Mechanical Systems and Signal Processing 111 (2018) 68–86

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

# Echo state kernel recursive least squares algorithm for machine condition prediction

Haowen Zhou [a,*], Jinquan Huang [a], Feng Lu [a], Jeyarajan Thiyagalingam [b], Thia Kirubarajan [c]

[a] *Jiangsu Province Key Laboratory of Aerospace Power Systems, College of Energy and Power Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China*
[b] *Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, United Kingdom*
[c] *Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada*

A R T I C L E   I N F O

*Article history:*
Received 2 April 2017
Received in revised form 24 January 2018
Accepted 23 March 2018

*Keywords:*
Kernel adaptive filter
Reservoir computing
Long-term prediction
Remaining useful lifeprediction
Prognostics

A B S T R A C T

Kernel adaptive filter (KAF) has been widely utilized for time series prediction due to its online adaptation scheme, universal approximation capability and convexity. Nevertheless, KAF's ability to handle temporal tasks is limited, because it is essentially a feed-forward neural network that lacks dynamic characteristics. Traditionally, a sliding widow that contains consecutive data points is constructed to deal with the temporal dependency between data points at neighboring time steps, but the restricted widow length may be incapable of capturing temporal patterns on a larger time scale. To manage this issue, a novel sequential learning approach called echo state KRLS (ES-KRLS) algorithm is proposed by incorporating a dynamic reservoir into kernel recursive least squares (KRLS) algorithm. The reservoir, consisting of a large number of sparsely interconnected hidden units, is treated as a temporal function that transforms the history of the time series into a high-dimensional reservoir state space. Subsequently, the spatial relationship between the reservoir state and the target output is effectively approximated by KRLS algorithm. With the utilization of the fixed reservoir, our novel method not only maintains the simplicity of the learning process but also leads to a significant improvement in the capability of modeling dynamic systems. Numerical results on benchmark tasks demonstrate the excellent performance of the novel method with respect to long-term prediction. Finally, an online prognostic method that combines ES-KRLS and a Bayesian technique is developed for tracking the health status of a degraded system and predicting remaining useful life (RUL). This prognostic method is applied to a turbofan engine degradation dataset to demonstrate its effectiveness.



## 1. Introduction

Kernel method provides a unified framework for pattern analysis and nonlinear signal processing, and as such they appear in numerous successful applications, including the support vector machine [1], kernel principal component analysis [2], and kernel regularization network [3]. The main idea of the kernel method is that a nonlinear mapping associated with a Mercer kernel is utilized to transform the data from the input space to a high-dimensional feature space with rich representations. However, the above-mentioned methods are formulated in a batch form. If the training samples arrive sequentially,

* Corresponding author.
*E-mail address:* zhouhaowen@nuaa.edu.cn (H. Zhou).

https://doi.org/10.1016/j.ymssp.2018.03.047
0888-3270/

these offline algorithms have to retrain the approximation model from scratch once a new training data is available. This may impose restriction on the applications of these algorithms to online scenarios, especially when real-time performance is emphasized. Therefore, a sequential learning method that updates the existing model incrementally would be a better choice for handling the data that arrives in a flow mode.

Online kernel-based learning (OKL) [4,5] provides an alternative to train the model recursively. As a subfield of OKL, kernel adaptive filters (KAFs) have gained widespread use because of their simple structure, universal approximation capability, and convexity. KAFs are the generalization of the well-established linear adaptive filtering algorithms, approximating a nonlinear function by reformulating the linear structure in a reproducing kernel Hilbert space (RKHS). The KAF family includes the kernel least mean square (KLMS) [6], kernel affine projection algorithm [7], KRLS [8], and extended kernel recursive least squares [9] algorithms, to mention a few.

Although KAF has been widely used for time series prediction, two drawbacks that remain to be solved. The first is the lack of sparseness. At each iteration, KAFs allocate a kernel unit for the new data point. Consequently, the network size grows linearly with the number of training samples, leading to a continuous increase in computational burden and memory requirements. To control the growth of the network, many sparsification techniques have been adopted. The basic idea behind these methods is to select only the informative training samples to train the network, according to some criteria, such as the approximation linear dependency (ALD) criterion [8], surprise criterion [10], novelty criterion [11], or minimum description length criterion [12].

The second drawback is that KAF is incapable of capturing temporal characteristics of nonlinear dynamic systems [13], because KAF is essentially a single-layer feed-forward neural network (FNN). An FNN aims to learn a static mapping where the outputs of the network depend solely on the current inputs, thereby neglecting the dependency of data points at neighboring time steps. As a remedy, often a fixed-size sliding widow storing consecutive values is constructed to transform the temporal correlation into the spatial correlation. Nevertheless, prior knowledge is required in order to select the appropriate time embedding dimension such that the latent dynamic characteristics of the systems unfold [14]. Apart from FNN, recurrent neural network (RNN) is another artificial neural network architecture where hidden units are interconnected. The existence of the recurrence connections between units makes RNN a dynamic system. The dynamic behaviors of RNN rely on the history of the inputs, and thereby achieve the necessary memory capability. RNN has been proven to be able to model any dynamic system with arbitrary precision [15]. The connection weights of RNN are mainly trained by different gradient-based algorithms, such as back propagation through time [16], extended Kalman filtering [17] and real-time recurrent learning [18]. Unfortunately, these algorithms often suffer from slow convergence and create a high computational burden, because the gradients tend to vanish or explode rapidly through propagation.

Aimed at avoiding the difficulty of adapting the connection weights of RNN, echo state network (ESN) and liquid state machine provide a new paradigm for RNN, called reservoir computing (RC) [19–21]. The RC framework contains a dynamic reservoir, where a large number of hidden units are sparsely connected, as well as a linear readout layer. The reservoir is randomly generated and stays fixed during the learning process. The reservoir can be considered as a spatiotemporal mapping that transforms the history of the inputs into a high-dimensional reservoir state space where the readout layer is learned by simple linear regression methods. As a result of its simple and effective learning procedure, ESN has become an appealing tool for time series prediction [22–26].

One fascinating property of ESN is that a temporal task is converted into a non-temporal task of learning a static mapping, because the dynamic reservoir has a fixed recurrent topology and thus only the output weights need to be solved. This facilitates the considerable flexibility of designing the training criterion. The direct method of training the readout layer is calculating the Moore-Penrose pseudo-inverse, which may suffer from ill-posed and over-fitting problems. Hence, regularized regression methods such as ridge regression and the lasso method are utilized to obtain the output weights of ESNs [27]. To alleviate the detrimental effect of noises and outliers on time series prediction, Gaussian process regression is combined with ESN to provide a robust prediction by deriving a posterior distribution of the network output [28,29]. In [22], the readout layer is treated as a linear support vector machine that is trained based on the epsilon-sensitive or Huber cost functions, such that the robustness to outliers is obtained. In [30], the learning procedure is formulated in a Bayesian framework, whereas the commonly used Gaussian distribution is substituted by a Laplace distribution, which is more insensitive to outliers. In order to make ESN suitable for online applications, the readout layer is traditionally constructed as linear adaptive filters, including the recursive least squares (RLS) and least mean square (LMS) [31] algorithms. The online adjustments to the output weight can also be conducted by various variants of the Kalman filter [32]. Furthermore, the readout layer is extended to a Volterra filter by using the nonlinear structure [33], but the computational complexity expands exponentially with the number of hidden units in the reservoir.

In this paper, a novel ES-KRLS algorithm that incorporates a dynamic reservoir into the KRLS algorithm is proposed. With the aid of the reservoir, the time series up to the current time step is mapped into a high-dimensional reservoir state space, rather than selected data points. Subsequently, another nonlinear mapping associated with a particular Mercer kernel is performed to transform the obtained reservoir state into the potentially infinite-dimensional RKHS, where the readout layer is constructed to rebuild the desired outputs. As a result of this two-step transformation, both the spatial and temporal dependencies between different data points are fully exploited for model approximation. As a matter of fact, the ES-KRLS algorithm is considered as an integration of ESN and KAF, created by reformulating the readout layer in RKHS. To avoid performing direct calculation in RKHS, the readout layer is expressed in the form of inner products, leading to a radial basis function network with more powerful generalization capability. Moreover, the novel method maintains the simplicity of the learning