Int J Adv Manuf Technol (2015) 76:513–521
DOI 10.1007/s00170-014-6299-0

ORIGINAL ARTICLE

Prognosis of the probability of failure in tool condition monitoring application-a time series based approach

Yue Wu · G. S. Hong · W. S. Wong

Received: 7 May 2013 / Accepted: 19 August 2014 / Published online: 5 September 2014

Abstract Tool condition monitoring (TCM) system serves as the link between the cutting tool condition and the maintenance decision. The recent prognostic system employs highly complex models, which might need long calculation time. This long calculation time is acceptable for machine health prognostics, as machines' maintenance interval is in the unit of month. However, a cutting tool's life varies between minutes to hours. The calculation time might be critical to achieve a valid prognosis. In this paper, a novel prognostic system is proposed for TCM prognostics. This system consists of two parts: (1) online cutting force prediction part and (2) tool wear estimation part. The first part predicts the future cutting force segmentation by projecting the embedded historical cutting force with function approximation methods. Three function approximation methods are compared in the aspect of prediction error and calculation time. It is found that the Saucer's local linear model could achieve the lowest prediction error (4.71 %) and calculation time (2.717 s) compared with global linear model and nonlinear model. The second part estimates the tool wear by inputting the predicted cutting force to a Bayesian-multilayer perceptron. It is found that this system can trace the progress of tool wear accurately (95 % successful rate has been achieved). Moreover, good generalization for different cutting conditions is also achieved.

Keywords Tool condition monitoring · Prognostics · Delay coordinate embedding · Bayesian multilayer perceptron

Y. Wu (✉) · G. S. Hong · W. S. Wong
Mechanical Engineering, National University of Singapore, Singapore, Singapore
e-mail: wuyuemailbox@gmail.com

G. S. Hong
e-mail: mpehgs@nus.edu.sg

W. S. Wong
e-mail: mpewys@nus.edu.sg

1 Introduction

In a manufacturing system, the failure of cutting tool will cause catastrophic destruction of the workpiece as well as a large machine downtime. Therefore, the maintenance process or the replacement decision, whose object is to make use of the knowledge of failures and accidents to achieve the possible safety of the cutting tool with lowest possible cost, is of significance to the manufacturing system.

Tool condition monitoring (TCM) system monitors the cutting tool condition in the manufacturing system. This system has been largely viewed as a link between tool condition and maintenance decision. However, in recent studies on TCM, little attention has been paid to contribute to the maintenance decision. The evidence is the vast number of diagnosis research that has been conducted to investigate the detailed faults of cutting tool. However, in order to make an optimal decision, the residual life of cutting tool and the probability of failure are more useful than diagnosing the tool wear value. Moreover, the diagnosis of the cutting tool failure is far too late for proper maintenance decision. Therefore, prognosis system, which could predict the future condition of cutting tool, is more important in TCM application.

Several researchers have prognosticated machinery health [1–3]. Lin and Makis [4] predict the probability of failure by using recursive filters. However, the calculation time is long and might not be applicable for TCM applications. Wang et al. [5] and Yam et al. [6] developed a system, which predicts the future condition of the engineering system. However, no confidence interval is studied in this study. Therefore, the result might not help to make decisions for maintenance.

In TCM prognostics, the literature is very few. Baruah and Chinnam [7] is the only group to study the prognostic problem of drilling process until 2011. Hidden Markov model (HMM) was applied in their study to build the prognostic system.

However, this model is difficult to generalize cutting conditions, which is not present in the training set [8].

Therefore, there are two challenges of prognostic probability of failure for TCM applications. The first one is the calculation time. Comparing machinery health monitoring and TCM, the machine life is in the unit of month or year, whereas the unit for a cutting tool is minutes or hours. Therefore, the calculation time for the TCM application is critical. The second is how to generalize the prediction results into new cutting condition.

In this study, a prognostic system is proposed and investigated. The system is constructed by adding an online force prediction part to a traditional TCM diagnostic system. The TCM diagnostic system is usually made up of signal processing, feature extraction, feature selection, and tool wear estimation. This process is illustrated in Fig. 1.

The prognosis purpose is achieved by online predicting the force signal and passes this predicted signal through the rest of the diagnostic system. This process is illustrated in Fig. 2.

In this paper, the cutting force is predicted instead of the tool wear. This is because the calculation process of tool wear might bring noise or error. The calculation process includes feature extraction; feature selection and multilayer perceptron (MLP) (see Fig. 2). These noise or error might increase the chance of mis-prediction. There are two sub-parts in the force prediction process: (a) reconstruct the time series by delay coordinates embedding and (b) prediction by function approximation methods.

In the first sub-part, delay coordinates embedding reconstructs the time series (cutting force signal) into a high-dimensional state space. The relationship between the delay coordinates of a point and the points that appear some time later in the state space can be used to infer the future value of the time series. This idea was first studied by Packard et al. [9]. It is found that delay coordinate embedding technique is more advanced than the conventional time series prediction technique in predicting the nonlinear time series. Therefore, it was wildly sued in time series prediction [10].

The second sub-part aims at studying the relationship between the delay coordinates. In this study, three function approximation approaches are investigated: (1) Sauer's [11] local linear regression approach, (2) global linear regression approach, and (3) nonlinear approaches. The Sauer's local linear approach is suitable for predicting the low-dimensional chaotic system. However, if the system has stochastic property, local linear regression might lead to under fitting. In this situation, the global linear regression might be the best approach. However, if the relationship between the coordinates in the state space is nonlinear, the nonlinear approach might perform better than those two approaches. The nonlinear approaches usually need a longer calculation time than the linear regression approaches, because the former usually have more unknown parameters. These three approaches have their advantages and disadvantages. Therefore, in this study, they are compared in the aspect of calculation time and average scale independent error (ASIE).

The tool wear estimation is achieved by Bayesian-multilayer perceptron (B-MLP).

Therefore, the objectives of this study are the following:

- Reconstruct the cutting force signal into a higher dimensional state space.
- Investigate the proper embedding dimension.
- Investigate the three function approximation approaches in the aspect of ASIE and calculation time.
- Establish the tool wear estimation part by B-MLP to get the prognosticated tool wear and its confidence interval.
- Compare and analyze the tool wear estimated by true cutting force and predicted cutting force.

It is found that the delay coordinate embedding with Sauer's local linear regression approach generate the smallest calculation time (2.7 s), and a reasonable predicting error (4.71 %). This indicates that delay coordinate embedding with Sauer's local linear model is promising in cutting force prediction. It is also found that the tool wear estimated by cutting force predicted by this approach well matched the tool wear estimated by true cutting force.

The paper is organized as follow: Section 2 introduces the theory background of our approach to prognosis, Section 3 describes the experiment verification process, Section 4 discusses the results, and Section 5 is the conclusion.

2 Theory background

2.1 Delay coordinate embedding

The technique of delay coordinate embedding is to reproduce the set of dynamical states of a system using vectors derived from a time series measured from the

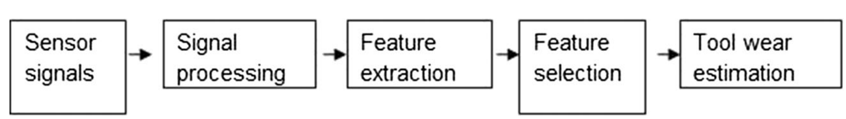

Fig. 1 TCM diagnostic system architecture

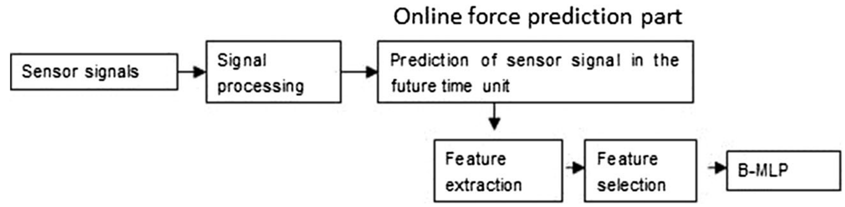

Fig. 2 The proposed prognostic process is achieved by adding online force prediction part to the diagnostic system

system. Consider a time series $\{x(1),x(2),x(3),\ldots,x(N)\}$. A delay coordinate vector is

$$b(t) = [x(t), x(t-\tau), \ldots, x(t-(m-1)\tau)] \quad (1)$$

where $t\in[1+(m-1)\tau,N]$, m is the embedding dimension, and τ is the time delay. It is called delay coordinate vector because its components consist of time-delayed versions of the observable system. $b(t)$ is a vector in a D-dimensional state space R^D. A trajectory B in R^D is defined as follows:

$$B = \left[b^T(t)\ b^T(t+1)\ldots b^T(t+m)\right. \quad (2)$$

In order to extract the behaviors of the time series in an efficient way, optimal values of m and τ have to be determined. In this study, m and τ are determined by false nearest neighbors [12].

2.2 Prediction

The known time series is referred as *training set*. For each reconstructed delay coordinate vector b in the training set, the value of the series t time units later is defined as its observation X. Our aim is to find the reconstructed vector $b*$ corresponding to the end of the series and use the knowledge of the training set to estimate the unknown observation $X* = P_t(b*)$.

2.2.1 Local linear approach

Local linear approach is achieved by choosing a small size of neighborhoods in the state space and using linear regression to approximate the function between the state space data and the corresponding observations. This approach is first studied by Casdagli [13]. In his study, it is found that local linear approach can give an accurate short-term prediction for low-dimensional chaotic system. This local linear approach used in this study is based on Sauer's idea [11]. In his study, the state space is constructed by passing original delay coordinate vector into a low-pass filter. The state space is defined as follows:

$$b = M[x(t),\ x(t-\tau), \ldots, x(t-(w-1)\tau)]^T \quad (3)$$

Where M is an $m\times w$ matrix of rank m, and $M = M_3M_2M_1$. The three compositions are the following linear operations:

- M_1=FFT of order w;

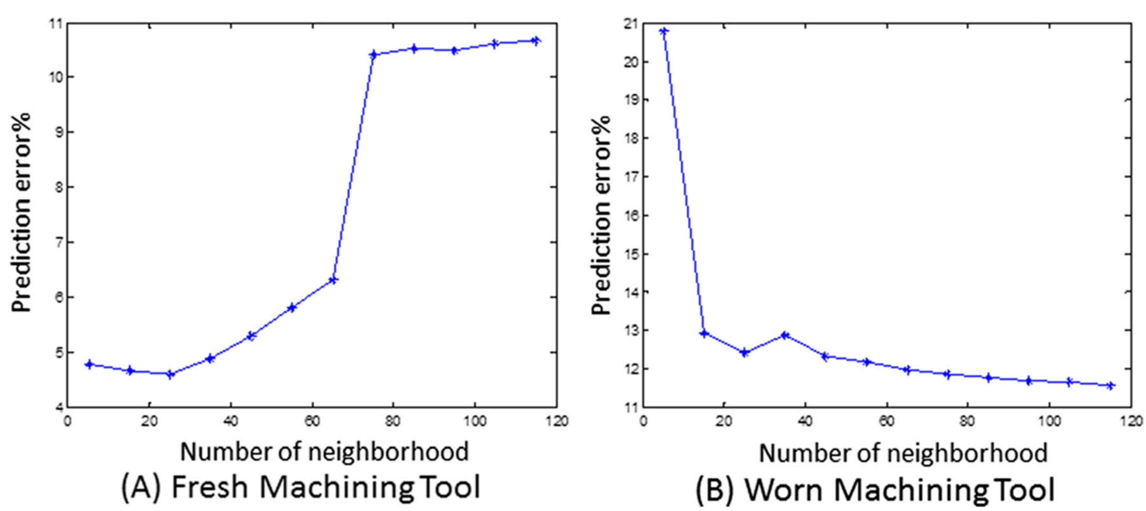

Fig. 3 The prediction error for **a** fresh machining tool and **b** worn machining tool, when the neighborhood increases from 5 to 120

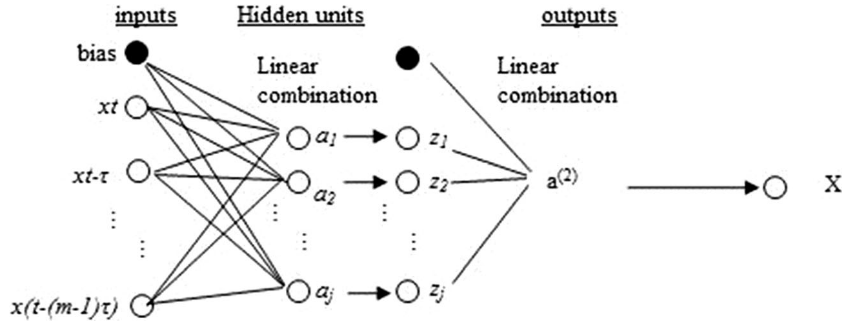

Fig. 4 The architecture of the multilayer perceptron. The delay coordinate vector is the input to this network. Their linear combination builds the hidden units. The observation X in t time unit later is the output

- M_2 sets to zero all but the lowest $m/2$ frequency contributions; and
- M_3=inverse FFT of order m, using the remaining $m/2$ frequencies.

After the low-pass embedding, the nearest neighborhoods are chosen. The distance between each delay coordinate vector is defined as the Euclid distance. Then, singular vector decomposition (SVD) is used to concentrate the state space information. Finally, weighted linear regression is applied to approximate the function between state space and the observation. However, if the underlying dynamics is indeed stochastic or chaotic of high dimension, then local models will give less accurate forecasts since the small neighborhoods give them the flexibility to fit the noise in addition to the signal. Figure 3 shows how the number of nearest neighborhood affects the prediction error for both fresh and worn cutting tool.

As shown in Fig. 3a, when the nearest neighborhood increases from 5 to 120, the prediction error increases from 5 to 10 % for fresh machining tool. However, as shown in Fig. 3b, the prediction error is higher (20.9 %) at small nearest neighborhood and lower (10.9 %) at larger nearest neighborhood. This indicates that the underline dynamic of the cutting force signal for worn cutting tool is chaotic of higher dimension than the fresh cutting tool. Therefore, global linear approach is proposed to compare with local linear approach.

Table 1 Selected features

Number	Feature	Notation
1	Average force	Fa
2	Maximum force level	fm
3	Total harmonic power	thp
4	Total amplitude of cutting force	fa
5	Amplitude ratio	ra
6	Standard deviation	std
7	Standard deviation of the force components in tool breakage zone	fstd

2.2.2 Global linear approach

Global linear approach is achieved also by using linear regression to approximate the function between state space and the corresponding observations. However, different from local linear approach, which uses small size of neighborhoods, global linear approach uses a large size of neighborhoods. This approach is suitable for modeling the linear and stochastic system [13] but will cause over fitting for the low-dimensional chaotic system. In this study, the global linear approach is also based on Sauer's idea, but the whole history state space is used for approximating the function between the state space and observations, instead of small size of neighborhoods.

2.2.3 Nonlinear approaches

Nonlinear approaches are also known as artificial neural network modeling. These approaches are wildly used in prediction of machine health monitoring index [5, 14–16]. However, the structure of artificial neural network is much more complex than the linear approaches (local linear approach and global linear approach). This indicates a longer calculation time, which is not expected in TCM applications. In this study, a multilayer perceptron (MLP), which is one of the artificial neural networks, is used to approximate the function between state space and observations. The network architecture is illustrated in Fig. 4.

As shown in Fig. 4, the input to this neural network is the delay coordinate vector. The output is the observation X in t time unit later. The network parameters were estimated by known delay coordinate vector and its corresponding observation X.

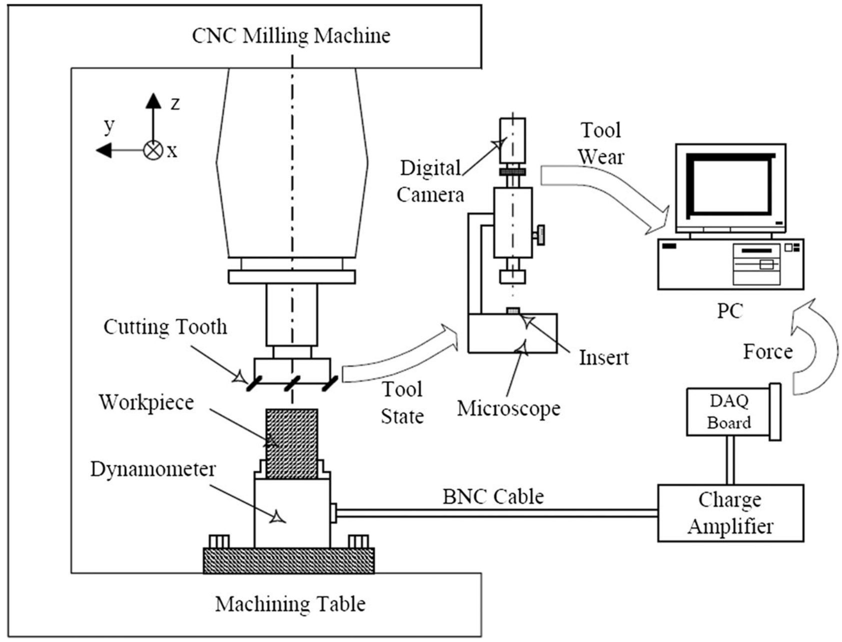

Fig. 5 Experimental setup

2.3 Bayesian-MLP for regression

The probability of failure could be extracted from the predicted cutting tool value and its uncertainty. Therefore, a diagnostic system, which estimates not only the tool wear value but also its uncertainty, is of great importance.

Bayesian-MLP (B-MLP) offers a reasonable solution for this diagnostic system. This model starts with a prior probability distribution before the data is observed. Once the data is observed, Bayes' theorem can be used to obtain the posterior probability and update the beliefs on the network parameters. In this process, Bayesian-MLP handles uncertainty in a natural manner.

The training of the neural network for regression can be conducted as follow:

1. Create and initialize the network by assuming the hyperparameters controlling the distribution of the network parameters, including the weights and biases.
2. Optimize the weights and biases through an optimization algorithm.
3. Update the hyperparameters by presenting the observation data.

Table 2 Experimental components

Components
Makino CNC milling machine with Funuc controller
EGD 4450R cutter with AC325 inserts
ASSAB718HH workpiece (206 mm×43 mm×106 mm)
Kistler 9265B Quartz 3-Component Dynamometer
Kistler 5019A Multi-channel Charge Amplifier
NI-DAQ PCI 1200 Board
Olympus microscope and Panasonic digital camera
Computer with Pentium |||600 MHz and 128 M SDRAM

Seven features are selected, which is listed in Table 1, as the input to the Bayesian-MLP [15]. The training set includes T3, T4, T5, T8, T10, and T12, and the rest of the tests are used as testing sets.

3 Experiment setup

Face milling tests are conducted as a case study. The schematic diagram of the experimental setup is illustrated in Fig. 5, and its components are listed in Table 2. The cutting force along the y direction of the machine (transverse force) was captured by the Kistler dynamometer in the form of charges and converted to voltages by the Kistler charge amplifier. The voltage signal was sampled by the PCI 1200 board at 2,000 Hz and directly streamed to the hard disk of the computer. Then, the signal went through a low-pass filter. The remaining signal is 500 Hz. The flank wear of each individual tooth was measured at an interval of five tool passes by the Olympus microscope, and at each time, an average was taken from all the teeth mounted on the cutter. The tool state was observed by a digital camera. Twelve experiments were conducted on the CNC milling machine, with the cutting conditions listed in Table 3.

Table 3 Cutting experiments

Test no.	Spindle speed (rpm)	Feed rate (mm/min)	Depth of cut (mm)	Insert number
T1	800	150	1	4
T2	1,000	100	1	2
T3	1,000	100	1	4
T4	1,000	200	1	2
T5	1,000	300	1	4
T6	1,200	150	1	2
T7	1,200	200	1	2
T8	1,200	300	1	4
T9	600	100	2	4
T10	600	200	2	4
T11	800	100	2	2
T12	1,000	100	1	4

4 Results and discussion

4.1 Force prediction

4.1.1 Embedding dimension and delay time

False nearest neighbors [12] is applied to select an appropriate embedding dimension. If the embedding dimension is too small, some points might be considered as the nearest neighbors. However, as the dimension increases to a certain number, the false nearest neighbors drop to zero. In that case, we have unfolded or embedded the time series in a high-dimensional Euclidian space (see Fig. 6).

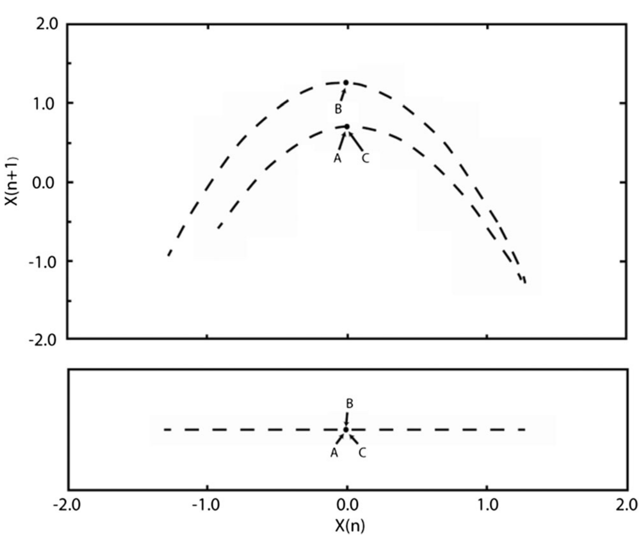

Fig. 6 The R1 and R2 embeddings of the *x* coordinate of the plane. The *points A* and *B* are false neighbors while the *points A* and *C* are true neighbors. Therefore, the false NN percentage for these three points in one dimension is 33.3 %

The percentage of false nearest neighbors (NN) for 269,995 data points from the entire tests is shown in Fig. 7. The smallest embedding dimension defined as the false NN drops until 0.01 and 0.001 are summarized in Table 4.

As shown in Table 4, the average embedding dimension for false NN percentage ≤0.1 is 17 and this value increase to 27 if the false NN percentage <0.001. It is suggested that projecting the cutting force time series into the higher dimension will have a positive effect in unfolding. Therefore, 27 is used as the embedding dimension in the following study.

4.1.2 Compare the three function approximation approaches

Three approaches, which are local linear approach, global linear approach, and nonlinear approach, are compared in the aspect of calculation time, which is the CPU time needed for predicting one revolution cutting force signal, and prediction results' ASIE, which is defined as average scale independent error:

$$\mathrm{ASIE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{D_0}\right|$$

where N is the prediction length, y_i is the true cutting force, and $\hat{y}_i$ is the predicted cutting force. D_0 is $\max(y_1, y_2 \ldots, y_N)$. The prediction horizon is 1 min in this comparison.

The ASIE and calculation time for both fresh tool and worn tool of Sauer's local linear approach, global linear approach, and nonlinear (MLP) approach are compared in Table 5. For fresh tool, as can be seen in

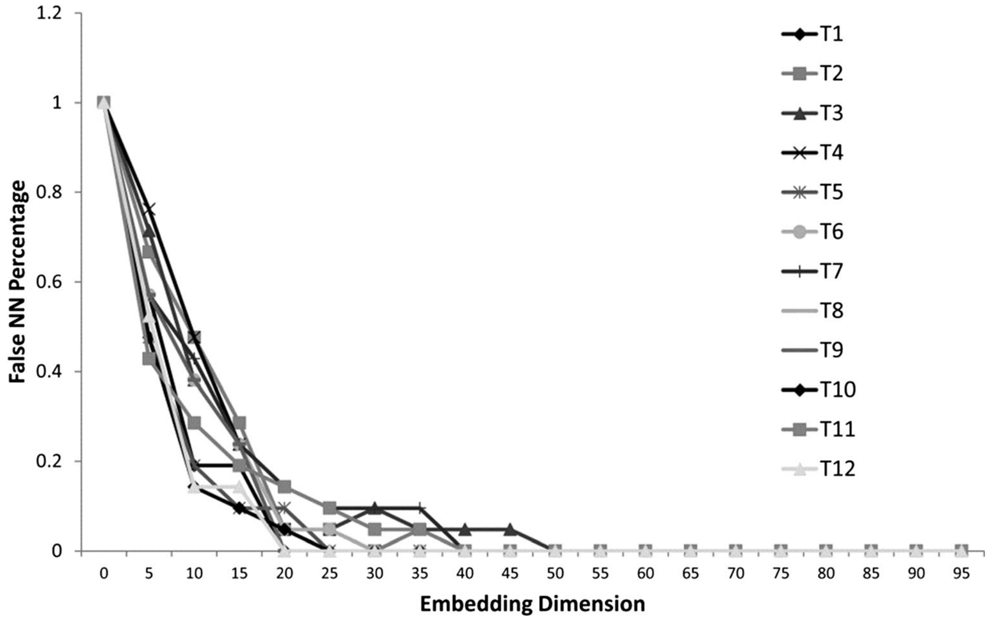

Fig. 7 The percentage of false nearest neighbors for the force signal of the entire tests

Table 5, the Sauer's local linear approach generates the smallest ASIE and the least calculation time. This indicates that the Sauer's local linear model is the most suitable approach for modeling the fresh cutting force signal. It is found that the MLP needs the longest calculation time (91.5 min). This indicates that this approach cannot achieve the prognostic purpose, as the prediction horizon is only 1 min. It is also found that global linear approach generates the highest ASIE (11.8 %). This ASIE is much higher than Sauer's local linear approach (4.71 %). This indicates that the cutting force for fresh tool signal follow the low-dimension and chaotic property, instead of stochastic property.

For the worn tool, it is found in Table 5 that for all three approaches, the ASIE are higher than the fresh tool. However, MLP generate the smallest ASIE. This indicates that the worn tool cutting force signal is nonlinear stochastic. Although the Sauer's local linear model generate a little bit higher ASIE for the worn cutting tool signal, this difference between three approaches is not significant. Therefore, for the tool wear estimation part, delay coordinate embedding with Sauer's local linear model is used for cutting force prediction. This predicted cutting force signal will undergo the diagnostic system to achieve the prognostic purpose.

Table 4 The smallest embedding dimension for the entire test and their average

	False NN percentage ≤0.1	False NN percentage <0.001
T1	17	20
T2	16	25
T3	16	50
T4	17	25
T5	13	25
T6	17	30
T7	24	40
T8	16	30
T9	17	20
T10	14	20
T11	24	20
T12	9	20
Average embedding dimension	17	27

Table 5 Compare the ASIE and calculation time for Sauer's local linear approach, global linear approach, and nonlinear (MLP) approach

		ASIE (%)	Time/s
Fresh tool (0–0.3 mm)	Sauer's local linear	4.71	2.717
	Global linear	11.80	3.2892
	MLP	6.86	91.4945
Worn tool (0.3–0.5 mm)	Sauer's local linear	13.64	2.7716
	Global linear	11	3.3028
	MLP	10.34	88.1626

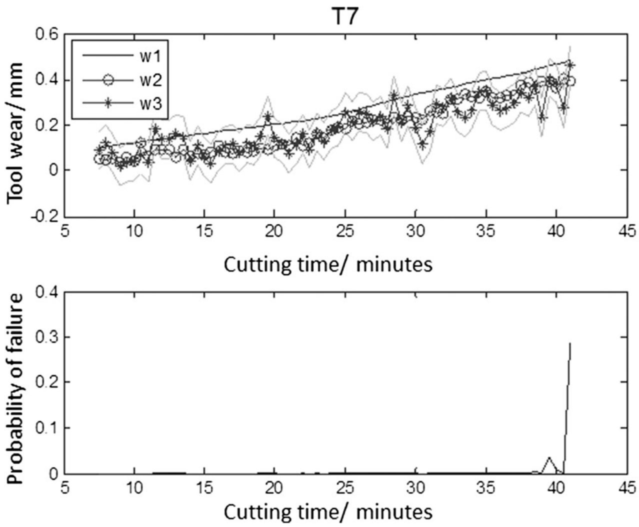

Fig. 8 Comparisons of true tool wear, predicted tool wear, and their confidence interval for T7. *w1* true tool wear, *w2* tool wear estimated by true cutting force, *w3* tool wear predicted by delay coordinate embedding

4.2 Tool wear prognostics

The tool wear for T7 estimated by the predicted cutting force and its 95 % confidence interval (w3), estimated by true cutting force (w2), are compared with true tool wear value in Fig. 8 (first subplot). The prediction horizon is 60 s. The prognosis results catch the trend of true tool wear well. The predicted probability of failure is shown in Fig. 8 (second subplot). The probability of failure increased rapidly at the 39th minute, which indicates high risk of tool failure. This result could help in determining the optimal tool replacement time to achieve highest profit rate.

If true tool wear falls between the 95 % confidence interval of the predicted tool wear, the prediction is considered as a success. The successful rates of all testing sets are listed in Table 6. As shown in Table 6, the smallest successful rate is 85 % (T7) and the highest is 95 % (T6). This result suggests that the delay coordinate embedding combine with B-MLP is capable of generalizing the prognostic results into different cutting conditions.

Table 6 Successful rate for testing sets

	Succeed percentage
T1	94
T2	86
T6	95
T7	85
T9	95
T11	86

5 Conclusions

TCM serves as a link between the tool condition and the replacement decision module. However, recent studies on TCM system mainly focuses on diagnostic of the different tool wear states and tool wear phenomenons, which are not helpful in making an optimal decision. However, the prognostic system in the literature tends to employ highly complex structures, which might need huge calculation time. Moreover, the existing TCM prognostic system is not able to be generalized into different cutting conditions. Therefore, a novel prognostic system has been proposed in this paper to solve these problems.

The proposed prognostic system consists of two major parts: (1) cutting force prediction and (2) tool wear estimation. In the first part, there are two sub-parts: (a) reconstruct the time series by delay coordinates embedding and (b) prediction by function approximation. In part (b), three function approximation approaches were compared: Sauer's local linear approach, global linear approach, and nonlinear approach. It is found that the cutting force predicted by Sauer's local linear approach generate the smallest ASIE (4.71 %) and calculation time (2.7 s). This result shows that the cutting force can be considered generated by a low-dimensional chaotic system. In the second part (tool wear estimation), it is found that the tool wear estimated by predicted cutting force (delay coordinate embedding with Sauer's model) well matched the tool wear estimated by true cutting force (more than 85 % successful rate). In conclusion, this study shows a promising prognostic system in TCM applications.

References

1. Qu J, Zuo MJ (2012) An LSSVR-based algorithm for online system condition prognostics. Expert Syst Appl 39(5):6089–6102. doi:10.1016/j.eswa.2011.12.002
2. Caesarendra W, Widodo A, Yang B-S (2011) Combination of probability approach and support vector machine towards machine health prognostics. Probabilistic Eng Mech 26(2):165–173. doi:10.1016/j.probengmech.2010.09.008
3. Liao W, Wang Y, Pan E (2012) Single-machine-based predictive maintenance model considering intelligent machinery prognostics. Int J Adv Manuf Technol 63(1–4):51–63. doi:10.1007/s00170-011-3884-3
4. Lin D, Makis V (2003) Recursive filters for a partially observable system subject to random failure. Adv Appl Probab 35:207–227, Copyright 2003, IEE

5. Wang WQ, Golnaraghi MF, Ismail F (2004) Prognosis of machine health condition using neuro-fuzzy systems. Mech Syst Signal Process 18:813–831, Compendex
6. Yam RCM, Tse PW, Li L, Tu P (2001) Intelligent predictive decision support system for condition-based maintenance. Int J Adv Manuf Technol 17:383–391, Copyright 2001, IEE
7. Baruah P, Chinnam RB (2005) HMMs for diagnostics and prognostics in machining processes. Int J Prod Res 43(6):1275–1293
8. Kunpeng Z (2007) Multi-categories tool wear classification in micro-milling. NUS, Singapore
9. Packard NH, Crutchfield JP, Farmer JD, Shaw RS (1980) Geometry from a time series. Phys Rev Lett 45:712–716, Copyright 1980, IEE
10. Zhang J, Shu-Hung Chung H, Wai-Lun L (2008) Chaotic time series prediction using a neuro-fuzzy system with time-delay coordinates. IEEE Trans Knowl Data Eng 20:956–964, Copyright 2008, The Institution of Engineering and Technology
11. Sauer T (1994) Time series prediction by using delay coordinate embedding (data set A). (Compendex):175-175
12. Kennel MB, Brown R, Abarbanel HDI (1992) Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys Rev A Stat Phys Plasmas Fluids Relat Interdiscip Top 45:3403–3411, Copyright 1992, IEE
13. Casdagli M, Weigend AS (1994) Exploring the continuum between deterministic and stochastic modeling (data sets A,B,D). (Compendex):347-347
14. Shao Y, Nezu K (2000) Prognosis of remaining bearing life using neural networks. Proc Inst Mech Eng I-J Syst 214(13):217–230
15. Wang P, Vachtsevanos G (2001) Fault prognostics using dynamic wavelet neural networks. AI EDAM Artif Intell Eng Des Anal Manuf 15:383–391
16. Chinnam RB, Baruah P (2004) A neuro-fuzzy approach for estimating mean residual life in condition-based maintenance systems. Int J Mater Prod Technol 20:166–179