

Anomaly Detection via Unsupervised Learning for Tool Breakage Monitoring

Chengming Shi, Bo Luo, Hongqi Li, Bin Li, Xinyong Mao, and Fangyu Peng

Abstract—Machining and manufacturing of mechanical equipment is developing towards the direction of high speed, precision and efficiency. The tool health condition can be reflected by the massive data which can promote tool condition monitoring into the big data field, so it is a new challenge for the field that mining the characteristics which can describes tool condition. As an unsupervised hard clustering method, the advantage of K-means clustering is to mine the information from the massive data sets and clustering it efficiently and rapidly. Meanwhile, with the latest achievement in the field of machine learning, we can combine deep learning with the K-means clustering and propose a method of monitoring the health condition of the tool with unsupervised learning. The method that carries out the unsupervised pre-training of the tool vibration signal to get rid of the disadvantage that we must rely on the artificial diagnosis experience has the advantage in adaptively extracting the fault features from the tool signals. Through the design experiments and results show that the method can realize the adaptively unsupervised extraction of tool fault characteristics and the accurately identification of tool condition under the condition of large sample and multi-tool condition.

Index Terms—K-means clustering, massive data, tool breakage, unsupervised learning.

I. INTRODUCTION

The study on monitoring tool conditions has been more and more popular and difficult, as one of main failure forms of cutting tools, tool breakage will seriously affect the quality of products, production costs and process efficiency, and the real-time monitoring of which is the key technology to ensure the stable operation of machining system. Hence the monitoring of tool breakage and fault diagnosis has been of great significance in practical application. The traditional method rely too much on artificial diagnosis experience and artificial selection, which is time-consuming, overworked and poor generalization performance. Therefore, the method of tool breakage monitoring based on *unsupervised learning* and

Manuscript received August 30, 2016. This work was supported in part by the U.S. Department of Commerce under Grant BS123456 (sponsor and financial support acknowledgment goes here).

Chengming Shi and Bo Luo are with School of Mechanical Science and Engineering, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan, Hubei Province, PR China (e-mail: 18771978627@163.com, hglobo@163.com).

Hongqi Liu, Xinyong Mao, and Fangyu Peng are with National NC System Engineering Research Center, School of Mechanical Science and Engineering, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan, Hubei Province 430074, PR China.

Bin Li is with State Key Laboratory of Digital Manufacturing Equipment and Technology, School of Mechanical Science and Engineering, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan, Hubei Province 430074, PR China.

unsupervised clustering is proposed in this paper and compared with the traditional method of signal processing and artificial characteristic extraction. Above all, the method proposed can satisfy the requirement of intelligent identification and monitoring based on tool massive data.

II. THE METHOD OF MONITORING TOOL CONDITIONS BASED ON SAE_s AND K-MEANS CLUSTERING

In the field of tool condition monitoring, because of the complex and alternating signal, a conventional single hidden layer neural network algorithm [1] will be applied to mining fault information and recognize tool condition, and the accuracy is far from enough. And in the training of the neural network with deep structure, deep learning has greatly improved the capable of extracting feature and recognizing fault through multi-layer greedy training algorithm. Firstly, deep learning is pre-training for DNN [2] in the way of unsupervised learning that DNN can reduce the data dimension and mining the fault feature of the signal effectively to improve the ability of monitoring and diagnosing. And then the cluster analysis of tool breakage fault signal will be accomplished through unsupervised clustering algorithm.

A. The Principle of Sparse Automatic Coding Network

The sparse automatic encoder (SAE_s) [3] is based on the alteration of the automatic encoder (AE) [4]. AE is mainly divided into two processes: coding (encoder) and decoding (decoder). The coding part is to reduce the input data which is high dimensional vectors into the coding vectors which is the low dimensional, and the decoding part is transform the coding vectors from a low dimensional into a high dimensional vectors which is output goal. The purpose of AE is that the input values should be equal as much as possible to the corresponding output values, that is $\left(\hat{x}_i^n = x_i^n \right)$.

Suppose that an unlabeled signal sample set is given, among that $x^j = (x_1^j, x_2^j, \dots, x_m^j)$. Through the nonlinear encoding function, the encoding part transform each sample x^j of the sample set into the coding vectors $h^j = (h_1^j, h_2^j, \dots, h_l^j)$ that is low dimensional, as shown in (1).

$$h_k^j = f(x_i^j) = N_f \left(\sum_{i=1}^m W_{ki} x_i^j + b_k \right) \quad (k=1, 2, \dots, l) \quad (1)$$

Thereinto, n names the sample numbers of the input

sample set, m names the dimensional that is the data of the input sample set, l names the dimensional of the coding vectors, weight matrix W and bias matrix b is the coding network parameters.

And the output vector is a reconstruction of the input samples, so \hat{x}^j is known as the reconstruction vector.

$$\hat{x}^j = g(h_k^j) = N_g \left(\sum_{k=1}^l W_{ik} h_k^j + b_i \right) \quad (i=1,2,\dots,m) \quad (2)$$

Thereinto, weight matrix W and bias matrix b is the decoding network parameters.

After completing the encoding and decoding process, we will obtain an output goal vectors which can reconstruct the original input data. There will have a difference between output and input, that is, the reconstruction error [5]

$$L(x^j, \hat{x}^j) = \left\| x^j - \hat{x}^j \right\|^2 \quad (3)$$

In this way, the loss function [6] of the sparse automatic encoding is shown in (4).

$$J_{SAE}(W, b) = \frac{1}{n} \sum_{j=1}^n L(x^j, g(f(x^j))) + \frac{\lambda}{2} \sum_{k=1}^l \sum_{i=1}^m (W_{ki})^2 + \beta \sum_{k=1}^l KL(\rho \| \rho_k) \quad (4)$$

Thereinto, β names the weight values and the penalty coefficient of the penalty item.

By calculating the loss function, back-propagating the reconstruction error, updating the weight values, applying the gradient descent method and making sure that the encoding vectors can reconstruct the original input vectors with the least difference. Finally, take the encoding vectors as the characterization of the high-layer feature of the original input.

B. The pre-Training Phase of the Depth of Sparse Automatic Encoding Network

Sparse automatic encoding network (SAE_s) [7] are composed of the input and hidden layer of various sparse automatic encoders. The core idea of the pre-training is of extracting the depth-layer feature of the input vectors by using unsupervised learning. The original known input layer is calculated by the coding function N_f of SAE₁ to get the hidden layer h^1 of SAE₁. And there have:

$$h^{j(1)} = N_f \left(\sum_{i=1}^m W^{(1)} x_i^j + b^{(1)} \right) \quad (5)$$

Then:

$$h^{j(2)} = N_f \left(\sum_{k=1}^l W^{(2)} h_k^{j(1)} + b^{(2)} \right) \quad (6)$$

And so on, finally the abstract coding vector $h^{j(M-1)}$ ($j=1,2,3,\dots,n$) with high level features is obtained, the specific implementation process is shown in Fig. 1.

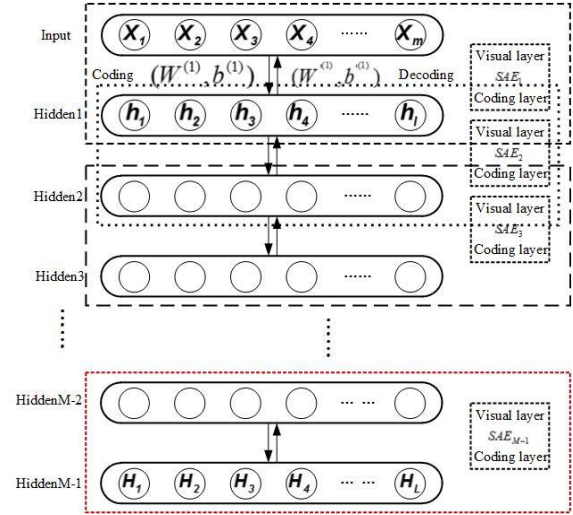


Fig. 1. Model structure of SAEs based on unsupervised learning.

C. The Theory of K-means Clustering Method

The clustering algorithm is an unsupervised clustering method, which can get together with the most similar samples of all the objects without having to know in advance which type samples belong to when in the cluster. We adopt K-means clustering algorithm [8], which will divide the training data into k classes in the case of being given the cluster class k in advance.

Suppose that the original input data set is $K = \{h^{j(M-1)}\}$ ($i=1,2,3,\dots,n$), and $h^{j(M-1)} = (h_1^{j(M-1)}, \dots, h_{l(M-1)}^{j(M-1)})$. The purpose of clustering is to class the sets K into k classes. And then the sets K are transformed into $Re = \{re_1, \dots, re_k\}$.

The main steps of K-means clustering algorithm are as follows:

a. Select k data from the data sets K randomly and take them as the centroid of k clusters respectively. And there have $G = \{G_1, G_2, \dots, G_k\}$.

b. Calculate the distance $d = \{d_1, d_2, \dots, d_k\}$ between each vector $h^{j(M-1)}$ of K and k centroids separately and the minimum value $h^j := \arg \min_i d_i$ of d , $h^{j(M-1)}$ is marked as the i -th class and stored in the cluster set re_i . After calculating all data of K , and then the clustering set $Re = \{re_1, re_2, \dots, re_k\}$ will be obtained.

c. Calculate the average value of each dimension of all elements of each class in the set Re :

$$G_i = \frac{\sum_{j=1}^n (h^j = i) \cdot h^{j(M-1)}}{\sum_{j=1}^n (h^j = i)} \quad (7)$$

And take G_i as the new centroid, meanwhile update all centroids $G = \{G_1, G_2, \dots, G_k\}$.

d. Repeat Step b and c, until the centroid of the cluster changes very little or no longer.

e. Export the clustering results.

So the most important of K-means clustering model is to minimize (8):

$$u = \sum_{j=1}^n \sum_{h^{(M-1)} \in Re_i} \left(h^{(M-1)} - G_i \right)^2 \quad (8)$$

Aims to avoid the situation that u value fall into the local optimum because of the improper selection of the initial centroids, different initial values can be selected for multiple K-means clustering, which can ensure that the final output Re and G is the global optimal values.

D. The Process of Monitoring Tool Breakage Based on SAEs and K-means Clustering

In this paper, we combined the characteristics of massive data of milling-tool with the advantage of the unsupervised deep learning and clustering. Unsupervised learning can complete the high-level features extraction of the fault signals of milling-tool data and unsupervised clustering is done to classify the fault signals based on the high-level features. The approach overcome the shortcomings of the generalization ability of the traditional methods being limited to extract low-level features to characterize failure feature.

III. THE DESIGN OF TOOL BREAKAGE EXPERIMENT AND EXPERIMENTAL SYSTEM

A. The Introduce of Experimental System and Experimental Design

The milling experiments are carried out on a MAG XS 321 horizontal machining center. Use current signal sensor produced by Honeywell International to collect data and turn number is 2000. In order to monitor the condition of the milling tool breakage, the sensors are combined with the spindle of the machine tool, sampling frequency is 2KHZ, data is collected and stored by Labview platform and its format is saved as binary. Finally the signal is sent into the Matlab for processing.

We have carried out the experiment on the test system to complete cylinder milling experiment, the experimental conditions are shown in Table I.

TABLE I: EXPERIMENTAL DESIGN OF TOOL BREAKAGE

Tool models	T22003
Tool diameter (m)	0.099
tooth number	13
Spindle speed (m/s)	0.00965
Tool life expectancy (s)	1.836×10^6
Work-piece material	Cast iron
Milling width (m)	Change over time

As shown in Fig. 2, the cutting tool will be severely deteriorated from normal to damage stage. In the process of analyzing the data, we did randomly select 400 samples as the input of SAEs network to complete the pre-training phase to stabilize the reconstruction error of the signal. Then, randomly select five different groups from the sample set into the SAEs network for K-means clustering and one of the group is consisted of 150 normal signal samples and 50

damage signal samples. Finally to verify the recognition effectiveness and ability of the proposed method. Among them, the number of nodes from the original input layer to the output coding vector layer is 3500, 1000, 150 and 15.

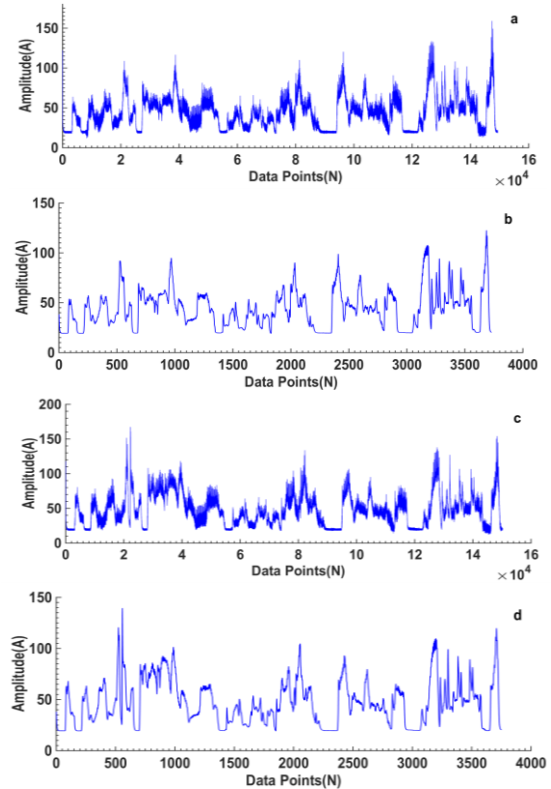


Fig. 2. The original signal and the signal after smoothing of three kind of tools: a. Original normal signal; b. Normal signal after smoothing; c. Original breakage signal; d. Breakage signal after smoothing.

IV. IDENTIFICATION AND VERIFICATION OF TOOL BREAKAGE

TABLE II: THE COMPARISON OF THE IDENTIFICATION ACCURACY RATE OF THE TWO METHODS BASED ON DIFFERENT NUMBER OF THE TRAINING ITERATIONS

method \ iterations	ITERATIONS				
	5	10	15	20	25
The proposed method	94%	94.5%	95%	98.5%	97.5%
SAE multi-layer network	25%	40%	75%	40%	75%
method \ iterations	ITERATIONS				
	30	35	40	45	50
The proposed method	97%	97.5%	97%	96.5%	98%
SAE multi-layer network	60%	25%	75%	60%	40%

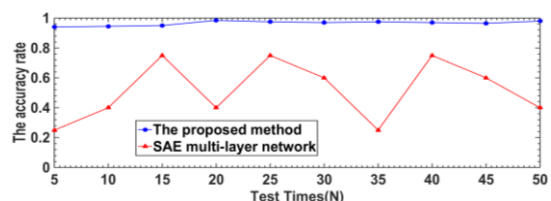


Fig. 3. The comparison of the accuracy rate curve of the two methods based on 10 trials.

According to the proposed method and SAE multi-layer network to analyze and compare the result of processing the data. There have 400 samples for pre-training based on the

proposed method and SAE multi-layer network, and the samples for testing are divided into five groups and each group have 200 samples. The experimental result is shown in Fig. 3 and Table II. We analyze the training results and the experimental effectiveness, the identification accuracy rate of SAE multi-layer network method in many experiments is low and with the change of the number of the iterations the accuracy of SAE multi-layer network is irregular and the floating range is relatively large; Contrary to the previous method, the identification accuracy rate of all the experiments based on the proposed method in this paper is very high, which is basically maintained above 94% and the highest value reached at 98.5%, and the floating range is nearly small.

TABLE III: THE COMPARISON OF THE EFFECTIVENESS OF THE TWO METHODS

The methods	Average accuracy	Accuracy variance	Range value	Maximum accuracy	Minimum accuracy
The proposed method	96.55%	0.00015%	4.5%	98.5%	94%
SAE multi-layer network	51.5%	3.6025%	50%	60%	40%

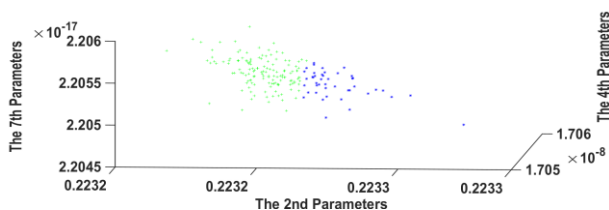


Fig. 4. Principal component of deep-layer feature based on the proposed method.

As to the other four test sample sets we can ensure the single variable principle through controlling the size and number of the samples and all network parameters, and then use the four sets to verify the identification accuracy of the proposed method. Clustering results are shown in Fig. 4 and Table IV. The above results show that the poor accuracy and stability have shown that the method based on SAE multi-layer network does not have a good ability to monitor tool breakage.

TABLE IV: THE IDENTIFICATION ACCURACY OF FIVE TEST SETS BASED ON THE PROPOSED METHOD

Iterations	10	20	30	40	50
The test sets					
1	94.5%	98.5%	97.5%	97%	98%
2	95%	96%	99.5%	97.5%	96.5%
3	96.5%	98.5%	99.5%	96%	99.5%
4	96.5%	96.5%	99.5%	99.5%	97.5%
5	96%	96.5%	99%	97%	99%

V. CONCLUSION

For the proposed method, its own network is able to extract the abstract deep-layer features characterizing tool breakage through unsupervised learning pre-training and K-means clustering can achieve the purpose of clustering the signal by using these features, and through different testing and training to make the accuracy rate of tool breakage identification reach above 99%. Finally, we use three different normal and broken sample sets to verify the effectiveness of the method. Experimental results show that the proposed method can be applied to mining adaptively tools broken features from the current signal, and it has achieved a high precision of tool condition diagnosis from the point of the classification effect

According to the experimental result of the two methods we carry out the quantitative analysis and comparison, the average identification accuracy rate of tool breakage based on SAE multi-layer network is 51.5%, the range value is 50% and the variance is 3.6025%; but the average identification accuracy rate of tool breakage based on the proposed method is 96.55%, the range value is 4.5% and the variance is 0.00015%, and the quantitative results of the two methods are shown in Table III. The results can fully indicates that the proposed method can accurately identify the normal and damaged condition of many kinds of tools in the situation that the interference of the machining environment and blank quality.

of normal and broken tool signal, which is further illustrated its efficiency and stability.

REFERENCES

- [1] D. Silver, A. Huang, C. J. Maddison *et al.*, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484-489, 2016.
- [2] Z. Y. Lu, D. Guo, A. B. Garakani *et al.*, "A comparison between deep neural nets and kernel acoustic models for speech recognition," *Respiratory Physiology and Neurobiology*, vol. 161, no. 2, pp. 214-217, 2016.
- [3] M. Ranzato, Y. L. Boureau, and Y. Lecun, "Sparse feature learning for deep belief networks," *Advances in Neural Information Processing Systems*, vol. 20, pp. 1185-1192, 2007.
- [4] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504-507, 2015.
- [5] Q. V. Le, J. Ngiam, A. Coates, A. Lahiri, B. Prochnow, and A. Y. Ng, "On optimization methods for deep learning," *International Migration*, vol. 7, no. 3-4, pp. 67-105, 2011.
- [6] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, no. 6, pp. 3371-3408, 2010.
- [7] M. Sun, D. Zhang, and J. Ren, "Brushstroke based sparse hybrid convolutional neural networks for author classification of Chinese ink-wash paintings," in *Proc. IEEE International Conference on Image Processing*, 2015, pp. 626-630.
- [8] M. U. S. Shameem, and R. Ferdous, "An efficient k-means algorithm integrated with Jaccard distance measure for document clustering," in *Proc. Asian Himalayas International Conference on Internet*, 2009, pp. 1-6.



Chengming Shi is a Ph.D. student in School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan City, Hubei Province, China. His main research field is in mechanical health condition monitoring, deep learning and data mining.



Bo Luo is a Doctor in School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan City, Hubei Province, China. His main research field is in mechanical health condition monitoring, dynamic analysis of machine tool and data mining.