Novel Selection Method of Drilling Condition Based on Data Mining of a Microdrill Catalog Database

Shunya Tanaka*, Toshiki Hirogaki, and Eiichi Aoyama

Department of Mechanical Engineering, Doshisha University, 1-3, Tatara Miyakodani, Kyotanabe, Kyoto, Japan
Email: tanashun0706@gmail.com (S.T.); thirogak@mail.doshisha.ac.jp (T.H.); eaoyama@mail.doshisha.ac.jp (E.A.)
*Corresponding author

Manuscript received April 10, 2023; revised May 24, 2023; accepted August 23, 2023 published March 7, 2024

***Abstract*—Printed circuit boards (PCBs) are circuits written on copper foil fixed to boards that are composed of electrically nonconductive glass fiber cloth and resin. Electrical products are becoming more miniaturized. As a result, drilling the PCBs has become increasingly difficult. Micro-drilling of PCBs is suitable for forming through holes. The drilling quality also affects the copper plating quality, which affects the reliability of the electrical connection. Thus, it is necessary to improve PCB drilling technology, develop drilling methods that increase productivity, and secure the reliability of the electrical connection. Many studies on the cutting force, temperature, and quality of drilled holes in PCB drilling have been conducted. Research on the cutting force in micro-drilling revealed that increasing the cutting force decreased the quality of the drilled holes and affected the breakage of microdrills. Currently, skilled engineers select the tools and drilling conditions, which is difficult for unskilled engineers. In micro-drilling, engineers must select drilling conditions such as the spindle speed and infeed rate. The system must support tool selection and drilling condition decisions based on open knowledge and data.**

***Keywords*—microdrill, drilling conditions, data mining, machine-learning, random forest, printed circuit boards**

I. Introduction

With the rapid development of science and technology in recent years, the performance of electronic devices has rapidly improved. Consequently, the holes in electronic circuit boards have become smaller and more sophisticated, thereby complicating the setting of hole processing conditions. Therefore, the drilling technology for Printed Circuit Boards (PCBs) must be improved, and a drilling method with high productivity has been devised [1]. However, it is difficult for unskilled engineers to set optimal drilling conditions, and the selection of tools and cutting conditions, which are important in the micro-drilling of PCBs, often depends on the knowledge and experience of skilled engineers, much of which is tacit knowledge. Therefore, it is necessary to utilize publicly available knowledge and data to determine the cutting conditions; however, there are few examples of research based on the formalization of tacit knowledge in micro-drilling.

This study focused on tool catalogs, which contain the knowledge and experience of skilled engineers working for tool manufacturers. Data mining methods were applied to tool catalog information (called "catalog mining") to construct a support system for setting cutting conditions and discovering knowledge useful for machining. In a previous study, a system was developed that could derive cutting conditions from tool information using machine-learning on an end mill tool catalog [2]. In the present study, the focus was on microdrill catalogs, and a machine-learning analysis was performed using tool catalogs for PCBs and metal plates. Moreover, an attempt was made to clarify the process of selecting cutting conditions by introducing a Partial Dependency Plot (PDP) to visualize the relationship between feature values and objective indicators.

II. Catalog-Mining Method

A. Data Mining and Its Basics

Data mining is a method for uncovering valuable information from large amounts of data accumulated in databases that contain useful patterns, rules, or noise. It is often used to discover and formalize valuable information that has gone unnoticed in background knowledge in unexplored fields with little prior research [3, 4]. This method, which consists of several statistical analysis methods, has been used in manufacturing, chemistry, and other fields, and its effectiveness has been demonstrated [5]. In catalog mining, tool catalogs already contain a large number of high-quality data based on the results of repeated experiments conducted by tool manufacturers; therefore, analysts do not need to spend time acquiring and selecting data. In addition, because much of the data in the catalogs is already displayed numerically, much of the data cleansing can be omitted.

B. Random Forest

Random forest is an ensemble learning method that uses multiple decision trees to form a forest for identification, etc. [6]. Fig. 1 shows a decision tree, which is a method for visualizing decision rules by performing if-then classification based on certain conditions. The group of data obtained through branching is called a "node," and the terminus is called a "leaf." The value of each leaf and node is the average value of the objective variable belonging to that node and leaf. Each decision tree is generated using the CART (classification and regression trees) algorithm, which is commonly used in machine-learning. In the case of regression binary trees, the binary tree is grown using the heterogeneity measure shown in Eq. (1) as the evaluation criterion, and branching is performed at the branching condition that maximizes the degree of improvement $\Delta R(t)$ shown in Eq. (2).

$$R(t) = \frac{1}{N}\sum_{n\in t}\{y_n - \bar{y}(t)\}^2 \tag{1}$$

$$\Delta R(t) = R(t) - R(t_L) - R(t_R) \tag{2}$$

DOI: 10.7763/IJMO.2024.V14.847

where t is an arbitrary node, N is the number of data points, y_n is the response, $\bar{y}(t)$ is the average value of the response at node t, t_L represents the node to the left after branching, and t_R represents the node to the right after branching. Although individual decision trees do not have high discriminative performance, a random forest can achieve high prediction performance by using multiple decision trees to compensate for each other's results. Fig. 2 shows a schematic diagram. As the figure shows, multiple decision trees are constructed by randomly selecting explanatory variables. Within each decision tree, the target objective variable is searched (the red-filled tree in Fig. 2), and the final output is the average of the outputs of each decision tree. The performance of the regression model was evaluated using the coefficient of determination R^2.

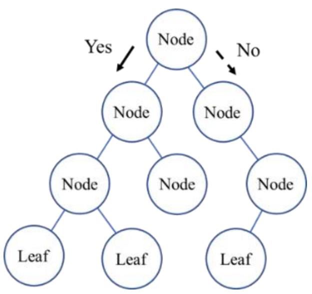


Fig. 1. Conceptual diagram of decision tree.

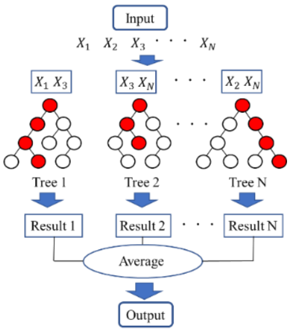


Fig. 2. Processes of random forest.

Random forest also makes it possible to quantify the importance of the explanatory variables. First, the prediction error *MSE* for each decision tree is obtained from Eq. (3). Next, only the variables whose importance is to be determined are randomly reordered, and the prediction error *MSE'* for each decision tree is obtained. The importance of the explanatory variables is then obtained using Eq. (4).

$$MSE = \frac{1}{N}\sum_{k=1}^{N}(f_k - y_k)^2 \tag{3}$$

$$lmp = \frac{1}{N_t}\sum_{i=1}^{N_t}(MSE_i - MSE_i') \tag{4}$$

where f_k is the kth training datum, y_k is the kth prediction datum, N is the number of data points, and N_t is the number of decision trees.

C. Partial Dependency Plot Method Theory

Machine-learning models, such as random forests, provide high prediction accuracy. However, they have an interpretability problem in that the relationship between inputs and outputs is not well understood. In this study, a PDP, which is a method for visualizing the relationship between a certain feature and a target indicator, is used. The relationship between the inputs and outputs is visualized by estimating the learned model $f(x)$ using Eq. (6).

$$\hat{f}_s(x_s) = E_C[\hat{f}(x_s, x_c)] = \int \hat{f}(x_s, x_c)p(x_c)dx_c \tag{5}$$

$$\bar{f}_s(x_s) = \frac{1}{N}\sum_{i}\hat{f}(x_s, x_c^{(i)}) \tag{6}$$

where x_s is the variable of interest, and x_c is the group of other variables.

III. DATA-SET

Four catalogs were used for catalog mining: PCB tool catalogs from Companies A and B and metal plate tool catalogs from Companies C and D. Table 1 shows the range of values for the variables listed in the PCB drill catalogs of Companies A and B. Table 2 shows the range of values for the variables listed in the metal plate drill catalogs of Companies C and D. Fig. 3 shows a magnified view of the drill shape and drill tip. The drill shape can vary depending on the tool manufacturer and work material. Fig. 4 shows a general shape.

Table 1. Variables of catalog (for PCBs)

Company	A	B
Data volume	2873	2150
Diameter *D* [mm]	0.1 - 6.5	0.105 - 6.5
Flute length *l* [mm]	1.3 - 12	1.5 - 12
Shank diameter D_s [mm]	3.175	3.175
Overall length *L* [mm]	38.1	38.1
Helix angle *α* [degree]	30	30 - 45
Point angle *β* [degree]	130 - 165	120 - 165
Stack height *Sh* [-]	1 - 10	1 - 10
Board thickness *Bt* [mm]	0.1 - 1.6	0.1 - 6.4
Spindle speed *S* [k rpm]	20 - 200	15 - 200
Infeed rate *F* [mm/min]	400 - 4400	500 - 2700
Cutting speed *V* [m/min]	37.7 - 408.4	41.2 – 306.3
Chip load *C* [μm/rev.]	5 - 106.5	5 - 63

Table 2. Variables of catalog (for metal plate)

Company	D	E
Data volume	214	1672
Diameter *D* [mm]	0.01 – 6	2 – 20
Flute length *l* [mm]	0.025 – 18	12 – 430
Shank diameter D_s [mm]	3 – 6	3 – 20
Overall length *L* [mm]	38 – 60	62 – 490
Helix angle *α* [degree]	20 – 30	25 – 30
Point angle *β* [degree]	120	140 – 160
Spindle speed *S* [k rpm]	1.6 – 40	0.7 – 12.7
Infeed rate *F* [mm/min]	1 – 1200	264 – 1280
Cutting speed *V* [m/min]	0.63 – 141.4	33 – 119.7
Chip load *C* [μm/rev.]	0.3 - 133	60 – 420

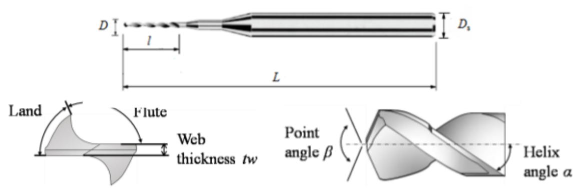


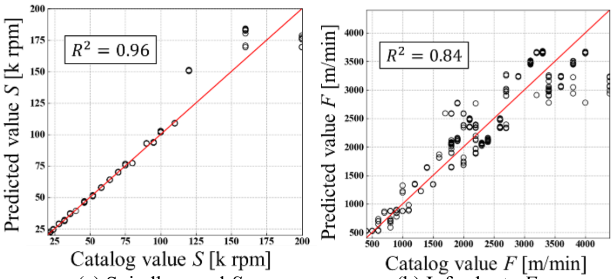
Fig. 3. Drill shape.

The tool catalogs contain information on drill geometry, such as shank diameter D_s [mm], overall length L [mm], diameter D [mm], flute length ℓ [mm], helix angle α [degree], and point angle β [degree], as well as data on workpiece materials, such as stack height *Sh* [-] and board thickness *Bt* [mm], in catalogs for tools for PCBs, and S35C and SS400 in catalogs for tools for metal plates. In addition, spindle speed *S* [rpm], spindle cutting speed *V* [m/min], spindle infeed rate *F* [mm/min], and chip load *C* [μm/rev] are listed as recommended cutting conditions for these materials.

IV. Experimental Results and Discussion

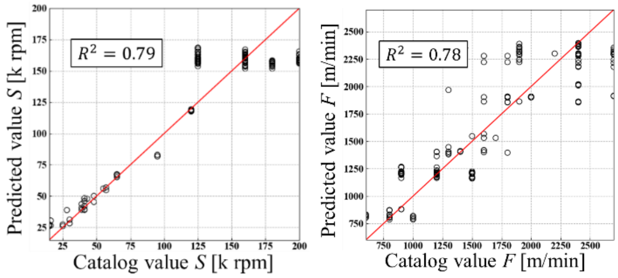
A. Prediction of Cutting Conditions by Random Forest

In addition to the drill shape variables (D, l, D_s, L, α, and β) and work material variables (*Sh* and *Bt*) listed in the PCB tool catalogs of Companies A and B shown in Table 1, D/D_s (the ratio of diameter to shank diameter), l/L (the ratio of flute length to overall length), and D/l (the ratio of tool diameter to blade length) were added as drill shape variables. Using 11 explanatory variables, the cutting conditions (spindle speed S and infeed rate F) were predicted using a random forest. Figs. 4 and 5 show the relationship between the catalog values of the cutting conditions and the predicted values for Companies A and B, respectively, and the value of the coefficient of determination R^2. Figs. 4 and 5 show that the coefficient of determination is high for Companies A and B, indicating that the cutting conditions (S and F) can be predicted.



(a) Spindle speed S (b) Infeed rate F

Fig. 4. Results of random forest (Co. A).



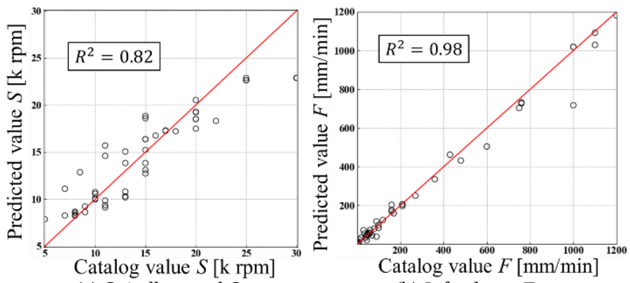
(a) Spindle speed S (b) Infeed rate F

Fig. 5. Results of random forest (Co. B).

Next, random forest was used to predict the cutting conditions (spindle speed S and infeed rate F) for the metal plate tool catalogs of Companies C and D, as listed in Table 2. The PCB tool catalogs contain work material variables, such as stack height *Sh* and board thickness *Bt*; however, these variables are not listed in the metal plate tool catalogs. Therefore, the focus was on the material properties of metals, and they were added as work material variables to the explanatory variables. The work material variables added in this report are Young's modulus E, thermal conductivity λ, tensile strength σ, and Vickers hardness H. Table 3 shows the types of work materials listed for Companies C and D and the values of the added work material variables, which are basically JIS steel and aluminum alloys except HAP7, NAK55 and NAK80, pre-hardened steels in Japanese steel companies. The cutting conditions (S and F) were predicted using 13 explanatory variables in the metal plate tool catalogs: drill shape variables (D, l, D_s, L, α, β, D/D_s, l/L, and D/l) and work material variables (E, λ, σ, and H) listed in Table 3. Figs. 6 and 7 show the relationship between the catalog and predicted values of the cutting conditions and the coefficient of determination R^2 for Companies C and D, respectively.

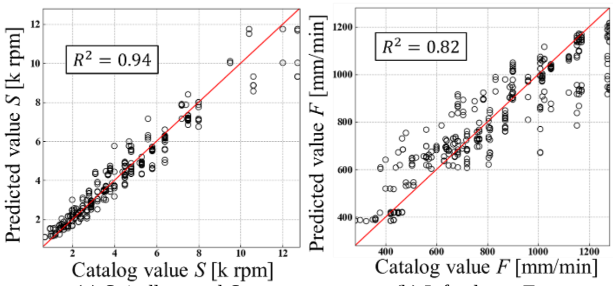
Table 3. Work material parameters

Work material	E [Gpa]	λ [W/m·K]	σ [N/mm²]	H [HB]
S10C	206	59	310	133
S35C	205	52	510	178
S50C	205	44	610	207
SS400	206	51.6	455	130
SCM440	212	42.7	980	319
SUS304	193	16.7	520	187
SUS430	200	26	450	183
A5052	68	137	260	68
ADC12	72	100	310	92
SKD61	206	30.5	1250	369
HPM7	208	34.3	974	303
NAK55	201	38.9	1265	369
NAK80	201	38.9	1255	369
FC250	115	50	250	241



(a) Spindle speed S (b) Infeed rate F

Fig. 6. Results of random forest (Co. C).



(a) Spindle speed S (b) Infeed rate F

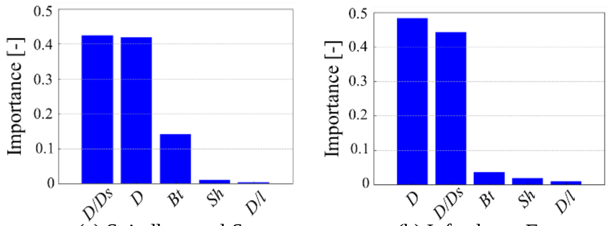
Fig. 7. Results of random forest (Co. D).

Figs. 6 and 7 show that the coefficient of determination is high for Companies C and D, indicating that the cutting conditions (S and F) can be predicted. This indicates that the random forest can be used to derive the optimal cutting conditions from the information on the drill geometry and workpiece material in the tool catalog and that it is easy to search for cutting conditions.

As described in this section, a random forest was used to predict the cutting conditions using a tool catalog, and accurate results were obtained. Therefore, the machine-learning model learned the information contained in the tool catalogs sufficiently. In the next section, the process of determining cutting conditions by evaluating the importance of explanatory variables and PDP is discussed, and the effects of the drill shape and work material on the cutting conditions are determined.

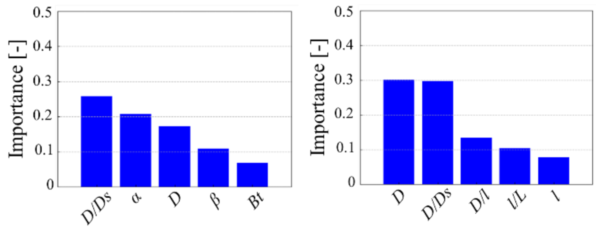
B. Discussion on the Importance of Explanatory Variables

Figs. 8 and 9 show the five most important explanatory variables for Companies A and B, respectively. The figures show that D and D/D_s are important for determining the cutting conditions (S and F) of drills for PCBs. Because the work material variables (Sh and Bt) are less important, the cutting conditions of drills for PCBs can be determined based on the drill diameter in many cases.



(a) Spindle speed *S* (b) Infeed rate *F*

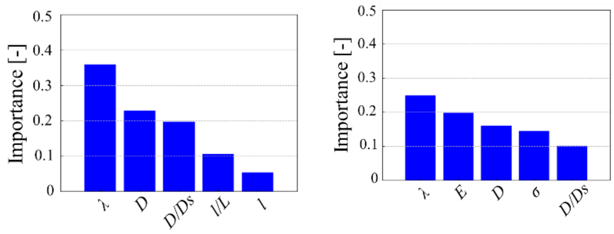
Fig. 8. Importance of each parameter (Co. A).



(a) Spindle speed *S* (b) Infeed rate *F*

Fig. 9. Importance of each parameter (Co. B).

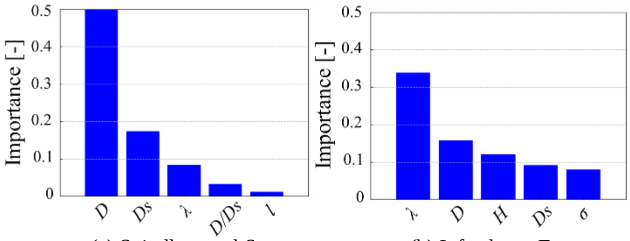
Figs. 10 and 11 show the five most important explanatory variables for Companies C and D, respectively. The figures show that the work material variable, λ, is significant in addition to D in the metal plate tool catalogs. Thermal conductivity indicates the diffusivity of heat; the larger the thermal conductivity, the higher the diffusivity of heat in the work material, and the less heat transferred to the drill, thus having a smaller effect on the temperature rise of the drill. Therefore, a work material with higher thermal conductivity can be machined at a higher spindle speed. Thus, it is likely that the spindle speed is determined based on the thermal conductivity of the work material.



(a) Spindle speed *S* (b) Infeed rate *F*

Fig. 10. Importance of each parameter (Co. C).

In addition to D and λ, E and H are important for predicting the infeed rate. Young's modulus expresses the rigidity of the work material, and the larger the value, the higher the rigidity and the lower the drill bite; therefore, it is necessary to reduce the infeed rate to promote the drill bite. In addition, because hardness affects tool wear and cutting resistance, the infeed rate must be reduced to lower the cutting resistance. Therefore, the infeed rate is likely to be determined by considering the Young's modulus and hardness of the work material.

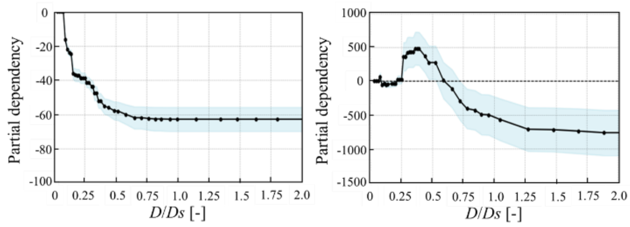


(a) Spindle speed *S* (b) Infeed rate *F*

Fig. 11. Importance of each parameter (Co. D).

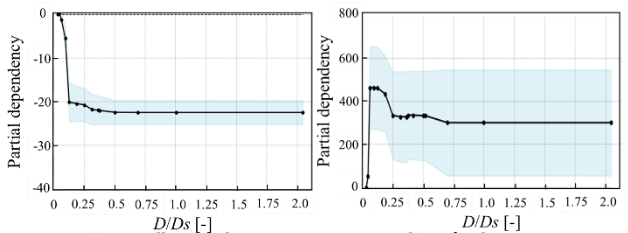
C. Consideration of Explanatory Variables Using PDP

In this study, the influence of explanatory variables on the prediction of the cutting conditions (S and F) was evaluated using PDP, a method for visualizing features and objective indicators. Therefore, the focus was on D/D_s and λ, which were variables of high importance in the results in the previous section. Figs. 12 and 13 show the PDP of D/D_s and the cutting conditions (S and F) for Companies A and B, respectively. Figs. 14 and 15 show the PDP of λ and the cutting conditions (S and F) for Companies C and D, respectively.



(a) Spindle speed *S* (b) Infeed rate *F*

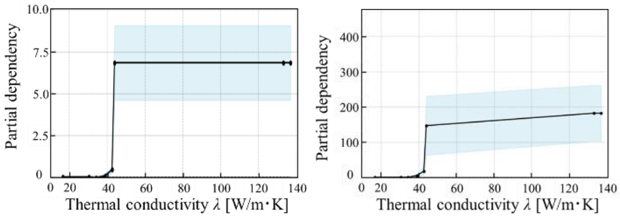
Fig. 12. Partial dependency plot (Co. A).



(a) Spindle speed *S* (b) Infeed rate *F*

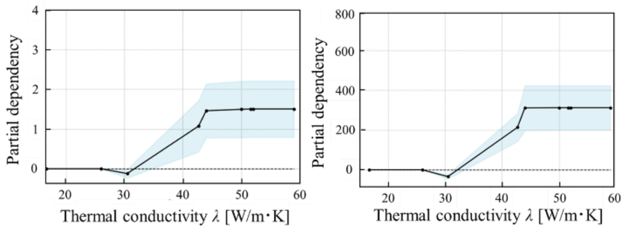
Fig. 13. Partial dependency plot (Co. B).

Figs. 12 and 13 show that the contribution of D/D_s to the spindle speed prediction increases in the negative direction (lower rotation speed) up to a value of 0.5, after which it remains almost constant. In other words, for drills with diameters less than half the shank diameter ($D/D_s < 0.5$), the diameter need not be considered significant in determining the spindle speed; however, as the drill diameter approaches half the shank diameter ($D/D_s = 0.5$), the degree to which the diameter must be considered gradually increases. For drills with diameters larger than half the shank diameter ($D/D_s > 0.5$) were considered to reduce the spindle speed by a constant value. Although there were differences between the PDPs of Companies A and B in predicting the infeed rate, they both had in common that the infeed rate was considered to increase based on a certain value of D/D_s and decrease thereafter because a peak was observed in the PDP.



(a) Spindle speed *S* (b) Infeed rate *F*

Fig. 14. Partial dependency plot (Co. C).



(a) Spindle speed *S* (b) Infeed rate *F*

Fig. 15. Partial dependency plot (Co. D).

Figs. 14 and 15 show that, in the metal plate tool catalog, the contribution of λ changes significantly at approximately 40 W/m-K in the prediction of the spindle speed and infeed rate. This is the thermal conductivity value for carbon steel for machine structural use (S50C) and chrome molybdenum steel (SCM440), and it is highly likely that these work materials are the criteria for determining the cutting conditions.

V. Conclusion

A random forest was used to predict the cutting conditions of microdrills for PCBs and metal plates using the drill shape and work material variables listed in the tool catalogs as explanatory variables and the cutting conditions (spindle speed and spindle infeed rate) as objective variables. As a result of random forest prediction, the following findings were obtained:

(1) The coefficient of determination was 0.7 or higher, and high accuracy was obtained; therefore, random forest is an effective means for deriving cutting conditions using drill tool catalog information.

(2) The drill diameter (D, D/D_s) was found to be important for determining the cutting conditions of the PCB drill. In addition to the diameter D, work material variables, such as thermal conductivity λ, Young's modulus E, and hardness H, are important for metal plate drills.

(3) By visualizing the relationship between explanatory and objective variables using a PDP, it was possible to determine the effects of the ratio of the diameter to the shank diameter, D/D_s, on the cutting conditions and the effects of the thermal conductivity, λ, on the cutting conditions.

(4) In the PCB tool catalogs, it was found that a ratio of 0.5 of the diameter to the shank diameter, D/D_s, is likely to be the criterion for determining the spindle speed. In the metal plate tool catalogs, it was found that the cutting conditions were likely to be determined by materials with a thermal conductivity λ of approximately 40 W/m-K, such as S50C and SCM440.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Shunya Tanaka conducted the research, analyzed the data and wrote the paper; all authors had approved the final version.

Acknowledgment

We gratefully acknowledge the work of past and present members of our laboratory.

References

[1] Y. Nakamura, "Technological trends of multilayer circuit board materials," *Panasonic Electrical Engineering Technical Report*, vol. 59, 2011.

[2] R. Miyajima, T. Hirogaki, E. Aoyama, and H. Kodama, "Radius end mill cutting condition prediction system by tool catalog mining applying machine-learning to data mining," *Precision Engineering Society Fall Lecture Series*, 2021, pp. 58–59.

[3] S. D'Oca and T. Hong, "A data-mining approach to discover patterns of window opening and closing behavior in offices," *Building and Environment*, vol. 82, 2013.

[4] R. Wang, X. Gong, M. Xu, and Y. Li, "Fault detection of flywheel system based on clustering and principal component analysis," *Chinese Journal of Aeronautics*, vol. 28, no. 6, 2015.

[5] Y. Takahashi, S. Fujishima, and H. Kato, "Chemical data mining based on structural similarity," *Journal of Computer Chemistry Japan*, vol. 2, no. 4, 2003.

[6] T. Sakuma, A. Asakura, K. Yamada, T. Hirogaki, E. Aoyama, and H. Kodama, "Proposal of data mining process for tool catalog data introducing machine-learning," *Transactions of the JSME*, vol. 85, no. 877, 201.