Original Research Paper Engineering

The Use of Decision Trees in Supporting Design of Technological Process

Izabela Rojek

Kazimierz Wielki University, Chodkiewicza 30, 85-064 Bydgoszcz, Poland

ABSTRACT

The article presents selection models that aid the design of technological processes using decision trees. The models aid the selection of the machine, the tools, and the machining parameters. They were developed using selected decision trees, such as C4.5, general classification model (C&RT), general CHAID models, reinforced classification trees, and a random forest. These models were prepared using the StatSoft Statistica Data Miner software.

KEYWORDS decision trees, machine tool, tool, machining parameters

INTRODUCTION

In machine manufacturing, the technological process is the basic part of the manufacturing process that directly involves changing the shape, dimensions, surface quality, and physico-chemical properties of the workpiece [11]. Over the years, the approach to the design of technological processes has changed. It is now common to use databases with ready-to-use technological processes and to use CAD and CAPP systems, as well as to integrate them (via shared databases and knowledge bases [3,12].

The first artificial intelligence technologies used in CAPP systems included expert systems with knowledge bases in the form of frameworks, decision rules, and semantic networks. Contemporary technologies take the form of machine learning (decision trees), fuzzy logic, neural networks [7-9], genetic algorithms, and hybrid systems [4]. Artificial intelligence in CAPP systems makes it possible to record the process engineer's experience in a knowledge base and to conduct inference similar to human reasoning during the design of the technological process.

This work constitutes continuation of the research that is partly described in article [6]. The current research expands the tool selection parameters. Research has been conducted on the development of the selection of the machine and the machining parameters selection models, and other classification trees have been used to develop the models. The models were prepared using the StatSoft Statistica Data Miner software.

DECISION TREE INDUCTION METHOD

A decision tree is a directed acyclic graph, whereby the edges are referred to as branches, the peaks with at least one edge protruding are referred to as vertices, and other peaks are referred to as leaves. Moreover, it is assumed that in such a graph there is only one path between different peaks. The structure of a tree can best be presented, based on a set of examples, in the form a recurrence algorithm that is started for each vertex in the tree. The argument selection technique is of key importance to the appearance of the decision tree because it is the order of attribute selection that has a major impact on the depth and the degree of expansion of the tree. The recurrent initiations of the algorithm that forms the tree must end sometime and the stop criterion is responsible for this. A tree is created using a selection criteria based on a measure of the increase of information, which is connected to the concept of measure of entropy used in the theory of information [5].

In this research, selected algorithms were used to build classification trees: C4.5, general models of classification trees (C&RT), general CHAID models, reinforced classification trees, and a random forest [1,2,10,13]. As far as prediction accuracy is concerned, it is difficult to recommend any particular method. This subject is still being studied. From a practical point of view, it is best to use different algorithms and to compare them and select the model that achieves the smallest prediction error [10].

CASE STUDY

The real data comes from a company that offers a broad range of products. This research focused on the machining of metals. During the manufacturing of products, basic technological operations are performed, such as cutting, laser cutting, milling, rolling, grinding, threading, etc. The article presents the milling process models. The models were evaluated using cross-validation based on a test sample and a 10-fold cross-validation.

For each type of tree (C4.5, C&RT, CHAID, reinforced classification trees, and a random forest) the parameters of the models were changed. In the case of the C&RT model, the following classification parameters were set: the costs of erroneous classifications, the quality of the match, and the a priori probability. The stop criterion contained the following stop rule: cut according to the variance; also, it contained the parameter of the minimum number of examples in the vertex. In the case of the C4.5 and CHAID model, the costs of erroneous classifications were set. The stop criterion pertained to the minimum number of examples in the vertex. In the case of the reinforced classification tree model, the following classification parameters were set: the costs of erroneous classifications and the a priori probability. The stop criterion contained the parameter of the minimum number of examples in the vertex. In the case of the random forest model, the following classification parameters were set: the costs of erroneous classifications and the a priori probability. The stop criterion contained the parameter of the minimum number of examples in the vertex.

The costs of erroneous classification relate to the distribution of examples among the classes. Reduction of the costs to a minimum corresponds to the reduction of the portion of cases classified as erroneous assuming a priori probabilities proportional to the size of the classes and costs of erroneous classification equal to each class. The quality of match refers to finding a division for each predictive variable in order to find a division for which the increase of the quality of match is the largest. There are three ways to measure the quality of match. The Gini measure was the preferred match quality measure of the authors of the C&RT program. Excellent match means excellent classification. A priori probability determines the probability, without any prior knowledge of the value of the predictive variables in the model, that a given case or item belongs to a specific class.

Machine selection models

In order to prepare the learning data for decision trees, an analysis of the machinery used at Bohamet has been performed, in particular in relation to the CNC machines: mills, mill-drills, grinders, and turning lathes. The machine was selected separately for each technological operation.

Based on the machine data and the selection criteria, a learning file was prepared and it contained: type of operation (e.g. rough, finishing); Xp – product length [mm]; Yp – product width/diameter [mm]; Zp – product height/diameter [mm]; X – size of the working space [mm]; Y – size of the working space [mm]; Z – size of the working space [mm]; max. diameter of the tool [mm]; length of the tool [mm]; cost of operation of the machine tool [PLN/h]; min. rotational speed [rpm]; max. rotational speed [rpm]; max. working range f [mm/min.], machine tool power (power at the spindle motor) [kW], and machine symbol.

The models were prepared in the form of classification trees: C4.5, C&RT, CHAID, reinforced trees, and random forest. Figure 1 shows example of random forest (a) and rules generated on the basis of that tree (b). Attributes Xp, cost and Z had the highest rank, and on its basis was built tree.

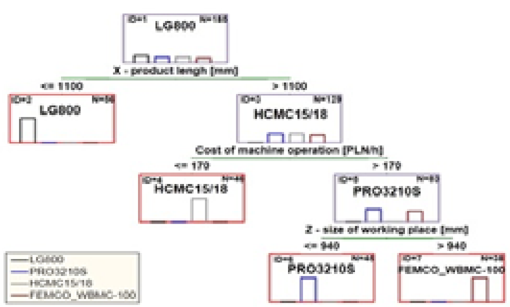
In the case of machine selection, the reinforced trees and random forest methods also turned out to be the best. The respective value of correct classification was 100%.

Tool selection models

In order to prepare the learning data for decision trees, an analysis was performed using Bohamet's tools, which were divided into drills, milling cutters, lathe tools, and grinding wheels. The tools were selected separately for each technological operation.

Based on the tool data and the selection criteria, a learning file was prepared and it contained: type of operation; type of machined surface; type of machined material; roughness; type of tool; type of milling cutter mounting; diameter of the milling cutter [mm]; shape of the milling cutter; number of blades; total length of the milling cutter [mm], milling speed v_c [m/min.]; milling depth a_p [mm], feed rate [mm/min.]; cost of operation of the tool [PLN/h]; milling width a_e [mm], and tool symbol.

Table 1 shows a comparison of the cost and assessment of risk and the standard error for the generated classification trees.



a)

rules
01: machine = 'LG800' if
    Xp <= 1100;
02: machine = 'HCMC15/18' if
    Xp > 1100,
    Cost <= 170;
03: machine = 'PRO3210S' if
    Xp > 1100,
    Cost > 170,
    Z <= 940;
04: machine = 'FEMCO_WBMC-100' if
    Xp > 1100,
    Cost > 170,
    Z > 940;

b)

Figure 1. Example of a) random forest, b) decision rules

Table 1. Comparison of milling cutter selection classifiers

Classifier type	Classifier evaluation	
	Cost/assessment of risk	Standard error
C4.5	0.392176	0.027591
C&RT	0.065217	0.010509
CHAID	0.201087	0.017060
Reinforced trees	0.000000	0.000000
Random forest	0.031746	0.012753

The best models turned out to be reinforced trees and random forest. The respective values of correct classification were 100% and 97.64%.

Machining parameter selection models

In order to prepare the learning data for the decision trees, an analysis was performed of the technological processes with regard to selection of the machining parameters for specific machines and tools at Bohamet. The machining parameters were selected separately for each technological operation.

Based on the machining parameter data and the selection criteria, a learning file was prepared and it contained: type of operation; type of machined material; symbol of selected tool; roughness; machining depth a_p [mm]; milling width a_e [mm]; target depth [mm]; machine symbol; and machining parameters: feed rate [mm/min.]; machining speed [m/min.]; duration of the operation [min.]; and tool service life [min.].

In the case of machining parameters selection, the reinforced trees and random forest methods also turned out to be the best, too (99,99%).

The technological process design system

Figure 2 shows system menu (a) and selected parameters for milling operation (b).

CONCLUSIONS

In the case of aided technological process design, due to the large quantity of input data in symbolic form, it is appropriate to use the decision tree induction method as the classification method. Decision trees demonstrate very good classification properties. Generation of rules based on decision trees enables compact recording of rules and significantly reduces the time needed for inference. The research that was performed demonstrated the usefulness of classification trees and their high effectiveness in aided technological process design. The comparison of C4.5, C&RT, CHAID, reinforced classification trees, and random forest models has also led to interesting research conclusions. The best models turned out to be reinforced trees and random forest. It does require a little more time to build these trees.

The technological knowledge acquisition system where the classification models are included can aid less experienced process engineers in the course of their design of technological processes.

Figure 2. Technological process design system – a) menu and b) selected parameters

REFERENCES

1. Breiman, L., Friedman, J. H., Stone, C. J., & Olshen, R. A. (1984). ***Classification and regression trees***. Belmont, CA: Wadsworth.
2. Breiman, L. (2001). Random forests. ***Machine Learning***, ***45***, 5-32.
3. Chlebus, E. (2000). ***Technika komputerowa CAx w inzynierii produkcji***. Warsaw: WNT.
4. Knosala, R. (2002). ***Zastosowania metod sztucznej inteligencji w inzynierii produkcji***. Warsaw: WNT.
5. Larose, D.T. (2005). ***Discovering Knowledge in Data: An Introduction to Data Mining.*** New Jersey: John Wiley&Sons.
6. Rojek, I. (2009). Classifier Models in Intelligent CAPP Systems. In: K.A., Cyran, S., Kozielski, J.F., Peters, U., Stanczyk, & A. Wakulicz-Deja (Eds.), ***Man-Machine Interactions, Advances in Intelligent and Soft Computing*** (311-319). Springer-Verlag.
7. Rojek, I. (2010). Neural networks as performance improvement models in intelligent CAPP systems. ***CONTROL&CYBERNETICS***, ***39(1),*** 55-68.
8. Rojek, I., & Jagodzinski M. (2012). Hybrid Artificial Intelligence System in Constraint Based Scheduling of Integrated Manufacturing ERP Systems. ***Lecture Notes in Artificial Intelligence***, ***7209***, 229-240.
9. Rojek, I. (2016). Technological Process Planning by the Use of Neural Networks. ***Artificial Intelligence for Engineering Design, Analysis and Manufacturing,*** 1-15. http://dx.doi.org/10.1017/S0890060416000147
10. StatSoft Statistica Internetowy podrecznik statystyki. http://www.statsoft.pl/textbook/stathome_stat.html?http%3A%2F%2Fwww.statsoft.pl%2Ftextbook% 2Fstclatre.html, Access: 10.01.2015.
11. Szatkowski, K. (2008). ***Przygotowanie produkcji***. Warsaw: PWN.
12. Weiss, Z. (2002). ***Techniki komputerowe w przedsiebiorstwie***. Poznan Wydawnictwo Politechniki Poznanskiej.
13. Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z-H., Steinbach, M., Hand, D. J., & Steinberg, D. (2008). Top 10 algorithms in data mining. ***Knowledge and Information Systems***, ***14(1),*** 1–37.