Mechanical Systems and Signal Processing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

An adaptive online learning approach for Support Vector Regression: Online-SVR-FID

Jie Liu [a], Enrico Zio [a,b,*]

[a] Chair on Systems Science and the Energetic Challenge, Foundation Électricité de France (EDF), CentraleSupélec, Grande Voie des Vignes, 92295 Châtenay-Malabry, France

[b] Energy Departement, Politecnico di Milano, Campus Bovisa, Via Lambruschini, 4, 20156 Milano, Italy

ARTICLE INFO

Article history:
Received 8 October 2015
Received in revised form
15 January 2016
Accepted 27 February 2016

Keywords:
Online learning
Support Vector Regression
Time series data
Pattern drift
Feature vector selection
Incremental and decremental learning

ABSTRACT

Support Vector Regression (SVR) is a popular supervised data-driven approach for building empirical models from available data. Like all data-driven methods, under non-stationary environmental and operational conditions it needs to be provided with adaptive learning capabilities, which might become computationally burdensome with large datasets cumulating dynamically. In this paper, a cost-efficient online adaptive learning approach is proposed for SVR by combining Feature Vector Selection (FVS) and Incremental and Decremental Learning. The proposed approach adaptively modifies the model only when different pattern drifts are detected according to proposed criteria. Two tolerance parameters are introduced in the approach to control the computational complexity, reduce the influence of the intrinsic noise in the data and avoid the overfitting problem of SVR. Comparisons of the prediction results is made with other online learning approaches e.g. NORMA, SOGA, KRLS, Incremental Learning, on several artificial datasets and a real case study concerning time series prediction based on data recorded on a component of a nuclear power generation system. The performance indicators MSE and MARE computed on the test dataset demonstrate the efficiency of the proposed online learning method.

1. Introduction

Various efforts of research on machine learning have been devoted to studying situations in which a sufficiently large and representative dataset is available from a fixed, albeit unknown distribution. The models trained for these situations can function well only for patterns within the representative training dataset [20].

In real-world applications, systems/components are usually operated in non-stationary environments and evolving operational conditions, whereby patterns drift. Then, to be of practical use the models built must be capable of timely learning changes in the existing patterns and new patterns arising in the dynamic environment of system/component operation. The up-to-date snapshot of the ongoing research in this area can be found in [6,7,13,20,21,22]. Different adaptive strategies have been proposed for neural networks [6], Markov Chain [7], fuzzy inference systems [21], feature extraction [13], and fault detection [22].

* Corresponding author at: Energy Departement, Politecnico di Milano, Campus Bovisa, Via Lambruschini, 4, 20156 Milano, Italy
E-mail addresses: jie.liu@centralesupelec.fr (J. Liu), enrico.zio@centralesupelec.fr (E. Zio).

http://dx.doi.org/10.1016/j.ymssp.2016.02.056
0888-3270/

Support Vector Regression (SVR) is a popular, supervised data-driven approach, which also must cope with the problem of changing environments and adaptation to pattern drifts, but has attracted relatively less attention for adaptive learning. This paper provides a strategy for feature extraction and adaptive learning with SVR.

Some approaches have been proposed in the literature for SVR to adaptively learn new data points [2,3,8,10,18,23]. In these approaches, the online learning of a trained SVR model is mostly guided by prediction accuracy and/or characteristics of the inputs of the data points: increasing data points are used to update the SVR model when the corresponding predictions are not precise and/or they belong to a less explored zone of the input space and, thus, contain new information. With respect to the consideration of the input characteristics for achieving model update, reference [18] proposes an approach based on an adaptive Kernel Principal Component Analysis (KPCA) and Support Vector Machine (SVM) for real-time fault diagnosis of High-Voltage Circuit Breakers (HVCBs). Bordes et al. [2] propose an online algorithm which converges to the SVM solution by using the τ-violating pair paradigm. Wang et al. [23] propose an online core vector machine classifier with adaptive Minimum-Enclosing-Ball (MEB) adjustment. Reference [10] uses a small subset of basis vectors to approximate the full kernel on arbitrary points. Engel et al. [8] present a nonlinear kernel-based recursive least squares algorithm, which performs linear regression in the feature space and can be used to recursively construct the minimum mean squared-error regressor. Csato and Opper [3] combine a Bayesian online algorithm with a sequential construction of relevant subsets of the training dataset and propose Sparse Online Gaussian Process (SOGP) to overcome the limitation of Gaussian process on large datasets.

The methods above consider only the characteristics of the inputs to update the model, not the prediction accuracy. Reference [4] proposes an online recursive algorithm to “adiabatically” add or remove one data point in the model while retaining the Kuhn-Tucker conditions on all the other data points in the model. Martin [17] further develops this method for the incremental addition of new data points, removal of existing points and update of target values for existing data points. But the authors provide only the “how” for model update, while the “when” and “where” to make such update are not presented, and adding each new point available can soon become quite time-consuming. Karasuyama and Takeuchi [11] propose a multiple incremental algorithm of SVM, based on the previous results. These above mentioned incremental and decremental learning approaches feed to the model all new points including noisy and useless ones, without bothering of selecting the most informative ones. Crammer et al. [5] propose online passive-aggressive algorithms for classification and regression, but considering only the prediction accuracy as the update criterion. Reference [12] considers using classical stochastic gradient descent within a feature space and some straightforward manipulations for online learning with kernels. The gradient descent method destroys completely the Kuhn-Tucker conditions, which instead are necessary for building a SVR model.

In this paper, the authors propose an online learning approach for SVR to adaptively modify the model when different types of pattern drifts are detected, providing a solution for “when” and “where” to modify the trained model. The proposed online learning approach is a compromise among prediction accuracy, robustness and computational complexity, obtained by combining a simplified version of the Feature Vector Selection (FVS) method introduced in [1] with the Incremental and Decremental Learning presented in [4], considering the characteristics of the inputs of new data points and the bias of the corresponding prediction. The method is hereafter called Online learning approach for SVR using FVS and Incremental and Decremental Learning, Online-SVR-FID for short. FVS aims at reducing the size of the training dataset: instead of training the SVR model with the whole training dataset, only part of it (the set of Feature Vectors (FVs) which are nonlinearly independent in the Reproduced Kernel Hilbert Space (RKHS)) is used and the mapping of the other training data points in RKHS can be expressed by a linear combination of the selected FVs. In this paper, FVS is simplified and used for the proposed adaptive online learning approach. According to the geometric meaning of FVS in RKHS, in this paper, each data point (input-output pair) is defined as a pattern and two types of pattern drifts are given: new pattern and changed pattern. A new data point is a new pattern (or new FV) if the mapping of its inputs in RKHS cannot be represented by a linear combination of the mapping of existing patterns (this is integrated in some papers), while it is a changed pattern if its mapping can be represented by such a linear combination but the bias of its predicted value is bigger than a predefined threshold. Once a new data point is judged as a new pattern, it is immediately added to the present model no matter the bias of its prediction is small or big, thus keeping the richness of the patterns in the model. A changed pattern is used to replace a selected existing pattern instead of adding it into the model, thus keeping the nonlinear independence in RKHS among all the data points in the model, which is critical for FVS calculation. When adding or removing a FV in the model, instead of retraining the model, Incremental and Decremental Learning can construct the solution iteratively. Two criteria are proposed to detect new and changed patterns, considering respectively the characteristics of the inputs and bias of the prediction. The proposed approach can efficiently add new patterns and change existing patterns in the model, to follow the incoming patterns and at the same time reduce the computational burden by selecting only informative data points. The two criteria proposed for verification of new patterns and changed patterns can also help avoiding the overfitting problem bothering SVR and reducing the influence of the intrinsic noise in the data.

The structure of the proposed Online-SVR-FID is similar to that of the method proposed in [23]. However, the method proposed in [23] is based on MEB whereas Online-SVR-FID is based on FVS. Another main difference between these two methods is that the method proposed in [23] adds only the new patterns (as defined previously) in the model, while in Online-SVR-FID, two kinds of drifts (new patterns and changed patterns) are identified. New patterns are added to the model directly as in [23] whereas the changed patterns are used to replace existing FVs to keep the nonlinear independence among FVs, which is critical during online learning.