^N}{\sqrt{def2\pi K_{\mathbf{X},\mathbf{X}} P[\mathbf{\Gamma}]}} \exp\{-C\sum_{\mathbf{x}_i \in \mathbf{X}} L_\varepsilon\left(y_i - a(\mathbf{x}_i)\right) - \frac{1}{2}\mathbf{a}(\mathbf{X})^T K^{-1}_{\mathbf{X},\mathbf{X}} \mathbf{a}(\mathbf{X})\}, \quad (2)$$

where $G(C, \varepsilon) = \frac{1}{2}\frac{C}{C\varepsilon + 1}$, and $K_{\mathbf{X},\mathbf{X}}$ is a shortened notation for the covariance matrix $K(\mathbf{x}_i, \mathbf{x}_j)$.

The Maximum A Posteriori (MAP) solution of Eq. (2) is found by finding the minimum of the following function

$$R_{GSVM}(a) = C\sum_{\mathbf{x}_i \in \mathbf{X}} L_\varepsilon\left(y_i - a(\mathbf{x}_i)\right) + \frac{1}{2}\mathbf{a}(\mathbf{X})^T K^{-1}_{\mathbf{X},\mathbf{X}} \mathbf{a}(\mathbf{X}) \quad (3)$$

Following the discussion in Mackay (1997), we can write the solution of the minimization problem associated to Eq. (3) in the following form

$$a^*(\mathbf{x}) = \sum_{\mathbf{x}_i \in \mathbf{X}} \beta_i K(\mathbf{x}_i, \mathbf{x}) \quad (4)$$

where $\beta_i = a_i - a_i^*$ is a combination of the Lagrange Multipliers associated to the optimization problem (Smola and Scholköpf, 2004). The a_i and a_i^* can be determined by a Quadratic Programming approach. According to Smola and Scholköpf (2004), $\forall i = 1, \ldots, n$, a_i and a_i^* lie in the interval $[0, C]$, and β_i consequently lies in the interval $[-C, C]$, which is the domain of the optimization problem.

2.2 Hyperparameters

The three hyperparameters C, ε, γ need to be determined.

Parameter C is the penalty factor. It controls the trade-off between complexity and the proportion of non-separable samples, and must be selected by the user. If it is too large, it will induce a high penalty for non-separable points, hence we may store too many support vectors and go towards over fitting. If it is too small, it may result in underfitting (Alpaydin, 2004). For what concerns the optimization process, C influences the computational burden of the regression: the bigger C is, the heavier the computational burden is.

Parameter ε controls the sparsity of the data. It has an effect on the smoothness of the SVM response and it affects the number of support vectors; so, both the complexity and the generalization power of the network depend on its value. By inspecting the ε-insensitive loss function in Eq. (1), we see that data points inside a tube of radius ε surrounding the predicted values are not considered in training the regression model.

Finally, parameter γ influences the width of the kernel, and hence the accuracy of the prediction and its variability.
Methods exist in the literature to determine these hyperparameters; among others, we recall VC-theory (Vapnik, 1995), Bayesian approaches (Mackay, 1991), Akaike's Information Criterion (Akaike, 1974), Network Information Criterion (Murata et al., 1994) and Maximizing Evidence Function (Kim et al., 2012). For our work, we adopt a novel interpolation method to obtain the best values of the three hyperparameters (C, ε, γ). For further details, see Liu et al., 2013.

2.3 Error Bar Estimation

In the Bayesian treatment of the prediction problem, error bars arise naturally from the predictive distribution. They are made up of two terms, one due to the a posteriori uncertainty (the uncertainty of $a(\mathbf{x})$) and the other due to the intrinsic noise in the data (Kim et al., 2012). Suppose that $\mathbf{x}$ is a test input vector, and that the corresponding value of the target is the random variable y, obtained adding to $a(\mathbf{x})$ an unknown noise δ with zero mean; then,

$$P[\boldsymbol{\Gamma}|\mathbf{a}(\mathbf{X})] \propto \exp(-C\sum_{i=1}^{n} L(\delta_i)). \quad (5)$$

We can also obtain the density of the noise δ

$$P[\delta] = \frac{C}{2(C\varepsilon+1)} \exp(-CL_{\varepsilon}(\delta)), \quad (6)$$

and the noise variance

$$\sigma_{\delta}^2 = \frac{2}{C^2} + \frac{\varepsilon^2(C\varepsilon+3)}{3(C\varepsilon+1)} \quad (7)$$

The conditional probability distribution of $a(\mathbf{x})$ given $\boldsymbol{\Gamma}$, can instead be written as

$$P[a(\mathbf{x})|\boldsymbol{\Gamma}] = \frac{1}{\sqrt{2\pi}\sigma_t} \exp\{-\frac{(a(\mathbf{x}) - a^*(\mathbf{x}))^2}{2\sigma_t^2}\}, \quad (8)$$

with

$$\sigma_t^2(\mathbf{x}) = K(\mathbf{x},\mathbf{x}) - K_{\mathbf{x}_M,\mathbf{x}}^T K_{\mathbf{x}_M,\mathbf{x}_M}^{-1} K_{\mathbf{x}_M,\mathbf{x}}. \quad (9)$$

Consequently, the error bar width of the prediction corresponding to the test input point $\mathbf{x}$ is

$$\sigma^2(\mathbf{x}) = \sigma_{\delta}^2 + \sigma_t^2(\mathbf{x}) = \frac{2}{C^2} + \frac{\varepsilon^2(C\varepsilon+3)}{3(C\varepsilon+1)} + K(\mathbf{x},\mathbf{x}) - K_{\mathbf{x}_M,\mathbf{x}}^T K_{\mathbf{x}_M,\mathbf{x}_M}^{-1} K_{\mathbf{x}_M,\mathbf{x}}. \quad (10)$$

The conditional probability distribution and the error bar are given in Eq. (8) and (10). See Gao et al. (2001) for more details on the calculations.

3. Case Study

A set of data from the Reactor Coolant Pump (RCP) of one of EDF's NPPs is used to test the efficiency and the accuracy of the PSVR modelling approach developed in our work. In the following, we describe the data and illustrate the pre-processing steps.

3.1 Data Description

The dataset concerns one scenario of increasing leak flow in the first seal of the RCP of a NPP (a variable denoted with IntVar 9). The dataset contains the values of seventeen different variables recorded by seventeen different sensors along a period of 406 days. The variables whose measurements concern sensors inside the RCP are hereafter called internal variables; the others are called external variables. The description of all the internal and external variables and their physical meanings are given in Table 1.
We note that: all the variables are time-dependent, and there are seventeen variables in total, hence leading to a multivariate problem; each variable is measured hourly, giving 9200 measurements for each variable, and hence making computations challenging; all the variables show a nonlinear behaviour, hence requiring a nonlinear model; the data need pre-processing, because there are many outliers and missing observations.

Table1: Physical meaning of each internal and external variable

Internal variables		External variables	
Name	Physical meaning	Name	Physical meaning
IntVar 1	T cold leg loop 1 [WR]	ExtVar 1	T by-pass hot leg loop 3
IntVar 2	T water seal #1 051PO	ExtVar 2	T seal injection line
IntVar 3	T stator winding motor 051PO	ExtVar 3	P primary amount file B [GL]
IntVar 4	T motor lower bearing 051PO	ExtVar 4	Debit general file A
IntVar 5	T lower thrust bearing 051PO	ExtVar 5	Debit general file B
IntVar 6	T motor upper bearing 051PO	ExtVar 6	T aval exchanges file A
IntVar 7	T motor upper thrust bearing 051PO	ExtVar 7	T aval exchanges file B
IntVar 8	Flow seal injection supply RCP051PO	ExtVar 8	Debit refrigeration GMPP 051PO
IntVar 9	Seal leak flow #1 RCP051PO		

3.2 Data Pre-processing

Outliers can be easily selected by deciding some constraints, and then by identifying those data points which lie outside the constraints. The data selected as outliers are then removed.

A possible way to deal with the reconstruction of missing data is the local polynomial regression fitting (Masry and Mielniczuk, 1999). It estimates the values of the internal and external variables when there are missing data points. For details on the reconstruction procedure, see Liu et al. (2013).

3.3 Model Identification

In order to select the most proper variables to be included as inputs in the PSVR model for improved prediction accuracy and reduction of the computational burden, a correlation analysis is carried out between the target variable IntVar 9 and the other internal and external variables. The inputs are chosen to be the variables maximizing their correlations with the target IntVar 9. Correlations are measured by the classical Pearson correlation coefficient (Rodgers and Nicewander, 1988). According to the correlations analysis, three external variables are the most related to the target and six internal variables have also a strong correlation with the target, with a correlation of more than 48 %. Hence, these six most related internal variables and the three most related external variables are included as inputs in the prediction model. IntVar 8 is also chosen as input, as suggested by expert judgment. Historical values of the target can also be exploited as inputs to improve the accuracy of the prediction. The number of previous values of the target to be used in the model is set equal to 3 after carrying out an autocorrelation analysis on the time series of the target values. The results are given in the next Section.

4. Case Study Results

We consider the scenario of increasing leak flow in the first seal of the RCP. The prediction target is the variable IntVar 9, and we focus on short-term (1-hour ahead) prediction. At time t, we want to predict the target value at time $t + 1$; for this, we use as inputs the historical values of the target itself till three time steps before t, the values of the three most correlated external variables at time t and the values of the six most correlated internal variables at time t. We select a portion of the scenario to build a PSVR model for prediction. A size of 200 data points (5600-5800 observations) are used for training the model; another set, form 5601 to 8000 observations, is used for testing. The hyperparameters selection has been carried out with the novel interpolation method of Liu et al. (2013): $(C, \varepsilon, \gamma) = (6309.6, 0.0032, 7)$.

The results are shown in Figure 1, where the prediction of the target with corresponding Prediction Interval (PI) of confidence level of 95 % is reported. The solid line is the target, the dot line is the point prediction, while the two dashed lines are the upper and lower bounds of the 95 % PI computed according to the predictive distribution. Hence, for each test point $\mathbf{x}$, the PI is $[a^*(\mathbf{x}) - 2\sigma(\mathbf{x}), a^*(\mathbf{x}) + 2\sigma(\mathbf{x})]$, where $a^*(\mathbf{x})$ is the predicted value according to Eq. (4) and $\sigma(\mathbf{x})$ is the variance associated to the prediction (error bar) given by Eq. (10). We remark that the predictive distribution of $\mathbf{x}$ is a Gaussian with mean $a^*(\mathbf{x})$ and variance $\sigma^2(\mathbf{x})$.

The prediction interval empirical coverage estimated on the whole testing set is 91.50 %. The MSE is $5.4332 \cdot 10^{-5}$. The relative error is smaller than 4 %. The results are satisfactory both from the methodological and engineering points of view.

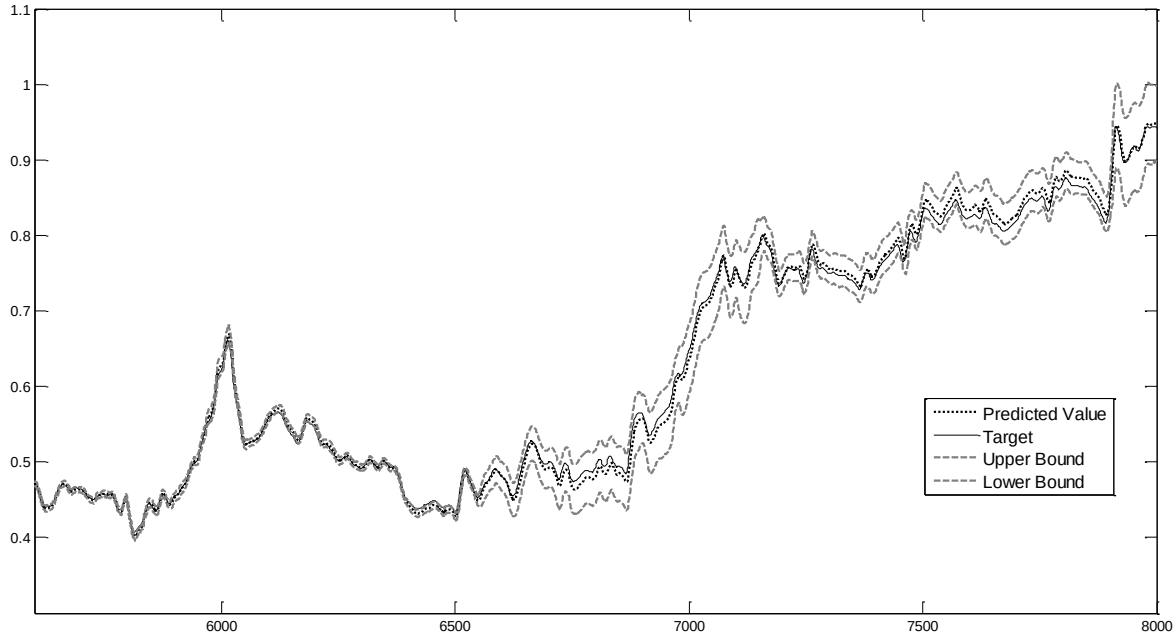

Figure 1: PSVR point predictions and PIs (both training and testing data)

5. Conclusion

In this paper, a short-term forecasting approach is proposed for the purposes of condition monitoring. It includes pre-processing for data reconstruction and model selection, and PSVR for estimation of the prediction interval and conditional predictive distribution of the target of interest. The results of the application to a real case study of leak flow in the first seal of a RCP are satisfactory. The coverage of the prediction interval is 91.50 % with a confidence level of 95 %.

The future work will focus on the development of a method to compute the Remaining Useful Life (RUL) on the basis of the prediction of the target parameter of interest. This entails propagating the uncertainties due to both the observed data and the model itself. Other failure scenarios, concerning the behavior of the same component in other plants, will also be included in the model. This may require the development of an ensemble of models. Finally, a model-based approach to prediction will also be inspected, for comparison.

References

Akaike H., 1974, A new look at the statistical model identification, IEEE Trans. Auto. Control. 19(6), 716–723

Alpaydin E., 2004, Introduction to Machine Learning, MIT Press. Cambridge, Massashusetss, London, England

Baraldi P., Canesi R., Zio E., Seraoui R. and Chevalier R., 2011, Genetic algorithm-based wrapper approach for grouping condition monitoring signals of nuclear power plant components, Integr. Comput. Eng. 18(3), 221-234

Chu W., Keerthi S. and Ong C. J., 2002, Bayesian Support Vector Regression Using a Unified Loss Function, Journal of Latex Class. Files, 1 (11), 1-14

Gao J.B., Gunn S.R., Harris C.J. and Brown M., 2001, A Probabilistic Framework for SVM Regression and Error Bar Estimation, Mach. Learn. 46(1-3), 71-89

Kim D.S., Kim J.H., Gyunna M.and Kim J.W., 2012, Uncertainty analysis of data-based models for estimating collapse moments of wall-thinned pipe bends and elbows, Nucl. Eng. & Tech. 44(3), 323-330

Liu J., Seraoui R., Vitelli V., Zio E., 2013, Nuclear Power Plant Components Condition Monitoring by Probabilistic Support Vector Machine, Ann. Nucl. Energy, in press, DOI: 10.1016/j.anucene.2013.01.005

MacKay D.J., 1991, Bayesian Modelling and Neural Networks, PhD thesis, California Institute of Technology, Pasadena, CA

MacKay D.J., 1997, Gaussian processes, a replacement for neural networks, NIPS tutorial 1997, Cambridge University

Masry E. and Mielniczuk J., 1999, Local linear regression estimation for time series with long-range dependence, Stoch. Process. & Appl. 82 (2), 173–193

Murata N., Yoshizawa S. and Amari S., 1994, Network information criterion–determining the number of hidden units for artificial neural network models, IEEE Trans. Netw. 5, 865–872

Parry G.W., 1996, The characterization of uncertainty in Probabilistic Risk Assessments of complex systems, Reliab. Eng. Syst. Saf. 54(2-3), 119-126

Rodgers J.L. and Nicewander W.A., 1988, Thirteen ways to look at the correlation coefficient, Am. Stat. 42 (1), 59-66

Smola A.J. and Schölkopf B., 2004, A tutorial on support vector regression, Stat. & Comp. 14(3), 199-222

Venkatasubramanian V., 2005, Prognostic and Diagnostic Monitoring of Complex Systems for Product Lifecycle Management: Challenges and opportunities, Comput. Chem. Eng. 29(6), 1253-1263

Vapnik V.N., 1995, The Nature of Statistical Learning Theory, Springer. New York

Vapnik V. N., Golowich S. E. and Smola A., 1996, Support vector method for function approximation, regression estimation, and signal processing, Proceedings of the 10th Neural Information Processing Systems (NIPS) Conference, Denver, Colorado

Yang D., Usynin A., Hines J.W., 2006, Anomaly-Based Instruction Detection for SCADA Systems, 5th Intl. Topical Meeting on Nuclear Plant Instrumentation, Control and Human Machine Interface Technologies (NPIC & HMIT 05), Albuquerque, NM, Nov 12-16

Zio E., 2012, Diagnostics and Prognostics of Engineering Systems: Methods and Techniques, Chapter 17, Engineering Science Reference, USA

Zio E., Gola G., 2006, Neuro-fuzzy pattern classification for fault diagnosis in nuclear components, Ann. Nucl. Energy, 33, 415-426