J Intell Manuf (2012) 23:227–237
DOI 10.1007/s10845-009-0356-9

An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring

Zhigang Tian

Received: 2 June 2009 / Accepted: 20 October 2009 / Published online: 4 November 2009

Abstract Accurate equipment remaining useful life prediction is critical to effective condition based maintenance for improving reliability and reducing overall maintenance cost. In this paper, an artificial neural network (ANN) based method is developed for achieving more accurate remaining useful life prediction of equipment subject to condition monitoring. The ANN model takes the age and multiple condition monitoring measurement values at the present and previous inspection points as the inputs, and the life percentage as the output. A function generalized from the Weibull failure rate function is used to fit each condition monitoring measurement series for a failure history, and the fitted measurement values are used to form the ANN training set so as to reduce the effects of the noise factors that are irrelevant to the equipment degradation. A validation mechanism is introduced in the ANN training process to improve the prediction performance of the ANN model. The proposed ANN method is validated using real-world vibration monitoring data collected from pump bearings in the field. A comparative study is performed between the proposed ANN method and an adapted version of a reported method, and the results demonstrate the advantage of the proposed method in achieving more accurate remaining useful life prediction.

Keywords Remaining useful life · Prediction · Artificial neural network · Accurate · Bearing

Z. Tian (✉)
Concordia Institute for Information Systems Engineering, Concordia University, 1515 Ste-Catherine Street West EV-7.637, Montreal H3G 2W1, Canada
e-mail: tian@ciise.concordia.ca

Introduction

Condition based maintenance (CBM) aims at achieving reliable and cost-effective operation of engineering systems such as aircraft systems, wind turbine generators, hydro power plants and manufacturing systems (Jardine et al. 2006). In CBM, condition monitoring data, such as vibration data, oil analysis data and acoustic data, are collected and processed to determine the equipment health condition; Future health condition and thus the remaining useful life (RUL) of the equipment is predicted; and optimal maintenance actions are scheduled based on the predicted future equipment health condition, so that preventive replacements can be performed to prevent unexpected failures and minimize total maintenance costs (Jardine et al. 2006; Levitin 2005; Vachtsevanos et al. 2006; Liao et al. 2006; Inman et al. 2005). Accurate health condition prediction is the critical to effective implementation of condition based maintenance.

Existing equipment health condition and RUL prediction methods can be roughly classified into model-based (or physics-based) methods and data-driven methods. The model-based methods predict the remaining useful life using damage propagation models based on damage mechanics (Vachtsevanos et al. 2006; Inman et al. 2005). Marble and Morton (2006) predicted the health condition of propulsion system bearings using the bearing spall propagation physical model and finite element model. Kacprzynski et al. (2002) developed an approach for gear health condition prediction based on gear tooth crack initiation and propagation physical models. Li and Lee (2005) proposed an approach for remaining useful life prediction of gears with a fatigue tooth crack based on a gear meshing stiffness identification model, a gear dynamic model and a fracture mechanics model. If properly used, physics-based models can greatly improve the RUL prediction accuracy. However, damage propagation processes are typically very complex, and authentic physics-based models are difficult to build for many components and systems.

Data-driven methods utilize collected condition monitoring data for RUL prediction. Jardine et al. developed the Proportional Hazards Model approach for CBM, where health condition indicators are predicted using the transition probability matrix (Jardine et al. 2006; Banjevic et al. 2001; Makis and Jardine 1992). However, in order to ensure the CBM optimization efficiency, each health condition indicator, or covariate, has to be divided into a small number of bands, which affects the health condition indicator prediction accuracy. Moreover, the state transition rates are hard to accurately predict if a large amount of condition monitoring data is not available. Dong and He developed hidden semi-Markov models based methods for equipment diagnosis and prognosis (Dong et al. 2006; Dong and He 2007a,b). Artificial neural networks (ANNs) have been considered to be very promising tools for equipment health condition and RUL prediction due to their adaptability, nonlinearity, and arbitrary function approximation ability (Tian and Zuo 2009; Rojas 1996; Huang et al. 2007; Tse and Atherton 1999). Neural network methods do not assume the analytical model of the damage propagation, but aim at modeling the damage propagation process, or degradation process, based on the collected condition monitoring data using neural networks and perform health condition prediction. Lee et al. (2006) proposed to extract an overall health indicator based on the collected condition data, and predict future health indicator values using the autoregressive moving average (ARMA) method and Elman neural networks. Gebraeel et al. developed ball bearing remaining life prediction methods based on feedforward neural networks (Gebraeel et al. 2004; Gebraeel and Lawley 2008). The output of the ANN model was a condition monitoring measurement, such as overall vibration magnitude. However, the failure threshold values are typically hard to clearly define in many practical applications. Wu et al. (2007) proposed another RUL prediction method based on ANN, where the ANN output was the life percentage, or in another word, one minus the remaining life percentage. The accuracy and robustness of this method, though, can be further improved.

In this paper, we propose a new ANN based method for achieving more accurate RUL prediction. The ANN model takes the age and multiple condition monitoring measurement values at discrete inspection points as the inputs and the life percentage as the output. The prediction accuracy is improved mainly by reducing the effects of the noise factors that are irrelevant to the equipment degradation in the condition monitoring data, and by utilizing the validation mechanism in the ANN training process. The remainder of this paper is organized as follows. The proposed ANN method is presented in section "The proposed ANN remaining useful life prediction method". Section "Case study" discusses the case study, in which the proposed ANN method is validated using real-world vibration monitoring data collected in the field from pump bearings, and a comparative study is performed. The final section presents the conclusions.

The proposed ANN remaining useful life prediction method

In this section, we first present a modified ANN method based on Wu's method (Wu et al. 2007), which can be used to deal with condition monitoring data collected at discrete inspection time points. A new ANN RUL prediction method is proposed next. The procedure of the proposed approach is discussed in details and illustrated using examples.

The modified Wu's method

In this section, we present a modified version of the ANN prediction method developed by Wu et al. (2007), and refer to it as the "Modified Wu's method". In the original version of Wu's method, they considered only one condition monitoring measurement, and developed a feedforward ANN model with three inputs and one output, with current time t, the current measurement and the measurement at $(t-1)$ as the inputs and the life percentage as the output. The life percentage is the ratio between the current age of the unit and its failure time. However, this model cannot handle many practical situations in which condition monitoring data is collected at discrete inspection time points that are not equally spaced, such as the vibration monitoring data for many pump bearings, and the oil analysis data for truck transmissions (Banjevic et al. 2001). Moreover, typically there are more than one measurement that are correlated with the degradation of the equipment, and they should be incorporated into the ANN model to produce more accurate RUL prediction results.

Bearing these in mind, we propose a modified version of Wu's model. The structure of the modified model is shown in Fig. 1. The ANN model has an input layer, an output layer and two hidden layers. The reason we use an ANN with two hidden layers instead of one is that we find it is able to produce more reliable results according to our experiments. The inputs to the ANN include the age values and the condition monitoring measurements at the current inspection point and those at the previous inspection point. We only consider two measurements in Fig. 1, but more measurements can be handled by adding more input nodes to the ANN model. Specifically, t_i is the age of the equipment at the current inspection point t, and t_{i-1} is the age at the previous inspection point $i-1$; z_i^1 and z_{i-1}^1 are the values of measurement 1 at the current and previous inspection points, respectively; z_i^2 and z_{i-1}^2

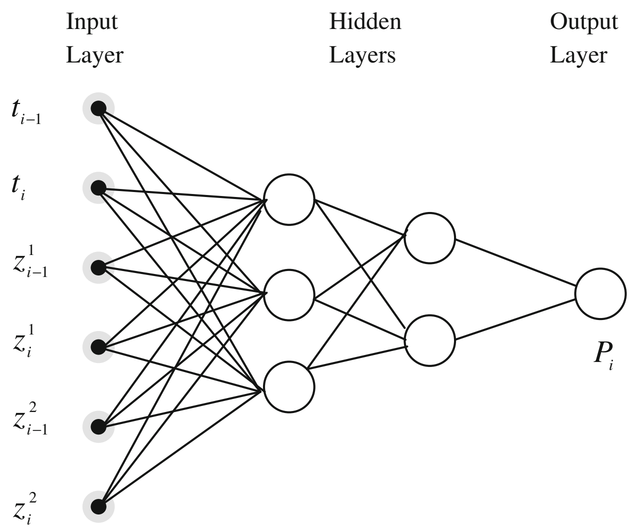

Fig. 1 Structure of the Modified Wu's ANN model

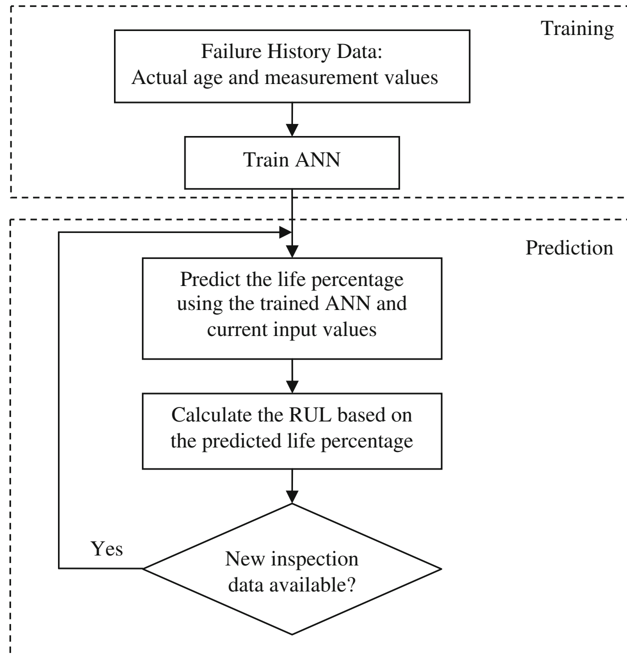

Fig. 2 Procedure of the Modified Wu's method

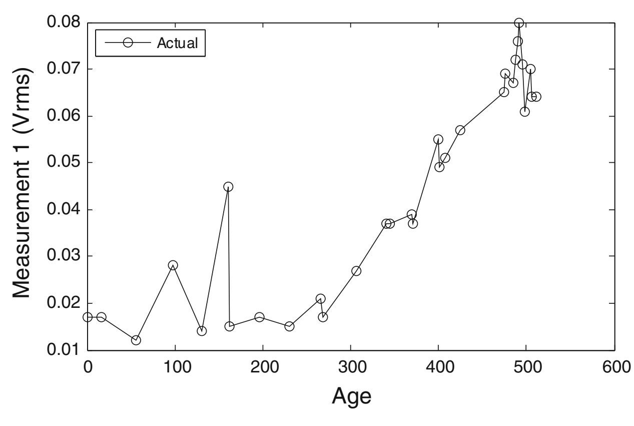

Fig. 3 An actual measurement series for a sample failure history (measurement 1: vibration magnitude in the horizontal direction in frequency band 5)

are the values of measurement 2 at the current and previous inspection points, respectively.

The Modified Wu's method can handle unequally spaced inspection points and multiple condition monitoring measurements. Adapted from Wu's original method (Wu et al. 2007), the procedure of the Modified Wu's method is shown in Fig. 2. Several key points that are worth noting are:

(1) Actual condition monitoring measurements are used to train the ANN model. The data is from the available failure histories in the training set.
(2) When predicting the RUL of the current monitored equipment using the trained ANN, actual condition monitoring measurements are used as the inputs.
(3) No validation process is used when training ANN (Wu et al. 2007).

Issues with using actual measurement values as the inputs to the ANN model

As described before, the Modified Wu's method uses actual condition monitoring measurement values as the inputs to the ANN model. However, when the measurements are collected at inspection points in practical applications, there are usually external noise factors that will affect the measurement values. As an example, Fig. 3 shows an actual measurement series for a bearing failure history, which was collected from a pump in the field. A history refers to the period of a unit from the beginning of its life to the end, failure or suspension, of its life, and the inspection data collected during this period. The bearing failed at age 511 days. It can be seen that although the measurement shows a generally increasing trend, there are large fluctuations at multiple places. On the other hand, the deterioration of the health condition of a component, such as the propagation of a spall in a bearing or the propagation of a root crack or the surface wear in a gear, is generally a monotonic process. Therefore, directly feeding the current and the previous actual measurement values to the ANN model and mapping it to life percentage, the health condition measure, will introduce noises into the ANN model, and compromise its capability to accurately represent the health condition of the equipment.

Function for fitting the actual measurement series

To address the above-mentioned issues with using actual condition monitoring measurements as the inputs to the ANN model, we propose to use an appropriate function to fit the measurement series first, and use the fitted measurement values as inputs to the ANN model, so as to better represent the deterioration of a piece of equipment. We propose a function which is generalized from the Weibull distribution failure rate function. In reliability analysis, it is the failure rate measure that indicates the health condition of a certain type of component at a given time. Weibull distribution is very powerful in representing various practical lifetime distributions, and flexible enough to represent distributions with different scales and shapes (Kuo and Zuo 2003). Thus, the following function generalized from the Weibull distribution failure rate function is used to fit the measurement series:

$$\hat{z}(t) = Y + K\frac{\beta}{\alpha^{\beta}}t^{\beta-1}, \tag{1}$$

where t is the age of the unit, $\hat{z}(t)$ is the fitted measurement value, and are the scale parameter and the shape parameter, respectively. $\left(\beta/\alpha^{\beta}\right)t^{\beta-1}$ is the failure rate function for the 2-parameter Weibull distribution. Parameter K is introduced to scale the fitted measurement values to any ranges, and parameter Y is used to indicate the value when the age is 0. We refer the function in Eq. 1 as the "Generalized Weibull-FR function". Thus, the Generalized Weibull-FR function has four parameters, which can be determined using least-square method based on the actual measurement series. In this work, we use genetic algorithm (GA) to find the optimal values for the four parameters because of the good global optimization performance of GA (Levitin 2005; Levitin et al. 1998). We have tested the Generalized Weibull-FR function using many actual measurement series collected from the field, and found that the function is capable of fitting all the tested measurement series very well.

Example 1 The function is used to fit the actual measurement series shown in Fig. 3, which is collected from a pump bearing in the field. There are totally 31 inspection points, the bearing failed when the age was 511 days. Using GA, we can find the optimal values for the parameters in the Generalized Weibull-FR function as shown in Eq. 1, so as to best fit the measurement series. The obtained function with the optimal parameters is given as follows:

$$\hat{z}(t) = 0.021 + 8.62\frac{3.99}{541.80^{3.99}}t^{3.99-1}, \tag{2}$$

The results are shown in Fig. 4, in which the fitted values are represented by "*". We can also observe that the fitted measurement series give a better indication of the degradation of the equipment.

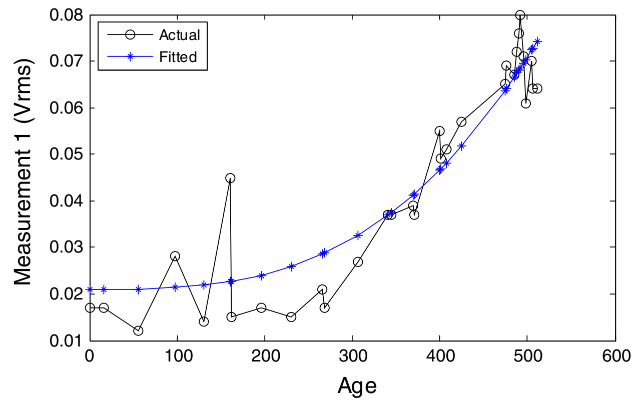

Fig. 4 An actual measurement series and the fitted measurement series (measurement 1: vibration magnitude in the horizontal direction in frequency band 5)

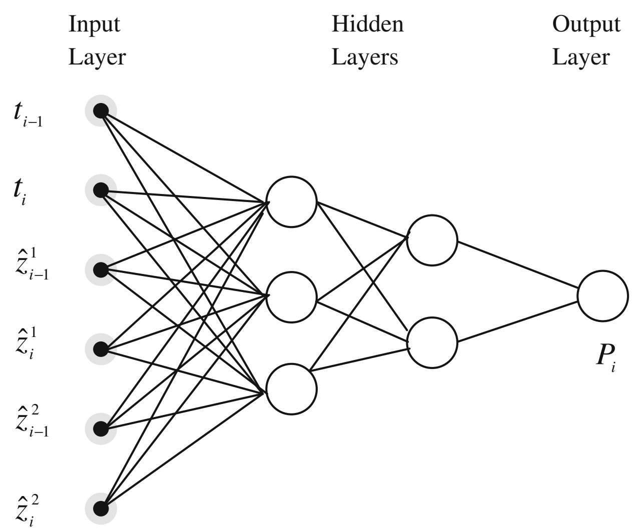

Fig. 5 Structure of the proposed ANN model

The proposed ANN model

The structure of the proposed ANN model is shown in Fig. 5. The proposed ANN model uses the fitted measurement values instead of the actual measurement values as the inputs. Specifically, t_i and t_{i-1} are the age values at the current inspection point i and the previous inspection point $i-1$, respectively; $\hat{z}^1_i$ and $\hat{z}^1_{i-1}$ are the fitted values of measurement 1 at the current and previous inspection points, respectively; $\hat{z}^2_i$ and $\hat{z}^2_{i-1}$ are the fitted values of measurement 2 at the current and previous inspection points, respectively. With these data at the current and the previous inspection points, we take into account the age and measurement values as well as the changes of the measurement values at these points, and these important information will be processed by ANN to estimate the remaining useful life. We only use data at the two inspection time points instead of incorporating data at more past inspection points, because ANNs with less input nodes have better generalization capability, and the experiments results in this work show that adding more input nodes will not improve the RUL prediction performance.

The output of the proposed ANN model is the life percentage, denoted by P_i, which is the same as that in the Modified Wu's method. As an example, suppose the failure time of a bearing is 511 days and, at an inspection point t, the age is 400 days, then the life percentage at inspection point i would be:

$$P_i = 400/511 \times 100\% = 78.3\%.$$

It is worth justifying the use of the life percentage to represent the health condition of a piece of equipment. It is usually hard to find a single indicator that generally changes monotonically with time to represent the health condition of a piece of equipment. Even if we can find such an indicator, it is very difficult to establish a failure threshold value for the indicator. However, it is true that the health condition of a piece of equipment deteriorates with time, and, without loss of generality, we can assume that the true inherent health condition index increases monotonically with time. Since it is hard to determine the true inherent health condition index based on the collected condition monitoring measurements, we can try to identify a measure that has a monotonic mapping relationship with the true inherent health condition index. We believe "life percentage" is a good option for this purpose: (1) the mapping between the inherent health condition index and the life percentage is monotonically non-decreasing, and (2) life percentage is also able to indicate when the failure occurs, that is, the failure occurs when the life percentage reaches 100%.

The remaining life prediction problem is a complex non-linear problem. The objective is to predict the equipment remaining useful life based on the available age and condition monitoring data. The relationship between the 6 inputs, the age and condition monitoring data, and the output, the life percentage, is very complex and non-linear. Because of the capability in modeling complex non-linear relationship, ANN is a powerful tool to be used to address the remaining useful life prediction problem in the proposed approach.

Training of the proposed ANN model

The training set for the proposed ANN, shown in Fig. 5, is formed by the age and fitted measurement values at the inspection points for the available failure histories. For example, suppose we have four failure histories, and they have 30, 30, 40, 40 inspection points, respectively. So for failure history 1, we have 29 training pairs, including the six inputs and the corresponding one output, since we need data at the current and previous inspection points in the input vector. Thus, the total number of training pairs is:

$$29 + 29 + 39 + 39 = 136.$$

For each condition monitoring measurement in a failure history, we first use the Generalized Weibull-FR function to fit the measurement series, and then use the fitted measurement values in the ANN training set.

A critical issue of using ANN is to avoid overfitting the network. If an ANN is overfitted, noise factors will be modeled in the network, which affects the generalization capability of ANN, and thus affects the prediction accuracy. Wu et al. (2007) did not consider this issue in their work, thus there is no mechanism as to when to stop training so as to achieve a trained ANN that can best model the mapping relationship between the inputs and outputs without overfitting. A widely used approach for avoiding overfitting is the use of validation set (Rojas 1996). That is, during the ANN training process, the mean square error for the training set and that for the validation set are calculated. Both of the mean square errors drop early in the training process because the ANN is learning the relationship between the inputs and the outputs by modifying the trainable weights based on the training set. After a certain point, the mean square error for the validation set will start to increase, because the ANN starts to model the noise in the training set. Thus, the training process can be stopped at this point, and the trained ANN with good modeling and generalization capability can be achieved.

Another question is how to construct the validation set. One method is dividing the available failure histories into two groups, one used as the training set and the other as the validation set. The percentage of histories used in the validation set is typically around 40% (Rojas 1996). This causes a problem if we do not have a lot of failure histories, which is the case in many practical applications. As an example, there are 11 failure histories in the problem to be investigated in the case study later. If we use this method to construct the validation set, only 6–7 histories can be included in the training set. The valuable information in the other 4–5 failure histories will not be directly used to modify the trainable weights, which will affect ANN's ability to model the relationship between the inputs and output based on available data. In this paper, we propose to construct the validation set using the actual measurement data for all failure histories, and, as mentioned before, use the fitted measurement data to construct the training set. In this way, data based on all failure histories can be taken advantage of for modifying the trainable weights in the ANN training process, and the validation process can help to avoid overfitting the network. Experiments based on practical condition monitoring data have shown that the use of such a validation set in the training process can lead to more accurate and robust predictions.

Based on the training set and the validation set, the ANN is trained using the Levenberg–Marquardt (LM) algorithm. The details of the LM algorithm for training feedforward neural networks can be found in Rojas (1996).

Because of the uncertainty in the ANN training algorithms such as the LM algorithm, with the same training set and validation set, typically we will not obtain the same neural network after training. In this work, we train the ANN five times, and select the one with the lowest training mean square error (MSE). This is actually another advantage of using the validation mechanism in the training process. Without the validation process, a lower training MSE does not necessarily means a better network, and in many cases it is not. Thus, we cannot select the best trained ANN based on the training MSE if we do not have the validation mechanism. However, if we do have the validation mechanism, as in the proposed ANN method, a lower training MSE will indicate a trained ANN with better modeling and generalization capabilities, and thus an ANN with better prediction capability.

Procedure of the proposed ANN method

The procedure of the proposed method is shown in Fig. 6. The explanation of the procedure is given as follows.

- Step 0: We start from the available failure history data, which includes the age values and actual condition monitoring measurement values at inspection points for each failure history.
- Step 1: Each measurement series for a failure history is fitted using the Generalized Weibull-FR function. The age values and the fitted measurement values at inspection points for all failure histories are used to construct the ANN training set.
- Step 2: The ANN validation set is constructed using the age values and the actual measurement values at inspection points for all failure histories.
- Step 3: Train the ANN model based on the training set and the validation set using the LM algorithm.
  After Step 3, the ANN training process is completed, and the trained ANN model is obtained for RUL prediction. For a piece of equipment which is currently being monitored, the following steps are for the RUL prediction for the equipment.
- Step 4: At a given inspection point, we first fit each measurement series based on the measurement values up to the current time. The fitted measurement values at the current and previous inspection points, as well as the age values at these two inspection points, are used as the inputs to the trained ANN model.
- Step 5: The predicted life percentage value at the current point is calculated using the trained ANN model.
- Step 6: The RUL is calculated based on the current age of the equipment and the predicted life percentage. For example, if the current age is 400 days and the predicted life percentage is 80%, the predicted failure time would be 400/80% = 500 (days), and the RUL would be 500 − 400 = 100 (days).

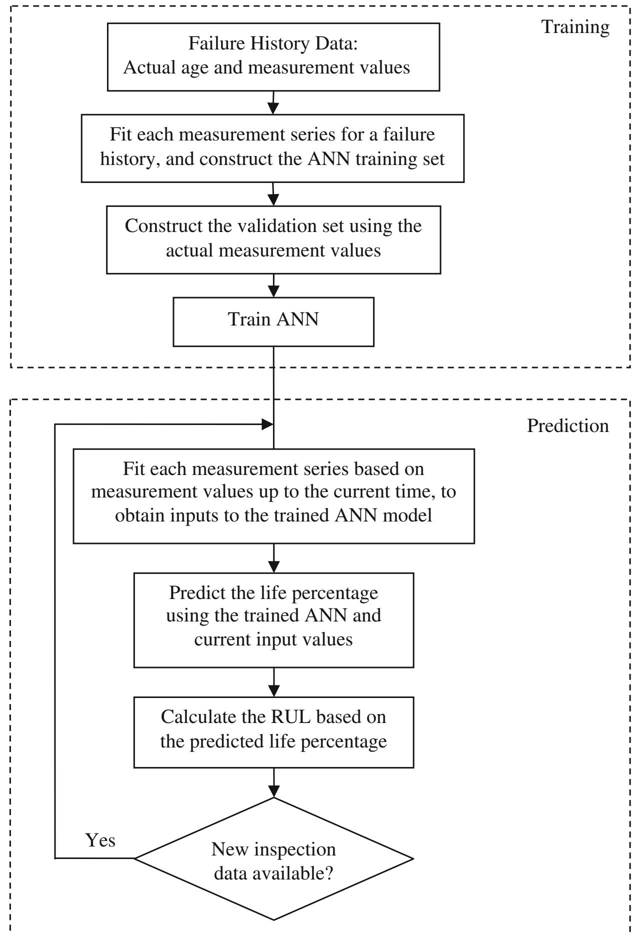

Fig. 6 Procedure of the proposed ANN method

- Step 7: When new inspection data is available, repeat Step 4 to Step 6 based on the available data, and update the RUL prediction.

Discussions of the proposed ANN method

The proposed ANN aims at producing more accurate equipment RUL prediction. The following two points are worth mentioning.

(1) The proposed ANN method predicts the life percentage of the unit. Thus, we do not have to define the failure threshold, which is hard to clearly define in many practical applications. The unit is failed when the life percentage reaches 100%.
(2) The proposed ANN method does not require the identification of the time when the incipient fault occurs. In practical applications, it is usually hard to identify the incipient fault time for an available failure history based on the collected condition monitoring measurements.

However, if the incipient fault time can be properly identified in some applications, the incipient fault times can be considered to be age 0 for the histories. The proposed ANN method can still be used, and its prediction accuracy is expected to be improved comparing to directly using the available failure histories.

Case study

Case study introduction

In this section, condition monitoring data collected from the field is used to validate the proposed ANN method for RUL prediction. The condition monitoring data were collected from bearings on a group of Gould pumps at a Canadian kraft pulp mill company (Stevens 2006). A comparative study is performed between the proposed ANN method and the Modified Wu's method to demonstrate the advantage of the proposed method.

The pump bearing subject to condition monitoring is shown in Fig. 7 (Stevens 2006), in which the arrows point to the location of the bearing. There are 11 bearing failure histories, and they are used in this work. Vibration monitoring data were collected from the pump bearings using accelerometers, and processed and recorded using data acquisition systems. The collected vibration measurements include velocities in three directions, axial, horizontal and vertical, and in each of these directions, the velocity spectrum is obtained in five frequency bands. In addition, overall velocity and acceleration are also measured in the three directions. However, not all the measurements are correlated with the degradation of the bearings. Significance analysis can be used to identify the significant condition monitoring measurements (Lin et al. 2006). The significance analysis capability built in the software EXAKT is utilized (Rojas 1996), and two measurements are identified to be significant: the velocity measurements in vibration frequency band 5 in the horizontal and the vertical directions, respectively. We refer to these two measurements as Measurement 1 and Measurement 2. There are totally 310 inspection points for these 11 bearing failure histories. The inspection points for a history are not equally spaced, and the number of inspection points for a history ranges from 8 to 45. As mentioned before, Fig. 3 actually shows the actual measurement series for Measurement 1 for one of the failure histories. The bearing failure time ranges from 63 days to 1,468 days, and the average failure time is approximately 790 days.

RUL prediction using the proposed ANN method

The bearing condition monitoring data mentioned above are used to test the RUL prediction performance of the proposed ANN method, which is described in Fig. 6. The following approach is used to test the RUL prediction performance of the proposed ANN method. When we apply the ANN prediction method in practical situations, we first use the available data, failure history data, to train the ANN, and then use the trained ANN model to predict the RUL of a new unit which has not yet failed. In this case study, we have 11 bearing failure histories. Thus, we can use 1 bearing failure histories to construct the ANN test set to test the prediction performance, and use the remaining 10 failure histories to construct the ANN training set and the validation set to train the ANN model. After that, we can use a different failure history to construct the ANN test set to test the prediction performance and use the remaining 10 failure histories to train the ANN. We can repeat this process until we go through all the failure histories. Using the approach described above, the test failure history is not involved in the ANN training process, and we can make full use of the data available to test the prediction performance of the ANN method at as many inspection points as possible to achieve accurate prediction performance evaluation.

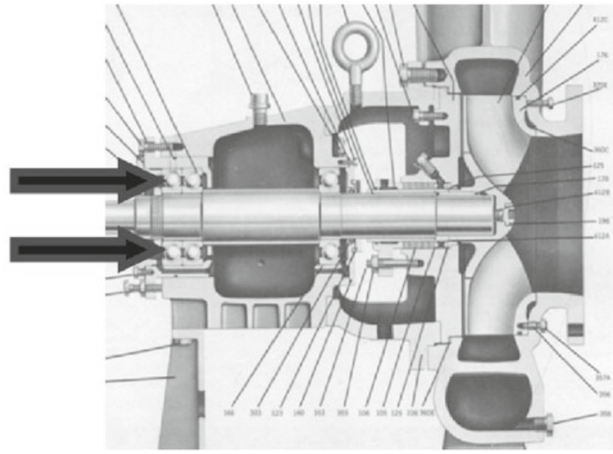

Fig. 7 The pump bearings subject to condition monitoring (Stevens 2006)

The Generalized Weibull-FR function is used to fit each of two measurements for the 11 failure histories. Results show that the function can fit each of the measurement series very well, although different measurement series have different length, different values ranges, and times that the measurement values start to show obvious increase. Figure 4 shows the actual measurement series and the corresponding fitted measurement series for Measurement 1 for one of the failure histories. For the 10 failure histories used for ANN training, the fitted measurement values at the inspection time points and the corresponding age values are used to construct the ANN training set, and the actual measurement values are used to construct the validation set. The ANN model we use has two hidden layers with three hidden neurons in the

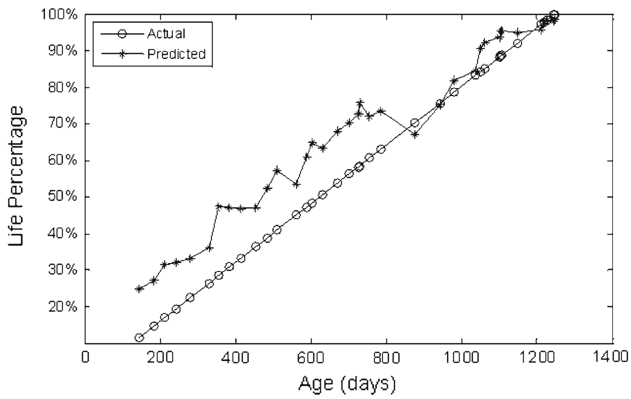

Fig. 8 Prediction results for a sample test failure history

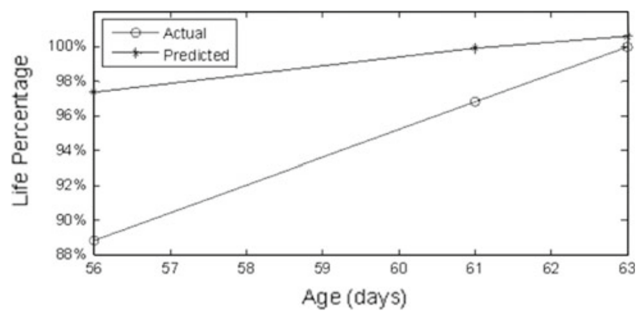

Fig. 9 Prediction results for the shortest test failure history

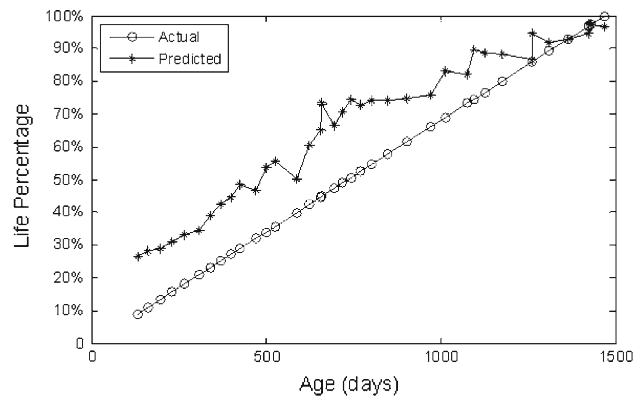

Fig. 10 Prediction results for the longest test failure history

first hidden layer and two hidden neurons in the second hidden layer. From our experiments, such ANN configuration is found to be able to produce better prediction results compared to other ANN configurations with different number of hidden neurons. The LM algorithm for the ANN training is run five times, and the trained ANN corresponding to the lowest training MSE is selected for the prediction performance testing. The algorithm is implemented in MATLAB.

The prediction performance of the trained ANN is evaluated using the test set constructed based on the remaining one failure history. We test the prediction performance starting from inspection point 6, since measurement values at the first several inspection points are needed to fit the measurement series and generate the fitted measurement values used as the inputs to the trained ANN model. The predicted life percentage values are obtained, and are compared with the actual life percentage values, which are calculated based on the age values at the inspection points and the bearing failure time. Each of the 11 failure histories is used to construct the test set once to test the prediction performance. The prediction results for a sample test failure history are shown in Fig. 8. The horizontal axis is the age of the bearing, and the failure time of the bearing is 1,246 days. The vertical axis is the life percentage. The actual life percentage values are represented by "o", and the life percentage values predicted using the trained ANN model are represented by "*". We can see that the predicted life percentage values are close to the actual values, and the prediction becomes more accurate when it is close to the failure time.

The prediction results for the shortest test failure history, with the duration of 63 days, are shown in Fig. 9. There are only three inspection points at which the prediction performance is tested, and the average prediction error is 4.2%. The prediction results for the longest test failure history, with the duration of 1,468 days, are shown in Fig. 10. The average prediction error considering all the inspection points is 14.6%, and that considering only the last five prediction point is 3.2%. From these two cases, we can see that the proposed ANN approach can produce accurate remaining life prediction results for both short test histories and long test histories.

The Average Prediction Error, denoted by $\bar{e}$, is used to quantify the prediction performance:

$$\bar{e} = \frac{1}{n} \cdot \sum_{k=1}^{n} |P_k - \hat{P}_k| \cdot 100\%, \tag{3}$$

where n is the number of inspection points for testing the prediction performance, P_k is the actual life percentage at inspection point k, and $\hat{P}_k$ is the predicted life percentage at inspection point k. First we calculate the Average Prediction Error over all the inspection points in the test sets. Since each of the 11 failure histories is used to construct the test set once, and the prediction performance is tested starting from inspection point 6 for each test failure history, we have totally 255 inspection points at which the prediction performance is tested. For the same test failure history, the resulting trained ANN model, and thus the prediction results, will not be the same every time we run the algorithm. Thus, for each test failure history, we repeat the ANN training and prediction process 10 times, and use the average prediction errors as the prediction error for the inspection points in the test failure history. Specifically, for an inspection point at which the prediction performance is tested, the absolute value of the prediction error is calculated as:

$$|P_k - \hat{P}_k| = \frac{1}{10}\sum_{r=1}^{10}|P_k - \hat{P}_{kr}|, \tag{4}$$

where k is the inspection point index, P_k is the actual life percentage at inspection point k, and $\hat{P}_{kr}$ is the predicted life percentage when we run the ANN training and prediction process the rth time. The value obtained from Eq. 4 can be used in Eq. 3 to calculate the Average Prediction Error. Considering all of the 255 inspection points in the test sets, the Average Prediction Error is 11.1%.

However, the prediction accuracy late in the life of the unit is more important than that early in its life, because this will more likely affect the decision on whether or not preventive replacement should be performed at the current inspection point. To investigate the prediction accuracy late in the unit life, we test the prediction performance at the last 5 inspection points for each test failure history (use all the inspection points in the test set if there are less than 5 inspection points in the test set for a test failure history). Totally, there are 53 such inspection points for this purpose, which accounts for approximately 21% of the total inspection points in the test sets, and the corresponding actual life percentage ranges from 87.3% to 100%. Considering only these 53 inspection points late in the life of the unit, the Average Prediction Error is 5.7%, which is significantly better than that when considering all of the inspection points in the test sets.

From the prediction results, we find that the prediction results for one of the failure histories are much worse, by an order of magnitude, than those for other failure histories. From the measurement series plots for this particular failure history, we can observe very large fluctuations late in the life of the bearing, and the measurement series do not generally follow the pattern exhibited by those in other failure histories. It is impossible to retrieve what occurred for this unit during the period. But it is very likely that some unusual events or factors affected the collected condition monitoring measurements, which makes this failure history an outlier. Thus, we decide to evaluate the prediction performance without considering this particular failure history. In this way, the Average Prediction Error considering all of the inspection points in the test sets is 10.6%, and the Average Prediction Error considering only the last 5 inspection points for each test failure history is 3.4%. These are satisfactory RUL prediction results, and can assist the condition based maintenance optimization. The complete RUL prediction results for the individual test histories are presented in Table 1. From Table 1, we can also obtain the same observation as we do from Fig. 8, that is, the remaining useful life prediction becomes more accurate when it is close to the failure time. For all the test histories except the outlier failure history #7, the prediction error at the last five inspection points is smaller than the overall prediction error considering all the test inspection points.

As another measure for the prediction accuracy late in the life of a unit, we can look at the Average Prediction Error for the prediction results where the predicted life percentage values are between 90% and 100%. Using the proposed ANN method, approximately 30% of the prediction results are covered by this criterion, and the Average Prediction Error is 3.65%, which is very satisfactory. The RUL prediction results are summarized in Table 2, where $\bar{e}_{\mathrm{All}}$ represents the Average Prediction Error considering all the inspection points in the test sets $\bar{e}_{\mathrm{L5}}$, represents the Average Prediction Error considering only the last 5 inspection points for each test history, and $\bar{e}_{90-100}$ represents the Average Prediction Error considering only the prediction results where the predicted life percentage values are between 90% and 100%.

Table 1 The RUL prediction results for different test histories using the proposed ANN method

Test history #	Average prediction error ($\bar{e}$)		
	$\bar{e}_{\mathrm{All}}$ (%)	$\bar{e}_{\mathrm{L5}}$ (%)	$\bar{e}_{90-100}$ (%)
1	6.4	4.9	1.4
2	15.6	2.2	2.7
3	9.4	5.4	7.2
4	10.5	3.3	3.4
5	8.3	5.4	5.1
6	4.2	4.2	4.1
7	16.0	29.0	27.7
8	10.9	3.1	4.8
9	17.5	1.5	1.9
10	8.9	0.9	1.7
11	14.6	3.2	3.1

Table 2 The RUL prediction results

	Average prediction error ($\bar{e}$)		
	$\bar{e}_{\mathrm{All}}$ (%)	$\bar{e}_{\mathrm{L5}}$ (%)	$\bar{e}_{90-100}$ (%)
The Proposed ANN method	10.6	3.4	3.65
The Modified Wu's method	12.5	4.9	8.28

Comparative study between the proposed ANN method and the Modified Wu's method

The Modified Wu's method, which can handle unequally spaced inspection points and multiple measurements, is used to perform RUL prediction based on the same bearing condition monitoring data used in this case study, and the prediction results are compared with those obtained using the proposed ANN method. To ensure a fair comparison, the same settings are used as much as possible in the Modified Wu's method. The ANN model used has two hidden layers with three hidden neurons in the first hidden layer and two hidden neurons in the second hidden layer. The LM algorithm is used for ANN training. The maximum training epoch is 500, and larger training epoch, say 1000, will not improve the prediction performance. The ANN training and prediction testing processes are also the same as those used in the previous section.

The RUL prediction results using the Modified Wu's method is presented in Table 2 as well. The Average Prediction Error considering all the inspection points in the test sets, $\bar{e}_{\mathrm{All}}$, is 12.5%, the Average Prediction Error considering only the last 5 inspection points for each test history, $\bar{e}_{\mathrm{L5}}$, is 4.9%, and the Average Prediction Error considering only the prediction results where the predicted life percentage values are between 90% and 100%, $\bar{e}_{90-100}$, is 8.28%. Comparing the results by the proposed ANN method and the Modified Wu's method, it can be seen that the proposed ANN method can produce more accurate prediction with respect to all of the three measures. Specifically, with respect to measure $\bar{e}_{\mathrm{All}}$, $\bar{e}_{\mathrm{L5}}$, and $\bar{e}_{90-100}$, the prediction results with the proposed ANN method are approximately 18, 44, and 127% better, respectively, than those with the Modified Wu's method. These results clearly demonstrate the advantage of the proposed ANN method over the Modified Wu's method.

Conclusions

Accurate equipment remaining useful life prediction is critical to effective condition based maintenance for improving reliability and reducing overall maintenance cost. This paper develops an ANN method for achieving more accurate remaining useful life prediction of equipment subject to condition monitoring. The ANN model takes the age and multiple condition monitoring measurement values at the present and previous inspection points as the inputs, and the life percentage as the output. The generalized Weibull-FR function is used to fit each condition monitoring measurement series for a failure history, and the fitted measurement values are used to form the ANN training set to reduce the effects of the noise factors that are irrelevant to the equipment degradation. When the trained ANN is used for RUL prediction, the inputs to the trained ANN are generated by fitting the available measurement values for the current unit using the generalized Weibull-FR function. The validation mechanism is introduced in the ANN training process to improve the prediction performance of the ANN model. In addition, the proposed ANN method does not require the definition of a failure threshold, which is hard to clearly define in many practical applications.

The proposed ANN method is validated using the condition monitoring data collected in the field from bearings on a group of Gould pumps. Experiment results show that the proposed ANN method can produce satisfactory RUL prediction results, which will assist the condition based maintenance optimization. A comparative study is performed between the proposed ANN method and the Modified Wu's method, and the results demonstrate the clear advantage of the proposed approach in achieving more accurate predictions.

Acknowledgments This research is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). We greatly appreciate the help from OMDEC Inc. for providing the condition monitoring data used in the case study. Comments and suggestions from anonymous reviewers are very valuable in the improvement of the quality of this paper.

References

Banjevic, D., Jardine, A. K. S., & Makis, V. (2001). A control-limit policy and software for condition-based maintenance optimization. *INFOR, 39*, 32–50.

Dong, M., & He, D. (2007a). A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology. *Mechanical Systems and Signal Processing, 21*(5), 2248–2266.

Dong, M., & He, D. (2007b). Hidden semi-Markov model-based methodology for multi-sensor equipment health diagnosis and prognosis. *European Journal of Operational Research, 178*(3), 858–878.

Dong, M., He, D., Banerjee, P., & Keller, J. (2006). Equipment health diagnosis and prognosis using hidden semi-Markov mod-

els. *International Journal of Advanced Manufacturing Technology, 30*(7–8), 738–749.
Gebraeel, N., & Lawley, M. A. (2008). A neural network degradation model for computing and updating residual life distributions. *IEEE Transactions on Automation Science and Engineering, 5*(1), 154–163.
Gebraeel, N., Lawley, M. A., & Liu, R. (2004). Residual life predictions from vibration-based degradation signals: A neural network approach. *IEEE Transactions on Industrial Electronics, 51*(3), 694–700.
Huang, R. Q., Xi, L. F., Li, X. L., Liu, C. R., Qiu, H., & Lee, J. (2007). Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods. *Mechanical Systems and Signal Processing, 21*(1), 193–207.
Inman, D. J., Farrar, C. R., & Lopes, V. (2005). *Damage prognosis: For aerospace, civil and mechanical systems*. NY: Wiley.
Jardine, A. K. S., Lin, D. M., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing, 20*(7), 1483–1510.
Kacprzynski, G. J., Roemer, M. J., Modgil, G., Palladino, A., & Maynard, K. (2002). Enhancement of physics-of-failure prognostic models with system level features. In *Proceedings of the 2002 IEEE aerospace conference*, Big Sky, MT, USA.
Kuo, W., & Zuo, M. J. (2003). *Optimal reliability modeling: Principles and applications*. Hoboken: Wiley.
Lee, J., Ni, J., Djurdjanovic, D., Qiu, H., & Liao, L. T. (2006). Intelligent prognostics tools and e-maintenance. *Computers in Industry, 57*, 476–489.
Levitin, G. (2005). *Universal generating function in reliability analysis and optimization*. London: Springer-Verlag.
Levitin, G., Lisnianski, A., Ben Haim, H., & Elmakis, D. (1998). Redundancy optimization for series-parallel multistate systems. *IEEE Transactions on Reliability, 47*(2), 165–172.
Li, C. J., & Lee, H. (2005). Gear fatigue crack prognosis using embedded model, gear dynamic model and fracture mechanics. *Mechanical Systems and Signal Processing, 19*, 836–846.
Liao, H. T., Elsayed, E. A., & Chan, L. Y. (2006). Maintenance of continuously monitored degrading systems. *European Journal of Operational Research, 175*(2), 821–835.
Lin, D., Banjevic, D., & Jardine, A. K. S. (2006). Using principal components in a proportional hazards model with applications in condition-based maintenance. *Journal of the Operational Research Society, 57*, 910–919.
Makis, V., & Jardine, A. K. S. (1992). Optimal replacement in the proportional hazards model. *INFOR, 30*, 172–183.
Marble, S., & Morton, B. P. (2006). Predicting the remaining life of propulsion system bearings. In *Proceedings of the 2006 IEEE aerospace conference*, Big Sky, MT, USA.
Rojas, R. (1996). *Neural networks: A system introduction*. Berlin: Springer.
Stevens, B. (2006). EXAKT reduces failures at Canadian Kraft Mill. www.omdec.com.
Tian, Z., & Zuo, M. J. (2009). Health condition prognostics of gears using a recurrent neural network approach. In *Proceedings of the reliability and maintainability symposium*, Fort Worth, TX, USA.
Tse, P. W., & Atherton, D. P. (1999). Prediction of machine deterioration using vibration based fault trends and recurrent neural networks. *Journal of Vibration and Acoustics, Transactions of ASME, 121*, 355–362.
Vachtsevanos, G., Lewis, F. L., Roemer, M., Hess, A., & Wu, B. (2006). *Intelligent fault diagnosis and prognosis for engineering systems*. NY: Wiley.
Wu, S. J., Gebraeel, N., Lawley, M. A., & Yih, Y. (2007). A neural network integrated decision support system for condition-based optimal predictive maintenance policy. *IEEE Transactions on Systems Man and Cybernetics Part A: Systems and Humans, 37*(2), 226–236.