




# A Reinforced $k$ -Nearest Neighbors Method With Application to Chatter Identification in High-Speed Milling

Fei Shi , Hongrui Cao , Xingwu Zhang, and Xuefeng Chen , *Senior Member, IEEE*

**Abstract**—Chatter is a kind of self-excited vibration which will destroy the manufacturing process badly. The detection or identification of chatter is attracting considerable interest for several years. In this article, a chatter identification method called reinforced  $k$ -nearest neighbors is proposed to realize both chatter identification and model self-learning. We conducted large amounts of experiments on a computer numerical control milling machine with different types of sensors in high-speed milling processes, where chatter occurs frequently. Signals from different sensors are compared and features are extracted by statistical methods. Then, a dimensional reduction method  $t$ -distributed stochastic neighbor embedding is used for extracting sensitive information and visualization. Finally, the proposed reinforced  $k$ -nearest neighbors is used for chatter identification under different cutting conditions and the experiment results show the effectiveness of the proposed method.

**Index Terms**—Chatter identification,  $k$ -nearest neighbor (kNN), machine learning.

## I. INTRODUCTION

HIGH-SPEED milling has been widely used in various areas such as mold and aerospace industries because of its high efficiency and accuracy. However, the so-called process damping decreases when the spindle rotates at a high speed, which will result in a kind of self-excited vibration chatter [1]. The occurrence of chatter will result in poor surface quality, unacceptable inaccuracy, excessive noise, disproportionate tool wear, etc. [2]. Therefore, chatter identification is crucial in preventing harmful effects in machining processes and is attracting considerable interest. In the history of researches on chatter, offline chatter prediction methods by analytical models and

lobe stable charts have been well established [3], [4]. However, chatter may still occur under the predicted stable cutting conditions determined by analytical dynamics models because of the low accuracy of the models and disturbance at industrial sites. In industrial applications, chatter identification with on-machine signals may be more feasible and useful. In order to reduce detrimental effects on the workpiece and machine tool components, chatter should be detected as early as possible [5].

Last two decades witnessed a huge growth in chatter identification techniques based on various kinds of signals like acceleration [6]–[8], cutting force [9], and acoustic signals [10], and acceleration is the most widely used signal for its high reliability and low economic cost. There are three main types of chatter identification methods. The first type is based on frequency-domain signal processing techniques such as the wavelet transform [6], [7], S-function transform [8], Hilbert–Huang transform [11], adaptive filter [12], coherence function [13], [14], and synchrosqueezing transform [15]. However, frequency-domain methods have poor resolution in the edge of time-axis where data are the most recent, which results in weak performance in real-time chatter identification. In addition, there also exist high-resolution estimators based on the beta divergence [16], [17], which has potential in chatter identification application. The second type is derived from statistics theory. Entropy methods, such as the permutation entropy [18], coarse-grained entropy rate [19], and approximate entropy [20], interpret chatter from randomness point of view. Nevertheless, since the weights for frequency bands and the threshold for the existence of chatter are empirical parameters, entropy-based approaches are still facing problems in industrial applications. The third type is based on pattern classification methods such as the artificial neural network [21], [22], fuzzy logic charts [23], and other machine learning techniques. These pattern classification methods have been used for chatter identification in various machining processes.

Most pattern classification methods are trained by data from many experiments for specific cutting conditions. In this way, they can accomplish classification tasks when new data come based on their knowledge. However, few researchers have addressed the problem of online learning. There are so many different cutting conditions in a cutting process, such as spindle rotating speed, cutting width, cutting depth, feed rate, cutting tools, workpieces, etc. What is more, the temperature and the

Manuscript received April 15, 2019; revised August 5, 2019 and October 9, 2019; accepted November 12, 2019. Date of publication January 1, 2020; date of current version August 18, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 11772244 and in part by the National Science Fund for Excellent Young Scholars under Grant 51922084. (Corresponding author: Hongrui Cao.)

The authors are with the State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: shifeis@stu.xjtu.edu.cn; chr@mail.xjtu.edu.cn; xwzhang@mail.xjtu.edu.cn; chenxf@mail.xjtu.edu.cn).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIE.2019.2962465

dynamics of the spindle may vary in a cutting process. So, it is hard to obtain all kinds of data for training the machine learning models and online learning is necessary.

The rapid development in the area of artificial intelligence brings us new inspiration in chatter identification. Reinforcement learning is an area of machine learning concerned with how the machine learning models ought to take actions with a known classification result to decrease the classification error. The recent applications of reinforcement learning show its strong ability in improving model performance and it can also be applied into chatter identification.

This article outlines a new approach called reinforced  $k$ -nearest neighbors (RkNN) for chatter identification in high-speed milling. The remainder of this article is organized as follows. The feature extraction and dimensional reduction processes are presented in Section II. Section III provides the proposed classification method RkNN. Section IV introduces the cutting experiments under different conditions and some signal processing results. In Section V, the implementation of the proposed method RkNN is described in detail to verify the effectiveness of the proposed method. Finally, Section VI concludes this article.

## II. FEATURE EXTRACTION AND DIMENSIONAL REDUCTION PROCESSES

### A. Feature Extraction

To monitor a high-speed milling process, several sensors are attached on both the spindle and the workpiece to obtain the vibration and the noise of a computer numerical control (CNC) milling machine, which will be stated in detail in Section IV. Signals from sensors are time series and are hard for classification. Therefore, feature extraction work should be done in advance.

A group of signals from a sensor in a complete milling process is represented by  $X_i$  and  $i$  represents the number of milling experiments.  $X_i$  is a time series and consists of sampling points  $(x_{it_1}, x_{it_2}, x_{it_3}, \dots)$ , where  $(t_1, t_2, t_3, \dots)$  corresponds to the sampling time. To achieve chatter identification, signals in its whole length cannot be used, since the complete signals hold a long time span and even if chatter is identified by them, it has already done harm to the workpieces and the cutting tools. Therefore, the complete signals in a milling process are divided into segments. To make it easier for feature extraction and ensure the features can be used under different cutting conditions, the segment length  $L$  is determined by

$$L = \frac{f_s}{RS/60} \quad (1)$$

where  $f_s$  is the sampling frequency and  $RS$  is the spindle rotating speed with a unit of r/min. After division, the original time series  $X_i$  is changed into groups of points with a constant length  $L$ , as  $(x_{it_1}, x_{it_2}, x_{it_3}, \dots, x_{it_L}), (x_{it_{L+1}}, x_{it_{L+2}}, x_{it_{L+3}}, \dots, x_{it_{2L}}), \dots$ . One group of points corresponds to the data collected within one rotating period of the spindle. In this way, features can be extracted only by data of one rotating period, which means that chatter can be identified timely and the method can be widely used despite the spindle rotating speed. To get rid of the influence

TABLE I  
EIGHT INDEXES FOR CHATTER IDENTIFICATION

Index	Description
a <sub>1</sub>	The proportion of points falling into the interval [0,0.125)
a <sub>2</sub>	The proportion of points falling into the interval [0.125,0.25)
a <sub>3</sub>	The proportion of points falling into the interval [0.25,0.375)
a <sub>4</sub>	The proportion of points falling into the interval [0.375,0.5)
a <sub>5</sub>	The proportion of points falling into the interval [0.5,0.625)
a <sub>6</sub>	The proportion of points falling into the interval [0.625,0.75)
a <sub>7</sub>	The proportion of points falling into the interval [0.75,0.875)
a <sub>8</sub>	The proportion of points falling into the interval [0.875,1]

of cutting depth, cutting width, and feeding rate, the groups of sampling points are normalized by

$$x_{it_j}^0 = \frac{x_{it_j} - x_{\min}^k}{x_{\max}^k - x_{\min}^k} \quad (2)$$

where  $t_j$  is the sampling time,  $x_{\max}^k$  and  $x_{\min}^k$  correspond to the maximum value and minimum value in the  $k$ th group, and  $x_{it_j}^0$  denotes the normalized point. In the milling process, the amplitude of the original vibration signals is related to the cutting conditions, such as cutting depth, cutting width, and feeding rate, but it has no direct relationship to chatter. Large amplitude of the vibration signals may correspond to normal condition. Small amplitude of the vibration signals may correspond to chatter condition. We use the normalization method in order to eliminate the influence of the amplitude of the vibration signals, and let the machine learning method learn the more specific features of chatter. We want to apply our chatter identification method in all kinds of machine tools and all kinds of cutting conditions, so we normalize the original vibration signals.

After normalization, the values of points are kept in the range of 0 to 1, which makes the strategy suitable for all cutting conditions. When chatter occurs, the vibration signals will have a new frequency component called chatter frequency. The energy of the new component is large, and the waveform is added a high-frequency fluctuation. To monitor this kind of change in waveform, eight kinds of indexes based on statistical methods are proposed as shown in Table I.

The indexes we employ are actually based on the characteristics of milling force. Chatter is mainly caused by the so-called regenerative effect. It can occur often because most metal cutting operations involve overlapping cuts which can be a source of vibration amplification. The cutter vibrations leave a wavy surface, which is shown in Fig. 1. When milling, the next tooth in cut attacks this wavy surface and generates a new wavy surface. The chip thickness and, hence, the force on the cutting tool vary due to the phase difference between the wave left by the previous teeth and the wave left by the current ones. This phenomenon can greatly amplify vibrations, become dominant, and build up chatter.

When one tooth comes into cut, the instantaneous thickness of chips first increases then decreases. Large instantaneous thickness causes large milling force. If the relative phase between the wave left by the previous teeth and the wave left by the current ones approximate  $\pi$ , the dynamic chip thickness variation can be very large. What is more, the largest instantaneous thickness

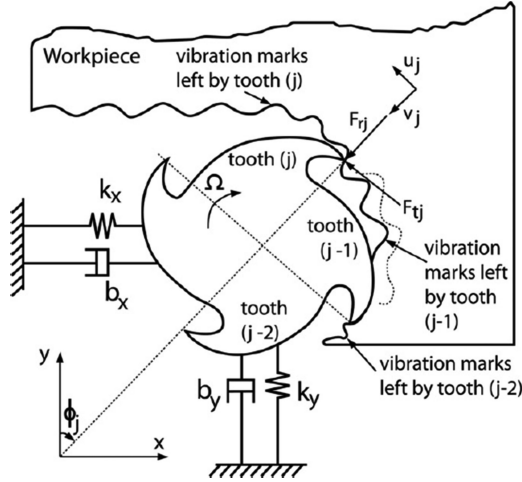


Fig. 1. Regeneration of waviness in a milling model with two degrees of freedom [24].

of previous and current chips will cause the largest variation of dynamic chip thickness variation, which will result in huge growth of the milling force and chatter. The signal we measure is vibration. Obviously, large milling force will cause excessive vibration, so large instantaneous thickness could cause chatter easier and new high frequency components will appear which are called chatter frequencies. In this way, points in the larger magnitude range of vibration will increase. Therefore, we use the proposed eight indexes in this article to realize chatter identification.

The eight indexes in Table I actually give the distribution of points of a given signal and they can reflect changes in waveshape when chatter occurs to some extent. In this way, a group of normalized points with the length in (1) can be represented by a vector  $\alpha_k$ , where  $k$  denotes the number of a group and

$$\alpha_k = (a_{1k}, a_{2k}, a_{3k}, a_{4k}, a_{5k}, a_{6k}, a_{7k}, a_{8k})^T. \quad (3)$$

## B. Dimensional Reduction

As stated in the previous section, eight indexes are extracted for chatter identification. However, among these eight indexes, not every index is sensitive to the occurrence of chatter and those insensitive indexes may do bad to the classification accuracy. What is more, a feature with eight dimensions is hard for visualization and cost much time for classification. Therefore, dimensional reduction work should be done before the feature applied to chatter identification.

In 2002, Hinton and Roweis [25] presented a dimensional reduction method called stochastic neighbor embedding (SNE), which converts the high-dimensional Euclidean distances between datapoints into conditional probabilities that represent similarities. Although SNE constructs reasonably good visualizations, it is hampered by a cost function that is difficult to optimize and by a problem referred to as the “crowding problem.” Afterwards, Maaten and Hinton [26], [27] present a new

technique called t-distributed stochastic neighbor embedding (t-SNE) that aims to alleviate these problems.

In this article, the dimensionality of the proposed feature is reduced from 8 to 2 by t-SNE in order to dig the most sensitive indexes, visualize the dataset, and reduce computation cost in classification processes. To realize dimensional reduction, the task is defined as minimizing a single Kullback–Leibler divergence between a joint probability distribution,  $P$ , in the high-dimensional space and a joint probability distribution,  $Q$ , in the low-dimensional space

$$C = KL(P||Q) = \sum_m \sum_n p_{mn} \log \frac{p_{mn}}{q_{mn}}. \quad (4)$$

Because we are only interested in modeling pairwise similarities,  $p_{mm}$  and  $q_{mm}$  are set to zero. What is more, the term  $p_{mn} \log \frac{p_{mn}}{q_{mn}}$  is set to zero when  $m$  equals  $n$ .

In (4), the similarity of datapoint to datapoint is the conditional probability,  $p_{mn}$ , that  $\alpha_m$  would pick  $\alpha_n$  as its neighbor if neighbors were picked in proportion to their probability density under a Gaussian centered at  $\alpha_m$  and with a variance of  $\sigma$ . The Gaussian distribution is an assumption by Hinton [23]. For nearby datapoints,  $p_{mn}$  is relatively high, whereas for widely separated datapoints,  $p_{mn}$  will be almost infinitesimal. Mathematically, the high-dimensional map  $p_{mn}$  is given by

$$p_{mn} = \frac{\exp(-\|\alpha_m - \alpha_n\|^2 / 2\sigma^2)}{\sum_{r \neq s} \exp(-\|\alpha_r - \alpha_s\|^2 / 2\sigma^2)}. \quad (5)$$

In the low-dimensional space, to alleviate the so-called crowding problem, a probability distribution that has much heavier tails than a Gaussian to convert distances probabilities is used. Using this distribution, the joint probability  $q_{mn}$  is defined as

$$q_{mn} = \frac{(1 + |\beta_m - \beta_n|^2)^{-1}}{\sum_{r \neq s} (1 + |\beta_r - \beta_s|^2)^{-1}} \quad (6)$$

where  $\beta$  is used to represent the feature vector in low-dimensional space.

After trained by vantage-point trees and Barnes–Hut algorithm, the feature  $\beta$  with a dimensionality of 2 can be obtained where

$$\beta_k = (b_{1k}, b_{2k})^T \quad (7)$$

and  $b_{1k}, b_{2k}$  are new indexes in the low-dimensional space which are more sensitive to chatter.

We use a Barnes–Hut algorithm to accomplish t-SNE task which is called Barnes–Hut–SNE. Barnes–Hut–SNE uses metric trees to approximate the distribution in high-dimensional space by a sparse distribution in which only  $\mathcal{O}(uN)$  values are nonzero, and approximates the gradients using a Barnes–Hut algorithm.

As the input similarities are computed using a Gaussian kernel, probabilities  $p_{mn}$  in the original manuscript corresponding to dissimilar input objects  $m$  and  $n$  are nearly infinitesimal. Therefore, we can use a sparse approximation to the probabilities  $p_{mn}$  without a substantial negative effect on the quality



of the final embeddings. In particular, we compute the sparse approximation by finding the  $\lfloor 3u \rfloor$  nearest neighbors of each of the  $N$  data objects. The nearest neighbor sets are found in  $\mathcal{O}(uN \log N)$  time by building a vantage-point tree on the data set.

To approximate the t-SNE gradient, we start by splitting the gradient into two parts as follows:

$$\begin{aligned} \frac{\partial C}{\partial \beta_m} &= 4(F_{\text{attr}} - F_{\text{rep}}) \\ &= 4 \left( \sum_{n \neq m} p_{mn} q_{mn} Z(\beta_m - \beta_n) - \sum_{n \neq m} q_{mn}^2 Z(\beta_m - \beta_n) \right) \end{aligned} \quad (8)$$

where  $F_{\text{attr}}$  denotes the sum of all attractive forces (the left sum), whereas  $F_{\text{rep}}$  denotes the sum of all repulsive forces (the right sum). Computing the sum of all attractive forces,  $F_{\text{attr}}$  is computationally efficient; it can be done by summing over all nonzero elements of the sparse distribution  $P$  in  $\mathcal{O}(uN)$ , where the term  $q_{mn}Z = (1 + |\beta_m - \beta_n|^2)^{-1}$  can be computed in  $\mathcal{O}(1)$ . However, a naïve computation of the sum of all repulsive forces,  $F_{\text{rep}}$ , is  $\mathcal{O}(N^2)$ . In our experiments, a Barnes–Hut algorithm is employed to approximate  $F_{\text{rep}}$  efficiently in  $\mathcal{O}(N \log N)$  [28].

In this kind of dimensional reduction way, low-dimensional representations  $\beta$  of individual datapoints is referred as map points, and the aim of dimensional reduction is to preserve as much of the significant structure of the high-dimensional data as possible in the low-dimensional map. Different from traditional dimensionality reduction techniques as principal components analysis and classical multidimensional scaling, which are linear techniques that focus on keeping the low-dimensional representations of dissimilar datapoints far apart. Linear dimensional reduction techniques mean that these techniques utilize linear transform to realize dimensional reduction, which limits the representation form of the data in low-dimensional space largely. For high-dimensional data that lies on or near a low-dimensional, nonlinear manifold, it is usually more important to keep the low-dimensional representations of very similar datapoints close together, which is typically not possible with a linear mapping. LASSO is a widely used and effective dimensionality reduction method which we also thought about. However, we finally abandon it because it is mainly used for regression problem and need a target to realize dimensionality reduction which is not suitable for chatter identification where the labels of new data are unknown.

Although there are many classification methods which can handle the classification task of the vectors with eight elements, we still insist on employing the dimensionality reduction technique in advance. The first reason is to get the most sensitive information. The eight indexes we proposed are based on statistical method and they reflect the whole characteristics of the signal segment. However, the characteristic of chatter may not exist in every magnitude range, which means not every index can contribute to chatter identification. Therefore, we employ t-SNE to explore the most sensitive ones based on the distance between them. The second reason is to visualization. While milling, if

---

**Algorithm 1: kNN.**


---

- 1.1: Decide the number of neighbors  $k$ .
  - 1.2: For  $\forall k = 1, 2, \dots, N$ , compute the distance  $d(\beta_t, \beta_k)$  between  $\beta_t$  and  $\beta_k$ .
  - 1.3: Arrange the distance  $d(\beta_t, \beta_k)$  in ascending order.
  - 1.4: Obtain the first  $k$  nearest distances and search the corresponding data as neighbors.
  - 1.5: Suppose  $k_c$  is the number of neighbors belonging to  $c(c = 1, 2, \dots, C)$ th class.
  - 1.6: Find the maximum  $k_c$ , and the data  $\beta_t$  belongs to corresponding  $c$ th class.
- 

the distribution of dataset and the current work condition can be seen on the screen of the machine tool, the workers can see the cutting condition directly. What is more, not only can they see the cutting condition of the machine tool, they also can watch the change path of current milling condition. The third reason is the representation of space. Space with high dimensionality need more data to represent, and low-dimensional space need less data. If we use dimensionality reduction technique in advance, we can employ less data in the original dataset and this can reduce the computation time which is important in chatter identification.

### III. RKNN FOR CLASSIFICATION

The  $k$ -nearest neighbors (kNN) rule is one of widely used machine learning algorithms in pattern recognition, data mining, and anomaly detection [29]. The kNN rule was first proposed to classify the data on the basis of the similarity with neighbors in the training dataset, and has been extensively studied such as its computation speed [30]. As shown in Section II, the feature of the  $k$ th group is represented by  $\beta_k$  in (7) of two dimensions. Supposing the labeled training data are  $(\beta_k, z_k)$ ,  $k = 1, 2, \dots, N$ , where  $\beta_k \in R^2$ ,  $z_k$  is the label of  $\beta_k$ , and  $N$  is the total number of groups. Assuming that the number of classes in the training dataset is  $c$ , then we can get  $z_i \in \{1, 2, \dots, c\}$ . For an unlabeled data  $\beta_t \in R^2$ , the detailed classification procedure of the kNN rule is shown as follows.

If the original dataset can cover all the conditions while data collecting, kNN will hold a high accuracy and a relatively short computation time. However, it is hard to acquire data in all conditions in a high-speed milling process, since the conditions in a milling process is quite complex. First, the cutting parameters can affect the extracted features, which means that features may get different when cutting parameters change such as spindle rotating speed, cutting width, cutting depth, feed rate, and so on. Second, different spindles and cutting tools have different dynamic characteristics, so vibration response may be different when using different spindles and cutting tools. Third, in a long-time high-speed milling process, the structure dynamic characteristics may vary caused by change of temperature and wear of cutting tool.

Therefore, the classical kNN is hard to utilize for chatter identification in a high-speed milling process for the lack of adaptivity. In this article, the thought of reinforcement learning

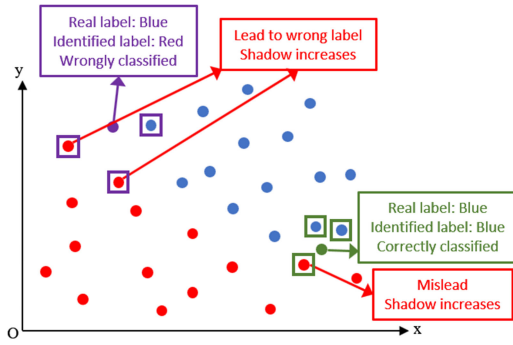


Fig. 2. Graphical interpretation for the two punishment processes.

is applied into kNN to keep the effectiveness of chatter identification model. We give each original data a new parameter called the Shadow. After a high-speed milling process, each data in the original dataset will be assessed by its own performance, where good performance will result in small Shadow. Shadow of each data will cause the change of another parameter called the Light. In next milling process, the classification process will not only depend on the displacement between the coming point and the dataset but is also related to the Light of each data.

For each low-dimensional data  $\beta_k$ , its Light is represented by  $L_k$  with an initial value of 0. Within a milling process, the real cutting condition is unknown, and it is classified by our model. After a milling process, the finished surface of the workpiece can be seen, so it can be judged if chatter occurs by whether there are obvious chatter marks [31]. By comparing the real milling condition and the judged condition by our model, the classification accuracy can be obtained as  $\nu$ . For the wrongly classified data with a number of  $N * (1 - \nu)$ , it is certain that there are neighbors around them which have the same label with their identified labels. These neighbors lead to wrong classification and low classification accuracy, so they should be punished. Therefore, for each wrongly classified data, the Shadow of their neighbors, which have the same label with their identified label is increased by  $\mu_1$ . For the correctly classified data with a number of  $N * \nu$ , there may also exist some neighbors around them which have a different label with their identified labels. Since the number of these neighbors is relatively small, the data are correctly classified. However, these neighbors still do bad to classification performance. Hence, for each correctly classified data, the Shadow of their neighbors, which have the different label with their identified label is increased by  $\mu_2$ . We recommend that  $\mu_1$  should be larger than  $\mu_2$ , because the data punished by  $\mu_1$  have already caused wrong classification. Small value of  $\mu_1$  and  $\mu_2$  can keep the stability of model but may reduce the adaptation for new environment, while large value of  $\mu_1$  and  $\mu_2$  can make the model more sensitive to the change of environment but easier affected by noises. To make it clearer, Fig. 2 is drawn to show both two punishment processes.

In Fig. 2, there are two categories, red and blue, and the number of reference neighbors is set as 3. The purple circle and the green circle are data to be classified. For the purple circle,

whose real label is blue, two of the nearest neighbors are red and one is blue, so the purple circle is identified as red which is wrong. The two red neighbors are those who lead to the wrong classification, so their Shadow increases. For the green circle, whose real label is blue, two of the nearest neighbors are blue and one is red, so the green circle is identified as blue which is right. The red neighbor misleads its classification, so its Shadow increases.

The Shadow of each data in the original dataset updates for each group of test data. Supposing the number of original datasets is  $N_a$  in each category, the number of test data in a milling process is  $N$ , and the classification accuracy is  $\nu$ , the Shadow of each data  $S_k$  will change into

$$S_k = S_{k_0} + \mu_1 N_a (1 - \nu) \sigma_{1k} + \mu_2 N_a \nu \sigma_{2k} \quad (9)$$

where  $S_{k_0}$  denotes the original Shadow value before updating,  $\sigma_{1k}$  denotes the percentage where the data  $\beta_k$  occur in the neighbors leading to the wrong label when the data are wrongly classified, and  $\sigma_{2k}$  denotes the percentage where the data  $\beta_k$  occur in the misleading neighbors when the test data are correctly classified.

In a high-speed milling process, the cutting condition is either normal or chatter. If we only use the punishment method in (9), only the data in the category different from the test data will be punished, which will cause only the Shadow of data in one category increases. So, a balance process needs to be done after (9). For the original data whose label is the same with the test data, supposing the data in the original dataset have same number in each category, their Shadow will change into

$$\begin{aligned} S_k &= \frac{1}{N_a} \sum_1^{cN_a} \mu_1 N_a (1 - \nu) \sigma_{1k} + \mu_2 N_a \nu \sigma_{2k} \\ &= \sum_1^{cN_a} \mu_1 (1 - \nu) \sigma_{1k} + \mu_2 \nu \sigma_{2k}. \end{aligned} \quad (10)$$

Data in the original dataset with the largest Shadow indicates that they do not match the current cutting conditions or the current milling machine dynamic characteristics. To make the dataset suitable for the current situation,  $N(1 - \nu)$  data with largest Shadow are deleted and the same amount of data are added into the dataset from the current cutting test.

As for the relationship between Light and Shadow, we refer to the Ebbinghaus' Forgetting Curve [32]. Ebbinghaus's classic experiment on forgetting from 1880 and 1885 is one of the most important early experiments in psychology. In RkNN, Shadow corresponds to the forgetting time in the Ebbinghaus' Forgetting Curve while Light corresponds to the remaining memory capacity. It is reasonable because the Shadow of data increases when misleading. If one datum misleads, it makes classification performance worse and should be forgot. Light influence the classification process which will be talked after, and it can also be described memory capacity since large memory capacity will have a larger effect on classification. In this way, Light of one datum  $L_k$  is obtained by the function of the Ebbinghaus'

Forgetting Curve as

$$L_k = (1 + 0.523S_k)^{-0.101}. \quad (11)$$

In kNN, the labels of test data are determined by their neighbors. Among the nearest neighbors, the label which has the most neighbors will be chosen. In RkNN, the labels of test data are also determined by their neighbors. Difference is the criterion. The label with largest Light will be chosen where the Light is the sum of every neighbor in each label. In this way, a sample with larger Light will have larger influence on classification.

To see how RkNN influences classification accuracy compared with kNN, let  $x$  denote the measurements on an individual and  $X$  the sample space of possible values of  $x$ . We shall refer to  $x$  as the observation. On the basis of  $x$ , a decision must be made about the membership of the individual in one of  $C$  specified categories.

For the purpose of defining Bayes risk, we assume  $f_1(x), f_2(x), \dots, f_M(x)$ , probability densities at  $x$  with respect to a  $\sigma$ -finite measure  $\nu$ , such that an individual in category  $i$  give rise to an observation  $x$  according to density  $f_i$ . Let  $L(i, j)$  be the loss incurred by assigning an individual from category  $i$  to category  $j$ .

Let  $\eta_1, \eta_2, \dots, \eta_M, \eta_i \geq 0, \sum \eta_i = 1$ , be the prior probabilities of the  $M$  categories. The conditional probability  $\hat{\eta}_i(x)$  of an individual with measurements  $x$  belonging to category  $i$  is, by the Bayes theorem,

$$\hat{\eta}_i = \frac{\eta_i f_i}{\sum \eta_i f_i}, i = 1, 2, \dots, c. \quad (12)$$

Thus, the random variable  $x$  transforms the prior probability vector  $\eta$  into the posterior probability vector  $\hat{\eta}(x)$ . If the statistician decides to place an individual with measurements  $x$  into category  $j$ , the conditional loss is

$$r_i^x = \sum_{i=1}^c \hat{\eta}_i(x) L(i, j). \quad (13)$$

For a given  $x$ , the conditional loss is minimum when the individual is assigned to the category  $j$  for which  $r_i(x)$  is the lowest. Minimizing the conditional expected loss obviously minimizes the unconditional expected loss. Thus, the minimizing decision rule  $\delta^*$ , called the Bayes decision rule with respect to  $\eta$ , is given by deciding the category  $j$  for which  $r_i$  is the lowest. Using  $\delta^*$ , the conditional Bayes risk  $r^*(x)$  is

$$r^*(x) = \min_i \left\{ \sum_{i=1}^c \hat{\eta}_i L(i, j) \right\}. \quad (14)$$

**Lemma** (Convergence of the Nearest Neighbor) [33]: Let  $x$  and  $x_1, x_2, \dots$  be independently identically distributed random variables taking values in a separate metric space  $X$ . Let  $x'_n$  denote the nearest neighbor to  $x$  from the set  $\{x_1, x_2, \dots, x_n\}$ . Then,  $x'_n \rightarrow x$  with probability one.

Therefore, it is possible to conclude that the  $k$ th nearest neighbor to  $x$  converges to  $x$  with probability one as the sample size  $n$  increases with  $k$  fixed. Since each of the nearest neighbors' casts conditionally independent votes as to the category of  $x$ , we may conclude, in the two-category case for odd  $k$ , that the

conditional kNN risk  $r_k(x)$  is given in the limit as  $n$  increases, by

$$r_k(x) = \hat{\eta}_1(x) \sum_{j=0}^{(k-1)/2} \binom{k}{j} \hat{\eta}_1^j(x) (1 - \hat{\eta}_1)^{k-j} \\ + (1 - \hat{\eta}_1(x)) \sum_{j=(k+1)/2}^k \binom{k}{j} \hat{\eta}_1^j(x) (1 - \hat{\eta}_1)^{k-j}. \quad (15)$$

Note that the conditional NN risk  $r_k(x)$  are monotonically decreasing in  $k$  (to  $\min\{\hat{\eta}_1(x), 1 - \hat{\eta}_1(x)\}$ ), as we might suspect. Observe that  $r_k$  is symmetric in  $\hat{\eta}_1$  and  $1 - \hat{\eta}_1$ . Thus,  $r_k$  may be expressed solely in terms of  $r^* = \min\{\hat{\eta}_1(x), 1 - \hat{\eta}_1(x)\}$  in the form

$$r_k = r^* \sum_{j=0}^{(k-1)/2} \binom{k}{j} (r^*)^j (1 - r^*)^{k-j} \\ + (1 - r^*) \sum_{j=\frac{(k+1)}{2}}^k \binom{k}{j} (r^*)^j (1 - r^*)^{k-j}. \quad (16)$$

In kNN, the  $k$  neighbors have equal effect on the classification of  $x$ . However, in RkNN, after learning from the results, each neighbor will have different effect on classification. Suppose that the neighbors leading to the wrong results are given Shadow once and neglect the balance procedure and the sample replacement procedure, the two terms in (14) will no longer divided by the middle stage. Let  $l$  denote the Light of punished samples. Then, the first term will change into the sum from  $j = 0$  to the largest integer smaller than  $\frac{l}{l+1}k$  and the second term will change into the sum from the smallest integer larger than  $\frac{1}{l+1}k$  to  $k$  as shown next

$$r'_k(x) = \hat{\eta}_1(x) \sum_{j=0}^{\lfloor \frac{l}{l+1}k \rfloor} \binom{k}{j} \hat{\eta}_1^j(x) (1 - \hat{\eta}_1)^{k-j} \\ + (1 - \hat{\eta}_1(x)) \sum_{j=\lceil \frac{1}{l+1}k \rceil}^k \binom{k}{j} \hat{\eta}_1^j(x) (1 - \hat{\eta}_1)^{k-j}. \quad (17)$$

Since  $r_k$  is symmetric in  $\hat{\eta}_1$  and  $1 - \hat{\eta}_1$

$$r'_k = r^* \sum_{j=0}^{\lfloor \frac{l}{l+1}k \rfloor} \binom{k}{j} (r^*)^j (1 - r^*)^{k-j} \\ + (1 - r^*) \sum_{j=\lceil \frac{1}{l+1}k \rceil}^k \binom{k}{j} (r^*)^j (1 - r^*)^{k-j}. \quad (18)$$

In this way, the conditional risk is reduced by

$$\Delta r_k = r_k - r'_k = r^* \sum_{j=\lfloor \frac{l}{l+1}k \rfloor}^{(k-1)/2} \binom{k}{j} (r^*)^j (1 - r^*)^{k-j} \\ + (1 - r^*) \sum_{j=\frac{(k+1)}{2}}^{\lceil \frac{1}{l+1}k \rceil} \binom{k}{j} (r^*)^j (1 - r^*)^{k-j}. \quad (19)$$

**TABLE II**  
INDEXES USED FOR SIGNAL EXTRACTION

Index	Expression
Root mean square	$X_{rms} = \sqrt{\frac{1}{L} \sum_{l=1}^L x_l^2}$
Square mean root	$X_s = \left( \frac{1}{L} \sum_{l=1}^L \sqrt{ x_l } \right)^2$
Mean absolute value	$X_a = \frac{1}{L} \sum_{l=1}^L  x_l $
Peak to peak value	$X_p = \max(x_l) - \min(x_e)$

The k-d tree is employed for the search of nearest neighbors. Since there are many possible ways to choose axis-aligned splitting planes, there are many different ways to construct k-d trees. The algorithm we employed for building a balanced k-d tree presort the data prior to building the tree, which has a worst-case complexity of  $\mathcal{O}(kn \log n)$ . This algorithm presorts  $n$  points in each of  $k$  dimensions using an  $\mathcal{O}(n \log n)$  sort to building the tree.

The nearest neighbor search algorithm aims to find the point in the tree that is nearest to a given input point. Finding the nearest point is an  $\mathcal{O}(\log n)$  operation on average in the case of randomly distributed points [34].

As for the two parameters we proposed and their updating processes as shown from (9) to (11), they are updated after each whole milling process. Because the two parameters of each data in the dataset are updated once after each whole milling process and the number of times is related to the number of data, the computational complexity is  $\mathcal{O}(n)$ .

Another problem in signal processing is to extract the signals in cutting stage. In a high-speed milling process, there are three stages in succession, where the first one is the idling stage, then the cutting stage, and last the idling stage again. Data from the idling stage is useless in chatter identification and model updating. In model updating, only data from cutting stage are needed, thus we should extract the data in cutting stage.

After cutting signals by (1), four commonly used indexes in time domain and k-means clustering technique are applied to realize signal extraction. Four indexes are root mean square, square mean root, mean absolute value, and peak-to-peak value as shown in Table II.

To sum up, the whole detailed RkNN-based chatter identification method is shown as follows.

#### IV. EXPERIMENTS AND DATA ACQUISITION

To verify the effectiveness of the proposed method, several milling experiments have been conducted on a CNC milling machine VMC-V5 as shown in Fig. 3.

In milling processes, five different sensors are placed on both the spindle and the workpiece. Two accelerometers (IMI 608A11, with sensitivity of 100 mV/g) are placed on the spindle in both  $x$  and  $y$  directions. Two accelerometers (PCB 333B50, with sensitivity of 1000 mV/g) are attached on the workpiece

#### Algorithm 2: RkNN-Based Chatter Identification Method.

- 2.1: Choose historical data from both normal condition and chatter condition as original dataset, decide the number of neighbors.
- 2.2: During a high-speed milling process, acquire real time data by the length in (1) and normalize real time data according to (2)
- 2.3: Calculate the eight indexes in Table I and reduce the real time data dimension to 2 by t-SNE
- 2.4: For  $\forall k = 1, 2, \dots, N$ , compute the distance  $d(\beta_t, \beta_k)$  between  $\beta_t$  and  $\beta_k$  and arrange the distance  $d(\beta_t, \beta_k)$  in ascending order.
- 2.5: Obtain the first  $k$  nearest distances and search the corresponding data in dataset as neighbors.
- 2.6: Calculate the total Light in normal condition and chatter condition, then classify the real time data to the label with largest Light, repeat from 2.2 until a whole milling process is finished.
- 2.7: After a whole milling process, get real cutting condition and cut signals by four indexes in Table II and k-means clustering
- 2.8: Update Shadow  $S_k$  of data in dataset according to (9) and (10), and then delete ill-suited data in dataset
- 2.9: Obtain updated Light  $L_k$  of data in dataset by (11)
- 2.10: Repeat from 2.2



Fig. 3. CNC milling machine VMC-V5 in experiments.

in both  $x$  and  $y$  directions. A microphone is used to acquire the sound signals. The data acquire device is ECON AVANTMI-7008 with the sampling frequency of 10 240 Hz. The arrangement of sensors is shown in Fig. 4.

A high-speed steel milling cutter with three teeth is used in the experiments. In high-speed milling, chatter is mainly influenced by two cutting parameters: the spindle rotating speed and the cutting depth. There are some research works that focus on the relationship between these two parameters and chatter. In the point of view of guarantee stable machining processes, these research works are called out-of-process strategies. However, in many cases, the SLD of the system cutting tool, machine tool, and workpiece is continuously changing, and it is difficult to make predictions in advance and schedule the correct parameters



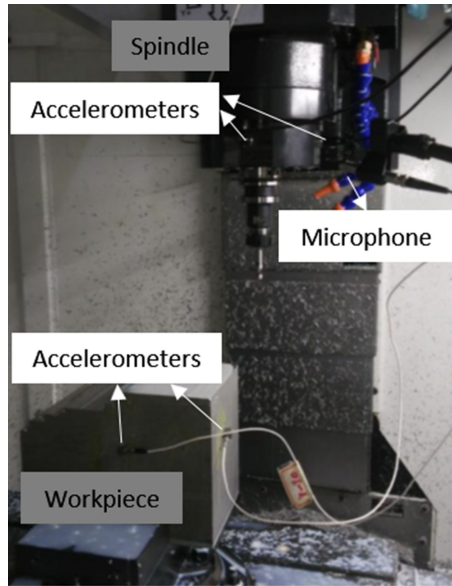


Fig. 4. Arrangement of five sensors in a CNC milling machine.

TABLE III  
CUTTING PARAMETERS IN MILLING EXPERIMENTS

Number	Rotating Speed /rpm	Depth /mm	Width /mm	Feed per tooth /mm
1-5	7000	1,3,4,5,7	0.6	0.05
6-10	8000	1,3,4,5,7	0.6	0.05
11-15	9000	1,3,4,5,7	0.6	0.05
16-20	10000	1,3,4,5,7	0.6	0.05
21-25	11000	1,3,4,5,7	0.6	0.05
26-30	12000	1,3,4,5,7	0.6	0.05
31-35	13000	1,3,4,5,7	0.6	0.05
36-40	14000	1,3,4,5,7	0.6	0.05
41-45	15000	1,3,4,5,7	0.6	0.05

to ensure stable operations [35]. In-process strategies focus on detecting chatter during the metal cutting process by monitoring a certain signal such as vibration sound power, etc., without SLD identification. What is more, thought of reinforcement learning is added in our identification process in order to deal with the change of cutting tools, machine tools, and workpieces.

In our experiments, the spindle rotating speed varies from 7000 to 15 000 r/min with a step of 1000 r/min and the cutting depth varies from 1 to 7 mm while the cutting width and the feed rate are kept unchanged which are 0.6 mm and 0.05 mm/tooth. Total 45 milling experiments are conducted with different milling parameters. The used groups of parameters are detailed in Table III.

Five sensors are used in the milling experiments, but they may not all have good performance on chatter identification. To find which signal reflect the characteristics of chatter better, we take Number 31 and 35 experiments, for example, where Number 31 is a normal cutting condition and Number 35 is a chatter condition. The time-domain signals of five sensors in these two experiments are shown in Fig. 5.

From Fig. 5, we can see the different performances toward chatter with different sensors. When chatter occurs in a milling

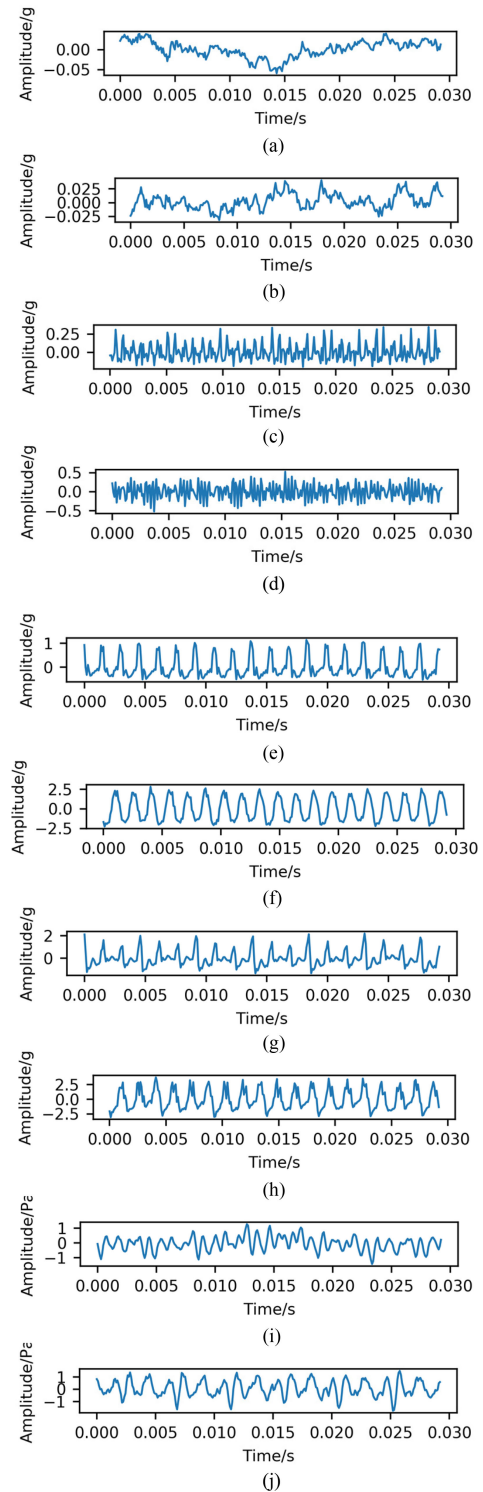


Fig. 5. Signal segments from the five different sensors. (a) Signal segment of the accelerometer on the spindle in x direction (normal). (b) Signal segment of the accelerometer on the spindle in x direction (chatter). (c) Signal segment of the accelerometer on the spindle in y direction (normal). (d) Signal segment of the accelerometer on the spindle in y direction (chatter). (e) Signal segment of the accelerometer on workpiece in x direction (normal). (f) Signal segment of the accelerometer on workpiece in x direction (chatter). (g) Signal segment of the accelerometer on workpiece in y direction (normal). (h) Signal segment of the accelerometer on workpiece in y direction (chatter). (i) Signal segment of the microphone (normal). (j) Signal segment of the microphone (chatter).



TABLE IV  
REAL CUTTING CONDITION OF EACH EXPERIMENT

Cutting condition	Experiment numbers
Normal	1,2,3,6,11,12,16,17,18,21,22,26,31,36,41
Chatter	4,5,7,8,9,10,13,14,15,19,20,23,24,25,27,28,29,30,32,33,34,35,37,38,39,40,42,43,44,45

process, the cutting force will have a new frequency component called chatter frequency, which relies on the dynamic characteristics of the spindle system. In Fig. 5(a)–(d), no clear periodical component caused by static cutting force can be seen in normal condition and no new high frequency component appears in chatter condition. Hence, we can say the accelerometer on the spindle cannot reflect the cutting condition and the vibration caused by the cutting force clearly. The reason may be that the stiffness of the spindle system is too large so the cutting force can hardly motivate the vibration of the spindle, and the spindle keep moving while cutting, which may result in larger noises in acceleration signals.

In Fig. 5(i) and (j), both components of cutting frequency and chatter frequency do not exist at all and in chatter condition, the amplitude of the signal does not increase obviously. So, signals from the microphone still cannot be used for chatter identification. This may be caused by the environment noise in the CNC milling machine.

Signals in Fig. 5(e)–(h) can reflect real milling condition better. In Fig. 5(e) and (g), the vibration of the workpiece caused by periodical cutting force can be seen clearly. When chatter occurs, from Fig. 5(f) and (h), besides the vibration caused by static cutting force, vibration caused by dynamic cutting force also exists, which accords with the characteristics of chatter. Compared with Fig. 5(f), components with high frequency are much more obvious in Fig. 5(h). In industry, we want to realize chatter identification with less sensors which can reduce economical cost. In this way, signals from the accelerometer on the workpiece in  $y$  direction are chosen for chatter identification.

## V. RESULTS AND DISCUSSION OF RKNN

After each experiment shown in Table III, the finished surface is examined to decide the real cutting condition this group of experiment is in. The real cutting condition or the label of each experiment is shown in Table IV.

We employ a low-pass filter to ban the environment noises. Since one sample for classification for classification last only one rotating period, it is hard to utilize filters on one sample to make identification of the components due to the static cutting force and the dynamic cutting force.

Signals when spindle rotating speed is 13 000 r/min are used as original dataset which is also shown in Section IV. The experiment numbers of these signals are 31–35, where 31 is in normal condition while 32–35 are in chatter condition as shown in Table IV. One thousand and two hundred samples are selected randomly and equally under normal condition and under chatter condition, which means  $N_a = 600$ . First, the effectiveness of the eight indexes in Table I and the dimensional reduction method t-SNE should be tested. Fig. 6 shows the dimensional reduction result.

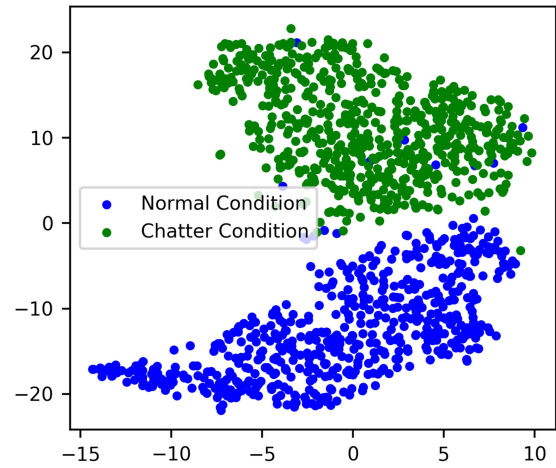


Fig. 6. Dimensional reduction result of original dataset.

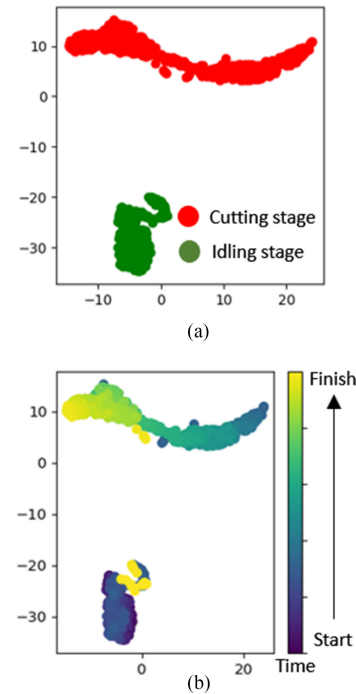


Fig. 7. Clustering result and its distribution towards time. (a) Clustering result of a milling process. (b) Distribution of samples toward time.

In Fig. 6, blue circle represents the samples under normal condition and green circle represents the samples under normal condition. Points with different labels are separated obviously except a few blue circles which may be caused by noises from environment or the sensor. Fig. 6 proves that the extracted eight indexes and t-SNE are efficient for chatter identification.

As stated in Section III, in a high-speed milling process, there are three states and what we need is the cutting stage. Therefore, four commonly used statistical indexes and k-means clustering are applied to get the signals in cutting stage. Since the dimension of each data is high, we still use the dimensional reduction technique for visualization. In this way, signals in cutting stage can be extracted for reinforcement learning procedure. Fig. 7 shows the clustering result.

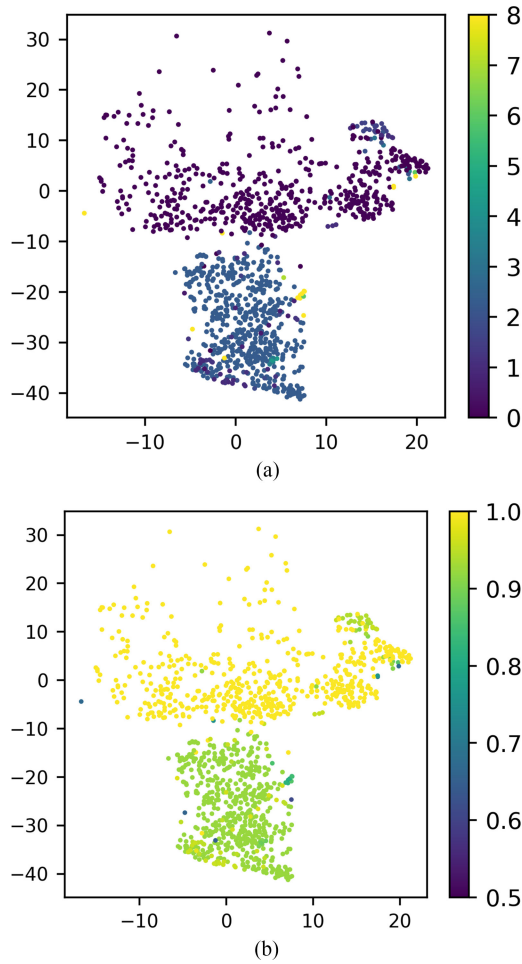


Fig. 8. Shadow and Light distribution of samples in the dataset. (a) Shadow value distribution of samples in the dataset. (b) Light value distribution of samples in the dataset.

Fig. 7(a) shows that the whole cutting process has been separated into two categories and the distance between these two categories is quite large. The green category is the idling stage and the red category is the cutting stage. In Fig. 7(b), as time goes on, the samples first are in the green category, then jump into the red category and at last back to the green category, which corresponds to the three stages in a milling process, which are idling, cutting, and idling.

Last is about the Shadow and the Light. When RkNN is used in real-time identification, the Shadow and the Light changes according to each sample's performance in last classification, which means that bad performance of a sample may result in large Shadow value. In our experiments,  $\mu_1$  and  $\mu_2$  are set as 1 and 0.5 separately, and identification processes begin from the largest spindle rotating speed. Fig. 8 shows the Shadow and the Light of the dataset after chatter identification based on RkNN.

In Fig. 8, the Shadow and Light distribution of samples in the dataset after chatter identification for the first three rotating speeds is shown. From Fig. 8(a), samples are with different Shadows. Samples with large Shadow means that they performs bad in the past classification tasks and they will have relatively small

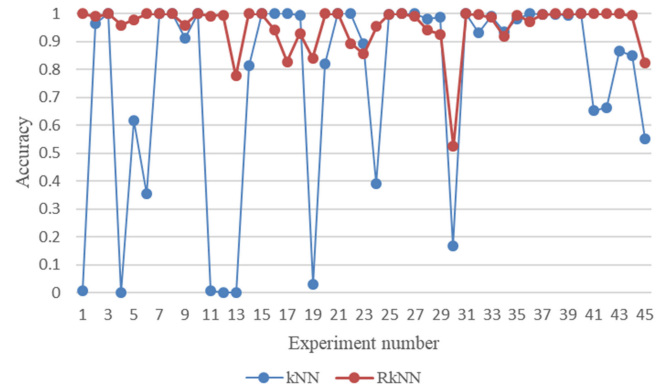


Fig. 9. Accuracy comparison between kNN and RkNN under all cutting conditions.

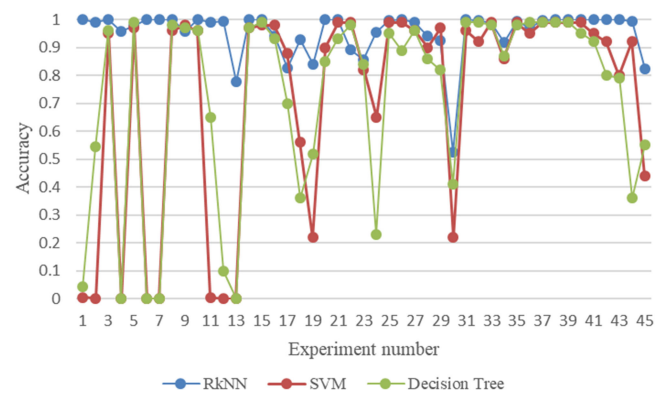


Fig. 10. Accuracy comparison between RkNN, SVM, and decision tree.

influence on the following classification and may be replaced in the future. In Fig. 8(b), samples have different Light according to their Shadow. Sample with larger Light will have more influence on the classification task.

Fig. 9 shows the overall classification accuracy between kNN and RkNN. In kNN, cutting conditions near the initial conditions have good performance, but it gets poor when goes to lower spindle rotating speed. After reinforcement learning, the dataset is more suitable for varies cutting conditions and have obviously better performance than kNN.

The proposed method is also compared with support vector machine (SVM) and decision tree, which are widely used in classification tasks. In these two classification methods, the proposed eight indexes are used without dimensionality reduction, because in SVM and decision tree, if t-SNE is used, the models need to be retrained when new data come. The results are shown in Fig. 10.

## VI. CONCLUSION

This article proposed a new chatter identification method based on t-SNE and RkNN. First, the signal segmentation and normalization method were introduced. Then, eight indexes were calculated for chatter identification and a day-ahead dimension reduction method t-SNE was used for feature

extracting. Next, a chatter identification method RkNN and its two specific parameters Shadow and Light were detailed. Finally, in the milling experiments, signals from different sensors were compared and the proposed chatter identification was examined. The results showed the effectiveness of the proposed method. The results and contributions of this article could be concluded as follows.

- 1) Milling experiments with different cutting parameters such as the rotating speed and the cutting depth were done which contain both normal and chatter conditions. Signals from five different sensors were collected and compared, and the accelerometer attached at y direction on the workpiece was found to be more sensitive to chatter.
- 2) The signals in the cutting stage were singled out successfully by four time-domain indexes and k-means clustering. To realize chatter identification, a new group of indexes based on statistical methods was proposed and a dimensionality reduction technique t-SNE was utilized in order to find the most sensitive features. The proposed indexes and the dimensionality reduction technique were shown to be efficient.
- 3) A new classification method RkNN was proposed based on the thought of reinforcement learning and two new parameters called Shadow and Light were proposed to realize RkNN. RkNN is confirmed to be effective in chatter identification under various milling conditions, which can accomplish chatter identification only by the signals when spindle rotates a round. What is more, the classification accuracy of RkNN is compared with kNN both theoretically and experimentally and compared with SVM and the decision tree experimentally.

However, we only apply RkNN on chatter identification task and verify its feasibility. We think that RkNN is not only suitable for chatter identification but can also be used in variety of fault diagnosis fields such as damage of bearing. In our future work, we want to improve the generality of RkNN and expand it to more fault diagnosis fields.

## REFERENCES

- [1] H. Cao, X. Zhang, and X. Chen, "The concept and progress of intelligent spindles: A review," *Int. J. Mach. Tools Manuf.*, vol. 112, pp. 21–52, 2017.
- [2] G. Quintana and J. Ciurana, "Chatter in machining processes: A review," *Int. J. Mach. Tools Manuf.*, vol. 51, no. 5, pp. 363–376, 2011.
- [3] L. Yuan, Z. Pan, D. Ding, S. Sun, and W. Li, "A review on chatter in robotic machining process regarding both regenerative and mode coupling mechanism," *IEEE/ASME Trans. Mechatron.*, vol. 23, no. 5, pp. 2240–2251, Oct. 2018.
- [4] H. Cao, B. Li, and Z. He, "Chatter stability of milling with speed-varying dynamics of spindles," *Int. J. Mach. Tools Manuf.*, vol. 52, no. 1, pp. 50–58, 2012.
- [5] Y. Sun and Z. Xiong, "An optimal weighted wavelet packet entropy method with application to real-time chatter detection," *IEEE/ASME Trans. Mechatron.*, vol. 21, no. 4, pp. 2004–2014, Aug. 2016.
- [6] D. Shi and N. N. Gindy, "Industrial applications of online machining process monitoring system," *IEEE/ASME Trans. Mechatron.*, vol. 12, no. 5, pp. 561–564, Oct. 2007.
- [7] Y. Sun, C. Zhuang, and Z. Xiong, "A scale factor-based interpolated DFT for chatter frequency estimation," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 10, pp. 2666–2678, Oct. 2015.
- [8] I. N. Tansel *et al.*, "Transformations in machining. Part 2. Evaluation of machining quality and detection of chatter in turning by using s-transformation," *Int. J. Mach. Tools Manuf.*, vol. 46, no. 1, pp. 43–50, 2006.
- [9] Y. Altintas and P. K. Chan, "In-process detection and suppression of chatter in milling," *Int. J. Mach. Tools Manuf.*, vol. 32, no. 3, pp. 329–347, 1992.
- [10] T. Delio, J. Tlustý, and S. Smith, "Use of audio signals for chatter detection and control," *J. Eng. Ind.*, vol. 114, no. 2, pp. 146–157, 1992.
- [11] H. Cao, Y. Lei, and Z. He, "Chatter identification in end milling process using wavelet packets and Hilbert–Huang transform," *Int. J. Mach. Tools Manuf.*, vol. 69, pp. 11–19, 2013.
- [12] L. Ma, S. N. Melkote, and J. B. Castle, "A model-based computationally efficient method for on-line detection of chatter in milling," *J. Manuf. Sci. Eng.*, vol. 135, no. 3, 2013, Art no. 031007.
- [13] K. M. Hynynen *et al.*, "Chatter detection in turning processes using coherence of acceleration and audio signals," *J. Manuf. Sci. Eng.*, vol. 136, no. 4, 2014, Art no. 044503.
- [14] X. Q. Li, Y. S. Wong, and A. Y. C. Nee, "Tool wear and chatter detection using the coherence function of two crossed accelerations," *Int. J. Mach. Tools Manuf.*, vol. 37, no. 4, pp. 425–435, 1997.
- [15] H. Cao *et al.*, "Chatter detection based on synchrosqueezing transform and statistical indicators in milling process," *Int. J. Adv. Manuf. Technol.*, vol. 95, nos. 1–4, pp. 961–972, 2018.
- [16] M. Zorzi, "A new family of high-resolution multivariate spectral estimators," *IEEE Trans. Autom. Control*, vol. 59, no. 4, pp. 892–904, Apr. 2014.
- [17] M. Zorzi, "Multivariate spectral estimation based on the concept of optimal prediction," *IEEE Trans. Autom. Control*, vol. 60, no. 6, pp. 1647–1652, Jun. 2015.
- [18] U. Nair *et al.*, "Permutation entropy based real-time chatter detection using audio signal in turning process," *Int. J. Adv. Manuf. Technol.*, vol. 46, nos. 1–4, pp. 61–68, 2010.
- [19] J. Gradišek *et al.*, "Automatic chatter detection in grinding," *Int. J. Mach. Tools Manuf.*, vol. 43, no. 14, pp. 1397–1403, 2003.
- [20] D. Pérez-Canales *et al.*, "Identification of dynamic instabilities in machining process using the approximate entropy method," *Int. J. Mach. Tools Manuf.*, vol. 51, no. 6, pp. 556–564, 2011.
- [21] E. Kuljanic, G. Totis, and M. Sortino, "Development of an intelligent multisensor chatter detection system in milling," *Mech. Syst. Signal Process.*, vol. 23, no. 5, pp. 1704–1718, 2009.
- [22] H. Cao *et al.*, "Early chatter detection in end milling based on multi-feature fusion and  $3\sigma$  criterion," *Int. J. Adv. Manuf. Technol.*, vol. 92, nos. 9–12, pp. 4387–4397, 2017.
- [23] A. Devillez and D. Dudzinski, "Tool vibration detection with eddy current sensors in machining process and computation of stability lobes using fuzzy classifiers," *Mech. Syst. Signal Process.*, vol. 21, no. 1, pp. 441–456, 2007.
- [24] Y. Altintas and E. Budak, "Analytical prediction of stability lobes in milling," *CIRP Ann.*, vol. 44, no. 1, pp. 357–362, 1995.
- [25] G. E. Hinton and S. T. Roweis, "Stochastic neighbor embedding," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2003, pp. 857–864.
- [26] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
- [27] L. van der Maaten, "Accelerating t-SNE using tree-based algorithms," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3221–3245, 2014.
- [28] L. Van Der Maaten, "Barnes-Hut-SNE," 2013, *arXiv:1301.3342*.
- [29] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, 2008.
- [30] Y. Ding *et al.*, "Yinyang k-means: A drop-in replacement of the classic k-means with consistent speedup," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 579–587.
- [31] L. Yuan, Z. Pan, D. Ding, S. Sun, and W. Li, "A review on chatter in robotic machining process regarding both regenerative and mode coupling mechanism," *IEEE/ASME Trans. Mechatron.*, vol. 23, no. 5, pp. 2240–2251, Oct. 2018.
- [32] J. M. J. Murre *et al.*, "Replication and analysis of Ebbinghaus' forgetting curve," *PLoS One*, vol. 10, no. 7, 2015, Art no. e0120644.
- [33] T. M. Cover, and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.
- [34] I. Wald and V. Havran, "On building fast kd-trees for ray tracing, and on doing that in  $O(N \log N)$ ," in *Proc. IEEE Symp. Interact. Ray Tracing*, 2006, pp. 61–69.
- [35] R. P. H. Faassen *et al.*, "Prediction of regenerative chatter by modelling and analysis of high-speed milling," *Int. J. Mach. Tools Manuf.*, vol. 43, no. 14, pp. 1437–1446, 2003.





**Fei Shi** was born in Hebei, China. He received the B.Sc. degree in mechanical engineering in 2015 from Xi'an Jiaotong University, Xi'an, China, where he is currently working toward the Ph.D. degree in mechanical engineering.

He is currently a Visiting Scholar with the University of California at Santa Barbara, Santa Barbara, CA, USA. His current research interests include vibration control, intelligent fault diagnosis, and deep learning.



**Xingwu Zhang** received the Ph.D. degree in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 2012.

He is currently an Associate Professor of Mechanical Engineering with Xi'an Jiaotong University. He is the leader of more than 4 projects, including National Natural Science Foundation of China, etc. He holds more than 30 patents and authored or coauthored more than 50 academic papers. His research interests include dynamic system modeling and analysis, structural health monitoring, and active vibration control.

Dr. Zhang is the member of IEEE Instrument and Measurement, and the Technical Committee Member of IEEE Instruments and Measurement Society TC-7 Signals and Systems in Measurement.



**Hongrui Cao** received the B.Sc. degree from the Harbin Institute of Technology, Harbin, China, in 2004, and the Ph.D. degree from Xi'an Jiaotong University, Xi'an, China, in 2010, both in mechanical engineering. From 2008 to 2010, he was a Visiting Ph.D. Student with The University of British Columbia, Vancouver, BC, Canada.

He is currently a Professor of Mechanical Engineering with Xi'an Jiaotong University. His current research interests include dynamics of

spindle bearing systems, cutting process monitoring, and fault diagnosis of rotating machinery.



**Xuefeng Chen** (M'12) received the Ph.D. degree in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 2004.

He is currently a Professor of Mechanical Engineering with Xi'an Jiaotong University. His current research interests include finite-element method, mechanical system and signal processing, diagnosis and prognosis for complicated industrial systems, smart structures, aero-engine fault diagnosis, and wind turbine system monitoring.

Dr. Chen is the Chapter Chairman of the IEEE Xi'an and Chengdu Joint Section Instrumentation and Measurement Society. He was a recipient of the National Excellent Doctoral Dissertation of China in 2007, the Second Award of Technology Invention of China in 2009, the National Science Fund for Distinguished Young Scholars in 2012, and a Chief Scientist of the National Key Basic Research Program of China (973 Program) in 2015.