Management and Production Engineering Review
Volume 3 • Number 2 • June 2012 • pp. 27–35
DOI: 10.2478/v10270-012-0013-1

SELECTION OF DATA MINING METHOD FOR MULTIDIMENSIONAL EVALUATION OF THE MANUFACTURING PROCESS STATE

Michał Rogalewicz, Maria Piłacińska, Agnieszka Kujawińska

Poznan University of Technology, Institute of Mechanical Technology, Poland

Corresponding author:
Agnieszka Kujawińska
Institute of Mechanical Technology
Piotrowo 3, 60-965 Poznań, Poland
phone: +48 61 665-2798
e-mail: agnieszka.kujawinska@put.poznan.pl

Received: 8 March 2011
Accepted: 10 May 2012

Abstract
The article deals with the issues involved in evaluating the process state on the basis of many measures, including: process parameters, diagnostic signals and events occurring during the process. These measures as well as those measurements traditionally used in the evaluation of process capability, offer a relevant source of information about the manufacturing process and the authors attempted to ascertain the most suitable method, or group of methods, for achieving this. They present the main criteria for the categorization division of the methods of the manufacturing process state evaluation and, from those identified, distinguish the traditional from Data Mining methods. The authors then specify some basic requirements regarding the desired method or group of methods and focus on the classification problem. A division and classification of the methods is made and briefly described. Finally, the authors specify the criteria for their selection of the Data Mining method type as being the most appropriate for the evaluation of the manufacturing process state and, from within this type, offer the most suitable groups of methods. Some directions for further research are discussed at the end of the article.

Keywords
quality control, process state evaluation, data mining methods, classification.

Introduction

A decisive factor in the competitiveness of a production company is its capability to support the client with the supply of products of a determined quality, at a fixed price, in good time. This capability is the result of the efficient management of the production process, which, for the most part, consists of a manufacturing process, which, in turn, consists of a number of operations. It is during the manufacturing process that added value and this is directly related to a change in the shape, dimensions, surface quality, physicochemical properties and appearance of the processed material or the mutual orientation of the component parts forming the product [1–4]. Crucial for controlling the manufacturing process (or more specifically, for controlling an individual manufacturing operation[1]) is the possibility of evaluating the current state of the process on line. This state can be described by a set of measures – process characteristics and, in the case of a manufacturing process, its state is often described as a set of deviations of the process from the model process [5]. In this paper, the process state is understood as the quality at a given moment (i.e. to what extent the process fulfils specific quality requirements) and the evaluation of the process state is one of the key tasks of a process engineer. It allows the prediction of the quality of the products (semi-fabricated products) at the output of the process and enables an efficient control of the process by taking possible corrective actions.

[1]In quality engineering, the name "process" is often assigned to a single operation, which involves resources on input and a ready product or processed material on output.

An evaluation of the quality of manufacturing processes using a classic approach is performed by the measurements of values of earlier defined properties (most frequently so-called critical features) of the product and a determination of so-called process capability indices (e.g. C_p, C_{pk}) and other measures of process performance (DPMO, ppm or others) on the basis of these values. The measurements can be performed during the process or after its termination and information about the quality level may influence the further realization of the process through a feedback loop. During the process, there is a process control and after termination only a post-operational inspection takes place. Such an approach allows for the evaluation of the process in regards to the fulfilment of the requirements for only one property of the product. Hence, for each characteristic relevant for the process, a separate capability index is determined and a separate control chart is maintained, etc.

Increasingly, researchers and engineers often make attempts at the development and implementation of tools for the evaluation of the process state which look at many features at the same time. The results of their work are statistical tools which allow the monitoring of process stability with a simultaneous consideration of several separate features. A multivariate control chart is an example of such a tool. However, they are inadequate when the aim is to:

- determine the optimal value (or value range) of an input parameter of the process (e.g. range of feed rate) in a way which obtains specific values for individual characteristics (e.g. not below or not above ...)
- determine a set of optimal values for input parameters (a machine, work environmental conditions and other parameters) in a way which obtains a specific level for a desired feature (e.g. the roughness of a shaft after turning).

The presented approaches to the evaluation of the manufacturing process state are illustrated in Figs. 1a and 1b.

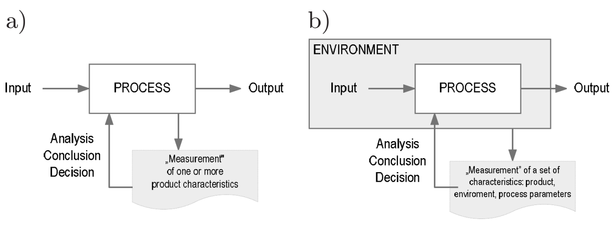

Fig. 1. a) Classical approach to the evaluation of the manufacturing process state b) Approach to the evaluation of the process state with a consideration of the process and its environment

The authors of this paper aimed to find methods which allow for the evaluation of the state of the manufacturing process on the basis of many measures, including:

- parameters of the process (e.g. injection moulding pressure, form temperature, time of cooling),
- diagnostic signals (e.g. vibrations, noise, ambient temperature),
- events occurring during the process (e.g. chipping of a cutting edge, operator error).

These measures (apart from measurement results traditionally used in the evaluation of process capability) are a relevant source of information about the manufacturing process. Simultaneously, they allow for the possible various factors related with the work environment, the applied manufacturing technology and the state of the machine or the processed material. An evaluation carried out on the basis of the variables mentioned above is performed on-line. A problem with the manufacturing process state evaluation obtained on the basis of many measures of state is presented in Fig. 2.

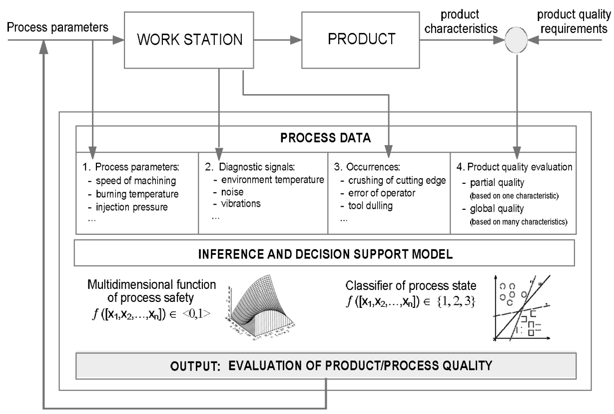

Fig. 2. The problem of evaluation of the manufacturing process state [6].

Because of the fact that at any given moment the process capability is influenced by many characteristics (measures of the process state), it is reasonable to treat the evaluation of the state of the process as a multidimensional problem. Such an approach prompted the authors to take a wide group of Data Mining[2] methods into consideration.

There are many traditional and Data Mining methods which can be used to evaluate the process state. Because of that there is a problem with choosing the proper one to use in manufacturing environment.

The main purpose of this paper is to make a classification of these methods and suggests the most

[2]Data Mining – process of extraction of hidden, potentially useful and previously unknown information from data sets [7].

suitable ones for estimation of the manufacturing process state.

The subsequent chapters of this work present the main criteria for the division of the methods of the manufacturing process state evaluation. Out of the available methods, the traditional and Data Mining methods were distinguished and the basic requirements regarding the desired method (group of methods) were specified. Then, a division of methods was carried out and they were briefly described to ultimately specify the criteria of selection of the most appropriate (in the authors' opinion) Data Mining methods for the evaluation of the manufacturing process state.

Criteria of division of methods of evaluation of the process state

There are many methods that allow the evaluation of the state of the process. The authors created a classification of these methods, taking selected criteria into account. Table 1 presents the most important criteria for the division of methods for the evaluation of the manufacturing process state.

Considering a specific problem, the desired features of the method which allow for a solution can be selected from the table. As has already been pointed out, the authors aimed to identify the methods which allow the evaluation of the state of the process, considering values of parameters, diagnostic signals and occurring events. There is a typical problem with many variables on the output. The result of the process is frequently influenced both by quantitative and qualitative variables and therefore the method of evaluation of the process state should process both types of variables. The authors also made the assumption that the desired approach should allow for the performance of an evaluation of the process state on-line, enabling a quick reaction and the taking of the appropriate corrective action by an operator or a process engineer.

Considering the aims of the method of process state evaluation and the general premises related to the possibility of the practical implementation of the presented solutions, the most important requirements were determined. It was acknowledged that the method to be selected should:

- allow the evaluation of the process state on the basis of many variables,
- process both quantitative and qualitative variables,
- deal appropriately with missing data values and redundant and irrelevant data,
- perform an 'online' evaluation (during the process run),
- be characterized by low user knowledge requirements of the mechanisms of the method,
- allow the obtaining of results in an easy to interpret form, e.g. diagrams or rules.

Table 1

Criteria of the division of methods for the evaluation of the process state.

Criterion	Example	
Number of input variables	- one variable	- evaluation of the grinding process on the basis of roughness of a surface layer
	- many variables	- evaluation of the turning process on the basis of roughness of a surface layer and dimensions of a two-stage shaft
Type of variables	- quantitative variables	- evaluation of the steel rod cutting process on the basis of rod length
	- qualitative variables	- evaluation of the casting process on the basis of the visual inspection of the casting (evaluation as "good" or "bad")
Time point of performing the evaluation	- on the production line evaluation	- monitoring of the cutting tool during the process of finishing shaft turning
	- evaluation after process termination	- evaluation of the grinding process on the basis of roughness of a surface layer
Character of the output variable	- specific value of quantitative variable	- evaluation of the casting process on the basis of the casting flexural strength, dependent on the chemical constitution of an alloy (e.g. using neural networks)
	- value of quantitative variable in probabilistic or fuzzy form	- finding a probability distribution for a number of broken seed on 1m^2 surface of a windowpane
	- specific class of qualitative variable	- assigning a casting to a class determining its quality on the basis of the chemical constitution of an alloy
	- qualitative variable class in probabilistic or fuzzy form	- assigning a casting to a class determining its quality on the basis of the chemical constitution of an alloy with indication of degree of affiliation to this class

In considering particular methods, the character of the output variables is a fundamental criterion for their classification. This classification is presented in the next chapter.

General classification of methods of evaluation of the manufacturing process state

As it has been pointed out, there are many methods for an evaluation of the process state. The authors focused on reviewing such methods from the Data Mining area as they allow an analysis of large data sets and enable the automatic discovery of relationships in these sets. A systematization of Data Mining methods was performed and examples were shown. They are presented in Fig. 3, alongside traditional methods (left branch of the tree), and represent a relevant group of tools for the evaluation of the manufacturing process state in terms of its quality. It is also noteworthy that the classification in Fig. 3 is not comprehensive.

The traditional methods of process quality evaluation comprise approaches affiliated with the Statistical Process Control approach, that is univariate and multivariate process capability indices and approaches in which the process evaluation is based on the number of faulty products obtained as a result of the process. The latter group consists of the following methods:

- DPMO (Defects per Million Opportunities),
- ppm (parts per million),
- fraction of faulty products,
- sigma level.

The second group of methods of the manufacturing process state evaluation are Data Mining methods, which are a particular object of interest of the authors. Two types of methods can be distinguished within the group: supervised and unsupervised learning methods. The principle of operation of the former group of methods is to create a model of dependency between input and output variables on the basis of a so-called training set, consisting of records containing input values matched with assigned output values. The supervised learning methods are mostly used for two tasks: classification and regression (numerical prediction). Each task has a different character output variable. During the classification on the basis of the input values, an algorithm assigns a new observation to the class (so an output variable is qualitative), while during the regression, an observation is assigned with a specific numerical value (the output variable is quantitative). The unsupervised learning is of a different character and is mostly used for a grouping task. A training set does not contain values of output variables and an algorithm itself divides the observation set into so-called clusters. The observations within a single cluster are very "similar"[3] and the distance between clusters is being maximized.

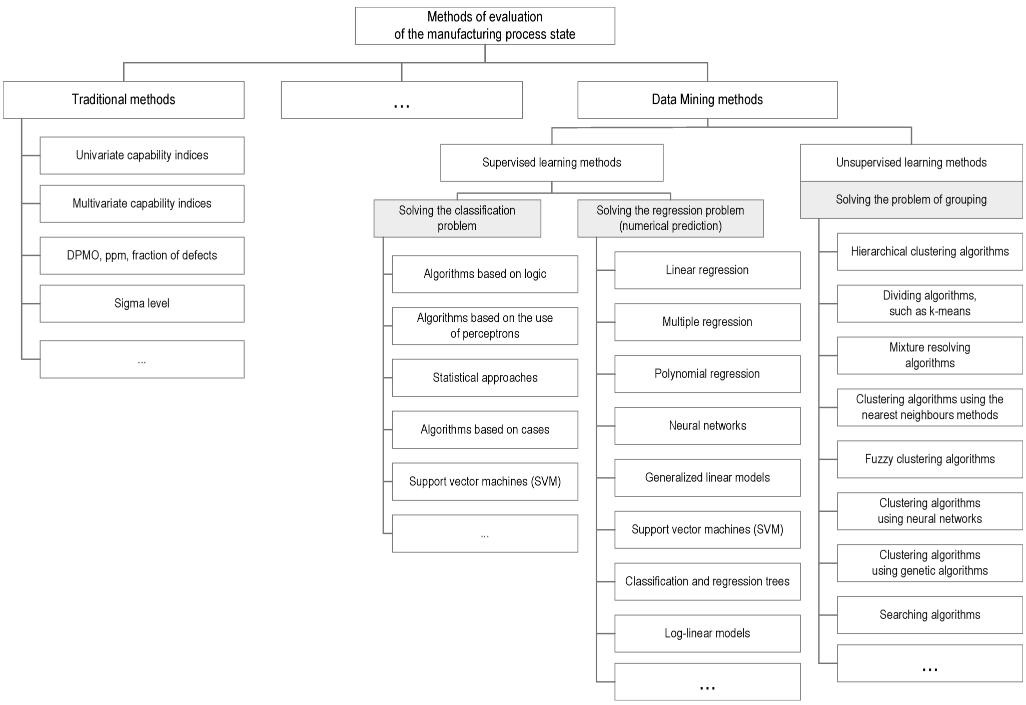

Fig. 3. Classification of methods of evaluation of the manufacturing process state (own work on the basis of [8–10]).

[3] There are many similarity measures, e.g. Euclidean distance in the feature space, Chebyshev distance or Manhattan distance.

In the considered problem, an evaluation of the manufacturing process state is performed through an evaluation of the quality of the obtained product. This quality is often expressed as a number in the ordinal scale. In such an approach, the problem of the state evaluation is reduced to the task of the classification of this state.

Consequently, a group of Data Mining methods solving the problem of classification was subjected to a thorough analysis. As has been mentioned, this classification is an example of supervised learning and its general concept is presented in Fig. 4.

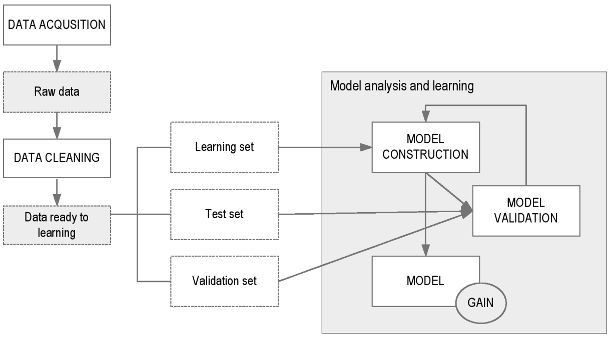

Fig. 4. A concept of supervised learning (own work on the basis of [11]).

As a result of the learning process, a classifier is acquired, which allows the prediction of the product quality during the manufacturing process (through the evaluation of the state of this process, expressed as a quantitative variable that can take values like: bad, average, good) and enabling an 'on-line' process control.

To create the classifier (model) of the manufacturing process state, a set of historical data describing the process run has to be obtained first. The data describing the specific realization of the manufacturing process consists of: values of parameters, diagnostic signals and events occurring during the process and also an evaluation of the quality of the obtained product. This evaluation is conducted by an expert (here: a process engineer or an operator), which allows for their knowledge (hidden, hard to extract) to be included in the procedure for creating a model of the process.

The next chapter briefly presents the Data Mining methods for realising the classification task.

Methods of classification

As has been mentioned, from the methods presented in Fig. 3, the classification methods are the subject of a detailed analysis and there are many divisions of methods for realizing the task of classification [12–14]. The division proposed by Kotsiantis [14] was accepted by the authors as the most systematized and comprehensive. Kotsiantis divides the methods of classification into:

1. **Logic based algorithms**
   a. Decision trees
   b. Learning set of rules
2. **Perceptron-based learners**
   a. Single layered perceptrons
   b. Multilayered perceptrons
   c. Radial basis function networks
3. **Statistical learners**
   a. Naive Bayes classifier
   b. Bayesian networks
4. **Instance-based learners**
   a. k nearest neighbours
5. **Support vector machine**

The chosen members of particular groups of classification methods are characterized below.

Logic based algorithms

The group of logic based algorithms is formed by decision trees and the learning of a set of rules. Decision trees are classification models in graphical form resembling a tree. An algorithm builds a tree, in each step searching for variables allowing it to divide the test set into subsets of records which are as homogenous as possible in terms of the value of the qualitative variable. In such a way, a structure consisting of a set of decision nodes (places of division of the data set) connected with branches propagating downwards from the root (first node dividing the whole set) to the finishing leaves (subsets which are no longer a subject of division and are a result of classification) is created. Decision trees are easy to interpret and allow a quick generation of decision rules and a classification of new instances [15].

Algorithms learning a set of rules search this set for a rule matched by as high a number of records as possible. Then they look for another rule for the remaining records etc. for as long as the rules do not cover the whole learning set. Genetic algorithms and methods related to inductive logic programming are examples of such algorithms. Inductive methods using rough set theory are also a part of this group

Perceptron based algorithms[4]

A perceptron is the simplest kind of artificial neural network. Its operation consists of processing the input signals (in the form of data records) inside

[4]Perceptron – unidirectional artificial neural network, consisting of neurons of non-linear activation function [11].

the network, comparing them with a defined threshold and then assigning output values. From perceptron based algorithms, three groups can be distinguished: single layered perceptrons, multilayered perceptrons and radial basis function networks.

Single layer perceptrons consist of two layers of connected neurons – input and output layers. They allow a classification of linearly separable observation sets [14]. Multilayered perceptrons are appropriately connected neurons, arranged in two or more layers. They allow the computation of much more complex functions than the single layer perceptrons [7]. Radial basis function networks realize the division of data record space by using circles/hyperspheres. They usually consist of three layers: an input layer, an output layer and a hidden layer and they enable the quick learning of a neural network and avoid the problem of local minimums (occurring in multilayered perceptrons) [16].

Statistical learners

Among many methods belonging to this group, naive Bayes classifier and Bayesian networks are the most frequently used. Naive Bayes classifier is based on the Bayes' theorem and assumes a very simple dependency between input variables and a qualitative variable class affiliation (assumptions made in this method are often very profound, hence the name "naive"). For a given set of input variables, *a posteriori* value of the probability of affiliation of a given object to one of the classes of qualitative variable is calculated. A class with the highest *a posteriori* probability is assigned to a given observation.

A Bayesian network takes the form of a graph. Its vertices often represent variables while the edges represent the dependencies between variables. An A node is a parent (a predecessor) of a node X and a node X is the descendant (a successor) of the node A, if there is a direct edge between A and X vertices. Each variable in a Bayesian network is conditionally independent of variables which are not offspring of the network for a given predecessor. A Bayesian network represents the total probability distribution, ensuring a defined set of assumptions concerning the conditional independence of variables and the tables of conditional probability associated with each vertex with fixed direct predecessors. The learning is simple when probability tables for every vertex are known. as it is then easy to calculate a total *a priori* and *a posteriori* probability [17].

Instance-based learners

The most popular method in this group is the method of k nearest neighbours. As the name suggests, a classification of a new instance is performed considering its k nearest (most similar) objects in the multidimensional space of variables. It means that there is no model built on the basis of a training set and all classified instances are directly used instead. In a practical application of this method, it is very important to adequately construct the set of observations and select appropriate distance metric and a relatively uniform distribution of observations in the space of variables [18].

Support vector machine

A support vector machine (SVM) is based upon the application of so-called Rosenblatt perceptron which searches for the straight line (hyperplane) dividing the record space into classes relating to the value of the qualitative variable. Such a straight line (hyperplane) has to meet certain conditions which allow its unequivocal definition. There is no need for all the observations for its determination, only these connected with a resulting line using vectors are necessary. These vectors are named support vectors and such a defined classifying algorithm is named a support vector classifier (SVC).

A SVC algorithm assumes only a linear division of classes and this is its significant limitation. A support vector machine is free of this assumption and is a generalization of SVC, allowing for non-linear divisions to be taken into consideration. Its main idea is a transformation of variables allowing the reduction of the problem into linear division. Original records become mapped into new space using kernel functions. It is possible to linearly separate two classes in this space, which allows the avoidance of a complex class boundary shape.

In the next chapter, the presented methods of classification are compared and analyzed in terms of the possibility of their application for an evaluation of the manufacturing process state.

Criteria of selection of classification methods for evaluation of the manufacturing process state

In [19], Kotsiantis made a comparison between the members of the individual groups of classification methods presented in the previous chapter of this paper. He included the following methods in his comparison: decision trees, neural networks, naive Bayes classifier, k-nearest neighbours method, support vector machine and learning sets of rules. The criteria used in this comparison can be divided into 4 groups:

- **basic performance of a classifier** – related to prediction capability and speed of learning and classifying of new instances,
- **robustness of a model** obtained using given classification method (robustness understood as capability of dealing with imperfections of a training set, e.g. lack or inconsistency of data or existence of irrelevant or redundant data),
- **additional performance of a classifier** – e.g. dealing with many types of input data and/or danger of overfitting,
- **user friendliness** – meaning ease of result interpretation, ease of understanding the classification mechanism and low requirements regarding user intervention in the algorithm operation (e.g. by supplying it with weightings or parameters of operation).

Kotsiantis applied a four-level grading scale for the evaluation of the methods on the basis of a specific criterion (1 – the worst grade, 4 – the best grade). The value of each grade was specified from evidence of existing empirical and theoretical studies. Out of the criteria analyzed in [19], the authors of this paper selected the most relevant (in their opinion) for a classification of the manufacturing process state and included them in Table 2. For the most part, they fit the requirements of the method desired, as presented in Sec. 2 of this paper. To emphasize the relations of the classifying algorithms analyzed in Table 2 with the requirements defined earlier, after the name of the specific feature (criterion), the number of its position on the requirements list (if applicable) is written in parentheses.

The authors considered the possibility of applying individual methods of classification in the evaluation of the state of the manufacturing process, taking their peculiarity and presented criteria into account. Algorithms with results hard to interpret or with too complex mechanisms of operation have been agreed to be excluded from further analysis. That is why neural networks and support vector machine methods were eliminated. Both these methods are accurate and allow quick classification but they are not "friendly" towards potential user and the sophisticated mathematical methods, being a base for the algorithms, often require the supply of many parameters not necessarily known to the user.

A low classification accuracy and a low tolerance to qualitative data are reasons why the naive Bayes classifier was abandoned. Although the predictive capability can often be improved by increasing data quality or eliminating irrelevant and redundant variables, the inability to process the qualitative data disqualifies the method from the considered application. A method of evaluation of the manufacturing process state must be able to manage both the quantitative and qualitative variables because in the manufacturing environment both the types characterize the process. The same reason stands behind the exclusion of the k nearest neighbours method from further analysis, with additional flaws of the method being the long classification time and the low tolerance to missing data.

Table 2

Comparison of selected classification algorithms (own work on the basis of [19]).

Feature	DT	NN	NB	kNN	SVM	LSR
1. Accuracy in general	2	3	1	2	4	2
2. Speed of classification	4	4	4	1	4	4
3. Dealing with a high number of variables (1)	4	4	4	4	4	4
4. Dealing with quantitative and qualitative data (2)	2	3	1	1	3	2
5. Ease of results interpretation (6)	4	1	4	2	1	4
6. User friendliness (5)	3	1	4	3	1	3
7. Dealing with missing values (3)	3	1	4	1	2	2
8. Tolerance to data redundancy (3)	2	2	1	2	3	2
9. Tolerance to irrelevant variables (3)	3	1	2	2	4	2

Legend of the table 2: DT - Decision trees; NN - Neural networks; NB - Naive Bayes classifier; kNN - k nearest neighbours method; SVM – Support Vector Machine, LSR - Learning set of rules.

Out of the examined methods of classification, decision trees and learning set of rules algorithms are the most suitable for an evaluation of the state of the manufacturing process in the opinion of the authors of this paper. They are characterized by easy interpretation and relatively simple mechanism of operation. They also do not require the user to provide many parameters and they perform a quick classification. According to Kotsiantis [19] learning sets of rules algorithms have, admittedly, a lower predictive capability, but frequently this can be improved by the better preparation of the training set of data, by choosing a different rule generation algorithm or by applying advanced methods for the selection of input variables.

Summary

Decisions regarding manufacturing process quality are presently taken on the basis of data supplied from 100% inspection, statistical acceptance inspection or statistical process control. A result of these decisions is mostly a statement indicating whether the process is stable or unstable (in a statistical sense). This evaluation is most often performed "post factum": after performing an operation, a measurement of a critical characteristic (or a few characteristics) is carried out and, on that basis, the evaluation is conducted.

In the era of the pursuit of *zero defect manufacturing*, an evaluation of the manufacturing process state cannot be limited only to the inspection of one (or even several) critical features of the product - during the ongoing process the state should be evaluated and the process should be corrected if necessary. Features of the process such as: parameters, diagnostic signals and events occurring during the process run are essential for consideration. A set of such data is a starting point to create a model of the process state and this allows the prediction of future process states.

During the first stage of the research work[5], the authors of this paper aimed at analyzing known methods from the Data Mining area, enabling a classification of the manufacturing process state and selecting that method which would best meet the criteria defined by the authors. Eventually, a method for an evaluation of the state of the manufacturing process will be implemented and applied in a production environment.

In the opinion of the authors, this is the reason why during the selection of an appropriate method of classification one should strive for the "golden mean" between predictive capability, classification speed, and ease of result interpretation and user friendliness.

The authors made some suggestions concerning choosing method of manufacturing process state evaluation:

- leaving out naïve Bayes and k-nearest neighbours classifiers because of its low classification accuracy and low tolerance to qualitative (this kind of data is frequently met in manufacturing environment),
- excluding algorithms giving results hard to interpret or having to complex mechanism of operation – lack of friendliness it's the reason why NN and SVM methods were eliminated,
- the most suitable Data Mining methods for the classification of the manufacturing process state are decision trees and learning set of rules methods. They are characterized by easy interpretation and relatively simple mechanism of operation.

Further research of the authors will focus on selecting the appropriate approach for these two groups of methods, possibly guaranteeing a higher accuracy of classification and a better handling with different data types.

References

[1] Durlik I., *Engineering management* [in Polish: *Inżynieria zarządzania*], PLACET 2004.

[2] Więcek-Janka E., *Games and decision* [in Polish], Publishing House of Poznan University of Technology, 2011.

[3] Feld M., *Basis of manufacturing processes design of typical machine parts* [in Polish: *Podstawy projektowania procesów technologicznych typowych części maszyn*], WNT, 2009.

[4] Pająk E., *Production management. The product, technology, organization* [in Polish: *Zarządzanie produkcją. Produkt, technologia, organizacja*], PWN, 2006.

[5] Korbicz J., Kościelny J., Kowalczuk Z., Cholewa W., *Diagnosis of processes. Models, methods of artificial intelligence, applications* [in Polish: *Diagnostyka procesów. Modele, metody sztucznej inteligencji, zastosowania*], WNT, 2002.

[6] Piłacińska M., Kujawińska A., Rogalewicz M., *Manufacturing process state evaluation with the use of rough set approach – method assumptions and description*, in: Fertsch M., Grzybowska K., Stachowiak A. [Eds.], "Efficiency of production processes. Monograph.", Publishing House of Poznan University of Technology, 2009.

[7] Rutkowski L., *Methods and techniques of artificial intelligence* [in Polish: *Metody i techniki sztucznej inteligencji*], PWN, 2006.

[8] Han J., Kamber M., *Data Mining. Concepts and techniques*, Morgan Kaufmann, 2006.

[9] Hastie T., Tibshirani R., Friedman J., *The elements of statistical learning. Data Mining, inference and prediction*, Springer, 2001.

[10] Jain A.K., Murty M.N., Flynn P.J., *Data clustering: A review, ACM Computing Surveys*, 31, 3, September 1999. //practiceovertheory.com/blog-

[5]Research is realized in the research grant no. 1988/B/T02/2009/37 financed by Polish Ministry of Science and Higher Education.

/2010/02/15/machine-learning-who-s-the-boss/, (2012-02-29)

[11] Goebel M, Le Gruenwald, *A survey of Data Mining and Knowledge Discovery software tools*, SIGKDD Explorations, 1, 20–33, 1999.

[12] Kantardzic M., *Data Mining – concepts, models, methods and algorithms*, Wiley Interscience, 2003.

[13] Kotsiantis S.B., Zaharakis I.D., Pintelas P.E., *Machine learning: A review of classification and combining techinques*, Artif Intell Rev, 26, 159–190, 2006.

[14] Perzyk M., Soroczyński A., *The comparison of selected tools to create engineering knowledge for the casting production* [in Polish: *Porównanie wybranych narzędzi do tworzenia wiedzy inżynierskiej dla produkcji odlewniczej*], Mat. Konf. Krzepnięcie i Krystalizacja Metali, Cedzyna, 2008.

[15] Wójtowicz P., *Introduction to neural networks* [in Polish: *Wprowadzenie do sieci neuronowych*] – http://www.statsoft.pl/czytelnia/neuron/pwersnn.-html#wprowadzenie, (2012-02-29)

[16] Larose T., *Models and methods of data mining* [in Polish: *Modele i metody eksploracji danych*], PWN, 2008.

[17] Pyle D., *Data preparation for Data Mining*, Morgan Kaufmann Publishers, 1999.

[18] Kotsiantis S.B., *Supervised machine learning: A review of classification techniques*, Informatica, 31, 249–268, 2007.