A novel method for accurately monitoring and predicting tool wear under varying cutting conditions based on meta-learning

Yingguang Li[1*], Changqing Liu[1], Jiaqi Hua[1], James Gao[2], Paul Maropoulos[3](1)

[1]College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China
[2]School of Engineering, University of Greenwich, Chatham Maritime, Kent, ME4 4TB, UK
[3]Manufacturing Technology Centre, Ansty Park, Coventry, CV7 9JU, UK

Monitoring and predicting tool wear is an important issue in dynamic process control under changing conditions, especially for machining large-sized difficult-to-cut materials used in airplanes. Existing tool wear monitoring and prediction methods are mainly based on given cutting conditions over a period of time. This paper presents a novel method for accurately predicting tool wear under varying cutting conditions based on a proposed new meta-learning model which can be easily trained, updated and adapted to new machining tasks of different cutting conditions. Experiments proved a substantial improvement in the accuracy of predicting tool wear compared with existing deep learning methods.

1. Introduction

With the advances in aviation industry, more difficult-to-cut materials such as titanium alloys and nickel-based super-alloys are used in aircrafts. Cutting tool wear is a serious problem in manufacturing such materials, because it has significant impact on the surface texture and precision of machined parts [1]. In practice, tools are changed more often than necessary because of the unknown state of actual tool wear. As reported by [2] and [3], over 40 cutting tools were needed to complete the milling of a nickel-based super-alloy part while only 50%~80% of tool life was actually used. On the other hand, in some complex cutting conditions such as corner milling, tool wear is more severe and tool life is shorter than expected, which may affect machining quality or even cause part failure. Therefore, accurately monitoring and predicting tool wear stage (initial, gradual or failure wear stage) remaining tool life in real-life complex and changing cutting conditions is important to ensure machining quality and significantly increase the usage of tool life (thus reducing costs).

Previous researchers using analytical and numerical modeling methods to predict tool wear made assumptions and approximations to simplify the complexity of machining process [4], which had limitations and inaccuracies. To improve the above methods, researchers attempted to make use of real-time monitoring data, such as sensor signals of cutting force, power and current that were closely related to tool wear, and developed data-driven methods for predicting tool wear [5]. The method reported in [6] could predict the stage of tool wear under stable cutting conditions by acquiring real-time monitoring data and extracting cutting force coefficients which were not sensitive to cutting conditions. However, the method has limitations in accurately predicting actual tool wear under changing cutting conditions.

Some researchers attempted to establish the relationships between sensor signals and tool wear in their methods based on machine learning, such as artificial neural network and support vector machine [7]. However, these methods had limitations in representing complex relationships, and can only predict the stage of tool wear (not the actual tool wear). To overcome the above limitations, deep learning was considered to establish the complex relationships between sensor signals and actual tool wear by its multi-layer representation [8, 9]. However, the main challenge in their methods is that a large number of samples of monitoring signals and corresponding tool wear are required to train the deep learning model for tool wear prediction under each given cutting condition over a period of time. In practice, cutting condition changes when any of the following factors change: workpiece (material, geometry, technical property), cutting tool (type, material, geometry and cutting edge), and cutting parameters (feed per tooth, spindle speed, cutting depth and width). Therefore, it is impossible to collect enough samples of data to train the deep leaning model corresponding to each of the almost unlimited number of cutting conditions. This paper presents a novel meta-learning (or learning to learn) method, which would significantly improve previously methods in predicting actual tool wear under varying cutting conditions.

2. The proposed tool wear prediction method based on meta-learning

2.1 The overall idea of the proposed meta-learning method

The aim of the meta-learning method is to train a model on a variety of learning tasks, so that it can solve new learning tasks using only a small number of training samples, i.e., it has the ability of learning to learn [10, 11]. The overall idea of the proposed meta-learning method is illustrated in Fig.1. A base-model is a fully connected neural network (FCNN) to predict tool wear under a specific cutting condition. Model-agnostic meta-learning (MAML) is used to update the parameters of the fully connected neural network to adapt different cutting conditions.

MAML can learn how the parameters are updated in different cutting conditions, and then quickly adapt the parameters of FCNN in a new cutting condition with very small number of training samples. For a given tool wear monitoring and prediction model, meta-learning method can learn how the model learns the correlation with tool wear and monitoring signals, which reflects the essential correlation of monitoring signals and tool wear. At the top of Fig.1, base-model t is the tool wear monitoring and predicting model under a specific cutting condition t. The gradient of loss function (∇L) is calculated to update the meta-parameter $\boldsymbol{\theta}$ during training the base-models.
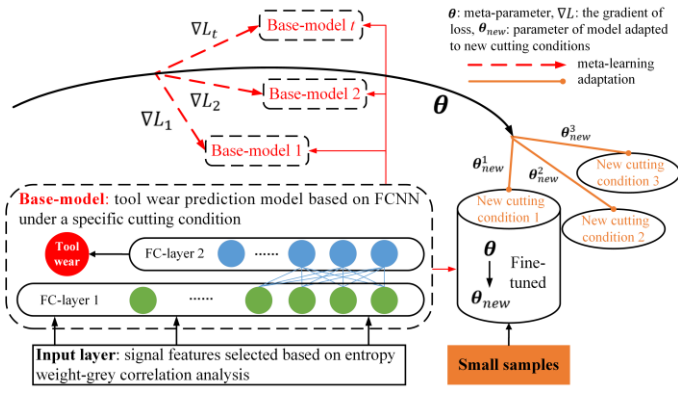
Fig. 1 Tool wear monitoring and prediction under varying cutting conditions based on meta-learning

After the meta-learning model is successfully trained, the model can be easily fine-tuned to obtain a θ_{new} with a very small number of samples or even just one sample so as to adapt to the new cutting condition. MAML is used as training approach of the meta-learning model [12], and a FCNN is used as the base-model of MAML, where the cutting tool wear rule can be implicitly learned. In order to further reduce the influence of cutting conditions and the reliability on the number of samples, the monitoring signals (cutting force, power and current) are preprocessed with an entropy weight-grey correlation analysis approach, where the signal features which are relatively more sensitive to tool wear while less sensitive to cutting conditions are selected as the input of the meta-learning model. Compared with existing signal processing methods [8, 9], the proposed method can significantly reduce the required input dimensional and sample numbers for meta-learning and improve the accuracy of tool wear prediction.

2.2 MAML modelling for tool wear prediction

Tool wear and cutting conditions have coupling effects on monitoring signals, resulting in the difficulty in predicting tool wear under varying cutting conditions. The mechanism of meta-learning is learning to learn, which can be quickly adapted to new machining tasks after being trained by different machining tasks. The adaption and learning ability are similar to human intelligence. The main advantage of meta-learning is to train the parameters of a meta-learning model during a meta-learning phase on a set of machining tasks so that a small number of gradient steps, or even one single gradient step, can produce good results for a new task. It can be viewed as establishing a general representation broadly adaptable to different tasks.

A base-model is built and denoted as f for convenience, which maps the signal feature input x to the tool wear output a. Each task T_i corresponds to a machining task under a specific cutting condition. There is no assumption on the form of the meta-learning model in MAML. The meta-learning model is parametrized by the meta-parameter θ. A different base-model parameter θ_i' under different cutting condition is calculated by gradient descent:

$$\theta_i' = \theta - \alpha\nabla_{\theta}L_{T_i}(f_{\theta}) \qquad (1)$$

where the learning rate α is fixed as a hyper-parameter, L is the loss function of mean-squared error (MSE). θ is trained by optimizing for the performance of $f_{\theta_i'}$. The meta-objective is:

$$\min_{\theta} \sum_{T_i \sim p(T)} L_{T_i}\left(f_{\theta_i'}\right) = \sum_{T_i \sim p(T)} L_{T_i}\left(f_{\theta - \alpha\nabla_{\theta}L_{T_i}(f_{\theta})}\right) \qquad (2)$$

The meta-optimization is performed over θ, whereas the objective is computed using updated θ_i'. In fact, the meta-learning model aims to optimize the base-model parameters such that a small number of gradient steps can produce accurate tool wear prediction under a new cutting condition. θ is updated by stochastic gradient descent (SGD) as:

$$\theta \leftarrow \theta - \beta\nabla_{\theta} \sum_{T_i \sim p(T)} L_{T_i}\left(f_{\theta_i'}\right) \qquad (3)$$

where β is the meta-learning rate. This step is essential for meta-learning, where the meta-parameter θ is optimized by learning the optimization process of the base-model parameter θ_i'. When it is adapted to a new task T_{new} with a new cutting condition, θ is fine-tuned to θ_{new}, which is updated as:

$$\theta_{new} = \theta - \alpha\nabla_{\theta}L_{T_{new}}(f_{\theta}) \qquad (4)$$

The samples for fine-tuning can be obtained in the beginning of a new machining task (under the new cutting condition), which are the tool wear values referred as label values. For the base-model in Fig. 1, a FCNN consisting of 4 layers is used to learn the abstract representation of input signal features. The input layer is the signal features selected based on entropy weight-grey correlation analysis and then two fully connected layers (20×20) are established. The output layer is a single unit to predict the current tool wear. Each base-model is trained under a specific cutting condition. The set of parameters of the base-model i can be represented by the base-parameter θ_i'.

3 Signal feature extraction and selection

3.1 Signal feature extraction based on statistical analysis and deep belief network

Tool wear can be reflected in the monitoring signals during machining, such as cutting force, power and current. Acquired signals need to be processed and analyzed to obtain features sensitive to tool wear. Features extracted by statistical analysis are artificially designed so that expert knowledge can be included. Deep learning is an automatic feature extraction method which is independent of manual experience and could extract high-level abstract features containing implicit information [13]. Statistical analysis methods analyze signals in terms of time domain, frequency domain and time-frequency domain. Features extracted in the time domain include mean, mean square, variance, peak coefficient and kurtosis coefficient. Features extracted in the frequency domain include mean square frequency, gravity center frequency and Empirical Mode Decomposition (EMD). Features extracted in the time-frequency domain include wavelet analysis and Fourier analysis.

Deep belief network (DBN) is capable of revealing deep patterns embedded in raw monitoring signals using densely connected restricted Boltzmann machines (RBMs) [14]. After training a RBM, the visible units, i.e., the original signals collected by sensors in a period of time, can be accurately reconstructed by the hidden units, which means that the original signals are precisely represented in a different dimension space without any information loss. The output of DBN can be taken as the features of the original signals extracted automatically. In order to reduce the computational cost and the number of required samples (monitoring data) for the training of DBN, its variant, i.e., deep feature selection (Deep-FS) [15] has been adopted in this research. Deep-FS aims to find a set of inputs with useful information, whilst the sample data without useful information about the input data are removed by the generative property of RBM. The number of finally selected signal points are reduced which can reduce the complexity of the network.

3.2 Signal feature selection based on entropy weight-grey correlation analysis method

In order to further improve the ability of the tool wear monitoring and prediction model under varying cutting conditions, the signal features which are more sensitive to tool wear and less sensitive to cutting conditions are selected. Entropy weight-grey correlation analysis is used to evaluate the sensitivity of different features with tool wear and cutting condition comprehensively. Correlation analysis is used to determine the weighing of the

relevance criterion of signal features with respect to tool wear and cutting conditions, represented as :

$$\rho_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2(y_i-\bar{y})^2}} \quad (5)$$

where x_i stands for the extracted feature. y_i stands for the corresponding tool wear value or cutting condition factors. $\bar{x}$ and $\bar{y}$ are the mean values, respectively. $|\rho_{xy}|$ represents the indicator of the correlation between the feature and impacting factor (tool wear and cutting conditions). The indicators are represented as:

$$A = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{m1} & \rho_{m2} & \cdots & \rho_{mn} \end{bmatrix} \quad (6)$$

where ρ_{mn} represents the correlation between feature Fea_m and influencing factor IF_n. The nth column represents correlation between each feature and tool wear. The first to (n-1)th columns represent correlation between each feature and each cutting condition. A reference indicator vector $\boldsymbol{\rho_0} = [0,0,0,\dots,1]$ is constructed. The ideal feature is when ρ_{in} is closer to 1 and $\rho_{i(1\sim n-1)}$ are closer to 0. The correlation coefficient, which represents the distance of extracted features and the ideal feature, is obtained as:

$$\zeta_i(j) = \frac{\min_j|\rho_{0j}-\rho_{ij}|+0.5\max_j|\rho_{0j}-\rho_{ij}|}{|\rho_{0j}-\rho_{ij}|+0.5\max_j|\rho_{0j}-\rho_{ij}|} \quad (7)$$

The correlation sequence r_i is calculated to rank the features by the rule of comprehensive indicators that are more sensitive to tool wear while less sensitive to cutting conditions, as below:

$$r_i = \frac{1}{n}\sum_{j=1}^{n} W_j \zeta_i(j) \quad (8)$$

where W_j is identified by entropy weight method. Generally, if the information entropy of a column is larger, it indicates more uncertainty in the column, which can be understood that the impacting factor (e.g., cutting depth) in the column has smaller impact on signal features, thus smaller weight should be given for the column. The information entropy of signal feature Fea_j is expressed as:

$$E_j = \frac{1}{\ln m}\sum_{i=1}^{m} \frac{\rho_{ij}}{\sum_{i=1}^{m}\rho_{ij}}\ln\frac{\rho_{ij}}{\sum_{i=1}^{m}\rho_{ij}} \quad (9)$$

The weight of feature Fea_j is expressed as:

$$W_j = \frac{1-E_j}{\sum_j(1-E_j)} \quad (10)$$

Consequently, the comprehensive indicator of each feature is obtained, and the top k features are selected, where k can be determined by experiments. Using the above method, signal features which are more sensitive to tool wear while less sensitive to cutting conditions can be abstracted and used as the input to the meta-learning model.

4 Meta-learning model training and validation

In order to train the proposed meta-learning model and validate its efficiency, 12 groups of machining experiments were designed with different cutting conditions, of which 8 groups were used for training base-models, and 4 groups were used for testing the adaptation to different cutting conditions. Part material used was titanium alloy and cutting tools used were high-speed steel end-milling tool with radius 6mm and 8mm. The cutting parameters change in the range of: spindle speed (n: 400r/min-600r/min), feed per tooth (fz: 0.06mm/r-0.1 mm/r), and cutting depth (ap: 1mm-3mm). The experiments were carried out on a DMG 80P machine tool, as shown in Fig.2 (a). The tool wear of each cutting tool ranged from 0.10mm to 0.40mm for each experiment and signal data were collected during the whole machining process (cutting force collected by Spike^T, current and power collected by OPCUA). The tool wear value labels were measured using a XK-T600V^T microscope (accuracy: 0.01mm), as shown in Fig.2 (b).

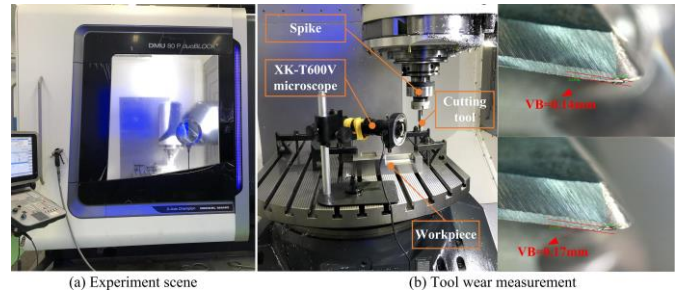(a) Experiment scene   (b) Tool wear measurement

Fig. 2 Experiment scene and tool wear measurement

During machining, the cutting tool wear was measured by fixing the microscope on the machine tool table after a certain cutting distance (flank wear of tool blade, the value of VB, and the largest wear value of the blades were taken, as shown in the right side of Fig. 2(b)). Each measure process took about ten minutes, and 20 days in total were used for the whole experiments. In order to obtain more labeled samples, each two near labeled samples of a cutting tool were discretized based on Fourier fitting algorithm, and 2916 labeled samples in total were obtained.

After the data were collected, feature abstraction was performed. The ranks of monitoring signal features were calculated using entropy weight-grey correlation analysis method. Fig. 3 (a) shows the weights of evaluated factors (fz, n, ap, VB) which were determined according to the entropy weight method. Fig. 3(b) shows the comprehensive evaluation of the signal features by grey correlation analysis with the evaluated weights of Fig. 3(a). The top 16 features (e.g., power spectrum skewness of current signal) with higher correlation coefficient value of r_i of 176 initial features were selected as the input of meta-learning, which were relatively more sensitive to tool wear and less sensitive to cutting conditions.

When training the meta-learning model, the learning rate α was fixed as 0.001, and Adam was used as the meta-optimizer [16]. In the proposed method, the prediction error on test sets with different number (1, 3, and 5) of samples for adaptation to new cutting conditions were compared. In Table 1, it can be seen that, the prediction error decreased as the sample number increased. Actually, the prediction accuracy was good enough for engineering applications, even the model was updated by 1 sample. The tool wear monitoring performance of the proposed method was also compared with other typical deep learning methods such as (i) Long Short Term Memory (LSTM) model using unselected signal features without meta-learning, and (ii) LSTM model using selected signal features in this research but without meta-learning. The results presented in Fig.4 show that the prediction values of methods (i) and (ii) have more deviations from the real measured tool wear values than the proposed method of this paper.

It can be seen that the proposed method, denoted by (iii) in Table 1, is obviously better than the two typical deep learning methods (i) and (ii). The prediction error of LSTM model without selected features or meta-learning was above 0.07mm or 0.06mm (more than 3 times of the proposed method). And the prediction accuracy of method (ii) with signal feature selection is 0.01-0.02 higher than that of method (i), which shows the effectiveness of the signal selection approach.
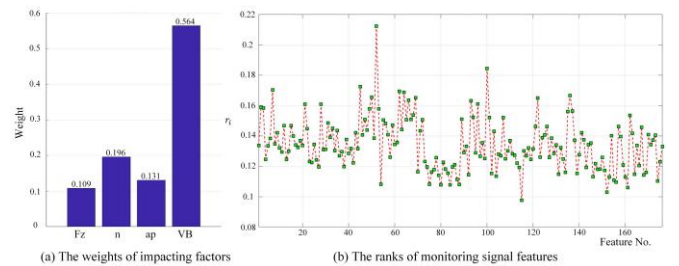

(a) The weights of impacting factors   (b) The ranks of monitoring signal features

Fig. 3 The result of entropy weight-grey correlation analysis
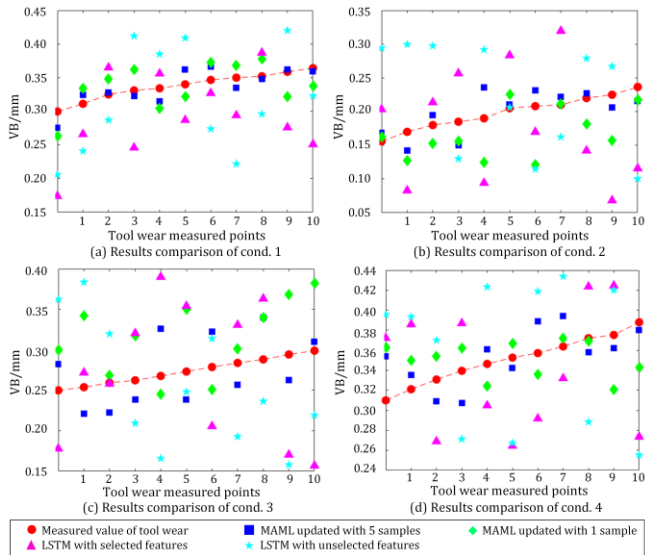
Fig. 4 Prediction results comparison

Table 1 Comparison of tool wear prediction accuracy

Method		Prediction errors (mean absolute error)/mm			
		Cond. 1*	Cond. 2*	Cond. 3*	Cond. 4*
(i)		0.07	0.09	0.08	0.08
(ii)		0.06	0.07	0.07	0.06
(iii)	1 sample	0.03	0.03	0.05	0.03
	3 samples	0.02	0.03	0.04	0.02
	5 samples	0.01	0.02	0.03	0.02

*Cond. 1: fz=0.1mm/r, n=400r/min, ap=3mm, tool radius (R) = 6mm;
*Cond. 2: fz=0.09mm/r, n=450r/min, ap=2mm, R= 6mm;
*Cond. 3: continuously changing cutting conditions with lines and corners, R= 6mm;
*Cond. 4: fz=0.1mm/r, n=400r/min, ap=3mm, R= 8mm (A different cutting tool).

In the experiment, 0.1mm tool wear took about 10 minutes of machining time, which means that the prediction error had a serious impact on the prediction of the remaining life of the tool. Both the meta-learning and signal feature selection methods improved the accuracy of tool wear prediction under different cutting parameters (including continuously changing cutting conditions) and cutting tools. The proposed method also had greater adaptation to new cutting conditions and achieved good results with a small number of samples or even just one sample.

Fig.5 (a) is the training process of MAML, the algorithm is converging within 2.5e-4 after about 2500 epochs. In order to illustrate why meta-learning has a great performance on tool wear prediction, the weights of FC-layers of base-models' parameters and meta-parameter have been studied using a t-distributed stochastic neighbor embedding (t-SNE) analysis method [17], which embeds a high-dimensional dataset into a low-dimensional space for visualization. As shown in Fig.5 (b), the distribution of weights in meta-parameter has high similarity with those in base-models' parameters under different cutting conditions. It demonstrates that the essential correlation of monitoring signals and tool wear are learned by meta-learning and the meta-parameter could be fine-tuned by a small number of samples to adapt to new cutting conditions because of the similarity.
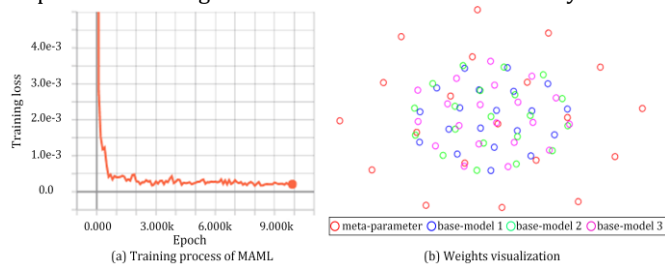

Fig. 5 Training process of MAML and weights visualization

5 Conclusions and further work

Compared with existing methods which are based on given cutting conditions over a period of time, the proposed meta-learning based method has significant advantages in both efficiency and accuracy for predicting tool wear in varying cutting conditions. When the meta-learning model is successfully trained, the model can be easily adapted to a new cutting condition with a very small number of samples or even one sample data under the new cutting condition. Special signal features are extracted and selected as the input of meta-learning model, which are relatively more sensitive to tool wear while less sensitive to cutting conditions. Experiments carried out in this research proved that the error of tool wear prediction can be reduced by over 3 times compared with existing typical deep learning methods reported by previous researchers.

The method provides an important contribution in the general area of data-driven intelligent manufacturing, particularly in monitoring and controlling machining processes under dynamic and changing cutting conditions. With the further use of the meta-learning model and learning process, the accuracy of tool wear monitoring and predicting could improve by self-evolution. The experiments were carried out using titanium alloys of aircraft structural parts. In future work, the meta-learning prediction model would be tested using different kinds of materials and different cutting tool types on different blades.

References

[1] Cuka B, Kim DW (2017) Fuzzy logic based tool condition monitoring for end-milling. *Robotics and Computer-Integrated Manufacturing* 47: 22-36.
[2] Zhu D, Zhang X, Ding H (2013) Tool wear characteristics in machining of nickel-based superalloys. *International Journal of Machine Tools & Manufacture* 64: 60-77.
[3] Tönshoff HK, Wulfsberg JP, Kals HJJ, König W, Van Luttervelt CA (1988) Developments and trends in monitoring and control of machining processes. *CIRP Annals - Manufacturing Technology* 37(2): 611-622.
[4] Rech J, Giovenco A, Courbon C, Cabanettes F (2018) Toward a new tribological approach to predict cutting tool wear. *CIRP Annals - Manufacturing Technology* 67(1): 65-68.
[5] Teti R, Jemielniak K, O'Donnell G, Dornfeld D (2010) Advanced monitoring of machining operations. *CIRP Annals - Manufacturing Technology* 59(2): 717-739.
[6] Nouri M, Fussell BK, Ziniti BL, Linder E (2015) Real-time tool wear monitoring in milling using a cutting condition independent method. *International Journal of Machine Tools & Manufacture* 89: 1-13.
[7] Liu C, Li Y, Zhou G, Shen W (2018) A sensor fusion and support vector machine based approach for recognition of complex machining conditions. *Journal of Intelligent Manufacturing* 29(8): 1739–1752.
[8] Shi C, George P, Luo B, Liu H, Li B, Lin X (2019) Using multiple feature spaces-based deep learning for tool condition monitoring in ultra-precision manufacturing. *IEEE Transactions on Industrial Electronics* 66(5): 3794 - 3803 .
[9] Zhao R, Yan R, Wang J, Mao K (2017) Learning to monitor machine health with convolutional bi-directional LSTM networks. *Sensors* 17(2): 273(1-18).
[10] Lemke C, Budka M, Gabrys B (2015) Metalearning: a survey of trends and technologies. *Artificial Intelligence Review* 44(1): 117–130.
[11] Wang JX, Kurth-Nelson Z, Kumaran D, Tirumala D, Soyer H, Leibo JZ, Hassabis D, Botvinick M (2018) Prefrontal cortex as a meta-reinforcement learning system. *Nature Neuroscience* 21: 860–868.
[12] Finn C, Abbeel P, Levine S (2017) Model-agnostic meta-learning for fast adaptation of deep networks. *Proceedings of the 34th International Conference on Machine Learning* (*ICML*). arXiv:1703.03400.
[13] Lecun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521: 436-444.
[14] Wang P, Gao RX, Yan R (2017) A deep learning-based approach to material removal rate prediction in polishing. *CIRP Annals - Manufacturing Technology* 66(1): 429-432.
[15] Taherkhani A, Cosma G, Mcginnity TM (2018) Deep-fs: a feature selection algorithm for deep boltzmann machines. *Neurocomputing* 322: 22-37 .
[16] Kingma D, Ba J (2015) Adam: a method for stochastic optimization. *The 3rd International Conference for Learning Representations (ICLR)*. arXiv:1412.6980.
[17] Maaten LVD, Hinton G (2008) Visualizing data using t-SNE. *Journal of Machine Learning Research* 9: 2579–2605.