

An efficient online learning approach for support vector regression

Jie liu, Valeria Vitelli, Redouane seraoui, Enrico Zio

In this paper, an efficient online learning approach is proposed for Support Vector Regression (SVR) by combining Feature Vector Selection (FVS) and incremental learning. FVS is used to reduce the size of the training data set and serves as model update criterion. Incremental learning can “adiabatically” add a new Feature Vector (FV) in the model, while retaining the Kuhn-Tucker conditions. The proposed approach can be applied for both online training & learning and offline training & online learning. The results on a real case study concerning data for anomaly prediction in a component of a power generation system show the satisfactory performance and efficiency of this learning paradigm.

Introduction

Support Vector Machine (SVM) is capable of estimating nonlinear relations among input-output variables by mapping the data points into a high dimensional Reproducing Kernel Hilbert Space (RKHS) with linear relations between their mapping. Training a SVM model involves solving a Quadratic Programming (QP) problem in a number of coefficients equal to the number of training examples. For very large datasets, this may become infeasible, so that a reduction in the size of the training dataset must be sought. Another problem with SVM (and other data-driven methods) is that when there is a drift of the assumed fixed distribution of the dataset, the model may no longer work well and would need to be updated. As retraining a SVM model from the beginning is time-consuming, a more efficient way is to update the model by adding / removing some Support Vectors (SVs) and / or changing the Lagrange multipliers and hyperparameters [1].

Some efforts have been made to solve these two problems. Most of these works tackle only one of the previous problems, and some of the solutions proposed do not work well in nonstationary environments. In this paper, we propose an online learning approach for regression using SVM (called SVR), by combining a simplified version of the FVS introduced in [3] with the incremental learning method presented in [4] to give a solution to the two problems mentioned above. The method is called Online Learning approach with FVS and Incremental Learning, and hereafter named OL-FVS-IL for short. FVS aims at reducing the size of the training dataset: only part of the training dataset is chosen (called Feature Vectors) to build the SVR model and the mapping of the other training data points in the RKHS can be expressed by a linear combination of the feature vectors. When adding a new FV in the model, instead of retraining the model, incremental learning constructs the solution recursively, while retaining the Kuhn-Tucker conditions on all previously seen data.

A real case study concerning a component nuclear power plant is carried out within the Probabilistic Support Vector Regression (PSVR) paradigm, which is a Bayesian treatment of SVR with Gaussian priors [2].

The rest of the paper is organized as follows. Section 2 gives some insights about the modified FVS and a brief recall of the incremental learning method; the proposed online learning approach is also detailed in this section. Section 3 describes the real case study and the experiments results. Some conclusions and prospects are presented in Section 4.

Online Learning approach with Feature Vector Selection and Incremental Learning (OL-FVS-IL)

In this paper, an efficient online learning approach is proposed for regression using SVM. The proposed method can reduce the number of data points used in building the model, thus reducing its computational burden with large datasets. It can also incrementally add new FVs in the model to improve its adaptivity and capacity for different patterns.

Feature Vector Selection

In [3], the authors define a global criterion to characterize the feature space in SVM. A number of Feature Vectors (FVs) are selected from a training dataset to represent the useful dimension of RKHS in these data points. Any data can be projected on these FVs and, then, application of classical algorithms for training and prediction can be carried

out. Several simplifications and developments are made for this FVS approach proposed in [3], after deliberate derivation. Constrained by the size of this document, sophisticated details of calculation cannot be listed and only important results are presented.

The aim of FVS is to represent all the data points with a linear combination of FVs in RKHS. Suppose (\mathbf{x}_i, y_i) , for $i = 1, 2, \dots, M$ are the training data points, and the mapping $\varphi(\mathbf{x})$ maps each input \mathbf{x}_i into RKHS with the mapping $\boldsymbol{\varphi}_i$, for $i = 1, 2, \dots, M$. $k_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$ is the inner product between $\boldsymbol{\varphi}_i$ and $\boldsymbol{\varphi}_j$.

In order to find a new FV, we just need to verify if the mapping $\boldsymbol{\varphi}_N$ of a new data point \mathbf{x}_N can be represented by a linear combination of the mapping of the existing FVs. Suppose the existing FVs are $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L\}$ and the corresponding mapping are $S = \{\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \dots, \boldsymbol{\varphi}_L\}$. The verification of the new FV is to find $\{a_{N,1}, a_{N,2}, \dots, a_{N,L}\}$, which gives the minimum of Eq. (1).

$$\delta_N = \frac{\|\boldsymbol{\varphi}_N - \sum_{i=1}^L a_i \boldsymbol{\varphi}_i\|}{\|\boldsymbol{\varphi}_N\|} \quad (1)$$

The minimum of δ_N can be expressed by an inner product, which is calculated by Eq. (2).

$$\min \delta_N = \left| 1 - \frac{K_{S,N}^t K_{S,S}^{-1} K_{S,N}}{k_{N,N}} \right| = J_{S,N} \quad (2)$$

where $K_{S,S} = (k_{i,j})$, $i, j = 1, 2, \dots, L$ is the kernel matrix of S and $K_{S,N} = (k_{i,N})$, $i = 1, 2, \dots, L$ is the vector of the inner product between $\boldsymbol{\varphi}_N$ and S . $J_{S,N}$ is called the local fitness of data point \mathbf{x}_N . If $J_{S,N}$ is 0, the new point is not a new FV, otherwise, it is added to S as new FV. Considering the noise in the data and the reduction of the computational complexity, a small value θ is compared with $J_{S,N}$ instead of 0.

We introduce also the global fitness J_S on the training data set with Eq. (3) for the following illustration.

$$J_S = \sum_{i=1}^M J_{S,i} \quad (3)$$

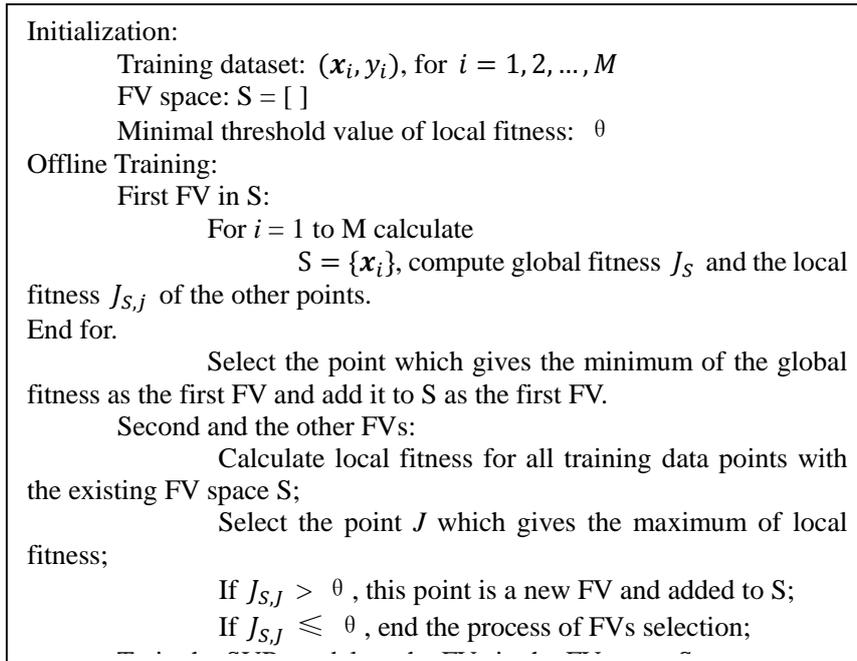
Incremental learning

The incremental learning method proposed in [4] provides a good solution for online learning SVR, with changing environmental and operational conditions. The key idea is to find the appropriate Kuhn-Tucker conditions for a new data by modifying its influence in the regression function while maintaining consistency in the Kuhn-Tucker conditions for the rest of the data used for learning. This method can “adiabatically” add a new point in the SVR model instead of retraining it from the beginning. Although this method is proposed for classification problems in [4], it can be expanded to regression problems.

In this section, we present the offline training and online learning procedures using FVS and the incremental learning method as shown in Figure 1. To keep within the limits of length of the paper, details on the procedure of computation are not provided; readers who are interested can refer to [4] for more details.

Online learning procedure

Figure 1. Procedure of OL-FVS-IL.



The offline training includes two parts. The first part is selecting the FVs in the available training dataset with FVS. Our aim is to find the feature space S formed by part of the training dataset which gives the minimum of the global fitness J_S on the training dataset. The procedure is an iterative process of sequential forward selection. For the first iteration, the point which gives the minimum of the global fitness J_S is selected as the first FV in the feature space S . The following iterations are the same: the next possible FV is the point which gives the maximum of local fitness with feature space S ; if the local fitness for this point is bigger than the predefined threshold value θ , it is added to the feature space S and we move to the next iteration; otherwise, the selection of FVs in the training dataset is finished. The next step is to train a SVR model on the selected FVs with a classical algorithm. The online learning is adding the new FVs in the model by the verification of the value of the local fitness of the new data with the existing feature space S .

Real Case Study

In the case study, a time series dataset from a sensor for measuring the leaking flow of the first seal of the reactor coolant pump in a nuclear power plant is used to test the performance of the proposed OL-FVS-IL approach applied within PSVR.

After the reconstruction of the raw data, the first 200 data points are selected as training dataset and the following 500 data points form the testing dataset.

A PSVR model is trained with FVs in the training dataset using the approach proposed in [5]. There are totally 26 FVs, which reduces greatly the computational burden of training compared to 200 points. The new FVs in the testing dataset are added in the model one by one with the proposed procedure. A total number of 48 new FVs are added in the model to improve the prediction performance on the testing dataset.

Figure 2. Prediction results before adding any new FVs (left) and after adding 48 new FVs (right).

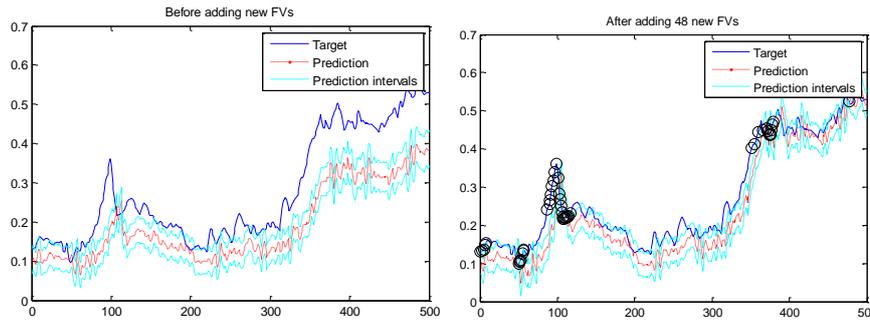


Figure 2 shows the prediction results before adding any new FVs and after adding 48 new FVs in the testing dataset. It is obvious that the performance is greatly improved with only 10% of all the test data points added to the model. In the right of Figure 2, the circles are the position of the new FVs in the training dataset. It is clear that a new FV is found when there is a change of the ongoing trend. This fact can prove that the proposed online learning approach can efficiently detect and learn the new patterns.

Compared to the prediction results using online learning approaches proposed in [4] and [6], OL-FVS-IL gives better performance on prediction accuracy with less computational complexity (Limited by the length of the paper, details cannot be presented).

Conclusions and prospective

In this paper, an efficient online learning approach called OL-FVS-IL is proposed for reducing the computational burden of offline training and online learning, and for updating the (P)SVR model by incrementally adding one data point in the model each time. The experimental results on a real case study show that the proposed approach can not only reduce significantly the number of data points used for training a (P)SVR model, but also make it appropriate for online prediction and incremental learning: the proposed approach can efficiently learn the new patterns in newly coming data and the prediction performance is improved.

References

1. M. Martin, "On-line support vector machines for function approximation," *tech. rep.*, Software Department, University Politecnica de Catalunya (2002).
2. J. B. Gao, S. R. Gunn, C. J. Harris and M. Brown, "A Probabilistic Framework for SVM Regression and Error Bar Estimation," *Mach. Learn.*, 46(1-3): 71-89 (2002).
3. G. Baudat and F. Anouar, "Feature Vector Selection and projection using kernels," *Neurocomputing*, 55(1-2): 21-38 (2003).
4. G. Cauwenberghs and T. Poggio, "Incremental and decremental support vector machine learning," *Fourteenth conference on advances in neural information processing systems, NIPS*: 409-415 (2001).
5. J. Liu, R. Seraoui, V. Vitelli and E. Zio, "Nuclear Power Plant Components Condition Monitoring by Probabilistic Support Vector Machine," *Ann. Nucl. Energy*, 56: 23-33 (2013).
6. J. Kivinen, A. J. Smola, R. C. Williamson, "Online learning with kernels," *IEEE Transactions on signal processing*, 52(8): 2165-2176, (2004).