Neurocomputing 000 (2017) 1–18

Contents lists available at ScienceDirect

Neurocomputing

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

Ensemble of optimized echo state networks for remaining useful life prediction

Marco Rigamonti[a], Piero Baraldi[a,*], Enrico Zio[a,b], Indranil Roychoudhury[c], Kai Goebel[d], Scott Poll[d]

[a] Energy Department, Politecnico di Milano, Via Ponzio 34/3, Milan 20133, Italy
[b] Chair System Science and the Energy Challenge, Fondation Electricité de France (EDF), CentraleSupélec, Université Paris Saclay, Grande Voie des Vignes, 92290 Chatenay-Malabry, France
[c] Stinger Ghaffarian Technologies, Inc., NASA Ames Research Center, Moffett Field, CA 94035, USA
[d] NASA Ames Research Center, Moffett Field, CA 94035, USA

ARTICLE INFO

Article history:
Received 1 March 2017
Revised 9 November 2017
Accepted 29 November 2017
Available online xxx

Communicated by H.R. Karimi

Keywords:
Echo state networks
Recurrent neural networks
Ensembles
Prediction uncertainty
Prediction Intervals
Differential Evolution

ABSTRACT

The use of Echo State Networks (ESNs) for the prediction of the Remaining Useful Life (RUL) of industrial components, i.e. the time left before the equipment will stop fulfilling its functions, is attractive because of their capability of handling the system dynamic behavior, the measurement noise, and the stochasticity of the degradation process. In particular, in this paper we originally resort to an ensemble of ESNs, for enhancing the performances of individual ESNs and providing also an estimation of the uncertainty affecting the RUL prediction. The main methodological novelties in our use of ESNs for RUL prediction are: i) the use of the individual ESN memory capacity within the dynamic procedure for aggregating of the ESNs outcomes; ii) the use of an additional ESN for estimating the RUL uncertainty, within the Mean Variance Estimation (MVE) approach. With these novelties, the developed approach outperforms a static ensemble and a standard MVE approach for uncertainty estimation in tests performed on a synthetic and two industrial datasets.

1. Introduction

Prognostics is the engineering discipline concerned with the prediction of the time left before the equipment will no longer perform its intended function, i.e., its Remaining Useful Life (RUL). Such prediction is typically performed starting from the current health state of the equipment and taking into account its past history and future operation. The capability of RUL prediction enables the possibility of identifying equipment problems at an early stage and timely performing maintenance to anticipate failures [41,65,68]. In this view, prognostics represents an important opportunity for industry, in terms of efficient and agile maintenance management, in principle providing the right part to the right place at the right time, with the necessary resources [23].

Prognostics require the availability of models capable of providing accurate RUL predictions and the associated uncertainty [46,88]. Prognostics models should take into account the different sources of uncertainty affecting RUL predictions [9,87]: (*i*) randomness in the equipment future degradation path, due to the intrinsic stochasticity of the degradation process and the unknown future operation and environmental conditions; (*ii*) inaccuracy of the prognostic model; (*iii*) measurement noise; and (*iv*) imperfect knowledge of the degradation initiation time.

Prognostic methods are typically classified as model-based, data-driven and hybrid [16]. Model-based methods use an explicit mathematical model of the degradation process to predict the future evolution of the degradation state and, thus, the RUL of the system [55]. In practice, even when the model of the degradation process is known, the RUL estimate may be difficult to obtain, since the degradation state of the system may not be directly observable and/or the measurements may be affected by noise and disturbances. In these cases, model-based estimation methods aim at inferring the dynamic degradation state and provide a reliable quantification of the estimation uncertainty on the basis of the sequence of available noisy measurements [27,28,44].

On the other side, data-driven methods are used when an explicit model of the degradation process is not available, but sufficient historical data have been collected. These methods are

* Corresponding author.
E-mail addresses: marcomichael.rigamonti@polimi.it (M. Rigamonti), piero.baraldi@polimi.it (P. Baraldi), enrico.zio@polimi.it, enrico.zio@ecp.fr (E. Zio), indranil.roychoudhury@nasa.gov (I. Roychoudhury), kai.goebel@nasa.gov (K. Goebel), scott.poll@nasa.gov (S. Poll).

https://doi.org/10.1016/j.neucom.2017.11.062
0925-2312/

based on statistical models that 'learn' trends from the data [80]. In this respect, Artificial Neural Networks (ANNs) are often used [16,33,66]; other examples are Autoregressive Moving Average techniques [74], Relevance Vector Machines [25,33,74], fuzzy similarity-based methods [98]. Finally, hybrid approaches combine physics-based models of the degradation process with the use of historical data collected from degrading components [80].

Among these data-driven approaches, ANNs have often been used for time series forecasting, due to their capability of approximating non-linear complex functions [33]. Feedforward ANNs have been used in prognostics for the prediction of rotating machineries [56] and Lithium-ion batteries [77] RUL. However, the use of feedforward ANNs in prognostics is limited by the fact that they are direct models characterized by oriented connections among neurons, without feedback and loop connections. Therefore, since the output of any layer does not affect the same layer, feedforward ANNs are not able to catch the system dynamic behaviors. An attempt to provide the system dynamics in input to feedforward ANNs has been proposed in [93], where the models receive in input the current and past signal values collected in a time window. The main limitations of this approach are the difficulty in identifying the proper lengths of the time window and the largely increased number of model inputs.

Local field NNs [37,92], whose dynamics is based on the local field states of their neurons, have also been exploited for time series predictions. Their local associative memory properties have been applied with success to the problem of predicting groundwater levels [96], but, at the best of the authors knowledge, they have not been applied to prognostic problems. Alternatively, Spiking NNs, based on the use of spiking neurons characterized by internal states which change with time, are attractive for inherent dynamic problems such as those typical of prognostics [57]. General shortcomings of SNNs are the computational burden and the sensitivity of gradient descent-based learning algorithms to the SNN initial state [57]. Although significant advances have recently been made, these issues have not been fully resolved [32].

An alternative solution to the problem of learning the system dynamic using ANNs is given by Recurrent Neural Networks (RNNs). Since the RNNs internal states are characterized by cyclic connections and feedbacks among neurons, they are capable of encapsulating into their neurons a nonlinear transformation of the input history [18,21,45,54]. This provides memory properties to RNNs, enabling them to handle sequential tasks, such as time series prediction [79]. RNNs have been applied to different prognostic problems, such as the prediction of machine deterioration evolution using vibration data [86] and of helicopter drivetrain system gearbox [75] and turbofan engines [35] RULs. However, the application of traditional RNNs in time series forecasting problems is limited by the difficulty of optimizing their numerous internal parameters and the significant computational effort of the training process [53]. To overcome this problem, we exploit the use of Echo State Networks (ESNs), a relatively new type of RNNs. An ESN is an RNN trained by using a Reservoir Computing (RC) method based on the random generation of a RNN, called reservoir, which remains unchanged during the training phase and is passively excited by the input patterns. Since the only weights of the ESN to be optimized are those of the connections among the reservoir internal states and the output, RNN training is computationally more efficient [53].

Although ESNs have been shown to provide good generalization capabilities [47], few applications of ESNs for RUL prediction have been proposed. In [58], an ESN-based approach for the prediction of the RUL of industrial Fuel Cells has been developed. In [29], a hybrid approach combining ESN and Conditional Restricted Boltzmann Machines (CRBM) for predicting the occurrence of railway operation disruptions has been proposed. A possible reason for the limited use of ESNs in prognostics is the difficulty of setting of the ESN architecture parameters, which heavily influence the ESN modeling capability and typically requires a high level of expertise.

Ensemble of models has been used in many application fields for prediction accuracy improvement and uncertainty quantification [7,14,22,42,52,69]. The basic idea is that the diverse models in the ensemble complement each other by leveraging their strengths and overcoming their drawbacks. Thus, the combination of the outcomes of the individual models in the ensemble improves the accuracy of the predictions compared to the performance of a single model [10,14,17,39]. Also, the distribution of the individual model outcomes provides information on the ensemble modeling error [42,97]. Different methods, such as ANN [11], Support Vector Machine (SVM) [51] and kernel learning [52], have been used with success to build the individual models. For example, it has been shown that the use of an ensemble of just-in-time kernel learning models has allowed reducing the root mean square error of the prediction of the crystal size distribution in crystallization processes by 26% with respect to an individual model [52].

With respect to improving accuracy in prognostics, an ensemble of feedforward Artificial Neural Network (ANN) has been embedded into a Particle Filter (PF) for the prediction of crack length evolution [9] and an ensemble of data-driven regression models has been exploited for the RUL prediction of lithium-ion batteries [91].

As far as the authors know, the only application of ensembles of RNNs in prognostics has been presented in [39], where the RUL predicted by various data-driven models, including a RNN, were aggregated considering three different methods, and the developed procedure was applied to RUL prediction of turbofan engines, power transformers, and cooling fans. RNN-based ensembles have also been used for time series forecasting in [6,83] and in [94], where the considered RNN models were ESN.

In this work, we develop an ensemble of ESNs, whose model architecture is optimized by Multi-Objective Differential Evolution (DE) [73]. Differently from [73], where the ESN architecture corresponding to the DE chromosome giving the most accurate RUL predictions is selected, a novelty of this work is the use of the ensemble of ESNs whose architectures are given by the chromosomes of the Pareto front reached at DE convergence. This allows increasing prediction accuracy and estimating prediction uncertainty [14].

Once the individual models of the ensemble have been generated, it is also necessary to define a strategy for the aggregation of their outcomes. Aggregation methods are typically classified into static or local [14]: a static ensemble assigns the same weight to each model, regardless of the input pattern under test, whereas a local ensemble assigns a dynamic weight to each model according to its local performance measured considering input patterns similar to that under test, such as the nearest neighbors in a validation set. The accuracy of local ensembles has been reported to be more satisfactory than that of static ensembles in several applications [14,52,53,63,85]. The challenge for the application of local aggregation methods to ensembles of recurrent models, such as ESNs, is the evaluation of the local performances of the individual models. Since the ESN output does not depend only on the current input pattern but also from the previous input pattern history, due to its memory property, the identification of similar patterns is not straightforward. For example, if an ESN has large memory, i.e., the current output depends on a large input history, the input time window to be considered for the identification of the test pattern nearest neighbors should be long. On the contrary, if an ESN has low memory, i.e., the current output depends on a short input history, the input time window to be considered should be