A Dynamic Weighted RBF-Based Ensemble for Prediction of Time Series Data from Nuclear Components

Jie Liu[1], Valeria Vitelli[2], Enrico Zio[3] and Redouane Seraoui[4]

[1]*Chair on Systems Science and the Energetic Challenge, EDF Foundation, CentraleSupélec, Paris, France*
jie.liu@ecp.fr

[2]*Department of Biostatistics, University of Oslo, Oslo, Norway*
valeria.vitelli@medisin.uio.no

[3] *Chair on System Science and the Energetic Challenge, EDF Foundation, CentraleSupélec, Paris, France. Energy Departement, Politecnico di Milano, Milano, Italy*
enrico.zio@ecp.fr

[4] *EDF R&D/STEP Simulation et Traitement de l'information pour l'exploitation des systemes de production, Chatou, France*
redouane.seraoui@edf.fr

ABSTRACT

In this paper, an ensemble approach is proposed for prediction of time series data based on a Support Vector Regression (SVR) algorithm with RBF loss function. We propose a strategy to build diverse sub-models of the ensemble based on the Feature Vector Selection (FVS) method of Baudat & Anouar (2003), which decreases the computational burden and keeps the generalization performance of the model. A simple but effective strategy is used to calculate the weights of each data point for different sub-models built with RBF-SVR. A real case study on a nuclear power production component is presented. Comparisons with results given by the best single SVR model and a fixed-weights ensemble prove the robustness and accuracy of the proposed ensemble approach.

1. INTRODUCTION

Combining various data-driven approaches into an ensemble has become a popular direction of research in the last decades, motivated by the aim of improving the robustness and accuracy of the final prediction. The models which compose the ensemble are called sub-models. Various strategies have been proposed for building sub-models, including error-correcting output coding, Bagging, Adaboost, and Boosting (Kim, Pang, Je, Kim & Bang, 2003; Hu, Youn, Wang & Yoon, 2012). Similarly, several

methods for aggregating the prediction results of the sub-models have been proposed, such as majority vote, weighted vote, Borda count, Bayes and probabilistic schemes, etc (Polikar, 2006).

Support Vector Machine (SVM) is a popular and promising data-driven method for prognostics. SVM-based ensemble models have been proposed for classification. Chen, Wang and Zuylen (2009) use ensemble of SVMs to detect traffic incidents. The sub-models use different kernel functions and parameters, and their outputs are combined to improve the classification performance. Acar and Rais-Rohami (2009) treat the general weighted-sum formulation of an ensemble as an optimization problem and, then, minimize an error metric to select the best weights for the sub-models of SVM. Kurram and Kwon (2013) try to achieve an optimal sparse combination of the sub-model results by jointly optimizing the separating hyperplane obtained by each SVM classifier and the corresponding weights of the sub-decisions. Valentini and Dietterich (2003) prove that an ensemble of SVMs employing bagging of low-bias algorithms improves the generalization power of the procedure with respect to single SVM. The ensemble of SVMs built with bagging and boosting can greatly outperform a single SVM in terms of classification accuracy (Kim *et al.*, 2003).

In this paper, we focus on the combination of multiple SVR sub-models (Liu, Seraoui, Vitelli & Zio, 2012) with Radial Basis loss Function (RBF). The case study considered to present the application of the method concerns the monitoring of the leak flow in the first seal of the Reactor

Coolant Pump (RCP) of a Nuclear Power Plant (NPP), using real data collected from sensors.

The high complexity of the NPP system and the catastrophic economic and environmental loss under accidents make the monitoring for failure prediction very important in NPP accident management (Lee, 1998; Hallbert and Thomas, 2014). RCP pumps the coolant into the reactor to transfer the heat to steam generator and to protect the nuclear material (In Soo and Kim, 2000). With large amount of leakage from RCP, NPP has the risk of melting down. Thus, it is critical to predict the leakage in the future. The prediction can be divided into long-term (months) and short-term (hours) prediction. Long-term prediction provides future information for the maintenance scheduling, while short-term prediction helps for emergency actions.

An ensemble of SVRs with RBF and dynamic weighting strategy is proposed in this paper. The elements of novelty of the method here proposed are various.

In the previously cited literature on ensembles of SVMs, the weights of the sub-models in the ensemble are calculated during training and kept fixed for testing. However, a sub-model may perform well only on a part of the dataset. Hence, the weights need to be updated considering the different datasets involved in the case study, and even different input vectors. In Fantoni, Figedy and Racz (1998), a dynamic strategy is integrated into a Neuro-Fuzzy Model. A dynamic weighting method is also used in Muhlbaier, Topalis and Polikar (2009), Yang, Yuan and Liu (2009) and Razavi-Far, Baraldi and Zio (2012), for adding new classifiers to the ensemble model, but the weights are not adjusted to the different input vectors.

A novel dynamic weighting strategy, based on local fitness calculation (Baudat & Anouar, 2003) is proposed in this paper.

To generate diversity in the sub-models, each of them is trained on a different dataset. In this respect, one can use bagging and boosting with possible overlapping between different datasets (Quinlan, 1996). In this work, the strategy to form the training dataset of each sub-model is based on the angle between different data points in the Reproducing Kernel Hilbert Space (RKHS), so as to reduce the computational burden.

Moreover, in order to be able to build ensembles of SVRs on very large datasets, FVS is used to select a smaller subset of the training data points of each sub-model, again to decrease the computational burden.

All the above novel strategies are tested on the case study concerning the prediction of leak flow of the RCP in a NPP.

The rest of the paper is organized as follows. Section 2 gives details about the proposed ensemble approach. Section 3 illustrates the case study, the available data and how the proposed ensemble model is constructed. Section 4 presents the experimental results from the SVR ensemble models and describes the comparison with a single SVR model and a fixed weighted ensemble. Finally, conclusions with some considerations are drawn in Section 5.

2. DYNAMIC-WEIGHTED RBF-BASED ENSEMBLE

The underlying strategy motivating the use of ensemble-based methods in prediction problems is to benefit from the strength of different sub-models by combining their outputs to improve the global prediction performance, if compared to the results of a single sub-model.

In this section, we give details about the proposed Dynamic-Weighted RBF-based Ensemble (named DW-RBF-Ensemble, in short).

2.1. Standard Support Vector Regression with RBF and ε-sensitive loss function

Suppose a set of training data points (x_i, y_i), for $i = 1, 2, \dots, T$ is available. The construction of an SVR model amounts to finding the best estimate function $f(x) = \omega x + b$ of the real underlying function. To this aim, the primal quadratic optimization problem is

$$\text{Minimize } \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{T}(\xi_i + \xi_i^*)$$

$$\text{Subject to } \begin{cases} y_i - \omega x_i - b \leq \varepsilon + \xi_i \\ \omega x_i + b - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (1)$$

where ξ_i and ξ_i^* are slack variables. The dual formulation of Eq. (1) is

$$L = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{T}(\xi_i + \xi_i^*) - \sum_{i=1}^{T}\alpha_i(\varepsilon + \xi_i - y_i + \omega x_i + b) - \sum_{i=1}^{T}\alpha_i^*(\varepsilon + \xi_i^* + y_i - \omega x_i - b) - \sum_{i=1}^{T}(\eta_i\xi_i + \eta_i^*\xi_i^*), \quad (2)$$

where $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$ are the Lagrange multipliers. By calculating the partial derivative of L with respect to the primal variable ω, the best estimate function can be written as

$$f(x) = \sum_{i=1}^{T}(\alpha_i - \alpha_i^*) * k(x_i, x) + b. \quad (3)$$

The values of α_i and α_i^* can be calculated by solving the Kuhn-Tucker conditions related to Eq. (2)

The kernel function $k(x_i, x)$ in Eq. (3) enables the mapping of an input vector in a higher-dimensional RKHS. By calculating pairwise inner products between mapped samples, the kernel functions return the similarity between different samples. In fact, only kernels that fulfill Mercer's Theorem (i.e. the kernel matrix must be positive semi-definite) are valid ones and, thus, can be used in SVM (Minh, Niyogi and Yang, 2006). The most common kernel functions include the linear kernel function, the polynomial kernel function and the RBF. In this paper, the ensemble approach is proposed to be built based on SVR with RBF.

2

For RBF, $k(x_i, x) = e^{-\frac{\|x_i - x\|^2}{2\sigma^2}}$ and a good property for RBF is that for each data point x, $k(x, x) = 1$, i.e., the data point in RKHS is a unit vector. The difference between different data points in RKHS is only the angle between them.

2.2. Feature Vector Selection

In Baudat and Anouar (2003), the authors propose a Feature Vector Selection (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.

Initialization:
    Training dataset: $T_r = \{(x_i, y_i)\}$, for $i = 1, 2, \ldots, T$
    Feature space: $S = [\ ]$
    Threshold of LF: τ
FVS:
    First FV in S:
      For $i = 1$ to T, calculate
        $S = \{x_i\}$, compute the global fitness J_S for all training data points with respect to
    the   present S.
      End for.
      Select the point which gives the maximum global fitness as the first FV and add it to S as the first FV.
      T_r is reduced to the complement of S in T_r, i.e. $T_r = T_r \backslash S$.
    Second and the other FVs:
      Calculate the local fitness for all data points in T_r with respect to the present feature space S.
      Select the data point k which gives the maximum of the local fitness:
        If $1 - J_{S,k} > \tau$, this point is a new FV and is added to S; and $T_r = T_r \backslash E$, with $E = \{(x_i, y_i) : 1 - J_{S,i} \leq \tau$ and $(x_k, y_k)\}$;
        If $1 - J_{S,k} \leq \tau$, end the process of FVS.

Fig. 1. Pseudo-code for FVS.

Suppose (x_i, y_i), for $i = 1, 2, \ldots, T$ are the training data points and the mapping $\varphi(x)$ maps each input vector x_i into RKHS with the mapping φ_i, for $i = 1, 2, \ldots, T$. The kernel $k_{i,j} = k(x_i, x_j)$ is the inner product between φ_i and φ_j. Suppose that the FVs selected from the training dataset are $\{x_1, x_2, \ldots, x_N\}$ and the corresponding mapping is $S = \{\varphi_1, \varphi_2, \ldots, \varphi_N\}$: the process for selecting the new next FV is to calculate $\{a_{new,1}, a_{new,2}, \ldots, a_{new,T}\}$ which gives the minimum of Eq. (4), with φ_{new} being the mapping of the new input vector x_{new}:

$$\delta_{new} = \frac{\left\| \varphi_{new} - \sum_{i=1}^{L} a_{new,i} \varphi_i \right\|^2}{\|\varphi_{new}\|^2}. \tag{4}$$

The minimum of δ_{new} can be expressed with an inner product, as shown in Eq. (5):

$$\min \delta_{new} = 1 - \frac{K_{S,new}^t K_{S,S}^{-1} K_{S,new}}{k_{new,new}}, \tag{5}$$

where $K_{S,S} = (k_{i,j}), i, j = 1, 2, \ldots, N$ is the kernel matrix of S and $K_{S,new} = (k_{i,N}), i = 1, 2, \ldots, N$ is the vector of the inner product between φ_{new}. The expression $J_{S,new} = \frac{K_{S,new}^t K_{S,S}^{-1} K_{S,new}}{k_{new,new}}$ is the local fitness of x_{new} with respect to the present feature space S. If $1 - J_{S,new}$ is zero, the new data point is not a new FV; otherwise, it is a new FV and is added to S. With the global fitness defined as in Eq. (6), the FVS procedure proceeds to select a subset of training data points with minimal size, which gives zero global fitness. The details for FVS is shown in Figure 1.

$$J_S = \sum_{i=1}^{T} J_{S,i} \tag{6}$$

2.3. Ensemble-Based Approach

An ensemble-based approach is obtained by training diverse sub-models and, then, combining their results following given strategies. It can be proven that this can lead to superior performance with respect to a single model approach (Bauer & Kohavi, 1999). A simple paradigm of a typical ensemble-based approach with N sub-models is shown in Figure 2. Ensemble models are built on three key components: a strategy to build diverse models; a strategy to construct accurate sub-models; a strategy to combine the outputs of the sub-models in a way such that the correct predictions are weighted more than the incorrect ones.

In the DW-RBF-Ensemble that we are proposing, the sub-models are built using a modified SVR model with RBF.
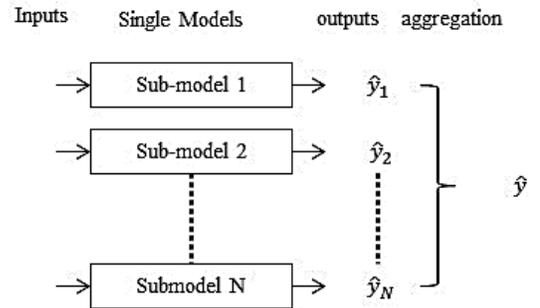
Fig. 2. Paradigm of a typical ensemble method.

A dynamic weighted-sum strategy is proposed to combine the outputs of the sub-models. As mentioned in the Introduction, different methods can be applied to calculate the weights for the sub-models. In the methods that can be found in the literature, the weights are normally fixed after

the ensemble model is built. They are only updated when new sub-models are added to the ensemble or when some sub-models are changed. In some real applications with fast changing environmental and operational conditions, the performance of the ensemble model may degrade rapidly. This degradation is not always caused by the low robustness or capability to adapt of the ensemble model, but can be due to the fact that the best sub-models are not given proper weights.

In this paper, a dynamic weighting strategy is thus proposed. The weights are no longer constant during the prediction, but dependent on the input vector. They are recalculated each time a new input vector arrives. Inspired by the work of Baudat and Anouar (2003) and considering the characteristics of SVR, a local fitness calculation is implemented in this paper to calculate the weights of the different sub-models for each input vector.

For training dataset $T = \{(x_i, y_i)\}, i = 1,2, \dots, T$, choose the number of clusters c. Initialize a random cluster centers vector $(v_1, v_2, \dots, v_c)$ chosen from the dataset T.
Repeat: for $l = 1,2, \dots$
  Compute the angle
    $D_{ik} = \arccos(k(x_k, v_i^{(l)})), 1 \le i \le c,$
$1 \le k \le T$. $k(\cdot, \cdot)$ is the RBF.　　(7)
  Select the points of minimal distances for a cluster, and they belong to that cluster. Suppose $\{(x_i, y_i)\}, i = 1,2, \dots, N_i$ are data points in cluster i.
  Calculate the cluster centers
    $v_i^{(l+1)*} = \frac{\sum_j^{N_i} x_j}{N_i}$.
  Choose the nearest data point to the calculated cluster center in the corresponding cluster to be the new cluster center.
    $D_{ik}^* = \arccos\left(k\left(x_k, v_i^{(l+1)*}\right)\right), 1 \le k \le$
N_i and $v_i^{(l+1)} = argmin_i(D_{ik}^*)$.
Until $|D^{(l+1)} - D^{(l)}| \le \tau, \tau$ is a small positive value and $D^{(l)} = \sum_{i=1}^{c} \sum_{k=1}^{T} D_{ik}$.

Fig. 3 Pseudo-code of angle-clustering algorithm.

2.3.1. Sub-datasets determination

Clustering methods are widely used in ensemble approaches for determining the sub-datasets for different sub-models.

In this paper, SVR models are trained with RBF. The difference between different data points in RKHS is only the angle between them, as the norm of all data points in RKHS is one. Thus, we can use the angular-clustering algorithm to divide the whole training dataset into several sub-datasets.

The pseudo-code is shown in Figure 3. As kernel function, RBF is the inner product of two vectors in RKHS and the angle between them can be expressed as Eq. (7) in the pseudo-code of Figure 3.

2.3.2. Train a RBF-SVR sub-model

With the angle-clustering method, the training dataset is divided into several clusters. But in the DW-RBF-Ensemble method, the data points in each cluster are not used directly to train a RBF-SVR. FVS is firstly used to select the FVs in each cluster and, then, the SVR model is trained on these selected FVs, in order to decrease the computational burden. The procedures for training a SVR model with FVs are not the same as shown in Sub-Section 2.1, as the estimate function in Eq. (2) is no longer a kernel expansion on all the training data points in one cluster, but only on the selected FVs.

Suppose that for the j-th cluster, the training data points are (x_i, y_i), for $i = 1, 2, \dots, T_j$ and the FVs selected by FVS are (x_i, y_i), for $i = 1, 2, \dots, N_j$; the estimate function of SVR for the i-th cluster is given in Eq. (8):

$$f(x) = \sum_{i=1}^{N_j}(\alpha_i - \alpha_i^*) * k(x_i, x) + b. \qquad (8)$$

In order to avoid the overfitting problem, the optimization still aims at finding the minimum of the objective function in Eq. (1) on all the training data points in the cluster. Thus, by replacing $\omega x_i + b$ in Eq. (2) with $\sum_{k=1}^{N_j}(\alpha_k - \alpha_k^*) * k(x_k, x_i) + b$, we can have the new, dual formulation of SVR. Classical methods can be used to estimate the unknowns in Eq. (8).

Such a process can efficiently decrease the risk of overfitting and guarantee the generalization performance of the sub-models.

2.3.3. Weights Calculation

In Section 2.2, FVS defines global and local criteria to characterize the feature space. The proposed local fitness can describe the linearity between the mapping of a new input vector and the mapping of all the Feature Vectors (FVs) of the model: if a linear combination of the mapping of the FVs can better approach the mapping of the new input vector, i.e. $1 - J_{S,new} \approx 0$ the model gives better approximation of the output of the new data point; otherwise, i.e. $1 - J_{S,new} \approx 1$, the model performs worse for this data point. Thus local fitness can be implemented to derive the weight of each sub-model for each input vector.

With Eq. (5), for a new coming data point at time t, we can calculate the local fitness $J_i(t)$ with respect to the FVs of the i-th sub-model. And the weight of the i-th sub-model for this data point is calculated as

$$\omega_i(t) = \frac{1/(1 - J_i(t) + \tau)}{\sum_{j=1}^{N} 1/(1 - J_j(t) + \tau)}, \qquad (9)$$

where τ is a very small value so that Eq. (9) works in the case $J_i(t) = 1$.

2.3.4. Combining Sub-Models Outputs

Figure 4 shows the paradigm of DW-RBF-Ensemble, where N is the number of sub-models, $\boldsymbol{x}(t)$ is a new input vector arriving at time t, $w_j(t)$ is the weight assigned to the j-th sub-model for the new input vector, $\hat{y}_j(t)$ is the predicted value for the j-th sub-model given by RBF-SVR and $\hat{y}(t)$ is the final output of the ensemble model.
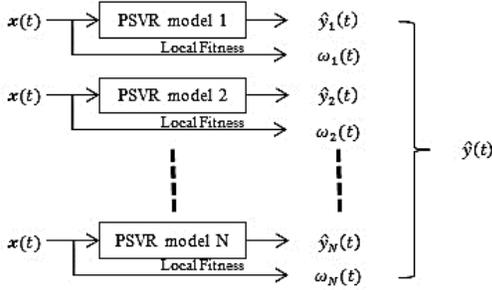
Fig. 4. Paradigm of the proposed DW-PSVR-Ensemble.

We can derive the fact that $\hat{y}(t) = \sum_{j=1}^{N} \omega_j(t)\hat{y}_j(t)$, if we assume sub-models results to be uncorrelated.

Note that all the sub-models weights and outputs are a function of t, which means that they are all dependent on the input vector of the ensemble model.

3. CASE STUDY DESCRIPTION

The study considered in this paper concerns the 1-day ahead prediction of leak flow from the first seal of the RCP of a NPP. RCP is a critical component in NPP, whose function is to circulate coolant into the reactor to transport the heat produced by nuclear fission to the steam generator. The leakage of coolant reduces such heat removal function, posing serious safety concerns. Short-term prediction can provide warning on a time horizon of hours and the lead time for deciding emergency actions. The time horizon of one day has been considered appropriate for NPP systems of interest, as indicated by the experts involved in this work.

In this section we describe the time series data and briefly recall the data pre-processing steps. We also detail the strategies to build the diverse sub-models of the ensemble.

3.1. Data Description and Pre-Processing

The data provided correspond to 9 scenarios of leak flow from different NPPs. Each scenario contains a time series data of the leak flow. They are named Scenario 1, Scenario 2, …, Scenario 9 in the following sections of the paper. These data are monitored every four hours. As these data are time-dependent and recorded within different time windows,

only scenarios coming from the same NPP have the same size. In some of the scenarios, there are missing data points and outliers.

Since the dataset we are going to analyze contains both missing data and outliers, we have to deal with both these issues. First of all, we must remove anomalous data, since their extreme values would affect the results of the analysis. Outliers can be detected with reference to some constraints, e.g. the limits $\bar{x} \pm 3 * \sigma_x$ where $\bar{x}$ is the mean of the data points values and σ_x is the standard deviation. These limits allow detecting the outliers, selected as those data points whose values are larger than $\bar{x} + 3 * \sigma_x$ or smaller than $\bar{x} - 3 * \sigma_x$, and subsequently removed. Some observations are in order with respect to the adopted procedure: i) In nonstationary time series, this outlier detection method should be carried out on local data and not on the whole scenario. We choose this method, not to delete any values which are possible indicators for changing conditions. ii) Given that the scenario is known, the strategy of outlier selection is chosen considering its overall development where there is no sudden changes, and the single values that are significantly outside the range of their neighbors are considered as outliers. iii) Note that we use those constraints, rather than the usual ones based on the median and the InterQuartile Range (IQR), to be more conservative in the outlier selection, due to the dependence among data (Brodsky, Lemmens, Brock-Utne, Vierra & Saidman, 2002).

Secondly, we want to reconstruct missing data. A possible way to deal with the reconstruction of missing data is local polynomial regression fitting (Masry, 1996). This local least squares regression technique estimates effectively the values of missing data points. Moreover, it can also be used to perform the smoothing of the available observations, in order to reduce noise. We will, thus, use this technique both to reconstruct data where missing, and to obtain a smoother and less noisy time series in all remaining time instances. All the time series data of all scenarios are, then, normalized from 0 to 1. All details on this pre-processing task can be found in Liu *et al.* (2012).

3.2. Strategies to Build Sub-Models

We have a time series dataset and we need to decide the best number of historical values to be used as inputs.

Suppose $a(t)$ represents an instance of the time series data of one scenario. For 1-day ahead prediction, the output $y(t)$ is $a(t + 6)$, because the signals are monitored every four hours. In order to decide the best H for selecting the input vector $\boldsymbol{x}(t) = (a(t - H + 1), \dots, a(t))$ most related to the output, a partial autocorrelation analysis is carried out, i.e. the correlation between the output values at current time and different temporal lags is computed. Figure 5 shows the results of this analysis on all the scenarios, where the x and y axis represent the temporal lag (a multiple of four hours)

and the corresponding empirical partial autocorrelation, respectively. The bounds of a 95% confidence interval are also shown with dashed lines in the Figure. The correlation decreases with the lag (although not linearly) and after a lag of 17 time steps it is no longer comparable with the values observed for lags smaller than 17, i.e. the best choice is $H = 17$. Although the autocorrelation is still high for t = 18 and t= 20, the results are not improved on the real case study, according to the numerical experiments. The selection of 17 is already a "conservative" choice: in fact the most important values appear to be the first 7 values.
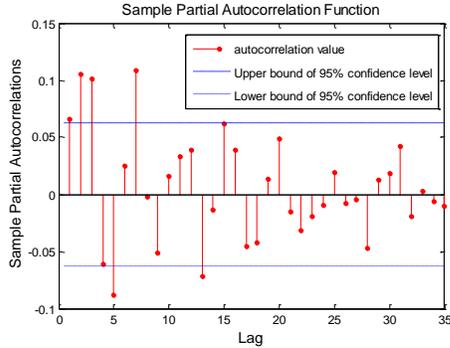
Fig. 5. Partial autocorrelation function with respect to time lags (multiples of four hours). Dotted lines are the bounds of the 95% confidence interval.

Then, the training dataset is divided into several sub-datasets for different sub-models using the angle-clustering algorithm described in sub-section 2.3.1.

3.3. Comparison of DW-RBF-Ensemble with Single SVR and Fixed Weights Ensemble

The ensemble model is expected to give better results than a single SVR model. To verify this claim, a comparison between a single SVR model and the proposed DW-RBF-Ensemble is carried out on the considered case study. A fixed weights ensemble (Kurram and Kwon, 2013) is also taken as a benchmark method to prove the benefit of using a dynamic weighting strategy.

Each time one out of 9 scenarios is chosen as the test dataset (named Observed Scenario) and the other 8 scenarios (named Reference Scenarios) from the training dataset which is used to construct the DW-RBF-Ensemble and the Fixed Weights Ensemble (FW-Ensemble). A SVR model is also trained on the training dataset for comparison (it is named Single SVR to be distinguished from the two ensemble models).

The steps for the comparison are the following:
1. Train a Single SVR model with all the training dataset.
2. The training dataset is divided into 6 clusters by the angle-clustering algorithm.
3. Train DW-RBF-Ensemble: FVS select the FVs in each cluster and a sub-model is trained on the selected FVs.

Weights of different sub-models for each data point are calculated with Eq. (9).
4. Train a FW-Ensemble: train a sub-model with all the data points in each cluster. The weight for each sub-model is decided by minimizing the MAE on the training dataset.
5. Calculation of Mean Absolute Error (MAE), Mean Relative Error (MRE) of the outputs of DW-RBF-Ensemble, FW-Ensemble and Single PSVR.
6. Compare prediction accuracy, computational burden and model robustness.

The results and comparisons among these models are presented in the next section.

4. RESULTS

In this section, the results from DW-RBF-Ensemble, FW-Ensemble and Single SVR are compared with respect to different aspects.
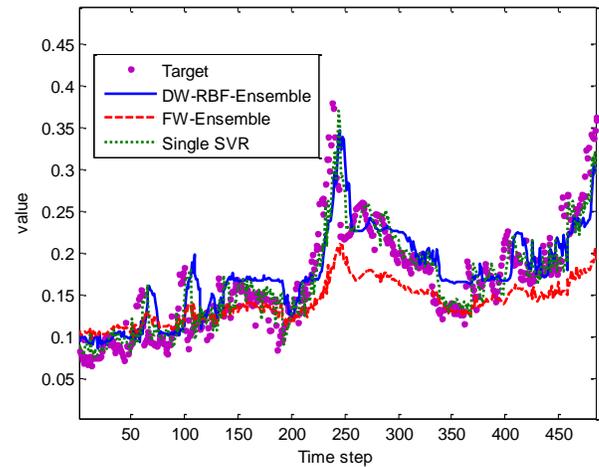
Fig. 6. Prediction results of ensembles and Single SVR, for the Scenario 1.
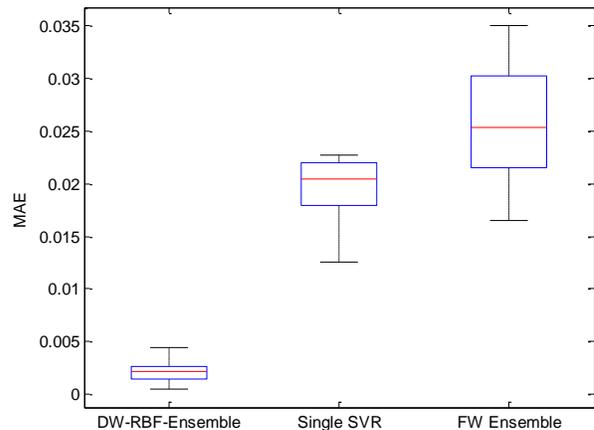
Fig. 7. MAE of prediction results of ensembles and Single SVR, for all 9 scenarios.
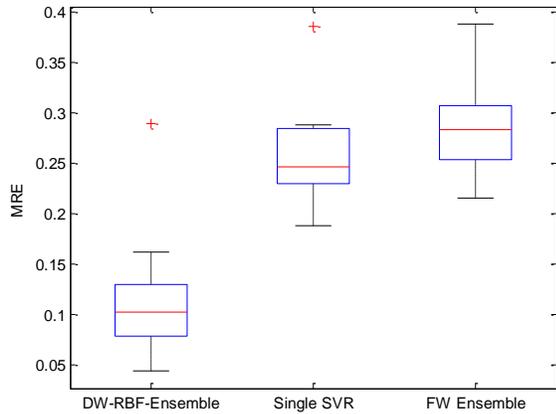
Fig. 8. MRE of prediction results of ensembles and Single SVR, for all 9 scenarios.

4.1. Prediction Accuracy

Figure 6 shows the prediction results of the ensembles (DW-RBF-Ensemble and FW-Ensemble) on the first scenario. Figures 7 and 8 report the prediction results of MAE and MRE obtained, respectively, by DW-RBF-Ensemble, FW-Ensemble and Single SVR. It is clear that DW-RBF-Ensemble gives best results in this case study, i.e. on average, the MAE and MAE values are smaller than for Single SVR and FW-Ensemble.

The bad results of the Single SVR are caused by the fact that the predictions are highly dependent on the training dataset. Moreover, the hyperparameters optimization is also critical to the performance of SVR. Well-chosen hyperparameters values can improve the performance of the SVR. However, the optimization method may converge to a local extreme, which results into a good performance at the beginning but bad at the end of the scenario. The ensemble approach can avoid such problem by combining the results from different sub-models.
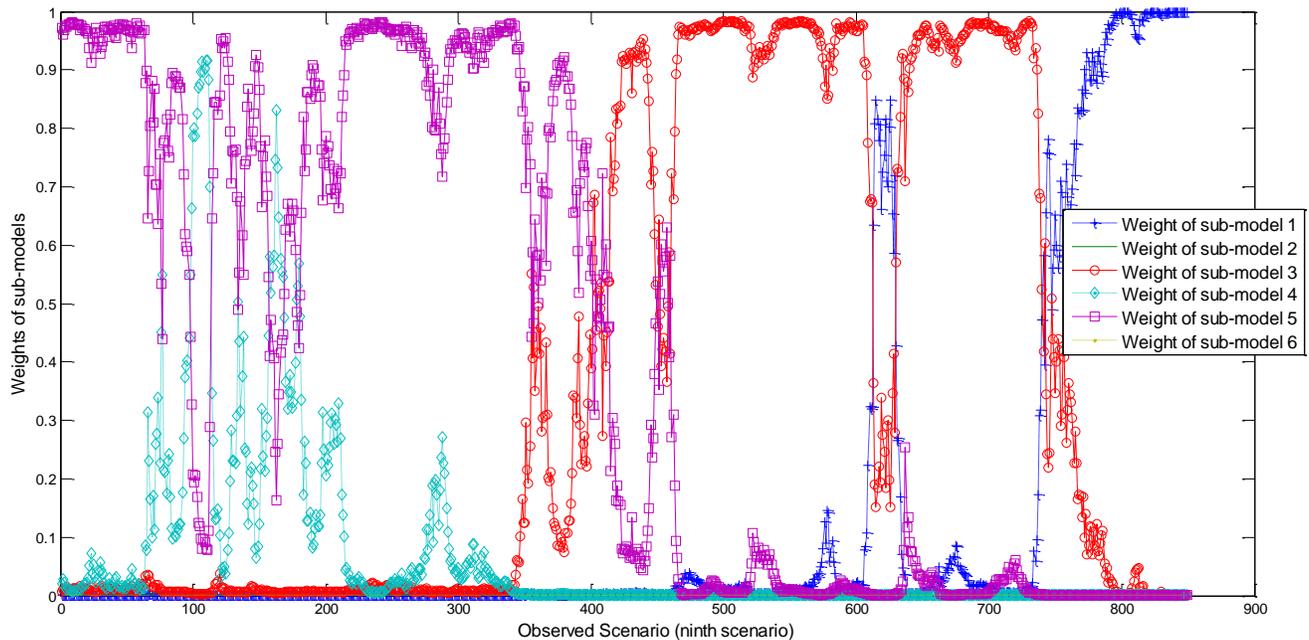
Fig. 9. Weights of different sub-models of DW-RBF-Ensemble for test data points of the Scenario 9.

These unstable results from the Single SVR prove the necessity of the ensemble approach for avoiding the limits of Single SVR in attaining the desired accuracy and robustness of the model.

In this case study, FW-Ensemble gives the worst results as the weights are fixed after training. Somewhat surprisingly, it even gives results worse than the single SVR but that is due to the fact that with the partitioning by angle clustering for the training of the ensemble sub-models, it turns out that for some data points, the best sub-model is not given the

most important weight. In this case, overlapping datasets for sub-models training would likely improve the FW-ensemble prediction performance. Figure 9 above shows the weights for different sub-models of DW-RBF-Ensemble in the case of selecting the ninth scenario as the Observed Scenario. It is clear that the weights of the sub-models change frequently to adapt to the ongoing data points.

The prediction results from DW-RBF-Ensemble confirm the practicability and efficiency of the proposed approach.

4.2. Robustness

From Figures 7 and 8, it is seen that the DW-RBF-Ensemble gives more stable prediction results compared to the Single SVR model and FW Ensemble. The Single SVR model cannot properly handle the noise in the data and it is difficult to find the global optimal values of the hyperparmeters. The weighted-sum ensemble models can decrease the influence of the noise by combining the prediction outputs of the sub-models. But the fixed weighting strategy cannot adapt to the changing environment and the weights of the sub-models are not changed adaptively. This is one reason for which DW-RBF-Ensemble model can give stable results, i.e. the DW-RBF-Ensemble model is more robust compared to the Single SVR and FW-Ensemble.

4.3. Computational complexity

Suppose the size of the training dataset is T; then, the computational complexities of the Single SVR for training and testing are T^3 and T, respectively. For very large datasets, the computational burden of the Single SVR model is very high and sometimes unacceptable. By dividing the training dataset into different sub-datasets, the total computational burden is decreased as $T^3 > T_1^3 + \cdots + T_N^3$, with $T_1 + \cdots + T_N = T$. With FVS, the size of the training dataset is further decreased for training and testing. Thus, the computational complexity of the DW-RBF-Ensemble approach is much smaller than the Single SVR trained on all the training dataset and the FW-Ensemble.

5. CONCLUSIONS

In this paper, we have proposed an innovative dynamic-weighted RBF-based ensemble approach for short-term prediction (1-day ahead prediction) with time series data. An angular-clustering algorithm is used to divide the training dataset into sub-datasets and FVS is used to decrease the size of the training data points by selecting only the representative data points in RKHS. Local fitness calculation is integrated to calculate the specific weights of the sub-models of the ensemble for each new input vector, without bringing too much computational burden.

The proposed ensemble approach has been shown to perform well in a real case study of signals recorded on a NPP component. Compared to the single SVR model and FW Ensemble, the proposed ensemble model outperforms them on prediction accuracy, computational burden, robustness and adaptability.

Further research needs to be carried out for optimizing the numbers of sub-models and the tuning of the hyperparameters. From the application point of view, the further developments of the method for long-term prediction will be investigated, for the purpose of remaining useful life prediction, i.e. prognostics.

REFERENCES

Acar, E., & Rais-Rohani, M. (2009). Ensemble of metamodels with optimized weight factors. *Structural and Multidisciplinary Optimization*, *37*(3), 279-294.

Bauer, E., & Kohavi, R. (1999). An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine learning*, *36*(1-2), 105-139.

Baudat, G., & Anouar, F. (2003). Feature vector selection and projection using kernels. *Neurocomputing*, *55*(1-2), 21-38.

Bonissone, Piero P., Feng Xue, and Raj Subbu. "Fast meta-models for local fusion of multiple predictive models." *Applied Soft Computing* 11.2 (2011): 1529-1539.

Brodsky, J. B., Lemmens, H. J., Brock-Utne, J. G., Vierra, M., & Saidman, L. J. (2002). Morbid obesity and tracheal intubation. *Anesthesia & Analgesia*, *94*(3), 732-736.

Chen, S., Wang, W., & Van Zuylen, H. (2009). Construct support vector machine ensemble to detect traffic incident. *Expert systems with applications*, *36*(8), 10976-10986.

Fantoni, Paolo F. "A neuro-fuzzy model applied to full range signal validation of PWR nuclear power plant data." *INTERNATIONAL JOURNAL OF GENERAL SYSTEM* 29.2 (2000): 305-320.

Gao, J. B., Gunn, S. R., Harris, C. J., & Brown, M. (2002). A probabilistic framework for SVM regression and error bar estimation. *Machine Learning*, *46*(1-3), 71-89.

Gurram, P., & Kwon, H. (2013). Sparse kernel-based ensemble learning with fully optimized kernel parameters for hyperspectral classification problems. *Geoscience and Remote Sensing, IEEE Transactions on*, *51*(2), 787-802.

Hallbert B. & Thomas K. (2014), Advanced Instrumentation, Information, and Control Systems Technologies Technical Program Plan for 2014, Idaho National Laboratory, INL/EXT-13-28055 Rev. 3, September 2014.

Hu, C., Youn, B. D., Wang, P., & Taek Yoon, J. (2012). Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life. *Reliability Engineering & System Safety*, *103*, 120-135.

Koo, In Soo, and Whan Woo Kim. "The development of reactor coolant pump vibration monitoring and a diagnostic system in the nuclear power plant." *ISA transactions* 39.3 (2000): 309-316.

Kim, H. C., Pang, S., Je, H. M., Kim, D., & Yang Bang, S. (2003). Constructing support vector machine ensemble. *Pattern recognition*, *36*(12), 2757-2767.

Lee, Terence. "Assessment of safety culture at a nuclear reprocessing plant." *Work & Stress* 12.3 (1998): 217-237.

Liu, J., Seraoui, R., Vitelli, V., & Zio, E. (2013). Nuclear power plant components condition monitoring by

probabilistic support vector machine. *Annals of Nuclear Energy*, *56*, 23-33.

Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. Journal of Time Series Analysis, 17(6), 571-599.

Minh, H. Q., Niyogi, P., & Yao, Y. (2006). Mercer's theorem, feature maps, and smoothing. In *Learning theory* (pp. 154-168). Springer Berlin Heidelberg.

Muhlbaier, M. D., Topalis, A., & Polikar, R. (2009). Learn. NC: Combining Ensemble of Classifiers With Dynamically Weighted Consult-and-Vote for Efficient Incremental Learning of New Classes. *Neural Networks, IEEE Transactions on*, *20*(1), 152-168.

Polikar, R. (2006). Ensemble based systems in decision making. *Circuits and Systems Magazine, IEEE*, *6*(3), 21-45.

Quinlan J R. Bagging, boosting, and C4. 5[C]//AAAI/IAAI, Vol. 1. 1996: 725-730.

Razavi-Far, R., Baraldi, P., & Zio, E. (2012). Dynamic Weighting Ensembles for Incremental Learning and Diagnosing New Concept Class Faults in Nuclear Power Systems. *Nuclear Science, IEEE Transactions on*, *59*(5), 2520-2530.

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

Sollich, P. (1999). Probabilistic interpretations and Bayesian methods for support vector machines. In *Artificial Neural Networks, 1999. ICANN 99. Ninth International Conference on (Conf. Publ. No. 470)* (Vol. 1, pp. 91-96). IET.

Valentini, G., & Dietterich, T. G. (2003, August). Low bias bagged support vector machines. In *ICML* (pp. 752-759).

Yang, X., Yuan, B., & Liu, W. (2009, November). Dynamic Weighting Ensembles for incremental learning. In *Proc. of IEEE conference in pattern recognition* (pp. 1-5).

BIOGRAPHIES

Jie Liu (B.Sc. in Mechanical Engineering, Beihang University, 2009; M.Sc. in Physics (at the same time, Engineer in Nuclear Energy), Ecole Centrale Pekin, Beihang University, 2012). He is now a PhD candidate of the Chair on Systems Science and the Energetic Challenge, European Foundation for New Energy - EDF, CentraleSupélec, France. His current research interests concern kernel-based pattern recognition methods for prognostics of nuclear component's Estimated Time To Failure (ETTF).

Valeria Vitelli received the Ph.D. degree in Mathematical Models and Methods for Engineering, with a focus on statistical models for classification of high-dimensional data, in May 2012. She worked as a postdoc researcher within the Chair on Systems Science and the Energetic Challenge, European Foundation for New Energy - EDF, CentraleSupélec, France from February 2012 to May 2013. She is currently a postdoc researcher in the Department of Biostatistics, University of Oslo, Norway. Her current research interests concern prediction and uncertainty quantification methods for complex data."

Enrico Zio (B.Sc. in nuclear engineering, Politecnico di Milano, 1991; M.Sc. in mechanical engineering, UCLA, 1995; Ph.D., in nuclear engineering, Politecnico di Milano, 1995; and a Ph.D., in nuclear engineering, MIT, 1998). He is the Director of the Chair on Systems Science and the Energetic Challenge, European Foundation for New Energy - EDF, CentraleSupélec, France, and the full Professor of Computational Methods for Safety and Risk Analysis in Politecnico di Milano, Italy. He is President of Advanced Reliability, Availability and Maintainability of Industries and Services (ARAMIS) ltd., and Chairman of the European Safety and Reliability Association, ESRA. His research topics are: analysis of the reliability, safety and security of complex systems under stationary and dynamic conditions, particularly by Monte Carlo simulation methods; development of soft computing techniques (neural networks, support vector machines, fuzzy and neuro-fuzzy logic systems, genetic algorithms, differential evolution) for safety, reliability and maintenance applications, system monitoring, fault diagnosis and prognosis, and optimal design. He is co-author of five international books and more than 200 papers on international journals, and serves as referee of more than 20 international journals.

Redouane Seraoui works for EDF in the Research & Development division. Following the graduation from the French engineering school "Ecole Superieure d'Electricite," he works now on a series of large research projects involving collaboration between many leading academic and industry partners. These projects include the development of tools and methods for the monitoring of power plants' critical components, applied to the nuclear industry but also to the fossil and renewable sides.