ARTICLE IN PRESS

Journal of Manufacturing Systems xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

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

Long short-term memory for machine remaining life prediction

Jianjing Zhang[a], Peng Wang[a], Ruqiang Yan[b], Robert X. Gao[a,*]

[a] Department of Mechanical and Aerospace Engineering, Case Western Reserve University, Cleveland, OH, 44106-7222, USA
[b] School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China

ARTICLE INFO

Keywords:
Prognostics
Health index
Deep learning
Remaining useful life

ABSTRACT

Reliable tracking of performance degradation in dynamical systems such as manufacturing machines or aircraft engines and consequently, prediction of the remaining useful life (RUL) are one of the major challenges in realizing smart manufacturing. Traditional machine learning algorithms are often constrained in adapting to the complex and non-linear characteristics of manufacturing systems and processes. With the rapid development of modern computational hardware, Deep Learning has emerged as a promising computational technique for dynamical system prediction due to its enhanced capability to characterize the system complexity, overcoming the shortcomings of those traditional methods. In this paper, a new approach based on the Long Short-Term Memory (LSTM) network, an architecture that is specialized in discovering the underlying patterns embedded in time series, is proposed to track the system degradation and consequently, to predict the RUL. The objectives of this paper are: 1) translating the raw sensor data to an interpretable health index with the aim of better describing the system health condition; and 2) tracking the historical system degradation for accurate prediction of its future health condition. Evaluation using NASA's C-MAPSS dataset verifies the effectiveness of the proposed method. Compared with other machine learning techniques, LSTM turns out to be more powerful and accurate in revealing degradation patterns, enabled by its time-dependent structure in nature.

1. Introduction

With the advancement in technology, the complexity of the machinery and system involved in today's modern manufacturing has dramatically increased over the years. To meet the demanding requirement for productivity, operational reliability and personnel safety, it is essential to have an intelligent management strategy that coordinates the scheduling and resources in a pro-active way to ensure the highest level of production while minimizing the maintenance cost [1]. Prognostics, defined as "an estimation of time to failure and risks of one or more existing or future failure modes" by the International Organization for Standardization [2], estimates the system performance degradation based on real-time analysis of its current health state, enabling condition-based inference of manufacturing system health status, and providing scientific basis for the prediction of its future physical behavior. It is playing the central role in today's manufacturing industry [3].

Over the years, the discipline of prognostics has evolved into an active research field. Recent developments in manufacturing prognostics have mainly been focused on two approaches: the Bayesian and the machine learning approach [1]. The Bayesian approach characterizes performance degradation of the manufacturing system as probability distribution and predicts its future physical behavior through the recursive steps of state prediction and update, given new sensor measurements [4,5]. Kalman Filter (KF) and Particle Filter (PF) are two main research areas in this category. As an example, a Switching Kalman Filter (SKF) is developed to infer the underlying gearbox bearing degradation process by applying the most probable filter [6]. PF is applied for tool life prediction and a novel adaptive resampling strategy is proposed for improved prediction accuracy [7]. PF is also integrated with total variation filter for better adaptation to the transient fault in heat exchanger performance tracking [8]. The machine learning approach establishes the predictive models by analyzing the related sensing data and numerically associating the discovered patterns to a specific learning task. For example, an auto-regressive integrated moving average-based (ARIMA) approach is developed to trend the vibration characteristic in rotating machinery [9]. The Quadratic Programming (QP) is also proposed to refine the fitting curve to the noisy sensing data for improved prediction accuracy [10]. Furthermore, Random Forest (RF) regression is explored to predict tool wear in milling operations to mitigate overfitting problems [11]. The methods with hybrid nature, which is based on the fusion of model-based and data-driven information, have also been reported. In [12], Dempster Shafer regression has been investigated for bearing prognosis by fusing the damage estimates from a physical fault

* Corresponding author.
E-mail address: robert.gao@case.edu (R.X. Gao).

https://doi.org/10.1016/j.jmsy.2018.05.011
Received 16 December 2017; Received in revised form 28 May 2018; Accepted 28 May 2018

Please cite this article as: Zhang, J., Journal of Manufacturing Systems (2018), https://doi.org/10.1016/j.jmsy.2018.05.011

propagation model, derived based on first principles which takes into account the factors such as damage mechanics, bearing geometry, lubrication status and material properties, and a data-driven model from the empirical fit of experimental data. The authors reported a more accurate and robust bearing damage prognostic result. In their follow-up work [13], the authors incorporated an Adaptive Neuro Fuzzy Inference System (ANFIS) for health assessment based on the data from vibration and debris monitoring. Kernel regression was used to fuse different damage estimates. Results from the experimental tests were found to be in close agreement with the prognostic estimates. In [14], the effectiveness of using features from the vibration data such as kurtosis to calibrate the helicopter intermediate gearbox pinion gear prognosis model has been confirmed, which has led to reduced uncertainties in gear remaining useful life estimation.

With the rapid development of modern computational resources for improved computational efficiency, deep learning has become one of the emerging research areas in prognostics and has attracted considerable attention recently due to its enhanced capability in complex system modeling [15]. Inspired by the biological brain architecture, deep learning refers to the supervised/unsupervised machine learning technique that automatically learns hierarchical patterns in deep structures [16]. Four major deep architectures exist [17]: *Auto-encoder* is a simple neural network capable of learning efficient representations of the input data in an unsupervised manner and is often used in network pre-training. The *Deep Belief Network* (DBN) is a feed-forward neural network with multiple hidden layers, capable of revealing deep data patterns. It consists of a stack of *Restricted Boltzmann Machine* (RBM) and a supervised perceptron [18]. *Convolutional Neural Network* (CNN) emerged from the research of human brain cortex. It is developed to extract abstract features by sequential operations of convolution and pooling [19]. The *Recurrent Neural Network* (RNN) is a deep architecture that retains the recent memories of input patterns. Its variant, the Long Short-Term Memory (LSTM) network further addresses the problem of capturing the long-term memory [20,21].

Recently, a growing number of deep learning research has been reported in manufacturing industry. An auto-encoder based prognostic method has been proposed to precisely identify the bearing degradation starting point [22], and the stacked denoising auto-encoder (SDA) is investigated for bearing fault identification [23]. In the category of DBN, for example, a DBN-based method is developed for material removal rate prediction in polishing [24]. In another study, a new regularization term is proposed for RBM to predict machine RUL [25]. As for CNN, it is constructed for fault inference in semiconductor manufacturing process [26], and used to learning features from time-frequency spectra directly for automatic fault related pattern recognition [27]. In addition, LSTM has been applied for predicting tool wear [28], fuel cell voltage output [29], lithium-ion battery cell capacity [30] and bearing health state [31].

Inspired by these prior research, this paper presents a bi-directional LSTM-based approach for characterizing system degradation behavior and subsequently predicting remaining useful life (RUL). The long-term dependency characteristic embedded in the LSTM structure is envisioned to capture inter-relationship of the time series data measured from the monitored system, leading to better prediction of its future behavior. Different from the reported research that only utilizes the LSTM in a forward manner, in the bi-directional network, each sequence is presented forwards and backwards in two separate LSTMs, allowing access complete information before and after each time step in each sequence. Furthermore, the reverse path LSTM further smooths the data and mitigates the noise impact.

The remainder of the paper is organized as follows. Section 2 outlines the concept of the proposed method for system health prognostics. Section 3 illustrates a case study using the NASA's C-MAPSS dataset to validate the effectiveness of the proposed method, followed by the data analysis and discussion in Section 4. Finally, the conclusions are drawn in Section 5.

2. Proposed prognostic method based on LSTM

In this section, a bi-directional LSTM is proposed for system performance degradation tracking and RUL prediction. First, the system degradation tracking and RUL prediction are mathematically formulated, followed by the theoretical background of the RNN and LSTM. Finally, the structure and training algorithm of the bi-directional LSTM network are presented.

2.1. Problem formulation

Given sensing data collected from n time steps $X = [\boldsymbol{x}_{(1)}, \boldsymbol{x}_{(2)}, \dots, \boldsymbol{x}_{(n)}]$ and the corresponding underlying system states $Y = [y_{(1)}, y_{(2)}, \dots, y_{(n)}]$, the system performance tracking is to find out the variation pattern associated with Y over time through exploring the variation of sensing data. Mathematically, it can be defined as:

$$\boldsymbol{x}_{(n)} = f\,[\boldsymbol{x}_{(n-1)}, \boldsymbol{x}_{(n-2)}, \dots, \boldsymbol{x}_{(1)}] \Rightarrow y_{(n)} = g\,[y_{(n-1)}, y_{(n-2)}, \dots, y_{(1)}] \tag{1}$$

Eq. (1) reveals: 1) current system performance relies on its preceding performance and 2) system performance evolution pattern is revealed by the sensing data variation. System performance can be described by physically defined parameters or artificially defined health index, denoted by system state Y in this paper. Then the tracking analysis of historical sensing data through either Bayesian inference or machine learning techniques is to find the two system evolution functions f and g, and subsequently, the future system performance can be predicted via the estimated function g:

$$Y' = [y'_{(n+1)}, y'_{(n+2)}, y'_{(n+3)}, \dots] \tag{2}$$

$$y'_{(n+1)} = g\,[y_{(n)}, y_{(n-1)}, \dots, y_{(2)}] \tag{3}$$

It should be noted the predicted system states are labeled as Y' to differentiate from the real values Y. Given a predefined failure threshold $y_{\text{threshold}}$, the RUL can be predicted as the first passage time when the future system state passes the failure threshold:

$$RUL_{predict} = \inf\{k\colon\, y'_{(n+k)} \le y_{threshold}\} \tag{4}$$

The main challenge in the system tracking process lies in capturing the non-linearity associated with the system variation and data uncertainty (e.g. due to environmental disturbance and sensor failure). To handle with the non-linearity, a deep learning-based system modeling technique is presented in this study. Specifically, a Long Short-Term Memory (LSTM) structure is investigated, as it is powerful in discovering the variation pattern underlying a time series. To deal with the data uncertainty, a bi-directional LSTM network is proposed, in which information flow through the LSTM cells forwards for prediction and backwards for ruling out the disturbance and smoothing the prediction.

2.2. Recurrent neural network and LSTM

Recurrent neural network (RNN) is composed of a series of recurrent neurons, as shown in Fig. 1 [17]. The subscript in the parenthesis indicates time step for the recurrent neuron. Different from the standard neuron, the output from a recurrent neuron is connected to the next

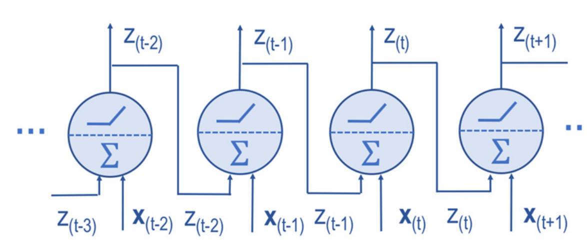

Fig. 1. Recurrent nature of RNN as each neuron output is dependent on the corresponding input as well as the output from the previous step.