Mechanical Systems and Signal Processing 95 (2017) 446–467

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Reduced kernel recursive least squares algorithm for aero-engine degradation prediction

Haowen Zhou [a,*], Jinquan Huang [a,b], Feng Lu [a,b,c]

[a] Jiangsu Province Key Laboratory of Aerospace Power Systems, College of Energy and Power Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
[b] Collaborative Innovation Center of Advanced Aero-Engine, Beijing 100191, China
[c] Aviation Motor Control System Institute, Aviation Industry Corporation of China, Wuxi 214063, China

ARTICLE INFO

Article history:
Received 4 July 2016
Received in revised form 29 March 2017
Accepted 31 March 2017
Available online 7 April 2017

Keywords:
Kernel recursive least squares
Sparse kernel method
Reduced technique
Aero-engine
Remaining useful life
Prognostics

ABSTRACT

Kernel adaptive filters (KAFs) generate a linear growing radial basis function (RBF) network with the number of training samples, thereby lacking sparseness. To deal with this drawback, traditional sparsification techniques select a subset of original training data based on a certain criterion to train the network and discard the redundant data directly. Although these methods curb the growth of the network effectively, it should be noted that information conveyed by these redundant samples is omitted, which may lead to accuracy degradation. In this paper, we present a novel online sparsification method which requires much less training time without sacrificing the accuracy performance. Specifically, a reduced kernel recursive least squares (RKRLS) algorithm is developed based on the reduced technique and the linear independency. Unlike conventional methods, our novel methodology employs these redundant data to update the coefficients of the existing network. Due to the effective utilization of the redundant data, the novel algorithm achieves a better accuracy performance, although the network size is significantly reduced. Experiments on time series prediction and online regression demonstrate that RKRLS algorithm requires much less computational consumption and maintains the satisfactory accuracy performance. Finally, we propose an enhanced multi-sensor prognostic model based on RKRLS and Hidden Markov Model (HMM) for remaining useful life (RUL) estimation. A case study in a turbofan degradation dataset is performed to evaluate the performance of the novel prognostic approach.

1. Introduction

Kernel method provides a unified framework for regression and classification in machine learning community and has gained a lot of successful applications, including support vector machine (SVM) [1], Gaussian process regression (GPR) [2], relevance vector machine (RVM) [3], kernel principal component analysis [4] and kernel regularization network [5]. The main idea of these applications is that a nonlinear mapping associated with a mercer kernel is utilized to transform the input data from the original space to a high-dimensional feature space, namely reproducing kernel Hilbert space (RKHS). The SVM model is typically derived based on the epsilon-sensitive cost function that just penalizes the error outside a predefined boundary. Hence, the model approximation boils down to a convex quadratic programming (QP) problem. Because of the inequality constraints introduced into the QP problem, the obtained solution is often sparse, which indicates that many weights are equal to

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

http://dx.doi.org/10.1016/j.ymssp.2017.03.046
0888-3270/

zero. Compared with normal SVM, least squares support vector machine (LSSVM) [6] that adopts the quadratic cost function provides a more computationally efficient alternative. With the utilization of equality constraints, the model solution is achieved just by solving a linear system. However, the main drawback of LSSVM is that its solution is not sparse, i.e., each training sample makes contribution to the learning process. In order to achieve sparseness, redundant data points are omitted according to various pruning schemes [7,8]. RVM is formulized in an identical function form to SVM, but utilizes a Bayesian inference method to obtain the solution. RVM is a sparse kernel method because much less basis functions are used to construct the approximated model. GPR is another kernel method that incorporates into the Bayesian learning framework. Both RVM and GPR are capable to provide probabilistic predictions. Since the covariance function is characterized by a kernel matrix, GPR is essentially equivalent to a RBF network. However, the above-mentioned algorithms are offline and derived in batch setting. If the training samples arrive sequentially, these batch algorithms have to retrain the network from scratch once a new sample is present. This imposes restriction on these algorithms' online applications especially when the real-time performance is emphasized. Hence, it would be a better choice to update the existing network incrementally. Online kernel learning (OKL) [9,10] provides an alternative to learn the data efficiently and recursively.

As a subfield of OKL, KAFs gain wide popularity due to universal approximation capability and convexity, which include kernel least mean square (KLMS) [11], kernel affine projection algorithm [12], kernel recursive least squares (KRLS) [13], extended kernel recursive least squares (Ex-KRLS) [14], KLMS with feedback [15,16], mixture KLMS [17], etc. These algorithms are kernel-based versions of the well-established linear adaptive filtering algorithms. The classic linear methods are extended to RKHS and then reformulated in the form of inner product. A technique known as "kernel trick" enables the inner product to be calculated by using a kernel function without making direct reference to feature vectors. Considering that the mean square error criterion adopted in the traditional KAFs is sensitive to large noises and outliers, mean p-power error criterion [18], maximum correntropy criterion (MCC) [19–21] and generalized MCC [22] in information theoretical learning have been utilized as robust cost functions to deal with non-Gaussian signals, especially the training samples contaminated by impulsive noises.

Compared with the offline learning methods, KAFs deal with the training samples in online mode at cost of less computations. However, the main drawback of kernel adaptive filters is that the network size keeps growing linearly with the number of the training samples, leading to the continuous increase in memory and time consuming. Aiming at this issue, online sparsification techniques such as novelty criterion (NC) [23], approximation linear dependency (ALD) criterion [13], coherence criterion [24], surprise criterion (SC) [25] and minimum description length criterion [26], were introduced into the kernel adaptive filters. In [27,28], a sparsity inducing l_1 norm penalty is integrated into the cost function, whereby the kernel units with negligibly small coefficients are pruned. These techniques restrain the dictionary growth by pruning the redundant samples. A new feature input is added into the center dictionary only when it satisfies a specific sparsification criterion combined with a pre-defined threshold. Hence, a minimal number of centers are selected to cover the area where the inputs are located.

Although these sparsification techniques gain a satisfying performance in controlling the structure growth, the exact network size can't be determined in advance based on the priori knowledge. It is necessary to fix the network size especially when memory size and computational consumption are given an exact bound. The idea that only the last N pairs of observed data are used for regression yields sliding-window KRLS algorithm [29]. Despite its simple structure, it displays good advantage when tracking abrupt changes and operating in non-stationary environment without additional memory requirements. In addition to constructing a sliding window, a pruning strategy was adopted in fixed-budget KRLS algorithm [30]. According to the pruning strategy in this algorithm, the center which makes the least contribution to minimizing the cost function, instead of the oldest center, is pruned from the existing dictionary at each iteration so as to keep the network size bounded.

Although the introduction of sparsification criteria reduces computation and memory effectively, there exists one problem that the training data which can't meet the aforementioned criteria are throw away directly and make no contribution to the learning process. As a result, the obtained network sustains accuracy degradation. Despite the fact that these discarded data are relatively useless for updating the structure, they can be employed to modify the corresponding coefficients. Hence, quantization techniques were introduced in quantized KLMS (QKLMS) [31], quantized KRLS (QKRLS) [32] and fixed budget quantized KLMS [33], modified quantized KLMS [34], density-dependent quantized KLMS [35] and quantized KLMS with desired signal smoothing [36]. Unlike conventional sparsification techniques, the information conveyed by these dispensable samples is used to update the coefficient of the closest existing center.

Motivated by the idea of taking full advantage of the redundant data, a novel sparsification technique is introduced into KRLS algorithm, yielding RKRLS algorithm. In this novel algorithm, the cost function is reformulated by considering the information conveyed by these redundant data. Our method is composed of two complementary steps, namely the structure update and the coefficient adjustment. To be specific, ALD criterion is employed to determine whether a new kernel unit is added for the new input so as to curb the network growth. Subsequently, the new data is predicted by the existing network. If the prediction error exceeds the pre-set threshold, this data will be used to modify the coefficients of the network. To sum up, the first step attempts to achieve a compact network, while the second step guarantees the desirable accuracy performance.

This paper is organized as follows. In Section 2, KRLS algorithm as well as ALD criterion is introduced briefly. The following section elaborates on the derivation of RKRLS algorithm. In order to test the efficacy of the presented RKRLS, experiments on the time series prediction and online regression are conducted. Then, RKRLS is employed for RUL estimation. Finally, Section 6 summarizes the conclusions.