Applied Soft Computing 64 (2018) 454–467

Contents lists available at ScienceDirect

Applied Soft Computing

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

Tool condition prognostics using logistic regression with penalization and manifold regularization

Jianbo Yu

School of Mechanical Engineering, Tongji University, Shanghai, 201804, PR China

A R T I C L E I N F O

Article history:
Received 27 August 2015
Received in revised form 5 December 2017
Accepted 27 December 2017
Available online 28 December 2017

Keywords:
Tool condition prognostics
Remaining useful life
Logic regression
Penalization regularization
Manifold regularization

A B S T R A C T

Appropriate and timely maintenance decision for tool health degradation (i.e., wear) is significantly required to prevent severe degradation in product processing quality. Multiple sensor signals (e.g., vibration, acoustic emission) collected from tools contain much valuable information about their health states. However, information fusion of multiple sensor signals for assessing and predicting the tool health presents a big challenge. In this paper, logistic probability (LP) generated by logistic regression with manifold regularization (LRMR) is used to serve as a comprehensible indication to assess tool health state. Prognostic features are selected firstly by logistic regression with penalization regularization (LRPR) to improve the performance of the proposed tool health prognostics system. Based on the health indication values (i.e., LPs) and tool ages, the LR model is further developed to online construct the interior relationship between the tool health state and its ages, and then predicts the remaining useful life (RUL) of tools subjected to condition monitoring. The proposed prognostics system provides an adaptive learning scheme for assessment and prediction of tool health, and hence is easier to use in real-world applications. The experimental results on a tool life test-bed illustrate the potential applications of the proposed system for tool health prognostics.

1. Introduction

Tool condition monitoring (TCM) is crucial to improve the quality of unmanned manufacturing systems where the tools are subjected to continuous wears. TCM aims to use tool health assessment and prediction techniques to identify and estimate the health state of the cutting tools, so as to reduce time and production loss brought about by tool wears or failures [1–3]. Basically, the tool wear procedure consists of several typical stages, i.e., an initial wear stage, a progressive wear stage, and a rapid wear stage. In the last stage, the presence of a large wear drastically increases the temperature and triggers some complicated changes in the tool, causing rapid deterioration (or breakages) [4]. Therefore, it is of prime importance to assess and predict accurately the health degradation (i.e., wear) of tools to prevent the sequent damages and to reduce the costly downtime.

Recently, some effective TCM systems have been developed and can be categorized mainly into two fundamental methods: direct and indirect methods [2,7,8]. Video-based and laser-based vision is often used for direct TCM [5,6]. But high cost and measurement inconsistency due to the big variation in illumination have prevented those direct methods from being performed widely in real-world applications. Although the direct methods could be more accurate than the indirect methods, some indirect methods that rely on changes in the signals of sensors associated with tool wears, are more commonly used for TCM. These measurement signals are often from cutting force [7], machine vibration [8], motor current [8], acoustic emission (AE) [9], or various combinations of these signals [10]. Thus, various methods such as support vector machine (SVM) [10], artificial neural network (ANN) [11], hidden Markov model (HMM) [12], adaptive Gaussian mixture model (AGMM) [13], time-frequency analysis [1,14], etc., using these sensor signals, have been developed for TCM.

However, recent attempts in designing TCM systems did not consider multiple sensor fusion techniques [4,15], since single particular sensor could not work well in quantifying the tendency of the tool health degradation. As for example, use of AE signal has been reported as not being that helpful for cutting processes [16,17]. Multiple sensor fusion [15,18] at various system levels is known to provide more accurate estimation of tool health (i.e., wear) than those methods using single sensor. Various sensor

E-mail address: jbyu@tongji.edu.cn

https://doi.org/10.1016/j.asoc.2017.12.042
1568-4946/

fusions, e.g., fusion of force and AE signals [19], fusion of force and vibration signals [10], fusion of force, vibration and sound signal [20] have been attempted and obtain some successes in some cases. These multi-sensor signals have been used as inputs of those machine learners, (e.g., ANN, SVM, GMM) to implement tool health prognostics [10,11,13]. Therefore, the study has occurred that efforts in future centers on fusing the most valuable information from the multiple sensor signals.

Mainly, the degrading behavior of aging component or system is reflected by those original features derived from various sensor signals. The features (or variables) affect the performance of the diagnosis and prognostics system significantly [21–24]. Thus, it is important to evaluate these original features by those appropriate methods to detect the degradation and then to predict the remaining life of tool. We aim to select those features with monotonic trends that properly correlate to degradation phenomenon and may lead to simple and accurate prognostics systems. Zhou et al. [22] proposed a method for diagnosis and prognosis of tool wears, where a time series prediction method based on the selected prognostic features was applied to estimate the tool wears. Yu [23] proposed a contribution analysis method to select the prognostic features to improve the effectiveness of machine health assessment system. Actually, it is still sparse for studying of prognostic feature selection in the construction procedure of TCM systems.

In recent years, health prognostics techniques that aim to realize near-zero downtime, maximum productivity and proactive maintenance, have received more and more attentions [25–27]. The research emphasis in this studying field has been centered on system (or machine) health assessment and prediction, so the final failure can be predicted and prevented. HMM-based prognostics system was developed to track and predict the evolution of health-states and remaining useful life (RUL) of cutting tools [11]. Lee et al. [28] proposed a modified HMM for online degradation assessment and adaptive fault detection of multiple failure modes in tools. Yu [12] proposed an AGMM for online monitoring of tool health. However, these proposed models generally do not consider the intelligibility of the proposed health indication (HI) to improve their effectiveness significantly. Logistic regression (LR) has shown its power in assessment of machine health because its logic curve (e.g., S-curve) is similar to the procedure of machine health degradation [29]. However, LR in applications of tool health prognostics (e.g., prognostic feature selection, health degradation assessment and prediction) is still worth studying.

Driving by the desire of improved production quality and near-zero breakdown productivity, we develop a novel prognostics system to implement on-line tool health assessment and prediction. In this system, LR with penalization regularization (LRPR) is developed to select prognostic features, and LR with manifold regularization (LRMR) is further proposed for tool health assessment and prediction online. Logistic probability (LP) derived from the baseline LR constructed by LRMR is developed to provide an HI based on failure risk probability once the tool is in degraded state. Based on LPs on time series flow, LR is constructed online to perform RUL prediction of tools. With the health assessment and RUL prediction, users will know clearly the tool health states, and then take correct and timely maintenance measurements to recover the tool health. Thus, the proposed prognostics system provides an effective methodology of enhancing the assessment and prediction of tool health: (1) A new LRPR-based prognostic feature selection method is developed for improving the prognostics performance of the proposed method, which is rarely considered in most tool prognostics methods; (2) A probability-based HI is developed to quantify tool health state, which provides some unique features, e.g., consistent range (0–1), very few false alarms, and quick calculation, to enable it to possess high applicability in real-world cases; (3) An LRMR-based model is developed to predict tool health state, which is capable of estimating the failure time point of the tool with high accuracy and calculation efficiency. A tool life test on a test-bed was implemented to collect multiple sensor data over the whole life time of tools. Experimental results validate the feasibility and validity of the proposed system for tool health prognostics.

The rest of the paper is organized as follows. Section 2 proposes a tool health prognostics system based on LRPR and LRMR. A real-world milling case is used to verify the effectiveness of the proposed prognostics system in Section 3. Conclusions are given in Section 4.

2. Methodology

This section proposes a tool health prognostics system (see Fig. 1), which includes two key parts, i.e., off-line system modeling and on-line tool health prognostics. We need to collect historic data to select the prognostic features by using LRPR and to construct a baseline LR by using LRMR for tool health assessment, and then to use LR with online learning scheme to implement tool RUL prediction. The proposed prognostics system consists of two parts, i.e., off-line modeling and on-line health prognostics. In the offline modeling phase, the data collected from the historic tools are used to construct the baseline LR, where the original feature generation and LRPR-based feature selection are firstly implemented. In the online health prognostics phase, LR-based LP is developed to quantify tool health degradation. Then, an LR with LRMR algorithm is established for modeling the dynamic propagation of the tool

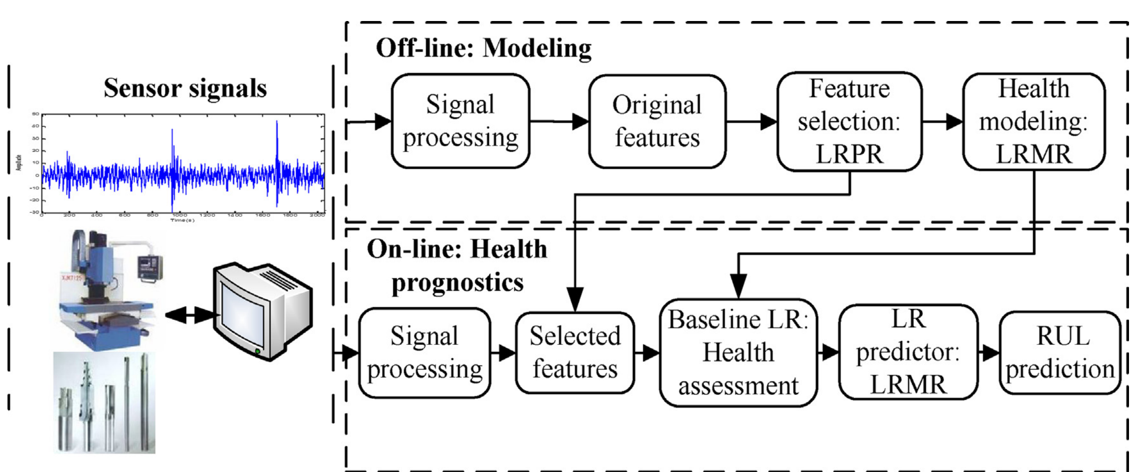

Fig. 1. System framework for tool health prognostics.

health, and then is used as the predictor for tool health prediction. This designing scheme for a tool health prognostics system intends to reduce needs of human intervention and then improve its applicability in real-world cases. A tool life test on a test-bed was implemented to illustrate the effectiveness of the proposed prognostics system. The following subsections present these key techniques used in this tool health prognostics system.

2.1. Logic regression-based tool health assessment

LR is a popular linear regression and classification method, which has been used successfully in many applications [30]. In statistics, LR is often applied for estimation of an event occurrence probability by fitting the data to a logistic curve. LR is very effective to seek the best fitting model that describes the relationship between the dependent variable and the independent variables. Its logistic probability function is:

$$P(x) = \frac{1}{1 + e^{-g(x)}} \tag{1}$$

where $g(x)$ is a logit model. Given the n training samples, $\{(x_i, y_i), i = 1, ..., n\}$, where each $x_i \in \Re^m$ is an m-dimensional feature vector, and $y_i \in \{0, 1\}$ is a class label, LR is defined as follows:

$$\log \frac{\pi_i}{1 - \pi_i} = \beta_0 + \sum_{j=1}^{m} \beta_j x_{ij} \Leftrightarrow \pi_i = \frac{1}{1 + e^{-(\beta^T x_i)}} \tag{2}$$

where π_i denotes $p(y_i = 1|x_i, \beta)$, and $\beta = (\beta_0, \beta_1, ...\beta_m)$ denotes the vector of regression coefficients including a constant or intercept β_0.

The prediction probability $P(x)$ of an event occurrence by LR (see Eq. (1)) is constrained between 0 and 1, which is very suitable as an HI for quantifying the tool health state, i.e.,

$$HI_t = P(x_t) \tag{3}$$

where $P(x_t)$ can be calculated by using Eq. (1). The S-shape function of the LR shown in Fig. 2 indicates a relative low probability of event occurrence until the threshold is reached, at which time the probability of occurrence increases rapidly. In this study, the LRMR algorithm (see Section 2.3) is used to construct LR to achieve the predicted LPs (i.e., HI_t) according to Eq. (3). In order to improve the sensitivity and reliability of the HI (i.e., LP) to the slight degradation of tool health, exponentially weighted moving average (EWMA) statistic [31] of the LPs is used as an improved HI of tool health,

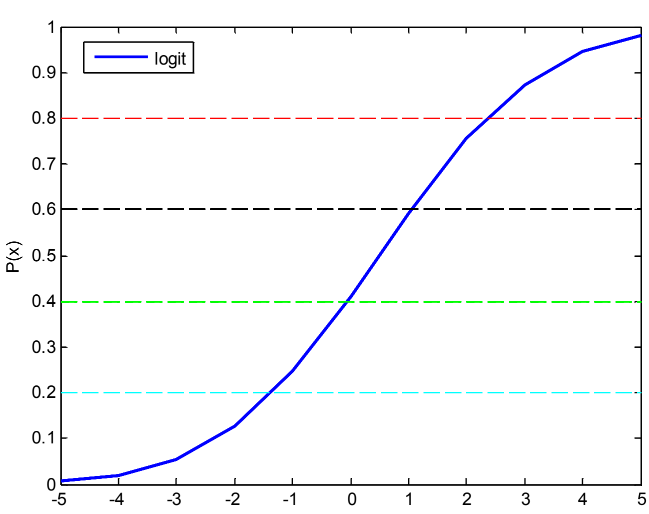

Fig. 2. Logistic function.

where a smoothing constant α (0.1) is used to get a good balance between the historic information from and current observations.

2.2. Logistic regression with penalization regularization

Given the n training samples, $\{(x_i, y_i), i = 1, ..., n\}$, the parameter estimation β in LR can be achieved by using the maximum likelihood method [32] based on the following negative log-likelihood probability (NLLP).

$$l(\beta) = \sum_{i=1}^{n} -y_i \beta^T x_i + \log(1 + \exp(\beta^T x_i)) \tag{4}$$

It is often necessary to impose a penalty on big fluctuations of the parameters β, which intends to avoid over-fitting and multicollinearity occurring generally in the LR model. Ideally we would prefer an LR based on a small selection of the most effective input variables, with the remaining being pruned in the learning procedure. Finally, an LR model with penalization is achieved in the following objective function through using a ridge penalty (i.e., l2 regularization) [33]:

$$l^*(\beta) = l(\beta) + \frac{\lambda}{2} \|\beta\|^2 \tag{5}$$

where λ is called the penalty parameter. The larger the λ, the stronger its influence is and the smaller the $\|\beta\|^2$ become. In this study, λ is setup as a value 0.1 by cross-validation on the training dataset in the experimental section.

Maximum likelihood estimation, $\hat{\beta}$, can be achieved by maximizing $l^*(\beta)$ with respect to β. We get the derivative of $l^*(\beta)$ with respect to β to zero and then use Newton-Raphson algorithm to iteratively update β. Finally, the vector of regression coefficients β can be achieved for the feature selection. It should be noted that other gradient-based optimization algorithms (e.g., fixed-Hessian, conjugate gradient) can be used to solve the above optimization problem. The advantage with Newton-Raphson algorithm is that it converges fast when approaching the optimal solution.

An important part in tool health prognostics is to select a small subset of original features that can effectively describe different health states of the tool. In LR, β gives the estimations of the regression coefficients and its absolute values indicate the relative importance of the input variables (i.e., features). With the penalization learning of β in LRPR, those features with the smallest absolute values in β are eliminated. Therefore, a smaller subset of the original features can be used for tool health prognostics. It should be noted that the l2 regularization used in LRPR tends to cut down the peaks and distribute the 'mass' across all the parameters. As a result, the parameters are not enough sparse for feature selection in some cases.

The health prognostics system aims to model the health degradation progression of the tool and then to predict its RUL. Feature selection is a key step in the procedure of diagnosis and prognostics system [21,23]. Mainly, the degrading behavior of aging component or system is reflected by the features derived from condition monitoring sensor data. Thus, it is important to evaluate the extracted features by those appropriate methods to detect the tool wears and then to predict its RUL. We aim to select the prognostic features with good tendency that properly correlate to degradation phenomenon.

In this study, an LRPR-based method is developed for discovering the effective physical features that are responsible for tool health degradation. Since LRPR automatically introduces penalization into the LR model, the non-zero β parameters help us in deciding the relevant features that contribute to recognizing the procedure of tool health degradation. Based on the parameter set β of LR with LRPR algorithm, we can sort smoothly the importance of the candidate features. The top ranked features will be selected and are then used along with the training set to design a model for the prognostic purpose. The LRPR-based feature selection algorithm named LRPR-FS is presented as follows:

Algorithm I. LRPR-FS

Input: a dataset $X = [x_1, ..., x_i, ..., x_n]$ and $Y = [y_1, ..., y_i, ..., y_n]$
Output: a feature subset
(1): $X \longleftarrow Norm(X)$//The dataset X is normalized
(2): $[\beta_0, \beta_1, ..., \beta_m] \longleftarrow LRPR(X, Y)$//Run LRPR to obtain the parameters β
(3): $\tilde{\beta} \longleftarrow Sort(abs(\beta))$//Sort all m features in a descending order of the absolute $[\beta_1 ..., \beta_m]$
(4): Select all the top ranked r features based on $\tilde{\beta}$.

2.3. Logistic regression with manifold regularization

In this section, we further propose the LRMR algorithm by exploiting the intrinsic geometry of the probability distribution in the LR model. The manifold regularization is integrated in the learning procedure of LR to improve its prediction performance. In the learning procedure, LR usually ignores the marginal distribution $p(x)$ and models the conditional distribution $p(y|x)$ directly. Although there is no certain assumption on marginal distribution, we can also augment the manifold assumption to obtain a more smooth conditional distribution, which aims to achieve more generalized parameter estimation. Recent work [34] on GMM was proposed to incorporate so-called Laplacian regularization based on manifold regularization technique [35] in the maximum likelihood function. Manifold regularization in the regular machine learners generally improves their performance due to the preservation of intrinsic information in the data [35].

The Laplacian regularization is based on spectral graph theory [36] and emphasizes on preserving locality of the intrinsic data geometry. Recent studies [34] on spectral and manifold learning theory have demonstrated that local geometric structure can be effectively constructed through a Laplacian graph on a scatter of data points. The corresponding nearest neighbor graph can be defined as follows:

$$S_{ij} = \begin{cases} 1, if x_i \in N_l(x_j) or x_j \in N_l(x_i) \\ 0, otherwise \end{cases} \tag{6}$$

where $N_l(x_j)$ denotes the set of l nearest neighbors of x_i, and the similarity between each pair of samples (x_i, x_j) is measured by using Euclidean distance. In this study, five nearest neighbors for each sample are used to generate the nearest neighbor graph, which is obtained through using a try-error method (i.e., testing the prediction errors on the training dataset by using the different numbers of nearest neighbors). Define $L = D - S$, where D is a diagonal matrix whose entries are column sum of $S(D_{ii} = \sum_j S_{ij})$, and L is Laplacian graph [36].

Intuitively, if two data points x_i and x_j are within the intrinsic geometry of $p(x)$, the corresponding conditional distributions $p(y|x_i)$ and $p(y|x_j)$ are similar. In other words, the conditional distribution $p(y|x)$ varies smoothly along the geodesics in the intrinsic geometry of $p(x)$. This is usually referred as a manifold assumption [35]. Based on this basic idea, LRMR is developed to involve the local intrinsic information from the probability space in the regular LR. They both introduce a regularization term into the objective function of LR to measure the smoothness of $p(y|x)$ on the nearest neighbor graph:

$$\begin{aligned} R_k &= \frac{1}{2}\sum_{i,j}^{n} S_{ij}(p(k|x_i) - p(k|x_j))^2 \\ &= \sum_{i=1}^{n} D_{ii} p(k|x_i)^2 - \sum_{i,j=1}^{n} S_{ij} p(k|x_i) p(k|x_j) \\ &= P_k^T (D - S) P_k \\ &= P_k^T L P_k \end{aligned} \tag{7}$$

where k is one type of the class labels y, $P_k = (p(k|x_1), \ldots, p(k|x_n))^T$, and $(p(k|x_i) - p(k|x_j))^2$ measures the similarity or proximity between two distributions according to some specific distance metrics (e.g., Euclidean distance). The smaller of R, the smoother of $p(y|x)$ on the graph that describes the intrinsic geometry of the data. Thus, the objective function of LR incurs a heavy penalty if the neighbor points x_i and x_j have different conditional probability distributions (i.e., $p(y|x_i)$ and $p(y|x_j)$). Incorporating the above smoothness term into the likelihood of the LR to obtain the objective function:

$$O(\beta) = l(\beta) + \frac{\gamma}{2}\sum_{k=1}^{K} R_k \tag{8}$$

where K is the number of classes, and γ is the regularization parameter that controls the smoothness of P_k. In this study, γ is setup as a value 0.1 to obtain a good balance between the prediction performance of LR and the smoothness of P_k. By considering the local geometric structure of the data manifold, LRMR can offer stronger discrimination than the original LR for many learning problems.

In the model fitting phase, we need to solve the minimization problem of (8). The first part $l(\beta)$ is as same as the case of the regular LR. We focus on calculating the gradient of the manifold regularization term. In the following, we discuss how to apply the iterative procedure to maximize the regularized LLP of LR. Instead of finding the globally optimal solution for $O(\beta)$ that minimizes (8), if suffices to find a better $O(\beta)$, the regularized term in Eq. (8) is as follows:

$$\frac{\gamma}{2}\sum_{k=1}^{K} R_k = \frac{\gamma}{4}\sum_{k=1}^{K}\sum_{i,j}^{n} S_{ij}(p(k|x_i) - p(k|x_j))^2 \tag{9}$$

Let β^{t-1} and β^t denote the parameters of the previous and current iteration, respectively. The used basic idea is to minimize $l(\beta)$ and $\sum_{k=1}^{K} R_k$ separately in hope that a decreasing of the current $O(\beta)$ in the iteration procedure. Notice that the regularization term only involves the posterior probability $p(k|x_i)$. We initialize $p(k|x_i)$ to the value of precious iteration, that is, $p(k|x_i, \beta^{t-1})$, and try to decrease $\sum_{k=1}^{K} R_k$ in the iteration procedure, which can be done by applying the Newton-Raphson algorithm. Given an objective function $f(x)$ and the initial value x_t, we repeat the Newton-Raphson steps, which results in the "iteratively reweighted ridge regressions" (IRRR) algorithm to decrease (or increase) [37] $f(x)$:

$$x_t = x_{t-1} - \eta \frac{f'(x_{t-1})}{f''(x_{t-1})} \tag{10}$$

where $0 \le \eta \le 1$ is the step parameter. The above iterative scheme in Eq. (10) can be generalized to a vector by replacing the derivative with the gradient, $\nabla_x f(x)$, and the reciprocal of the second derivative with the inverse of the Hessian matrix, $Hf(x)$. One obtains the iterative scheme.

$$x_t = x_{t-1} - \eta[Hf(x_{t-1})]^{-1}\nabla f(x_{t-1}) \quad (11)$$

where $\nabla f(x)$ is the derivative of $f(x)$ with respect to x, and H is a Hessian matrix and $H_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$. By taking the first and second derivatives of R_k with respect to $p(k|x_i)$, the updating scheme for $p(k|x_i)$ is:

$$p(k|x_i)^t \longleftarrow (1-\eta)p(k|x_i)^{t-1} + \eta\frac{\sum_{j=1}^{n} S_{ij}p(k|x_j)^{t-1}}{\sum_{j=1}^{n} S_{ij}}(i = 1, \ldots, n) \quad (12)$$

It is easy to show that the graph L is positive semidefinite, and each updating step will decrease R_k. After $p(k|x_i)$ is optimized and then $\sum_{k=1}^{K} R_k$ is fixed in $O(\beta)$, we can minimize the objective function $O(\beta)$ through using the following updating scheme for β. We get the derivative of $l(\beta)$ with respect to β to zero and then use Newton-Raphson algorithm to iteratively update β as follows:

$$\beta^t = (X^T WX)^{-1} X^T Wz \quad (13)$$

$$z_i = \frac{y - p(\beta^{t-1}, x_i)}{p(\beta^{t-1}, x_i)(1 - p(\beta^{t-1}, x_i))} + \ln\frac{p(\beta^{t-1}, x_i)}{1 - p(\beta^{t-1}, x_i)} \quad (14)$$

where X is the training data, $p(\beta^{t-1}, x_i)$ is the $p(k|x_i)$ with the parameters β^{t-1}, and W is the $n*n$ weight matrix with entries $w_i = p(\beta^{t-1}, x_i)(1 - p(\beta^{t-1}, x_i))$ on the diagonal matrix. We summarize the LRMR algorithm as follows:

Algorithm II. LRMR

Input: The training data $\{(x_i, y_i), i = 1, \ldots, n\}$, vector of regression coefficients $\beta = (\beta_0, \beta_1, \ldots\beta_m)$ initialized as a zero vector, regularization parameter γ, and termination condition value δ.

Output: β

(0) Set the initial value $\eta = 0.9$ to obtain a balance between the learning rate and the learning result.

(1) Construct a local graph with the weight matrix S

(2) $t \longleftarrow 1$

(3) While (true)

(4) $p(k|\beta^t, x_i) = \frac{1}{1+\exp(-\beta^t X)}$

(5) Smooth $p(k|x_i)$ until convergence (i.e., we try to decrease $\sum_{k=1}^{K} R_k$ separately in the iteration by applying the Newton-Raphson algorithm, see Eqs. (10)–(12) for details):

$$p(k|x_i)^t \longleftarrow (1-\eta)p(k|x_i)^{t-1} + \eta\frac{\sum_{j=1}^{n} S_{ij}p(k|x_j)}{\sum_{j=1}^{n} S_{ij}}(i = 1, \ldots, n, k = 1, \ldots, K)$$

(6) Compute the parameters β^t (we minimize $l(\beta)$ separately in the iteration):

$$w_i = p(\beta^{t-1}, x_i)(1 - p(\beta^{t-1}, x_i))$$

$$z_i = \frac{y - p(\beta^{t-1}, x_i)}{p(\beta^{t-1}, x_i)(1 - p(\beta^{t-1}, x_i))} + \ln\frac{p(\beta^{t-1}, x_i)}{1 - p(\beta^{t-1}, x_i)}$$

$$\beta^t = (X^T WX)^{-1} X^T Wz$$

(7) Evaluate the regularized log likelihood:

$$O(\beta^t) = l(\beta^t) + \frac{\gamma}{2}\sum_{k=1}^{K} R_k$$

(8) If $O(\beta^t) < O(\beta^{t-1})$

$$\eta \longleftarrow 0.9\eta, \text{goto}(5) \quad (9)$$

(10) If $O(\beta^{t-1}) - O(\beta^t) < \delta$

(11) Break

(12) $t \longleftarrow t + 1$

(13) Return β^t

2.4. LRMR-based tool health prediction

This section aims to predict the tool health change at the time t once the HI (i.e., LP) exceeds the predetermined threshold of slight degradation (i.e., δ_s), and the tool RUL can be further predicted with the predetermined threshold of tool failure (i.e., δ_f).

LR is used to construct the relationship between the health states (i.e., LPs) and the ages of the tool, and then it is used to implement tool health prediction. When implementing tool RUL prediction at the time t, the LR using LRMR learning algorithm is constructed firstly by using the historical HIs, i.e., LPs (LP_1, LP_2, ..., LP_t) and the corresponding tool ages (1, 2, ..., t) as its outputs and inputs, respectively. Once the LR is constructed at the time t, it will estimate the future HI values, i.e., $\hat{LP}_{t+h}$ for the future time point $t+h$. While the case for $h = 1$ is referred as one-step-ahead prediction, the prediction is generally called multi-steps-ahead prediction when $h > 1$. Eq. (15) is used to perform the h-steps-ahead prediction of the LR:

$$\hat{LP}_{t+h} = LR(t + h) \quad (15)$$

where the tool age (i.e., $t+h$) is input to the constructed LR, and then it outputs the LP prediction $\hat{LP}_{t+h}$ by using Eq. (15). With the constructed LR at time t, we will input the tool ages $t+1, \ldots, t+h$ into the LR predictor$LR()$ and then calculate the LP predictions $\hat{LP}_{t+1}, \ldots, \hat{LP}_{t+h}$ (i.e., the h-steps-ahead prediction is performed using Eq. (15)) until the predicted $\hat{LP}_{t+h}$ exceeds the predetermined failure threshold (i.e., δ_f). Finally, the tool RUL will be achieved based on the prediction $\hat{LP}_{t+h}$ and the prediction time point t. Under this method, thus, it repeatedly performs one-step-ahead predictions until the desired horizon is reached.

In this tool RUL prediction scheme, LR need adapt to the tool health changes and then to provide an accurate long-term prediction. When the time point moves ahead, LR with LRMR learning algorithm will be re-constructed by using those historic LPs, and then it is used to predict the tool RUL. Each such learning procedure is called an epoch. It is desired that LR is constructed adaptively based on the newest trend of the prediction variable, which is particularly true in the case of long-term prediction. Such learning can adapt to the basic trend of tool health to perform an effective RUL prediction online. Due to the high effectiveness of LR learning procedure, the time cost is very low for each construction of LR at the prediction time point.

It should be noted that the LRMR algorithm is mainly used for tool health assessment in the proposed method. The causes for why we use LR with manifold regularization for tool health prediction consist of the following: (1) The effectiveness of LR with LRMR algorithm for regression prediction should be further evaluated; (2) The consistence for applications of the LRMR algorithm on tool health assessment and RUL prediction can be persisted in the proposed method; (3) Although the adjacency matrix S_{ij} (see Eq. (6)) would be in this case very simplistic, the manifold regularization is mainly to preserve the smoothness of the corresponding conditional distributions $p(y|x)$ on the nearest neighbor graph (see Eq. (7)).

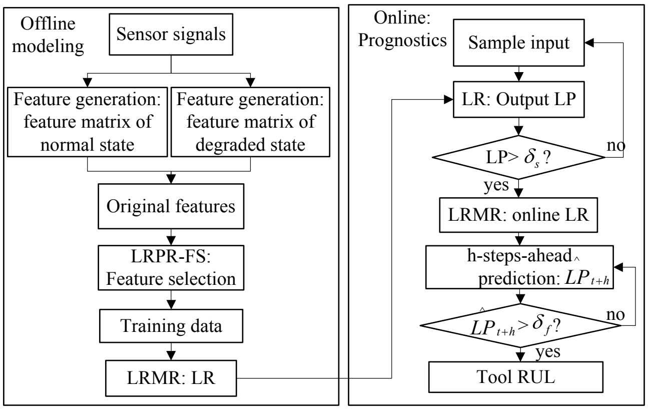

Fig. 3. The proposed tool health prognostics system.

Finally, the detailed procedure for using of the tool health prognostics system is presented in Fig. 3 and is further summarized as follows:

Step 1: Collect the sensor signals from the historic tools and extract the normal and degraded data to form the training dataset.

Step 2: Generate the original features from time and time/frequency domain of the sensor data and then normalize them.

Step 3: Use LRPR-FS to select the prognostic features from the original feature set.

Step 4: Use the training data with the selected features to construct the baseline LR by using LRMR algorithm, which fuses the multi-sensor data effectively.

Step 5. Input the testing sample x_t online to the baseline LR to generate the HI (i.e., LP) to assess the tool health state.

Step 6. If the current LP indicates that the tool is in degraded state ($LP_t > \delta_s$), the LR modeling for long-term-prediction will be triggered online based on the collected historic data of this tool.

Step 7. Run the LR recursively online to predict the tool health state, and then give the RUL prediction of the tool once the predicted LP exceeds the failure threshold ($\hat{LP}_{t+h} > \delta_f$).

3. Experiment and result analysis

In order to verify the effectiveness of the proposed tool health prognostics system, a dataset collected on a milling machine (Matsuura machine center MC−510 V) with various operating conditions was employed to perform this experiment [38,39]. The magnitudes of tool wears were recorded in a regular cut as well as entry and exit cuts. Six sensors including two vibration sensors (one on the table and spindle, respectively), two AE sensors (one on the table and spindle, respectively), and two motor current sensors (DC spindle motor current and AC spindle motor current) were setup on the milling machine to collect the signal data. A high speed acquisition board (MIO-16 from National Instrument company) connected all sensors for collecting all sensor signals. For more detailed information about the setup of the experiment, the readers can refer to the literature [39].

A 70 mm face mill having 6 inserts using two types of inserts (called Kennametal K420 and KC710) was used for roughing operations. KC710 was coated with multiple layers titanium carbide, titanium carboitride and titanium nitride (YiC/TiC-N/TiN) in sequence. The flank wear VB (i.e., the distance from the cutting edge to the end of the abrasive wear on the flank face of the tool) was measured with the help of a microscope when each insert was taken out of the tool.

The milling was implemented on workpieces (483 mm × 178 mm × 51 mm) material cast iron by using the following key parameters: cutting speed 200m/min, two different cut depths 1.5 mm/min and 0.75 mm/min, and two different feeds 0.5 mm/rev and 0.25 mm/rev. All experiments were done a second time with the same parameter a second set of inserts. The machining conditions of the eight tool cases (named #1–#8) are presented in Table 1. The number of all milling runs in the full life of each tool is also presented in Table 1. A milling run of the tool is finished when each insert is taken out of the tool.

Table 1
Experimental condition of the eight tool cases.

Tool	Depth of Cut	Feed	Runs
#1	1.5	0.5	17
#2	0.75	0.5	13
#3	0.75	0.25	14
#4	1.5	0.25	9
#5	1.5	0.5	9
#6	1.5	0.25	10
#7	0.75	0.25	7
#8	0.75	0.5	15

In this experiment, the operating conditions were varying with the different cutting parameters: i.e., the cut depth and feed rate. Thus, the proposed prognostic system will be insensitive to the varying operating conditions that often happen in real-world cases, which is an important for effective utility of TCM in those complicated applications.

For implementing LRPR-based prognostic feature selection and LRMR-based tool health assessment in the following subsection 3.1, the sensor data from the seven tools of the all eight tools were used as the training dataset to construct the baseline LR, and those of the rest one were used to test the health assessment performance of the proposed prognostics system (see the eight tools in Table 1). This generation of the training and testing datasets is appropriate because we generally collect the representative historic data to construct the model and then use it to implement health prognostics of the new tools online. It should be pointed out that the LR will be constructed by using the data from the first three runs (normal class with class label 0) and the last three runs (severe degradation class with the class label 1) of each historic tool, and then the baseline LR is used for online monitoring in the full life of the new tool. Therefore, the training dataset includes 42 runs, i.e., 3 normal runs + 3 severe degradation runs of each tool × 7 tools for constructing each LR with the LRPR or LRMR algorithm. The testing data include all runs of the rest one tool of the eight tools. In this study, the four tools (Tool #1, #3, #5 and #6) were used as the representative tools for testing the effectiveness of the prognostics system in their full life. Thus, there are four baseline LRs that will be constructed corresponding to the four tools (Tool #1, #3, #5 and #6) to implement tool health assessment (see Section 3.1), respectively.

It should be pointed out that, only the historic data from one tool before a prediction time point were used to construct the LR with LRMR learning algorithm for implementing the tool RUL prediction in the following Subsection 3.2. The detailed modeling procedure can be found in Section 2.4. The RUL prediction will be performed on the four representative tools (i.e., Tool #1, #3, #5 and #6).

The AE sensor signal and vibration sensor signal on table (i.e., AE-table and vib-table) were used to implement multiple sensor fusion. Various original features were firstly extracted from multiple sensor signals before constructing the prognostics system. A wavelet ‘db5’ from Daubechies family wavelets in 4 levels was used to generate wavelet energy (WE) of the first two levels as two original features. Time domain approaches usually involve statistical features that could be sensitive to tool wear, such as root mean square (RMS), kurtosis, skewness, crest factor, variance, peak–peak and square root mean (SRM) [20,28]. Therefore, this study takes a multidomain method by considering the typical features from

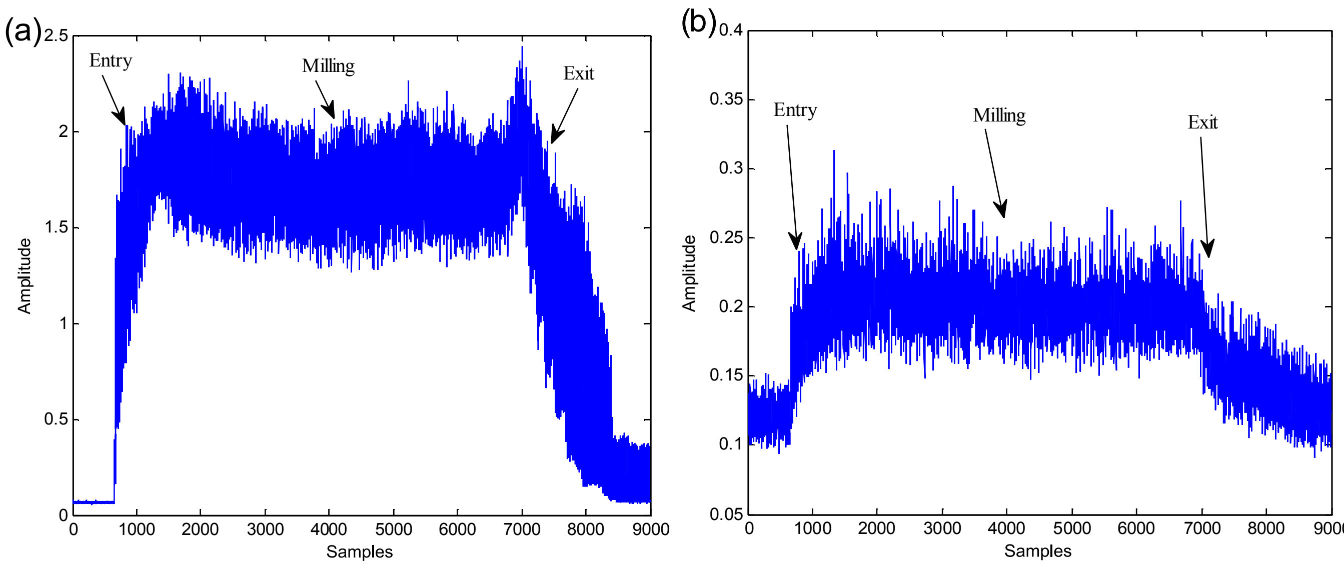

Fig. 4. Signals of one run (entry-milling-exit cutting procedure) from the two sensors: (a) AE sensor on table, and (b) Vibration sensor on table.

Table 2
Original features for tool health assessment (x is signal and N is numbers of samples).

Domain	Features	Formula
Time domain	RMS (F_{rms})	$RMS = \left(\frac{1}{N}\sum_{i=0}^{N-1} x_i^2\right)^{1/2}$
	Kurtosis (F_k)	$\text{Kurtosis} = \frac{1}{N}\sum_{i=0}^{N-1}(x_i - \bar{x})^4/\sigma^4$
	Skewness (F_s)	$\text{Skewness} = \frac{1}{N}\sum_{i=0}^{N-1}(x_i - \bar{x})^3/\sigma^3$
	Crest Factor (F_{cf})	$\text{Crest factor} = \frac{\max\|x_i\|}{RMS}$
	Peak-to-Peak (F_{pp})	$\text{Peak-to-Peak} = x_{\max} - x_{\min}$
	Variance (F_{var})	$Variance = \frac{1}{N}\sum_{i=0}^{N-1}(x_i - \bar{x})^2$
	Square mean root (F_{smr})	$SMR = \left(\frac{1}{N}\sum_{i=0}^{N-1}\|x_i\|^{1/2}\right)^2$
Time-frequency domain	Wavelet energy	$WE_j = \frac{1}{N_j}\sum_{k=1}^{N_j}[d_{j,k}^n(t)]^2$

the time and time-frequency domain as inputs to the prognostic model. Finally, an initial set of 18 features including 14 statistical features and 4 wavelet energy features from two sensors was compiled (as shown in Table 2). Before implement model construction, the normalization on the training and testing data were firstly performed.

In the procedure of data collection, approximately 9000 points were achieved for each milling run with entry-milling-exit cutting procedure. Fig. 4 (a) and (b) present the AE and vibration signals under no wear situation, respectively. For each run, about 1024*6 sample points were extracted during the steady stage. Thus, 6 observation vectors with 1024 points were generated for each milling run.

3.1. Tool health degradation assessment

The feature selection method (i.e., LRPR-FS) was firstly implemented on the training data to further improve the prognostic performance of LR. The top 9 ranked features by LRPR-FS were used as the inputs of LR. The full life cycle data were input to the baseline LR constructed by the LRMR algorithm to implement tool health degradation assessment. The LP calculation was performed by the baseline LR, and then the whole degradation propagation of each tool (i.e., the four representative tools Tool #1, #3, #5 and #6) is presented in Fig. 5. It can be observed from Fig. 5 that: (1) The tool health degradation procedure from health, slight degradation, to severe degradation has been quantified effectively in their full life, in which the failure probability (i.e., LP) increases consistently from near zero to near 1 as tool health deteriorates continuously; (2) It is generally a short period from slight degradation to severe degradation occurrence of incipient defect. This means that we should take some effective maintenance measurements to respond prior to catastrophic failure once the slight degradation is beginning in the tool; (3) In addition, although the life is different from each tool, the HI (i.e., LP) with a range from 0 to 1 still consistently depicts the tool degradation behavior in the whole running procedure. Fig. 6 presents the recorded wear magnitudes (some missing values were replaced with the mean of the previous and afterward wear values) of the four tools. It is obvious that the wear magnitudes of the four tools have a big discreteness, which often results in a challenge for the consistent health monitoring. However, LP provides an effective monitoring in the full life of tools based on the comprehensible quantification value and its good tendency.

To further verify the effectiveness of these features selected by LRPR-FS, LRMR with all original features was also implemented for health degradation assessment of tools and their results are presented in Fig. 7. It can be observed that the LR with the selected features provides the better monitoring results than the LR with all features. Firstly, Fig. 5 presents the better tendency of tool health degradation in the full life of the tools in comparison with Fig. 7. The HI, i.e. LP changes gradually from near zero value to near one value (see Fig. 5) when the tool starts wear down gradually in

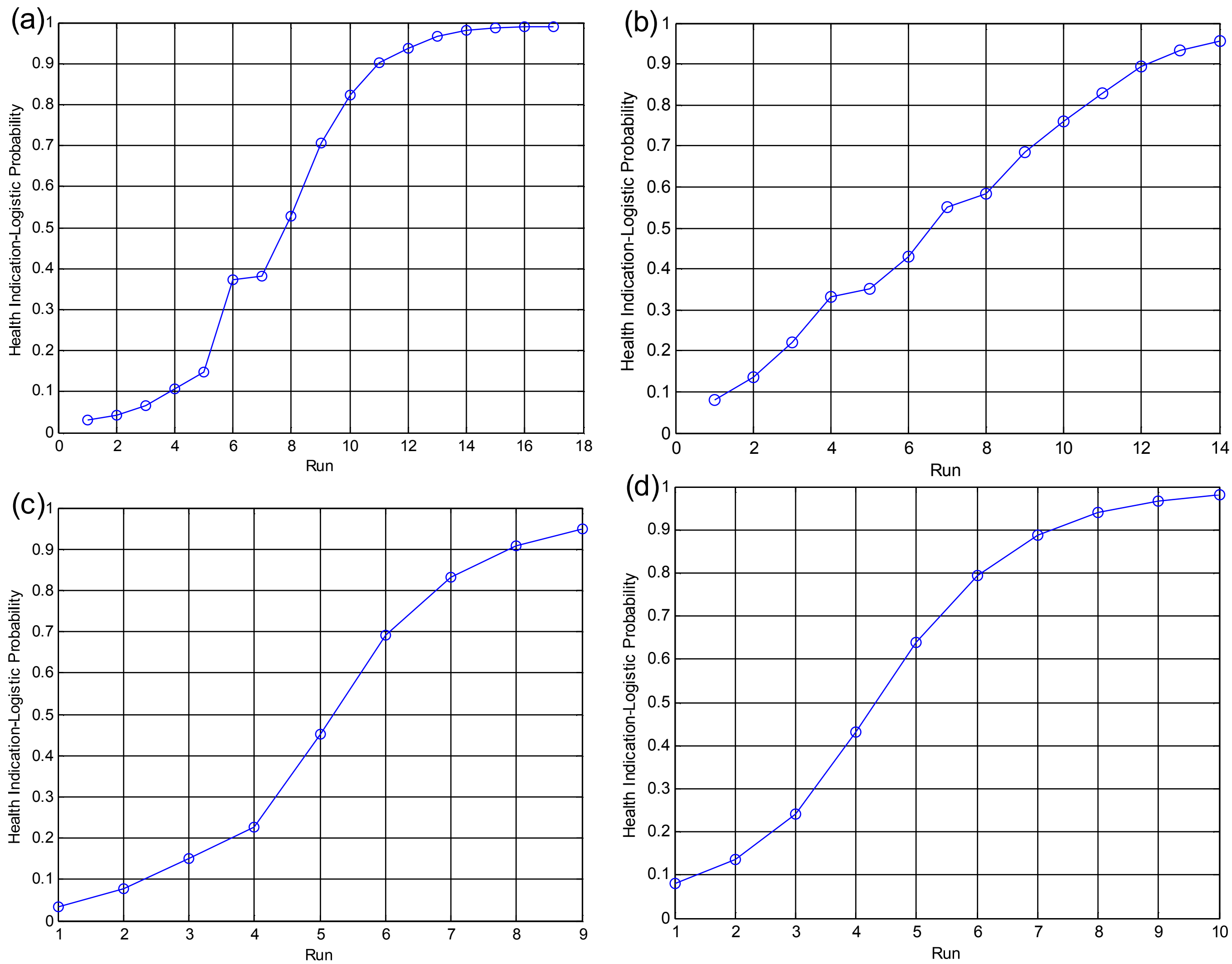

Fig. 5. Health degradation assessment of the representative four tools on their whole life time: (a) Tool #1, (b) Tool #3, (c) Tool #5, and (d) Tool #6.

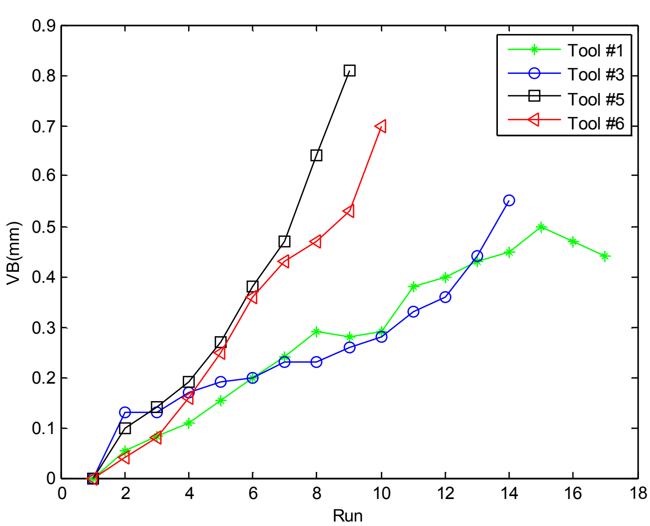

Fig. 6. The recorded wear magnitudes of the four tools.

their whole life (see Fig. 6). However, the LP changes gradually from near 0.2 to near 0.7 or 0.8 in Fig. 7. In addition, although the life and failure modes are different from each test and each tool, LP still consistently depicts the tool degradation behavior in the whole run-to-failure test (see Fig. 5), i.e., for the healthy and severe degradation states of all the four tools, LP is about near 0 and 1, respectively, which is very important in the real-world applications because it provides a consistent degradation assessment scheme for different tools working under different environments. In particular, the LR with the selected features provides the better results in the severe degradation phase of tools because the LPs approach 1 when the tools are in the last phase of the life, which is important for operators to determine whether this tool should be replaced. This comparison indicates that the prognostic feature selection based on LRPR-FS plays a key role to improve the effectiveness of LP for tool health monitoring.

For the purpose of comparison with the proposed HI (i.e., LP), the two typical features RMS and WE1 (the wavelet energy of the first level) on the whole life of the four representative tools are presented in Fig. 8. The general trend of the RMS from AE-table sensor and Vib-table sensor shows a decrease and increase over time, respectively. For WE1, these are an unclear downward trend of AE-table, and an upward trend of Vib-table. However, RMS and WE1 show strong inconsistent degradation pattern for these different tools. In conclusion, none of these typical features is capable of consistently providing a good trend of the degradation propagation (i.e., wears) with predictive or prognostic indication of an imminent tool failure. Therefore, it would be very difficult to construct a typical feature-based monitoring model to effectively

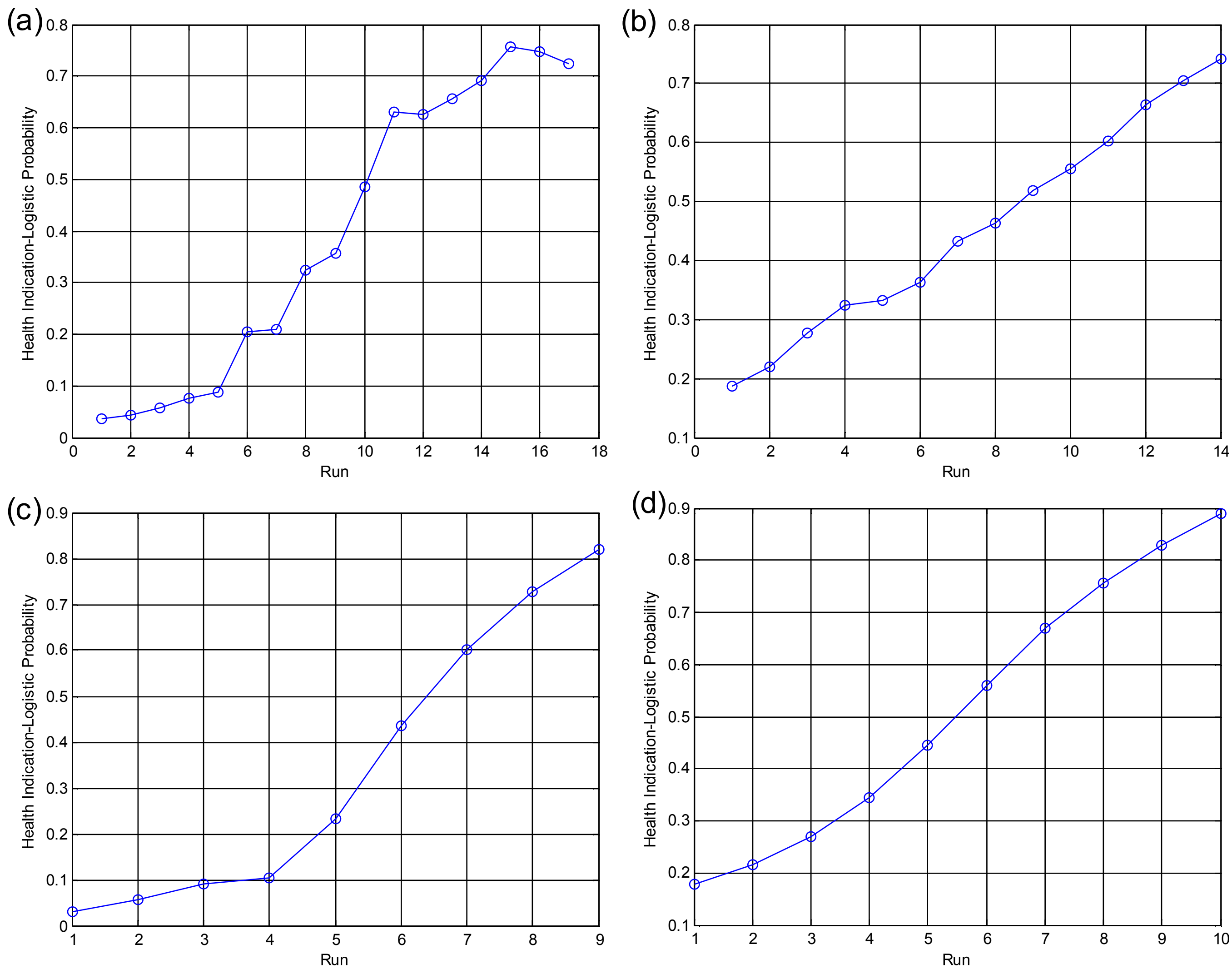

Fig. 7. Health degradation assessment of the representative four tools on their whole life time: (a) Tool #1, (b) Tool #3, (c) Tool #5, and (d) Tool #6.

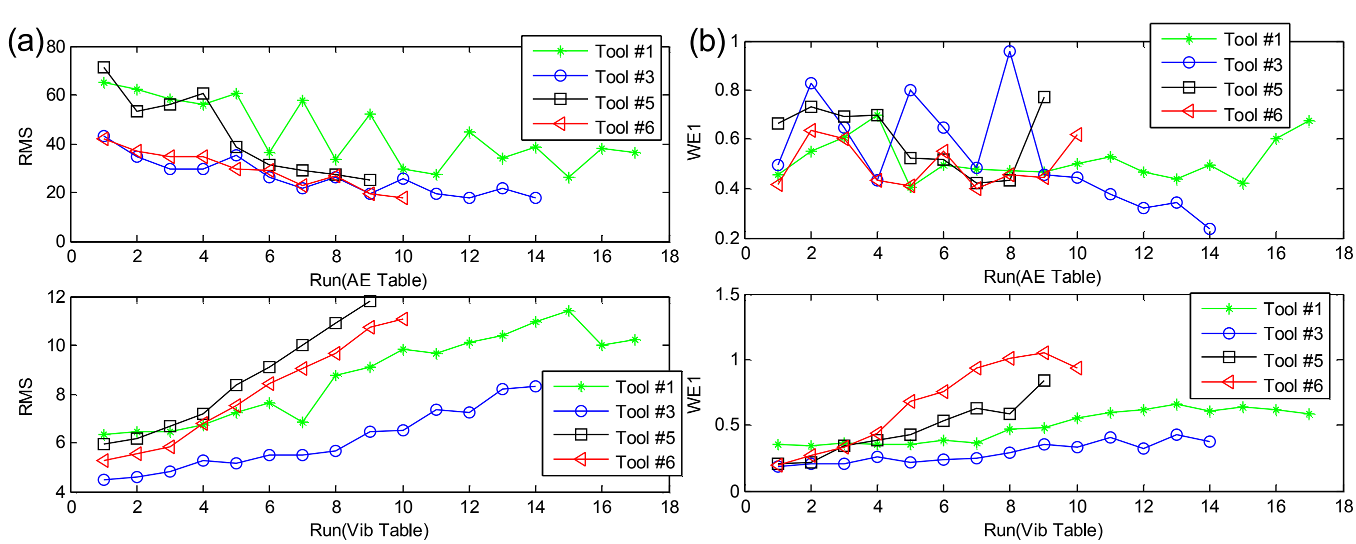

Fig. 8. Inconsistent degradation patterns of RMS and WE1 of AE and vibration sensors on table for the four tools, (a) RMS, and (b) WE1.

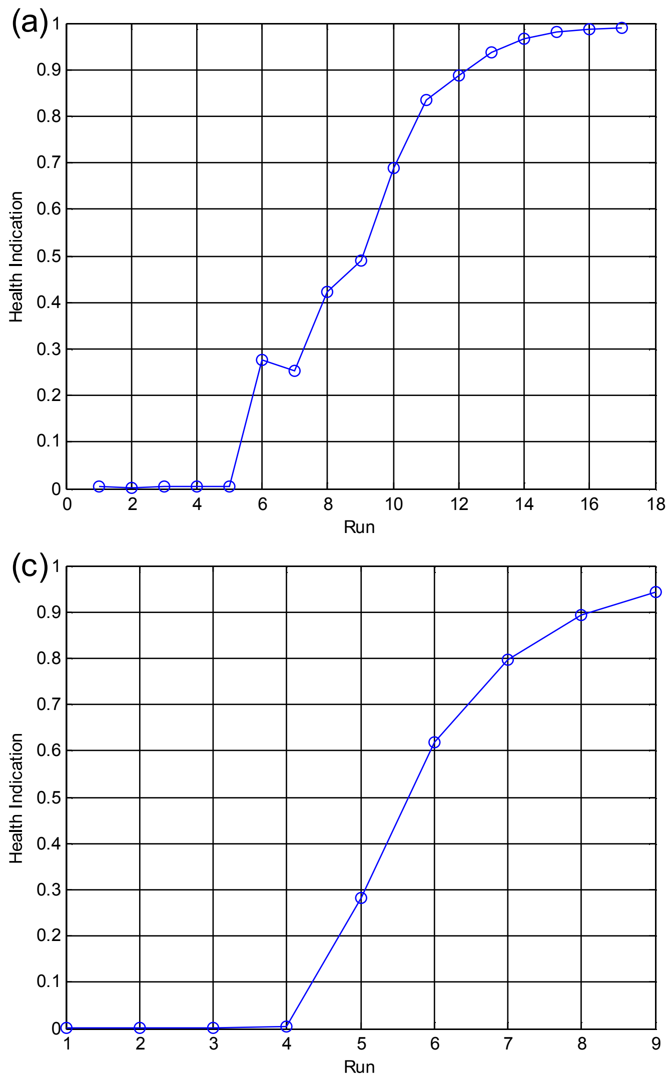

Fig. 9. Health degradation assessment of the representative four tools on their whole life time by using BPN: (a) Tool #1, (b) Tool #3, (c) Tool #5, and (d) Tool #6.

quantify tool wears. Therefore, an effective method to estimate the tool degradation should consider integration of health degradation trend and in-process condition symptoms based on multiple sensor-information fusion.

Furthermore, we provide a comparison between the LR and the back-propagation network (BPN) for tool health assessment to further demonstrate the effectiveness of the LR in the proposed prognostics system. BPN is a typical stochastic model and has been used for machine health assessment [40]. Thus, it is appropriate to compare the proposed method with the BPN for tool health assessment. The BPN structure consists of an input layer with 9 nodes for the 9 selected input features by LRPR-FS, one hidden layer with 15 hidden nodes, and an output layer with 1 node for the HI. The number of iterations is 500. The hyper tangent (tansig) and sigmoid (logsig) functions were used as the activation functions of the hidden and output layer, respectively. The same training dataset of the LR was also used to train the BPN. Fig. 9 presents the HI values of the BPN for the four representative tools. It is clear that the LR outperforms the BPN for tool health assessment. The comparison between Fig. 5 and 9 indicates that the LP indication provides a better tendency of the health degradation assessment for the full life of these tools in comparison with that of the BPN.

3.2. Multi-steps-ahead prediction

Based on the obtained LPs on the whole life of each tool, the two thresholds ($\delta_s = 0.4$ and $\delta_f = 0.95$) can be setup for triggering slight degradation and severe degradation(or failure) alarm, respectively. As the health degradation develops over time flow, LP will exceed or approach generally the severe degradation threshold after the slight degradation alarm is triggered. Fig. 10 presents the three-steps-ahead prediction of LR on the representative tools. A detection line (i.e., slight degradation) is plotted in Fig. 10, and the prediction is performed once the LP exceeds the detection line. The three-steps-ahead prediction is implemented when the time point moves ahead with three time-steps after the slight degradation alarm is triggered. It can be observed from Fig. 10 that the predictions are close to the actual values, and the prediction becomes more accurate when it is close to the final failure time point. However, the prediction accuracy late in its life of the unit is more important than that early in its life, this will more likely affect the decision on whether or not the preventive replacement should be performed.

For the purpose of comparison, relevance vector machine (RVM) [41] was used to implement multi-steps-ahead prediction. RVM is a Bayesian form representing a generalized linear model of identical

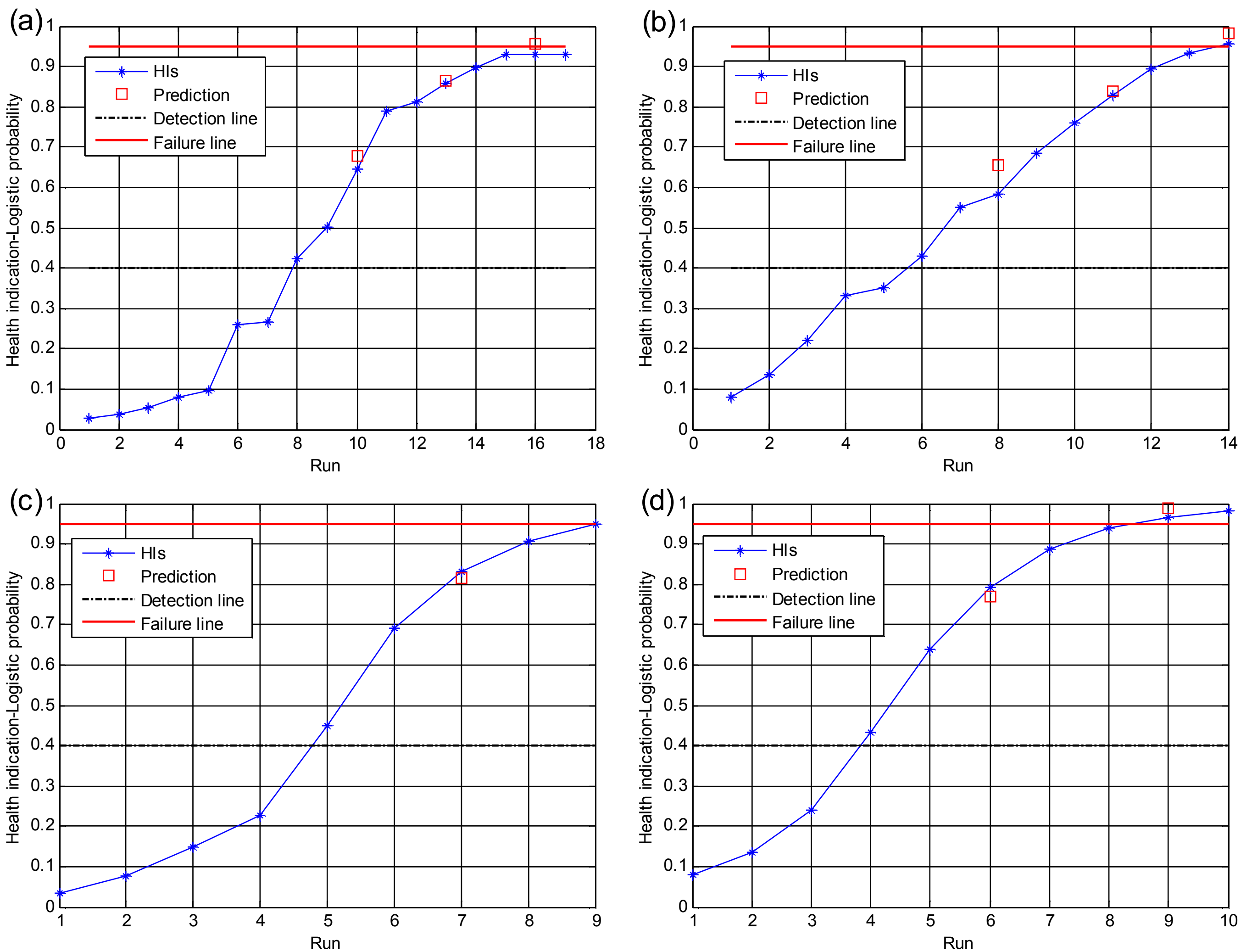

Fig. 10. Three-steps-ahead prediction of LR on the four tools, (a) Tool #1, (b) Tool #3, (c) Tool #5, and (d) Tool #6.

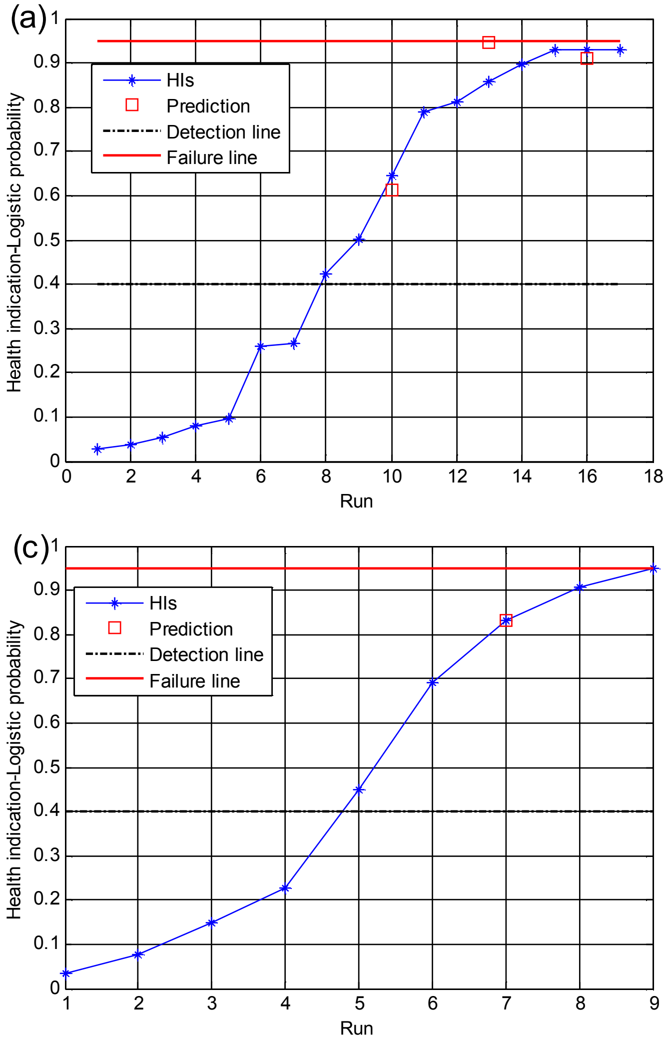

Fig. 11. Three-steps-ahead prediction of RVM on the four tools, (a) Tool #1, (b) Tool #3, (c) Tool #5, and (d) Tool #6.

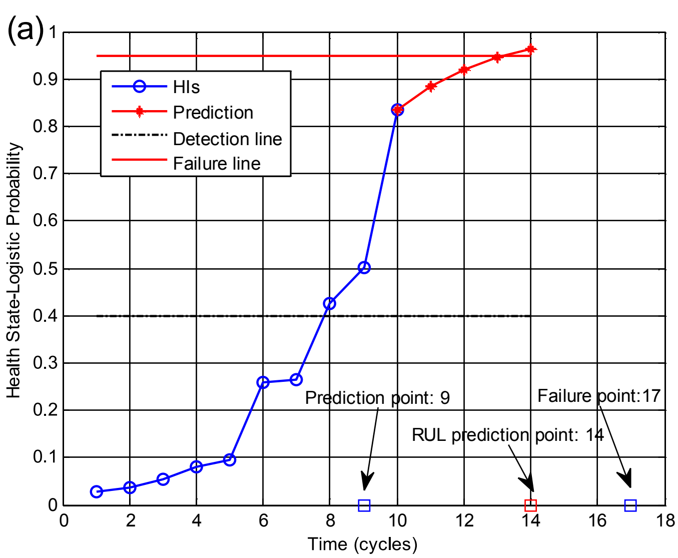

Fig. 12. Health state prediction for Tool #1 at different life phases: (a) Run 9, and (b) Run 13.

functional form of SVM. RVM has been applied recently in machine RUL prediction [42]. The modeling scheme for LR was also used for RVM, i.e., the tool ages and historic LPs of a tool were used to construct RVM. The Gaussian kernel function with the length scale (10) was used in RVM (the length scale can be setup by using try-error method to test the prediction performance on the training data). The prediction results of RVM for the four tools are shown in Fig. 11. In comparison with Fig. 11, it is clear from Fig. 10 that LR

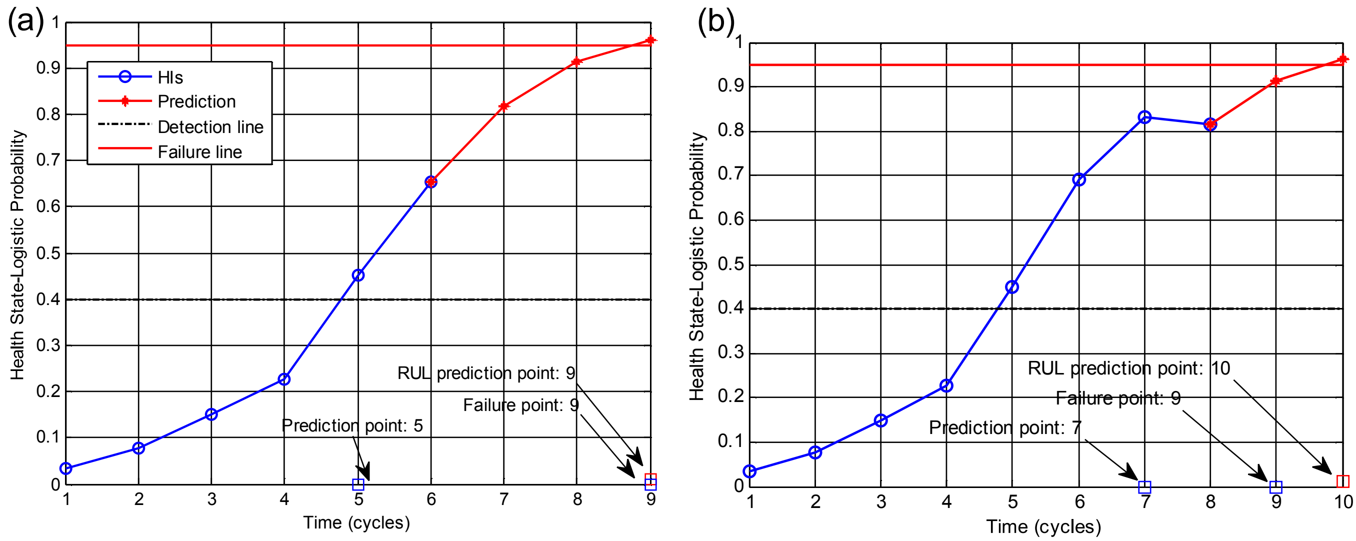

Fig. 13. Health state prediction for Tool #5 at different life phases: (a) Run 5, and (b) Run 7.

outperforms RVM for this multi-steps-ahead prediction. Moreover, the training of RVM spends more time cost than that of LR due to using of kernel function. Thus, LR is more effective for tool health prediction than RVM.

3.3. Tool RUL prediction

In order to further illustrate the prediction performance of the LR-based RUL predictor, the final failure time points (i.e., RUL) of Tool #1 and #5 were predicted at different life phases. In this study, the end point of the whole life time is regarded as the failure point of the tested tools. For online prediction of LR, the time point when the LR prediction exceeds the severe degradation threshold (i.e., δ_f) was determined as the predicted failure time point. In this study, we setup $\delta_f = 0.95$ according to the test results of the failure time points of the tools used for training. It should be pointed out that the threshold setup must be based on the requirements of real-world applications.

Fig. 12(a) and (b) show the prediction results for Tool #1 at Run 9 and 13, respectively. Fig. 13(a) and (b) show the prediction results for Tool #5 at Run 5 and 7, respectively. The prediction point, the predicted failure point and the actual failure point are also marked in Fig. 12 and 13. It can be observed from Fig. 12 and 13 that the LR traces the gradual nonlinear degradation of the tool health in the different phases of the whole life. In the early phase of tool life, the prediction results of the proposed method could have a relative big gap with the real RUL (see Fig. 12). Since the health changes of the tool are small when it is approaching the end of the life, the prediction starting at the late life of tools produces more accurate results based on these new arrival measurements. This can be attributed to the adaptive learning of the LR predictor, which assumes neither a regular pattern nor the stability of the underlying degradation process.

4. Conclusions

This paper proposes a novel tool health prognostics system base on LRPR and LRMR that consider penalization and manifold regularization, respectively. In this system, an LRPR-based feature selection method is developed to select effective prognostic features. Information fusion from multiple sensor signals is further performed through using LRMR. An LP-based failure probability indication is developed for tool health assessment in the full life of tool. The LR-based RUL predictor is further developed in the proposed prognostics system. The experimental results verify that a consistent and pragmatic degradation indicator with high comprehensibility can be effectively extracted by the LP derived from the LR model. With effective inputs using LP, LR is capable of performing tool RUL prediction online. The experimental results based on the full life datasets from a tool test-bed indicate that the proposed scheme is effective for tool health prognostics. The improvement of performance and applicability of LR with penalization and manifold regularization would increase the ability of the proposed system to implement tool condition prognosis successfully. It should be pointed out that there is a lot of ways to work with the used data of the tool full life [38,39] in future work including signal processing, prognostic feature extraction, tool failure detection, and tool wear prediction. With further extension and improvement of the proposed model, thus, it should consider to use l1 regularization in LRPR to select effective features, and to online learn the tool health propagation to improve its applicability in real-world cases.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (No. 51375290, 71777173), the Fundamental Research Funds for Central Universities, and Shanghai Science and Technology Commission's Innovative Scientific and Technological Action Plan (No. 17511109204). The authors would like to express sincere appreciation to the anonymous referees for their detailed and helpful comments to improve the quality of the paper.

References

[1] K. Zhu, Y.S. Wong, G.S. Hong, Wavelet analysis of sensor signals for tool condition monitoring: a review and some new results, J. Manuf. Sci. Eng.–Trans. ASME 49 (7–8) (2009) 537–553.
[2] J.T. Roth, D. Djurdjanovic, X. Yang, L. Mears, T. Kurfess, Quality and inspection of machining operations: tool condition monitoring, J. Manuf. Sci. Eng.–Trans. ASME 132 (4) (2010) 041015–041030.
[3] P.C. Khiang, J.H. Zhou, H.C. Yan, PDF and breakdown time prediction for unobservable wear using enhanced particle filters in precognitive maintenance, IEEE Trans. Instrum. Meas. 4 (3) (2015) 649–659.
[4] A.N. Khajavi, R. Komanduri, On multisensory approach to drill wear monitoring, Ann. CIRP 41 (1993) 71–74.
[5] S. Kurada, C. Bradley, A review of machine vision sensors for tool condition monitoring, Comput. Ind. 4 (1) (1997) 55–72.
[6] S. Dutta, S.K. Pal, S. Mukhopadhyay, R. Sen, Application of digital image processing in tool condition monitoring: a review, CIRP J. Manuf. Sci. Technol. 6 (3) (2013) 212–232.
[7] S.D.E. Dimla, P.M. Lister, On-line metal cutting tool condition monitoring I: force and vibration analysis, Int. J. Mach. Tools Manuf. 40 (5) (2000) 739–768.

[8] D.R. Salgado, F.J. Alonso, An approach based on current and sound signals for in-process tool wear monitoring, Int. J. Mach. Tools Manuf. 47 (14) (2007) 2140–2152.
[9] D.E. Lee, I. Hwang, C.M.O. Valente, J.F.G. Oliveira, D.A. Dorfeld, Precision manufacturing process monitoring with acoustic emission, Int. J. Mach. Tools Manuf. 46 (2) (2006) 176–188.
[10] B. Kaya, C. Oysu, H.M. Ertunc, H. Ocak, A support vector machine-based online tool condition monitoring for milling using sensor fusion and a genetic algorithm, Proc. Inst. Mech. Eng. Part B: J. Eng. Manuf. 226 (11) (2012) 1808–1818.
[11] J. Yu, Online tool wear prediction in drilling operations using selective artificial neural network ensemble model, Neural Comput. Appli. 0 (4) (2011) 473–485.
[12] F. Camci, R.B. Chinnam, Health-state estimation and prognostics in machine processes, IEEE Trans. Autom. Eng. 7 (3) (2010) 581–597.
[13] J. Yu, Machine tool condition monitoring based on an adaptive Gaussian mixture model, J. Manuf. Sci. Eng. – Trans. ASME 134 (3) (2012) 031004–031012.
[14] T. Kalvoda, A cutter tool monitoring in machining process using Hilbert–Huang transform, Int. J. Mach. Tools Manuf. 50 (5) (2010) 495–501.
[15] W.H. Wang, G.S. Hong, Y.S. Wong, K.P. Zhu, Sensor fusion for online tool condition monitoring in milling, Int. J. Prod. Res. 5 (21) (2007) 5095–5116.
[16] K. Jemielnaik, O. Otman, Tool failure detection based on analysis of acoustic emission signals, J. Mater. Process. Technol. 76 (1–3) (1998) 192–197.
[17] C.H.R. Martins, P.R. Aguiar, A. Frech, E.C. Bianchi, Tool condition monitoring of single-point dresser using acoustic emission and neural networks models, IEEE Trans. Instrum. Meas. 3 (3) (2014) 667–679.
[18] R.C. Luo, C.C. Yih, K.L. Su, Multisensor fusion and integration: approaches, applications, and future research directions, IEEE Sens. J. 2 (2) (2002) 107–119.
[19] K. Jemielniak, P.J. Arrazola, Application of AE and cutting force signals in tool condition monitoring in micro-milling, CIRP J. Manuf. Sci. Technol. 1 (2) (2008) 97–102.
[20] C. Aliustaoglu, H. Metin Ertunc, H. Ocak, Tool wear condition monitoring using a sensor fusion model based on fuzzy inference system, Mech. Syst. Signal Process. 23 (2) (2009) 539–546.
[21] J. Yu, A hybrid feature selection scheme and self-organizing map model for machine health assessment, Appl. Soft Comput. 1 (5) (2011) 4041–4054.
[22] J.H. Zhou, C.K. Pang, Z.W. Zhong, F.L. Lewis, Tool wear monitoring using acoustic emissions by dominant-feature identification, IEEE Trans. Instrum. Meas. 0 (2) (2011) 547–559.
[23] J. Yu, Health condition monitoring of machines based on hidden Markov model and contribution analysis, IEEE Trans. Instrum. Meas. 1 (8) (2012) 2200–2211.
[24] E. Zio, P. Baraldi, G. Gola, Feature-based classifier ensembles for diagnosing multiple faults in rotating machinery, Appl. Soft Comput. 8 (2008) 1365–1380.
[25] S. Wang, J. Yu, E. Lapira, J. Lee, A modified support vector data description based novelty detection approach for machinery components, Appl. Soft Comput. 3 (2) (2013) 1193–1205.
[26] O. Geramifard, J.X. Xu, J.H. Zhou, X. Li, A physically segmented hidden Markov model approach for continuous tool condition monitoring: diagnostics and prognostics, IEEE Trans. Ind. Informat. 8 (4) (2012) 964–973.
[27] J. Sanz, R. Perera, C. Huerta, Gear dynamics monitoring using discrete wavelet transformation and multi-layer perceptron neural networks, Appl. Soft Comput. 12 (9) (2012) 2867–2878.
[28] S. Lee, L. Li, J. Ni, Online degradation assessment and adaptive fault detection using modified hidden Markov model, J. Manuf. Sci. Eng.–Trans. ASME 132 (2) (2010), 021010-(1-11).
[29] W. Caesarendraa, A. Widodob, B.S. Yang, Application of relevance vector machine and logistic regression for machine degradation assessment, Mech. Syst. Signal Process. 24 (4) (2010) 1161–1171.
[30] D.W. Hosmer, S. Lemeshow, Applied Logistic Regression, 2nd ed., Wiley, New York; Chichester, 2000, ISBN 0-471-35632-8.
[31] J.M. Lucas, M.S. Saccucci, Exponentially weighted moving average control schemes: properties and enhancements, Technometrics 2 (1) (1990) 1–12.
[32] W.M. Houston, D.J. Woodruff, Empirical Bayes Estimates of Parameters from the Logistic Regression Model, ACT Research Report Series, 1997.
[33] A.E. Hoerl, R.W. Kennard, Ridge regression: application to nonorthogonal problems, Technometrics 7 (1) (1970) 69–82.
[34] X. He, D. Cai, Y. Shao, H. Bao, J. Han, Laplacian regularized Gaussian mixture model for data clustering, IEEE Trans. Knowl. Data Eng. 3 (9) (2011) 1406–1418.
[35] M. Belkin, P. Niyogi, V. Sindhwani, Manifold regularization: a geometric framework for learning from examples, J. Mach. Learn. Res. 7 (2006) 2399–2434.
[36] F.R.K. Chung, Spectral graph theory, AMS, in: Regional Conf Series in Mathematics, Vol. 92, 1997.
[37] M.Y. Park, T. Hastie, Penalized logistic regression for detecting gene interactions, Biostatistics 9 (1) (2008) 30–50.
[38] K. Goebel, Management of Uncertainty in Sensor Validation, Sensor Fusion, and Diagnosis of Mechanical Systems Using Soft Computing Techniques, Ph.D Thesis, Department of Mechanical Engineering, University of California at Berkeley, 1996.
[39] A. Agogino, K. Goebel, Mill Data Set, BEST, lab, UC Berkeley, NASA Ames Prognostics Data Repository, NASA, Ames, Moffett Field, CA, 2007 http://ti.arc.nasa.gov/project/prognostic-data-repository.
[40] J. Yan, C. Guo, X. Wang, A dynamic multi-scale Markov model based methodology for remaining life prediction, Mech. Syst. Signal Process. 25 (4) (2011) 1364–1376.
[41] M.E. Tipping, The relevance vector machine Advances in Neural Information Processing Systems, vol.12, MIT Press, Cambridge, MA, 2000, pp. 652–658.
[42] F.D. Maio, K.L. Tsui, E. Zio, Combining relevance vector machine and exponential regression for bearing residual life estimation, Mech. Syst. Signal Process. 31 (8) (2012) 405–427.