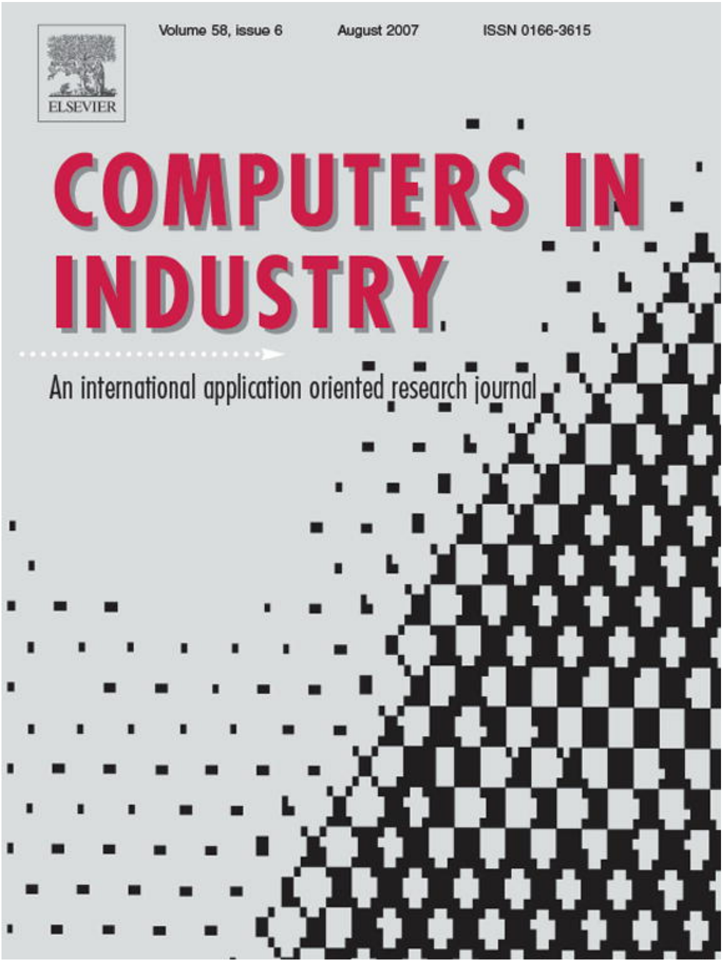

Volume 58, issue 6
August 2007
ISSN 0166-3615
ELSEVIER
COMPUTERS IN INDUSTRY
An international application oriented research journal

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Computers in Industry 58 (2007) 558–566
COMPUTERS IN INDUSTRY
www.elsevier.com/locate/compind

Similarity based method for manufacturing process performance prediction and diagnosis

Jianbo Liu [a], Dragan Djurdjanovic [a,*], Jun Ni [a], Nicolas Casoetto [a], Jay Lee [b]

[a] University of Michigan, Department of Mechanical Engineering, 2350 Hayward Street, Ann Arbor, MI 48109-2125, United States

[b] University of Cincinnati, Department of Mechanical, Industrial and Nuclear Engineering, 598 Rhodes Hall, P.O. Box 210072, Cincinnati, OH 45221-0072, United States

Received 18 February 2006; received in revised form 3 October 2006; accepted 23 December 2006
Available online 30 January 2007

Abstract

Full realization of all the potentials of predictive maintenance highly depends on the accuracy of long-term predictions of the remaining useful life of manufacturing equipments. In this paper, we propose a new method that is capable of achieving high long-term prediction accuracy by comparing signatures from any two degradation processes using measures of similarity that form a match matrix (MM). Through this concept, we can effectively include large amounts of historical information into the prediction of the current degradation process. Similarities with historical records are used to generate possible future distributions of features indicative of process performance, which are then used to predict the probabilities of failure over time by evaluating overlaps between predicted feature distributions and feature distributions related to unacceptable equipment behavior. The analysis of experimental results shows that the proposed method can yield a noticeable improvement of long-term prediction accuracy in terms of mean prediction errors over the Elman Recurrent Neural Network (ERNN) based prediction, which was shown in the past literature to predict well behavior of highly non-linear and non-stationary time series.

Keywords: Match matrix; Predictive maintenance; Time series prediction; Failure prediction

1. Introduction

Reducing downtime cost and achieving near zero downtime is the ultimate goal of predictive maintenance. However, it is impossible to realize all the advantages of predictive maintenance without accurate predictions of the remaining useful life before the actual failure occurs. The inaccurate predictive information may result in either unnecessary maintenance, such as early replacement of components, or production downtime because of unexpected machine failures. Therefore, the accuracy of remaining useful life prediction, particularly the long-term prediction, which gives sufficient time to prepare for a maintenance operation, plays an essential role in the full realization of the potentials of predictive maintenance.

The degradation process cannot be directly observed or measured in general. It can only be observed indirectly through the time series of features extracted from available process measurements, such as vibrations and forces. Extrapolating these time series in time can help us to predict risks of failure or unacceptable behavior of the process over time. This places great significance on one's ability to accurately and reliably predict the feature time series. A variety of techniques have been used in the past for time series modeling and prediction. Parametric linear prediction techniques, such as Auto-Regressive Moving Average (ARMA) [1,2] or Kalman filtering [3], may work well only for short-term predictions because of their assumption that the considered time series is generated from a linear process. These linear prediction techniques are well interpretable but with limited capabilities for predicting real world problems which are usually complex and non-linear. Variety of approaches for predicting non-linear time series, such as fuzzy time series and clustering [4,5], multi-resolution wavelet models [6–8] and neural networks [9], has been extensively studied in the literature. Without a priori knowledge

* Corresponding author. Tel.: +1 734 763 9975; fax: +1 734 936 0363.
E-mail address: ddjurdja@umich.edu (D. Djurdjanovic).

0166-3615/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.compind.2006.12.004

about the time series under consideration, selecting an appropriate non-linear model and its structure is a difficult task. Among these non-linear prediction technique, neural networks, such as Radial Basis Function (RBF) networks [10] Multi-Layer Perceptron (MLP) neural network [11] and Recurrent Neural Networks (RNN) [12–14], maybe the most extensively applied techniques for complex non-linear time series predictions because of their capability to approximate non-linear functional and dynamic dependencies [9]. Unlike feed-forward networks such as RBF and MLP, which have limitations of identifying temporal relationships in the time series, RNN takes into account temporal dependencies through local or global feedback connections in the network. As a result, RNN is able to approximate a wide class of non-linear dynamical systems [11]. However, the commonly used gradient descent algorithms for RNN training exhibit certain problems during training, such as having difficulty dealing with long-term dependencies in the time series [12], which in turn limits their capability of achieving accurate long-term predictions. In addition, finding a suitable number of hidden neurons and appropriate RNN structure remains a challenging problem.

In many manufacturing facilities, large amounts of historical records of past equipment behavior are available and can be used to enhance and reinforce the equipment performance prediction. The goal of the method pursued in this paper is to increase long-term prediction of signatures depicting equipment performance through incorporation of historical records into the prediction process, while at same time capturing the dynamic changes of the signatures as the process changes.

The following terminology will be used throughout the paper:

- The term *feature vectors* will be used for signatures describing the current state of the machine/process and containing a number of features which are considered to be correlated to the process degradation. The evolution/dynamics of those feature vectors over time is then essentially the characteristic of the degradation process. Once appropriate sensors and adequate features are selected, one can estimate the current degradation state and conduct prediction based on the evolution of the feature vectors.
- The time interval between two consecutive maintenance cycles will be referred to as *a run*. The maintenance operation may be a component replacement, part repair, etc. From the set of sensor readings in each past run, a time-ordered sequence of feature vectors can be extracted to represent the degradation states of the process. The time ordering of the features is necessary in order to be able to explore the temporal evolution of the feature vectors and characterize the process degradation over time.

The rest of the paper is organized as follows. In Section 2, the new match matrix (MM) based prediction method is introduced. The newly proposed approach has been tested on predicting tool degradation in a boring process and the results are shown in Section 3. Section 4 provides conclusions.

2. Match matrix based prediction method

2.1. Concept of a match matrix

We assume that the degradation process is described by a time series of signatures extracted from relevant sensors. The complexity of the degradation process such as tool wear and component aging, often makes the resulting time series so irregular that it cannot be modeled accurately using linear parametric techniques. In order to improve the accuracy the long-term prediction of the irregular behavior of such time series produced by non-linear dynamic systems, a method is needed to effectively utilize the existing historical information about process/equipment behavior over a long period of time. This will involve the comparison of time series from a potentially large number of runs and synthesis of that information into the prediction of the current degradation process.

The Mahalanobis distance is commonly used to evaluate the distance between multidimensional feature vectors whose components are quantities that have different ranges and amounts of variations. The Mahalanobis distance between two n-dimensional feature vectors $\vec{f}^{\,i} = [f_1^i, f_2^i, \ldots, f_n^i]^{\mathrm{T}}$ and $\vec{f}^{\,j} = [f_1^j, f_2^j, \ldots, f_n^j]^{\mathrm{T}}$ can be calculated as follows [15]

$$d_{i,j} = (\vec{f}^{\,i} - \vec{f}^{\,j})^{\mathrm{T}} \Sigma^{-1} (\vec{f}^{\,i} - \vec{f}^{\,j}) \tag{1}$$

where Σ^{-1} is the inverse of the covariance matrix which can be estimated from samples of feature vectors. The similarity between feature vectors $\vec{f}^{\,i}$ and $\vec{f}^{\,j}$ can then be expressed as

$$s_{i,j} = \exp(-d_{i,j}) \tag{2}$$

One can note that, if the Mahalanobis distance between two feature vectors is small, then $s_{i,j}$ is approximately 1, while as the Mahalanobis distance between the feature vectors grows, the similarity $s_{i,j}$ between them approaches 0.

If historical records of past signatures collected from several identical machines operating under identical conditions are available to describe various degradation mechanisms, the comparison of two different realizations of the degradation processes involves assessing the similarity between any pair of feature vectors from those two runs. Then, any time a new feature vector $\vec{f}^{\,j}_{\text{current}}$ from the current run is extracted, it can be compared with all the feature vectors in the past run by computing the similarities defined by (2). Based on this intuitive concept of feature comparison between the past and the current runs, one can introduce the notion of a *match matrix* as follows.

2.1.1. Definition of a match matrix

Let us assume that there are P feature vectors in the feature vector sequence $\{\vec{f}^{\,1}_{\text{current}}, \vec{f}^{\,2}_{\text{current}}, \ldots, \vec{f}^{\,P}_{\text{current}}\}$ describing the current run, and Q vectors in the feature vector sequence $\{\vec{f}^{\,1}_{\text{past}}, \vec{f}^{\,2}_{\text{past}}, \ldots, \vec{f}^{\,Q}_{\text{past}}\}$ describing the past run. One can compare the degradation patterns of the past and the the current run is being compared through a $Q \times P$ matrix S in

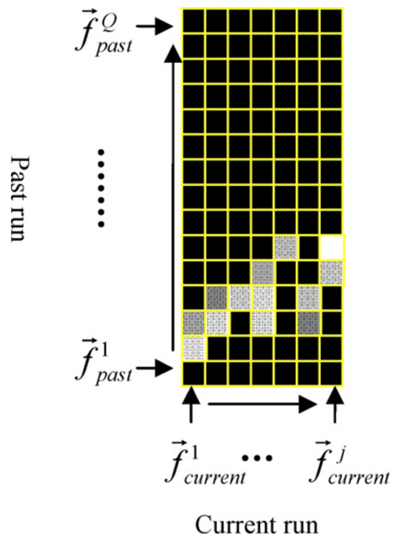


Fig. 1. Creation of the match matrix. The darker the color the smaller the value of the similarity.

which each element $s_{i,j}$ is the similarity between the jth feature vector $\vec{f}^{\,j}_{\text{current}}$ of the current run and ith feature vector $\vec{f}^{\,i}_{\text{past}}$ of the past run, as described by Eqs. (1) and (2). This matrix will be referred to as the *match matrix*. Fig. 1 illustrates the procedure of creating a match matrix between two runs.

If there are signatures from M past runs in the historical records, it is possible to create M match matrices. Every time a new feature vector in the current run is extracted from sensor readings, its comparison with all the feature vectors based on (2) will involve addition of a column of similarity measures to each of the M match matrices, which will involve $\sum_{m=1}^{M} K_m$ comparisons, where K_m represents number of feature vectors in mth past run.

The underlying principle can be illustrated using a simple example of three time series x_1, x_2 and x_3 containing two-dimensional feature vectors as shown in Fig. 2.

Fig. 2(a) shows time series x_1 and x_2, which evolve in different ways and represent two different degradation processes.

Fig. 2(b) shows a feature vector sequence x_3 that exhibits a similar degradation trend as x_1. Fig. 3(a) shows the match matrix created between time series x_2 and x_3, while Fig. 3(b) shows the match matrix between time series x_1 and x_3. The distinction between the underlying degradation processes represented by time series x_2 and x_3 is displayed by the fact that the values of the similarity measures at the best matches quickly drop down as the degradation progresses. On the other hand, the match matrix between x_1 and x_3 shown in Fig. 3(b) reveals that the highest matches in this case always occur around the main diagonal of the match matrix, and the values of similarity measures at the best matches remains high. This pattern of highest matches indicates that time series x_1 and x_3 have highly similar dynamics inside

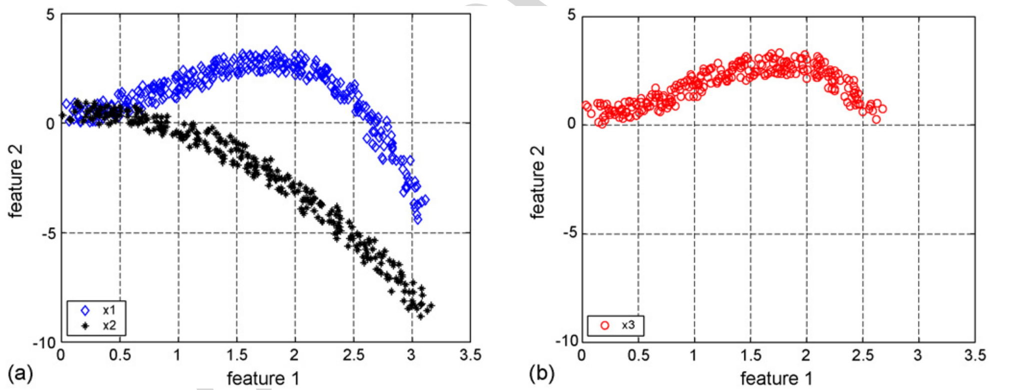


Fig. 2. Sample time series to illustrate the concept of the match matrix. Feature time series evolve from left to right.

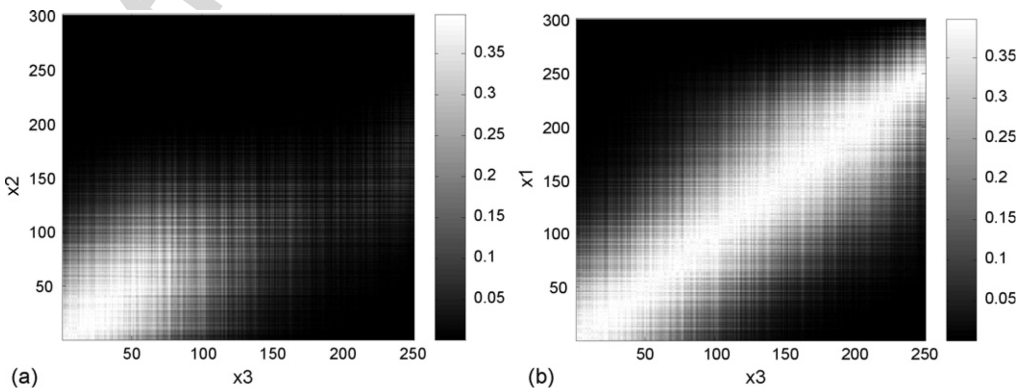


Fig. 3. Match matrix between time series x_2 and x_3 (a), and time series x_1 and x_3 (b).

them and time series x_1 can be used to predict the behavior of the time series x_3 (or *vice versa*).

2.2. Analysis of match matrix

If all the signals are collected from identical machines operating under identical operating conditions, one would expect that the changes in the time series of high similarity indices corresponding to each signal in the current run (i.e., corresponding to each column of the match matrix) will yield an indication of the evolution of the feature vectors over time so that the future behavior of signatures of the current run can be predicted. Hence, the analysis of the match matrices aimed at predicting future signatures of the current degradation run will focus on the time series made of *best match indices* in match matrices between the current run and the past runs.

For the jth feature vector $\vec{f}^{\,j}_{\text{current}}$ of the current run, the best match index in the match matrix between the current and a past run, is the index of the feature vector in the past run that is the closest to it in the sense of the Mahalanobis. This index actually corresponds to the row index of the match matrix that has the highest similarity index in the jth column of the match matrix.

Nevertheless, our experience with features extracted out of real-life processes suggests that very frequently, the feature vectors contain a considerable level of noise, which makes the highest similarity indices distribute widely among the indices of the feature vectors in any specific past run. In order to reduce the variation of the best match index time series obtained, the mean index of similarities is calculated from each column j of any match matrix as

$$I_j = \frac{\sum_{i=1}^{Q} i s_{i,j}}{\sum_{i=1}^{Q} s_{i,j}} \tag{3}$$

where Q denotes the number of feature vectors in the past run. The time series of mean best matches makes it possible to apply the linear parametric prediction technique based on ARMA modeling [1] to predict the future values of the time series of mean best match indices between the current and the past runs. Furthermore, ARMA prediction also allows one to analytically evaluate the uncertainty of prediction of the mean best match indices.

2.3. Match matrix based prediction

The prediction procedure based on the match matrix concept can be accomplished as follows. One match matrix between performance signatures in the current run whose behavior is being predicted and signatures in each of M past runs can be constructed (i.e., one will have M match matrices), with each match matrix yielding a time series of the mean indices of the best match. Future values of mean best match indices can now be predicted using a linear parametric technique, such as ARMA model based prediction, where one possible way for determining the adequate structure of the ARMA model is the F-test based procedure described in ref. [1]. ARMA modeling method also allows analytical expression for the variance of prediction errors and can therefore yield uncertainty/confidence intervals for the prediction of the mean best match indices. The predicted value of the best match index in a given match matrix is the index of the feature vector in the corresponding past run and is more likely to be similar to the feature vector of the current run in the future. This results in the predictive capability of this method (Fig. 4).

Based on the prediction confidence interval in each match matrix, a distribution of feature vectors can be obtained through random sampling and subsequent linear combination. Each sample in the prediction confidence interval of the time series extracted from any given match matrix corresponds to a feature vector in the corresponding past run. The M feature vectors (one from each past run) can be linearly combined to form a predicted feature vector. The weights are determined based on the similarities embedded in the corresponding match matrices. A set of predicted feature vectors can be generated by repeating this process. In this paper, the weights w_m associated with the feature vector from the past run m are calculated as

$$w_m = \frac{\bar{s}_m}{\sum_k \bar{s}_k}, \quad m = 1, 2, \ldots, M \tag{4}$$

where M is number of the past runs with which match matrices are calculated and $\bar{s}_m$ denotes the average similarity between the latest several feature vectors in the current run with the feature vectors in the mth past run. Such weighting places stronger emphasis on the features observed in runs that display

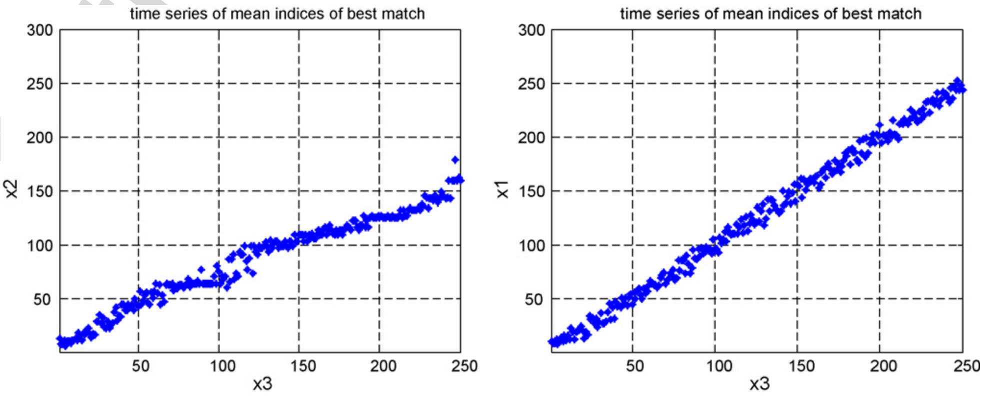


Fig. 4. Time series of mean indices of best match extracted from the matrices shown in Fig. 3. They evolve from left to right.

higher similarity with the run whose behavior we are trying to predict (the current run).

The weighting strategy results in an advantageous property of this method in that even if ARMA predictions of the time series of mean indices of the best match go astray in one of the match matrices, the accuracy of predictions will not be affected significantly. The reason for this is that such a move away from the actual degradation process causes the corresponding similarities in the match matrix to become much smaller, which consequently will assign the feature vectors obtained from this match matrix very small weights. Thus only the most similar degradation processes contribute to the prediction process. Consequently, long term predictions based on the match matrix method have a better chance of being more accurate than those based on purely stationary techniques, such as ARMA modeling.

Random sampling of the predicted best match indices and fusion of the corresponding predicted feature vectors obtained from each match matrix will result in a multitude of predicted process/equipment behavior features. This multitude of signatures represents the predicted feature distribution for each step ahead in time. In this paper, mixtures of multivariate Gaussians are used to describe the predicted distributions of performance signatures

$$f(\vec{x}) = \sum_{k} p_k N(\vec{\mu}_k, \Sigma_k) \tag{5}$$

where p_k are the weights for the kth mixture and $N(\vec{\mu}_k, \Sigma_k)$ denotes a multivariate Gaussian distribution with mean vector $\vec{\mu}_k$ and covariance matrix Σ_k. Mixture Gaussians can be used to describe any probability density function with arbitrarily small approximation error and the parameters p_k, $\vec{\mu}_k$ and Σ_k of the mixture density model can be estimated using, for example, the Expectation-Maximization (EM) method [15]. Based on these predicted feature distributions, one can use the feature vector corresponding to the highest value of the probability density function as the most likely future feature vector, or the gravity mean of the distribution of predicted feature vectors to predict the expected value of the future feature vector.

Match matrix based prediction of the distribution of possible feature vectors over time allows one to estimate the future probabilities of failure or unacceptable process behavior over time, provided that a set of signatures describing unacceptable process behavior exists and is available for one to describe the corresponding feature distribution. This can be accomplished by calculating the overlap of the predicted feature distributions with that corresponding to the process failure or unacceptable behavior. The overlap between two mixtures of Gaussians can be calculated analytically based on the correlation coefficient in the space of square integrable functions [16]

$$\frac{\int f_1(\vec{x}) f_2(\vec{x})\,\mathrm{d}\vec{x}}{\sqrt{\int (f_1(\vec{x}))^2\,\mathrm{d}\vec{x}}\sqrt{\int (f_2(\vec{x}))^2\,\mathrm{d}\vec{x}}} \tag{6}$$

where $f_1(\vec{x})$ and $f_2(\vec{x})$ are mixtures of multivariate Gaussians. Similar concept can be used to predict the proximity of future feature vectors to those describing normal behavior of the system by overlapping the predicted feature distributions with the feature distribution corresponding to the normal process behavior. Overlap between the predicted performance features and those describing normal process behavior will be referred to as the predicted performance Confidence Values (CVs), following [17].

3. Application example and results

The newly proposed method of predicting sequences of feature vectors representing signatures of equipment/process performance are experimentally tested on the in-process spindle load signals collected from a boring process in a major, domestic automotive manufacturing facility. Fig. 5 shows a typical load signal from one operation cycle.

The process signatures were extracted by first processing the spindle load signals into their time-frequency distributions [18], after which time-shift invariant moments of time-frequency distributions were extracted as suggested in refs. [19,20]. Dimensionality of the feature vectors was reduced using Principal Component Analysis (PCA) that resulted in principal components that described 99% of the variational energy in the time-frequency moments during nominal process behavior. For more details on the PCA, one may refer to [21] and references therein. In addition, the average signal energy is also added as a feature to depict the degradation process.

The spindle load sensor readings were collected over a 2 weeks period with 3000 boring process cycles recorded. During that period, the manufacturer changed the boring tool 25 times due to the tool wear-out, resulting in the total of 25 complete runs of spindle load features between consecutive tool changes. Fig. 6 shows the evolution of the total energy of the spindle load signals over several runs. One should note that sharp drops in the energy of the spindle load signals are due to tool changes when worn tools were replaced by new, sharp tools.

One run is selected to be the current run whose features need to be predicted and a group several past runs that were different

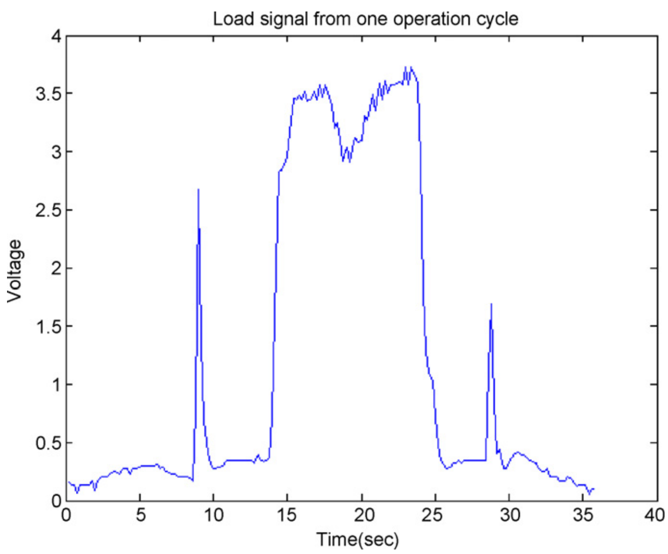


Fig. 5. Spindle load signal from one operation cycle.

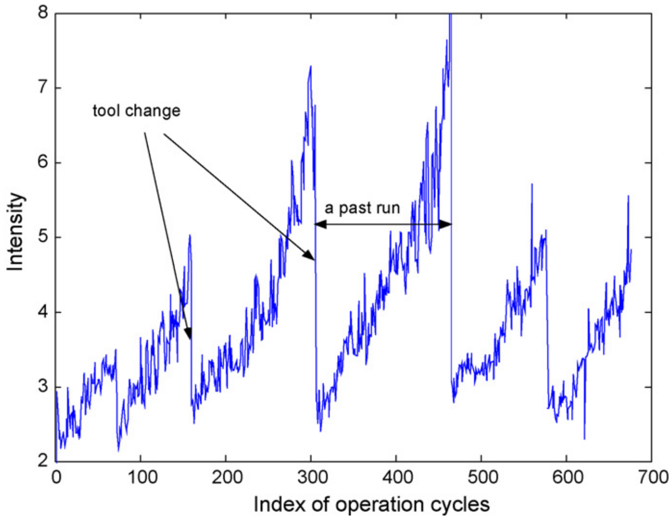


Fig. 6. Evolution of the average signal energy (intensity) over several runs, where sharp drops in total signal energy correspond to tool changes.

from the current run are identified for reference. A match matrix was formed between each of the past runs and the current run. A time series of mean indices of the best match was extracted from each match matrix. For each of the time series of mean best match indices, an adequate ARMA model was fitted and the mean best match indices were predicted up to 20 steps (working cycles) ahead. In addition, the distributions of the predicted features for each step ahead were evaluated using the past run features corresponding to the mean best match indices obtained by random sampling, as described in Section 2. The feature vectors obtained from random sampling in all match matrices were linearly combined, where the weights are calculated as the average of the match matrix similarities between the most recent five feature vectors in the current run and the corresponding best matches in the past runs. Fig. 7 shows the predicted sample feature vectors at 20th step ahead.

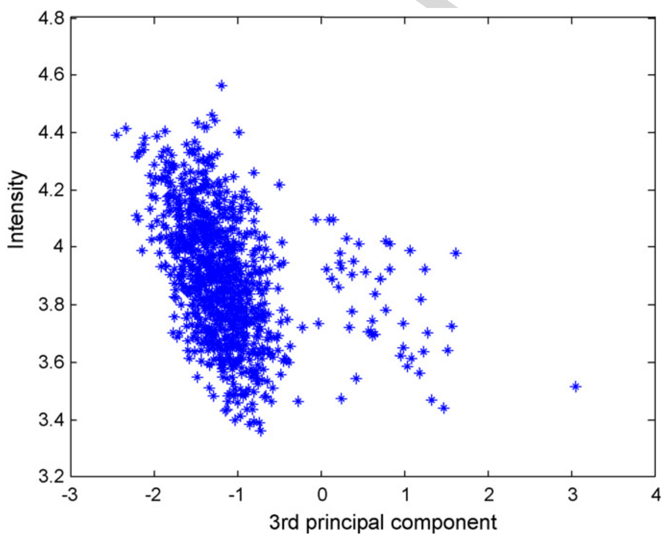


Fig. 7. Samples of 20th step ahead predicted features.

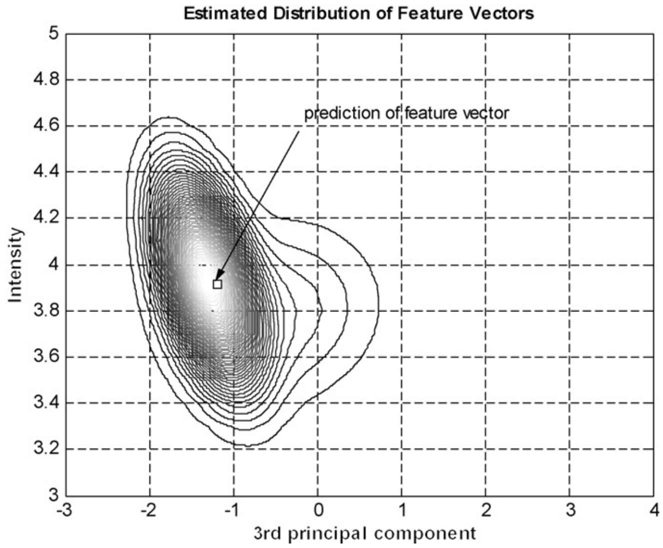


Fig. 8. Feature distributions estimated from the samples shown in Fig. 7 using a mixture of two multivariate Gaussians. The prediction of the feature vector at 20th ahead is also indicated.

Fig. 8 shows a 2-dimensional projection of the multivariate mixture-Gaussians fitted to the 20th step ahead predicted feature vectors shown in Fig. 7. The most-likely predicted feature vector can be identified from the predicted feature distribution as the feature corresponding to the maximum of the probability density function (pdf) of the predicted features, which is also indicated in Fig. 8.

The newly proposed match matrix based prediction method was evaluated in this data set through the comparison with ARMA based prediction and ERNN based prediction [22]. Recurrent neural networks have the ability to capture dynamics of non-stationary and non-Gaussian time series due to their ability to store the time dependencies of previous process realizations through the recurrent connections and have been employed for the prediction of such time series [14,23,24]. The structure of the ERNN, such as the number of hidden layers and the number of neurons, was determined following procedure reported in ref. [22], within which the whole training set (six past runs) was partitioned into a training and validation subset and the ERNN structure was chosen so that it minimizes the mean squared prediction error on the validation set. A sliding window ARMA model is used to perform multi-step prediction and the sliding window here is used to deal with the potential non-linearity of in the time series of interest. The order of the ARMA model is selected using the F-test criterion [1]. Fig. 9 shows the comparison of mean squared prediction errors obtained from ARMA, ERNN and match matrix based methods for predicting features up to 20 steps ahead. We can observe from Fig. 9 that for long-term prediction, match matrix achieves noticeably smaller prediction errors than ERNN and ARMA, while ARMA achieves slightly better short-term predictions than matrix and ERNN.

Another advantage of match matrix based prediction method is that it can explicitly provide the probability density distributions of the multiple step predictions. Fig. 10

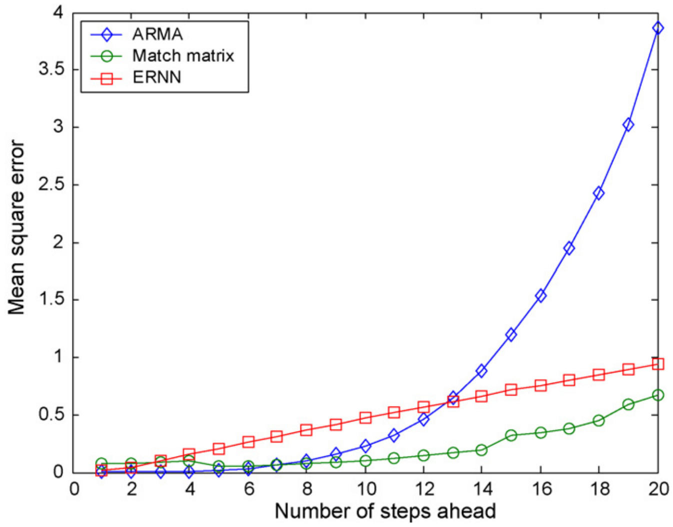


Fig. 9. Prediction errors for ARMA, ERNN and match matrix based prediction.

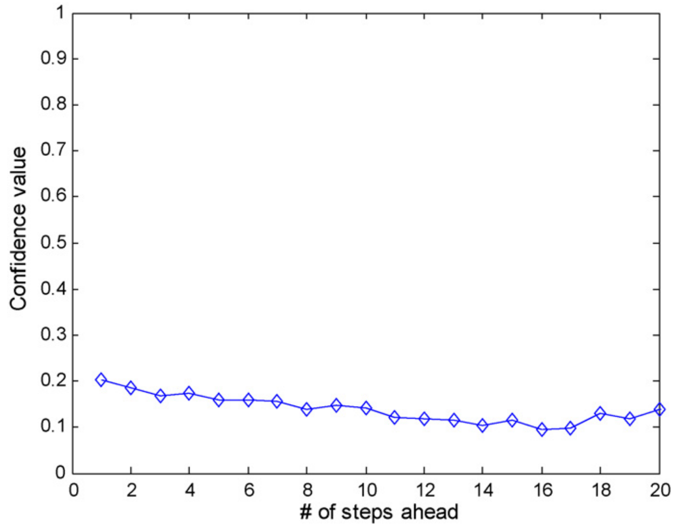


Fig. 11. Predicted performance confidence values.

schematically shows the overlap between the match-matrix predicted feature distribution and the distribution describing normal boring process behavior, while Fig. 11 shows the evolution of predictions of the performance confidence values for up to 20 steps ahead. The decreasing trend of the predicted performance CVs over time indicates an intuitively plausible result that the process is destined to drift away from the normal condition and the machine will degrade in time. However, even though these predicted performance CVs can be viewed as insightful process performance indices, performance CVs can only indicate that the process is deviating away from the normal condition, without providing any information about when unacceptable behavior will occur or what kind of failure/unacceptable behavior will occur in the future. To identify the risks of a failure or unacceptable behavior over time and what kind of failure/unacceptable behavior will occur in the future, the overlap between the predicted feature distributions with the feature distributions corresponding to various failure modes are needed.

The only mode of unacceptable behavior observed in the data set considered in this paper corresponded to the wearout of the boring cutting tool. Signatures collected during cutting with a worn cutting tool were used to describe the feature distribution corresponding to this mode of unacceptable behavior. Fig. 12 shows up to 20 steps ahead predictions of the probability of failure at that step, calculated as the overlap between the predicted feature distributions and the feature distribution corresponding to the worn boring cutting tool. Each point in this graph denotes the predicted probability that the cutting process will display unacceptable behavior at that particular step ahead. It should be noted that if there were signatures corresponding to several failure modes in this process, then inspecting the predicted probabilities of failure over time for different failure/unacceptable behavior modes can also determine what kind of failure is going to happen. Based on this predictive and diagnostic information, we can determine when and what kind of maintenance action we should take before an actual failure happens.

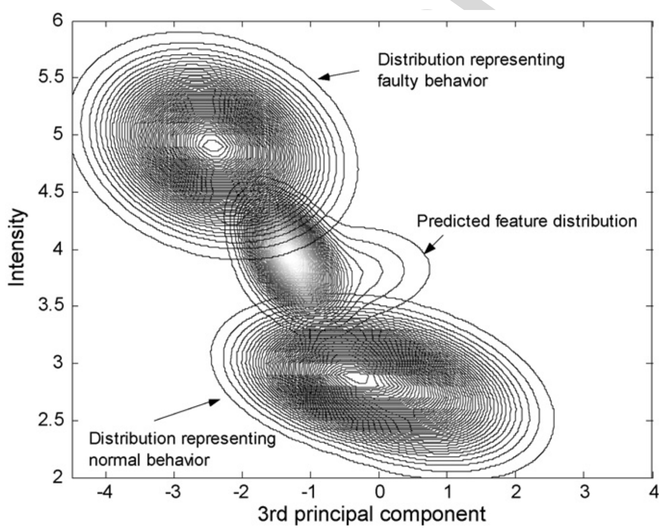


Fig. 10. Overlap between distributions.

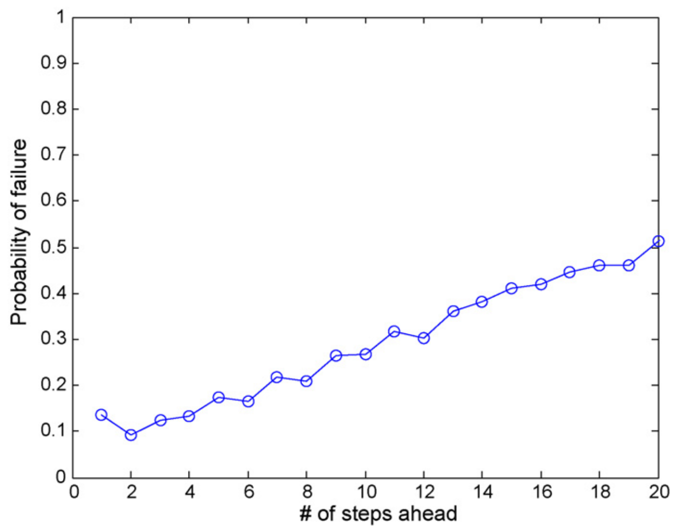


Fig. 12. Predicted probabilities of failure.

4. Conclusions and future work

A novel prediction algorithm capable of dealing with a long term prediction of non-stationary multivariate time series is presented in this paper. The method is based on the concept of match matrices for comparison of signatures describing the current degradation process with those observed in the past degradation processes of the same machine/process. It provides one with a natural way to derive the predicted feature distributions from the prediction uncertainties, which could be used to obtain information about predicted probabilities of failure or unacceptable behavior. The new method was tested in predicting signatures extracted from spindle load profiles of a boring process in an automotive manufacturing plant and results showed that the newly proposed prediction method based on match matrices yields noticeably smaller mean-squared errors of prediction, compared with ARMA based prediction and Recurrent Neural Network (RNN) based prediction.

In the case that a large number of past runs are present in historical records, comparing the current run with all the past runs and conducting random sampling within each match matrix may be computationally demanding and inefficient. Significant reduction in the computational load can be achieved if one could observe groups of similar degradation dynamics expressed through similar performance features evolution, and evaluate match matrices only within the group that is the most similar to the degradation process in the currently observed run. A Hidden Markov Model (HMM) based technique as reported in ref. [25] can be possibly used to select the most similar runs for creating match matrices with the current run in order to accomplish the prediction of feature evolution in the current run. Nevertheless, this problem is outside the scope of this paper.

Acknowledgments

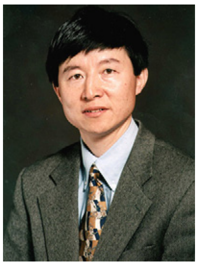
This research was supported in part by the NSF sponsored Industry/University Cooperative Research Center for Intelligent Maintenance Systems (IMS) at the University of Cincinnati and the University of Michigan.

References

[1] S.M. Pandit, S.-M. Wu, Time Series and System Analysis, with Applications, Wiley, New York, 1983.
[2] G.E.P. Box, G.M. Jenkins, Time Series Analysis: Forecasting and Control, revised ed., Holden-Day, San Francisco, 1976.
[3] R.E. Kalman, A new approach to linear filtering and prediction problems, Transactions of the ASME, Journal of Basic Engineering 82 (1960) 34–45.
[4] A.B. Geva, Non-stationary time-series prediction using fuzzy clustering, in: Presented at Proceedings of NAFIPS-99: 18th International Conference of the North American Fuzzy Information Processing Society, New York, NY, USA, June 10–12, 1999.
[5] C.-H.L. Lee, A. Liu, W.-S. Chen, Pattern discovery of fuzzy time series for financial prediction, IEEE Transactions on Knowledge and Data Engineering 18 (2006) 613–625.
[6] K. Papagiannaki, N. Taft, Z.-L. Zhang, C. Diot, Long-term forecasting of Internet backbone traffic, IEEE Transactions on Neural Networks 16 (2005) 1110–1124.
[7] O. Renaud, J.-L. Starck, F. Murtagh, Wavelet-based combined signal filtering and prediction, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 35 (2005) 1241–1251.
[8] H.L. Wei, S.A. Billings, Long term prediction of non-linear time series using multiresolution wavelet models, International Journal of Control 79 (2006) 569–580.
[9] J.-L. Chen, S. Islam, P. Biswas, Nonlinear dynamics of hourly ozone concentrations: nonparametric short term prediction, Atmospheric Environment 32 (1998) 1839–1848.
[10] K.B. Cho, B.H. Wang, Radial basis function based adaptive fuzzy systems and their applications to system identification and prediction, Fuzzy Sets and Systems 83 (1996) 325–339.
[11] S.S. Haykin, Neural Networks: A Comprehensive Foundation, second ed., Prentice Hall, Upper Saddle River, NJ, 1999.
[12] M. Han, J. Xi, S. Xu, F.-L. Yin, Prediction of chaotic time series based on the recurrent predictor neural network, IEEE Transactions on Signal Processing 52 (2004) 3409–3416.
[13] H. Leung, T. Lo, S. Wang, Prediction of noisy chaotic time series using an optimal radial basis function neural network, IEEE Transactions on Neural Networks 12 (2001) 1163–1172.
[14] J.T. Connor, R.D. Martin, L.E. Atlas, Recurrent neural networks and robust time series prediction, IEEE Transactions on Neural Networks 5 (1994) 240–253.
[15] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, Wiley, New York, 2001.
[16] G.J. McLachlan, Discriminant analysis and statistical pattern recognition, Wiley, New York, 1992.
[17] J. Lee, Measurement of machine performance degradation using a neural network model, Computers in Industry 30 (1996) 193–209.
[18] L. Cohen, Time-Frequency Analysis, Prentice Hall, Englewood Cliffs, NJ, 1995.
[19] D. Djurdjanovic, S.E. Widmalm, W.J. Williams, C.K.H. Koh, K.P. Yang, Computerized classification of temporomandibular joint sounds, IEEE Transactions on Biomedical Engineering 47 (2000) 977–984.
[20] B. Tacer, P. Loughlin, Time–frequency based classification, SPIE Proceedings 2846 (1996) 186–192.
[21] J.E. Jackson, A User's Guide to Principal Components, Wiley, New York, 1991.
[22] G. Yu, H. Qiu, D. Djurdjanovic, J. Lee, Feature signature prediction of a boring process using neural network modeling with confidence bounds, International Journal of Advanced Manufacturing Technology 30 (2006) 614–621.
[23] T. Senjyu, H. Takara, K. Uezato, T. Funabashi, One-hour-ahead load forecasting using neural network, IEEE Transactions on Power Systems 17 (2002) 113–118.
[24] J.-N. Hwang, E. Little, Real time recurrent neural networks for time series prediction and confidence estimation, in: Proceedings of the 1996 IEEE International Conference on Neural Networks, ICNN (Part 4 of 4), Washington, DC, USA, June 3–6, 1996.
[25] P. Smyth, Clustering sequences with hidden markov models, Advances in Neural Information Processing Systems 9 (1997) 648–654.

Jianbo Liu was born 1977 in PR China. He received the master's and PhD degree in mechanical engineering from the University of Michigan at Ann Arbor in 2003 and 2007, respectively. Having particular interest in identification, modeling and control, he has devoted a major portion of his PhD research to behavior modeling of dynamic systems using the available sensor readings for purpose of anomaly detection and fault diagnosis.

Dragan Djurdjanovic obtained his BS in mechanical engineering and in applied mathematics in 1997 from the University of Nis, Serbia, his MEng in mechanical engineering from the Nanyang Technological University, Singapore in 1999, and his MS in electrical engineering (systems) and PhD in mechanical engineering in 2002 from the University of Michigan, Ann Arbor. His research interests include quality control, intelligent proactive maintenance techniques and applications of advanced signal processing in biomedical engineering. He co-authored 30 journal and 15 conference publications, and is the recipient of several prizes and awards, including the 2006 Outstanding Young Manufacturing Engineer Award from the Society of Manufacturing Engineers (SME), Nomination for the Distinguished PhD Thesis from the Department of Mechanical Engineering, University of Michigan in 2002, and The Outstanding Paper Award at 2001 SME North American Manufacturing Research Conference.



Jun Ni is currently a professor in the department of mechanical engineering at the University of Michigan, USA. He is the director of the S.M. Wu Manufacturing Research Center. He also serves as the deputy director of a National Science Foundation sponsored Engineering Research Center for Reconfigurable Machining Systems, and the co-director of a National Science Foundation sponsored Industry/University Cooperative Research Center for Intelligent Maintenance Systems. Professor Ni's research and teaching interests are in the area of manufacturing science and engineering, with special focuses on precision machining, manufacturing process modeling and control, statistical quality design and improvement, micro/meso systems and manufacturing processes, and intelligent monitoring, maintenance and service systems.

Dr. Jay Lee is Ohio Eminent Scholar and L.W. Scott Alter Chair professor in advanced manufacturing at the University of Cincinnati and is founding director of National Science Foundation (NSF) Industry/University Cooperative Research Center (I/UCRC) on Intelligent Maintenance Systems (IMS www.imscenter.net) which is a multi campus NSF Center of Excellence between the University of Cincinnati (lead institution), the University of Michigan, and the University of Missouri-Rolla in partnerships with over 40 global companies including Rockwell Automation, Boeing, Caterpillar, DaimlerChrysler, ETAS, Festo, GE Aviation, Harley-Davidson, Honeywell, ITRI (Taiwan), Komatsu, Omron, P&G, Samsung, Toshiba, Toyota, United Technologies, etc.