10

APPLICATION OF COMPUTATIONAL INTELLIGENCE METHODS IN CONTROL AND DIAGNOSIS OF PRODUCTION PROCESSES

10.1 INTRODUCTION

It is a well known fact that manufacturing costs, along costs of materials, are the main cost components of products. As the quality of manufacturing has become a decisive factor in competing in a global market, a proper control of production processes, especially related to process fault management, has become one of the key issues in a plant operation. It has been a subject of extensive research and development for many years. This concerns both the discrete processes, characteristic of the parts industries such as electronics, cars, aircrafts, household products etc. and to the process industries such as chemical, textile, food etc, in which the continuous processes are typical. As indicated in [25], a fault or problem in a process does not have to be the result of a failure of an equipment component, or even involve specific hardware. A problem might be defined as non-optimal operation or off-spec product. For example, in a process plant, root causes of non-optimal operation might be hardware failures, but problems might also be caused by poor choice of operating targets, poor feedstock quality, poor controller tuning, sensor calibration errors, human errors etc.

The process fault management includes fault detection, its diagnosis, i.e. finding the root causes of the fault and fixing it. The right control procedures of a production process should support the prevention of the faults as well as provide the means for necessary corrections.

The process control procedures and the hardware controllers often utilize data-driven (empirical) models correlating the process inputs with the outputs. Many advanced techniques are currently employed in research and new developments in this field. The applications based on Data Mining (DM) become commonly used. Apart from statistical tools, the learning systems such as Artificial Neural Networks (ANNs), Decision Trees (DTs), various logic rule systems and other methods of Computational Intelligence (CI) are engaged. They can be applied in the control and fault diagnosis of hardware elements or systems, utilized in many areas including manufacturing (e.g. neuro-fuzzy controllers), as well as in the control and fault diagnosis of the production process treated as a whole, which is the scope of the present work.

The rest of this paper is organized as follows. In the next section the main types of control systems and fault diagnosis are briefly reviewed, together with the actual applications of the CI-based and advanced statistical techniques in these areas. In Section 3 the methodologies of significance analysis of process parameters, useful in building the data-driven models for process control as well for the fault diagnosis, based on the previous works of the present authors, are discussed. In Section 4 we present some evaluations of predictive capabilities offered by the time-series analysis from the standpoint of some industrial process control. In Section 5 the results of research aimed at evaluation of learning systems capabilities to detect out-of-control patterns of points observed in process run charts, are presented.

10.2 CONTROL AND FAULT DIAGNOSIS SYSTEMS USING ADVANCED DATA-DRIVEN METHODS

10.2.1 Types of control systems used in manufacturing industries

There are two main approaches to control of manufacturing processes: Statistical Process Control (SPC) and Engineering Process Control (EPC). Quality engineers employ SPC techniques to monitor the processes, whereas control engineers utilize EPC techniques to regulate them. Originally, SPC was first applied in the parts industry whereas EPC comes form the process industry. However, the discrepancies between these industries and the methods of controlling process variables are becoming smaller. Both control strategies are aimed at reduction of process variability, however, they seek to accomplish this objective in different ways [14]. SPC assumes that the process output can be described by statistically independent observations fluctuating around a constant mean and is intended to detect signals which represent the special (assignable) causes of external disturbances increasing the process variation. EPC actively counteracts the process disturbances by making adjustments to process variables in order to keep the output quality parameter on target. The output may shift or drift away from the desired quality target due to disturbances. These disturbances are usually not a white noise but exhibit a dependence on past values, i.e. they are auto-correlated. Hence, it is possible to anticipate the process behavior based on past observations and to control the process and outputs by adjusting the input variables [8].

To make an appropriate selection between the two approaches in practice, it is important to identify disturbance structures and strengths of the two control methods to influence the process. In the in-depth study [8] the following recommendations are made. If a process is not correlated, there is no need to employ EPC schemes and traditional and SPC control charts should be used for identifying assignable cause variations. When data are correlated, the possibility of employing EPC techniques should be examined. If appropriate controllers are available, EPC control schemes can be employed to compensate for the auto-correlated disturbance. To identify and understand the cause of process changes, a unified control framework should be applied to regulate a process using feedback control while using the diagnostic capability of SPC to detect unexpected disturbances to the process.

An example of such integrated approach is the Run-to-Run (R2R) or Lot-to-Lot control system [8]. The controller provides recipes (inputs) based on post-process measurements at the beginning of each run, updates the process model according to the measurements at the end of the run, and provides new recipes for the next run of the process. The R2R control originates from semiconductor industry where obtaining real-time information during a run is difficult and frequent changes of inputs to the process may increase the variability of the process outputs.

It should be noticed that application of EPC combined with SPC can cause some problems as the EPC feedback compensation mechanism affects out-of-control detection by SPC and degrades the output quality once suddenly assignable causes are removed. Further discussion on integration of the two control approaches can be found in [4].

10.2.2 The role of CI and advanced statistical methods in process control and fault diagnosis

The most important SPC tools are control charts, widely and successfully used for detection of appearance of unexpected abnormalities of the process in the form of excessive variations of quality characteristics of the products. However, they are not capable of identifying the root causes of the process instability and to provide the means for optimum control of the process.

Various types of models linking the potential causes with the process outputs, particularly the product characteristics, can be used for that purposes. Qualitative models, such as the well known Cause-and-Effect diagrams (also known as Ishikawa or "Fishbone" diagrams) are widely used. However, because of the obviously limited capabilities of the qualitative models, various quantitative data-driven models, based on DM approach and utilizing CI methods, become more common in industrial practice. Since the year 2000 a vast growth of various applications, aimed at supporting and improvement of manufacturing processes, has been observed. Some important reviews or systemic approaches can be found in [3, 9, 10, 11, 16, 24, 28, 29, 30].

SPC control charts assume that the process is not auto-correlated, i.e. the current observation does not depend on the previous observations. The presence of auto-correlation, e.g. the process mean's trend, makes it difficult either to recognize a state of statistical control or to identify departures from the in-control state [8]. To avoid this, the Special Cause Charts (SCC) have been proposed and developed [32]. The idea is to apply the time-series analysis to the original auto-correlated data (for example the ARIMA model), remove the trends and periodicity components and plot the control chart (SCC) using the residual data.

Application of control charts is based on finding characteristic patterns of points, indicating that the process variation is due to a special cause. Usually seven typical sequences of points are distinguished, which can be interpreted in terms of probable type of the special cause or probable further behavior of the process (see, e.g. [26]). Searching for the patterns is usually made 'manually', however, the advanced learning systems can be also useful for quick finding typical sequences, identifying more complex or subtle out-of-control cases or application of control charts for the auto-correlated processes (see, e.g. [2, 6, 7, 22, 33]). Some assessments of ANNs and DTs capabilities to detect these typical patterns of points, made by the current authors, are presented in Section 5.

In manufacturing industry the most commonly used type of EPC is probably the feedback control. It uses deviations of the output from the target to calculate the amount of adjustment. EPC always requires a process model in a form of input-output relationship which, for the feedback control, can utilize the time-series analysis tools (see, e.g. [8]). Predictive capabilities of the time-series models based on regression modeling of residual data, applied to some foundry processes by the present authors, are presented in Section 4.

Building the above mentioned qualitative and quantitative models used for identifying the root causes of the process instability as well application of EPC requires a careful selection of the process input variables. Therefore, a significance analysis of these variables, made from the viewpoint of the process output, would be of great interest since selection only significant variables allows quality engineers to obtain relatively simple and valuable models. In general, those variables which are found to be the most significant for a given process fault, e.g. increasing percent of defective parts, could be regarded as the first candidates for root causes of the fault. Finding the most significant input variables also allows selecting them as the most efficient process inputs in EPC, i.e. having the largest gain coefficients, being a measure of the impact of input control to process outputs. Identification of the most significant process variables can also facilitate establishing optimal inspection procedures in the manufacturing process, enabling engineers to concentrate on selected process variables and thus avoiding unnecessary costs. Various examples of the use of significance analysis for manufacturing process parameters can be found in a number of works, e.g. [1, 5, 12, 15, 21, 23, 29, 31].

Relative significances of process input variables can be found using different approaches, which assume different definitions of the significance and using various tools, including CI methods. In Section 3, the summary of the present authors' research on the methodologies utilized in the significance analysis of process variables, will be presented.

10.2.3 Examples of company systems

In this section two examples of company systems using CI tools in control and fault diagnosis of manufacturing processes are presented. The first example is a complete quality improvement system which expands traditional SPC capabilities through DM techniques [24]. The study is based on data which represent discrete and correlated process. The second example shows how useful ANNs can be in the process diagnosis, especially when conventional methods are difficult to apply [27]. In this case, a continuous production process was considered.

In the exhaustive study [24] the Continuous Quality Improvement System developed and tested in a LED packaging company was described. LED (Light Emitting Diode) is a very popular electric component used in a wide range of products from automobiles through computers to cell phones. Several facts indicate that traditional SPC methods are not an optimal tool for quality control and management in this case. Large lot sizes, high level of automation and fast production are characteristic for the LED packaging process. The process cannot be stopped due to effectiveness decrease so all potential problems should be detected, diagnosed and fixed during production, which cannot be achieved with the conventional approach. Data related to LED packaging operations are accumulated with high speed. From the quality control and improvement perspectives it is crucial to discover knowledge in this fast growing database. Furthermore, LED packaging process includes a set of operations which final goal is to bond the dice with PCB (Printed Circuit Board) and to enclose it in an epoxy compound. This should be treated as a discontinuous material flow. Every step may affect quality on the next stage. For that reason, the whole process and its parameters are strongly correlated, which makes a serious constraint in application of traditional SPC techniques. DM provides tools and methods which are used to expand SPC capabilities.

The first step in every process analysis is a description and parameterization. In this particular case, the complete list of all kinds of quality issues was made. The potential defects were assigned to related technology operations. It was discovered that for each step in this process a parameter which is critical to quality can be determined. A product dimension, which is specific for different operations, was accepted as a control factor. Through measurement of this parameter, yield rate and defect rate are monitored and controlled both for single operation and for the process overall. Predefined CTQ (Critical-To-Quality) dimensions are tracked with a SPC control chart. Every out-of-control signal is a trigger for DM analysis. A large-scale defect database and process data warehouse are considered as a source of rules and knowledge and DM methods are used to extract it. It is important to highlight that each result is certified whether it is reasonable or not before application. DM techniques allow to identify the root cause of out-of-control signal in a short time.

Rapid adjustments are essential in continuous quality control and improvement system. A SPC chart is re-applied after correction. Fig. 10.1 describes the model of continuous quality improvement system with central position of the data warehouse. Historical data and real time data are used to control and improve the quality. Efficiency measures are process yield and defect rate.

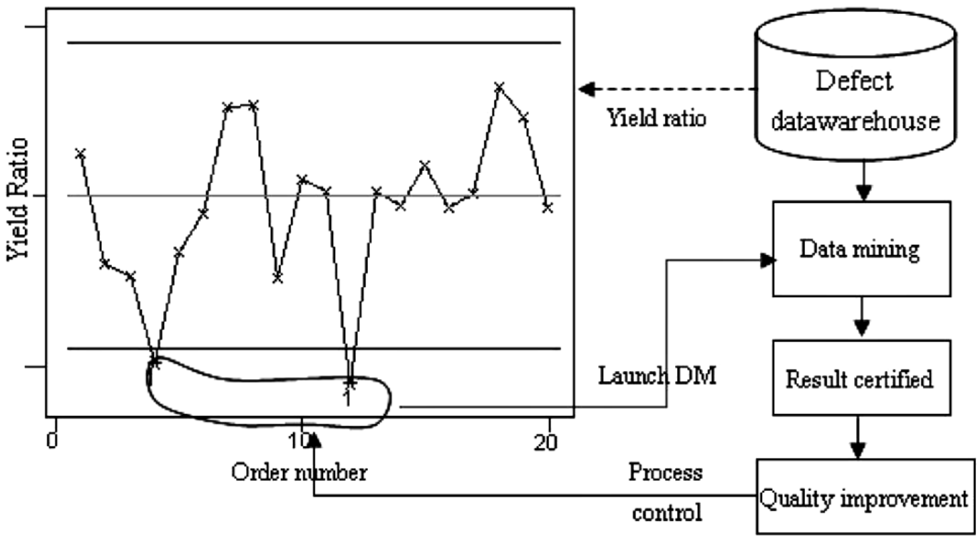


Fig. 10.1 Scheme of continuous quality improvement system [24]

A well-designed system should be helpful to the whole operating personnel. It should also collect data needed to identify the batch and include all the factors which are related to the production process. Not only parameters of each technological operation are taken into account but also an environment, understood as a machines, operators, materials and environment status. In this task, the large-scale data warehouse is used which contains all the information related to material flow as well as an additional part dedicated to defect description, if any occurred. The infrastructure of the system consists of three layers responsible for different tasks and connected to each other, shown in fig. 10.2.

The first layer is responsible for data collecting and is dedicated for operators. With the use of that initial layer it is possible to prepare daily reports and simple queries. The second layer is essential for quality management and continuous improvement. In this layer data from all the sources (process and data warehouse) are analyzed. Quality engineers may use that layer as a source of all kinds of information. Traditional SPC techniques, like control charts, are expanded by DM techniques (DT) to create a comprehensive quality management tool. This layer is responsible for on-line process control, diagnostic, and knowledge discovery necessary to continuous quality improvement. The third layer is dedicated to top managers to support a decision-making process. Reports created at this level visualize overall production efficiency.

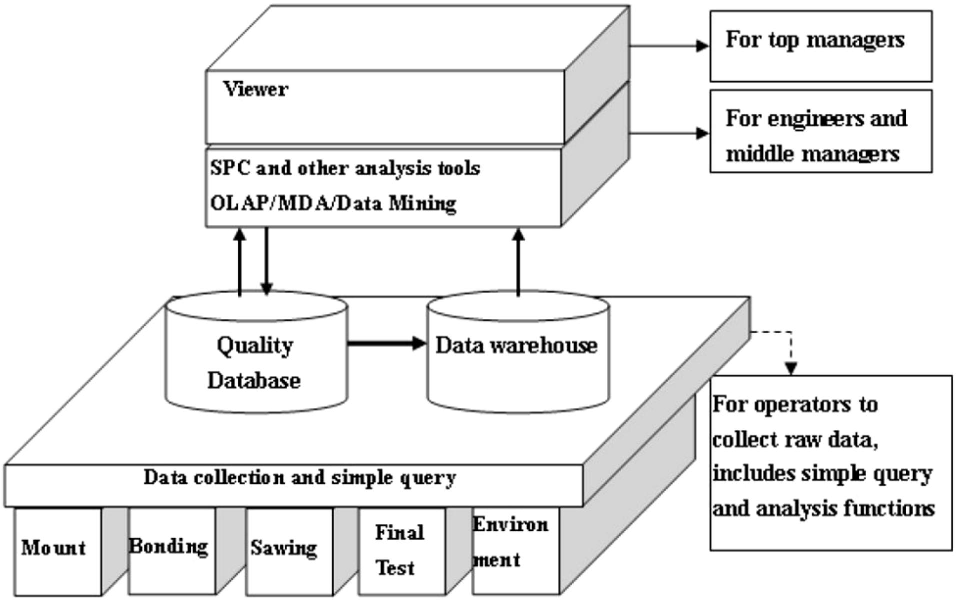


Fig. 10.2 Three-layer system infrastructure [24]

The parallel control of multiple parameters allows to apply three dimension SPC. The first element is a control chart which monitors the CTQ factor within the operation. At the same time, this operation yield is controlled by the other chart. The same system monitors the overall production efficiency. In that way it is possible to establish easily each disturbance on manufacturing process outcome.

The whole system was found to be useful and helpful in the process control and allows a continuous quality improvement.

The second example concerns multi-stage continuous process applied in the chemical industry [27]. In such conditions traditional statistical methods are ineffective, but combined with DM techniques may create a valuable tool for production control and management. This unique tool development was an answer to unexplained fluctuations in defect measurements observed in a chemical processing plant. The system objectives were to take full advantage of the site infrastructure and to increase production efficiency by quality control. In order to complete that task, a model capable of predicting defects and identifying casual relations within the process has to be developed. A production process should be reflected in measurable parameters. Then, those factors should be analyzed to detect potential defects just by observation of conditions in the plant. From the range of DM techniques and methods, ANNs of MLP type have been chosen. In this particular case, production involves a number of large capacity stages responsible for mixing and processing of the liquid product. Material flow is continuous, there are no batches, raw materials are delivered and the final product is collected incessantly. The production system runs twenty-four hours a day, 365 days a year. Routine maintenance operations are conducted without breaks and so fast that it is considered unlikely for them to have significant impact on production parameters. The quality control is based on laboratory measurements of selected parameters taken at regular time intervals.

Throughout the process more than four hundred sensors are installed and utilized for measurements of standard parameters like temperature and pressure. It is worth noticing that due to continuous material flow and frequent mixing operations there is no possibility to track exact raw material distribution in the production system. The authors of the system had access to almost all the data related to the production but some of them were not available because of their confidential character. A network design team also received information about operators and shift assignments. However, a traditional statistic approach indicated no impact of the human factor on a defect rate.

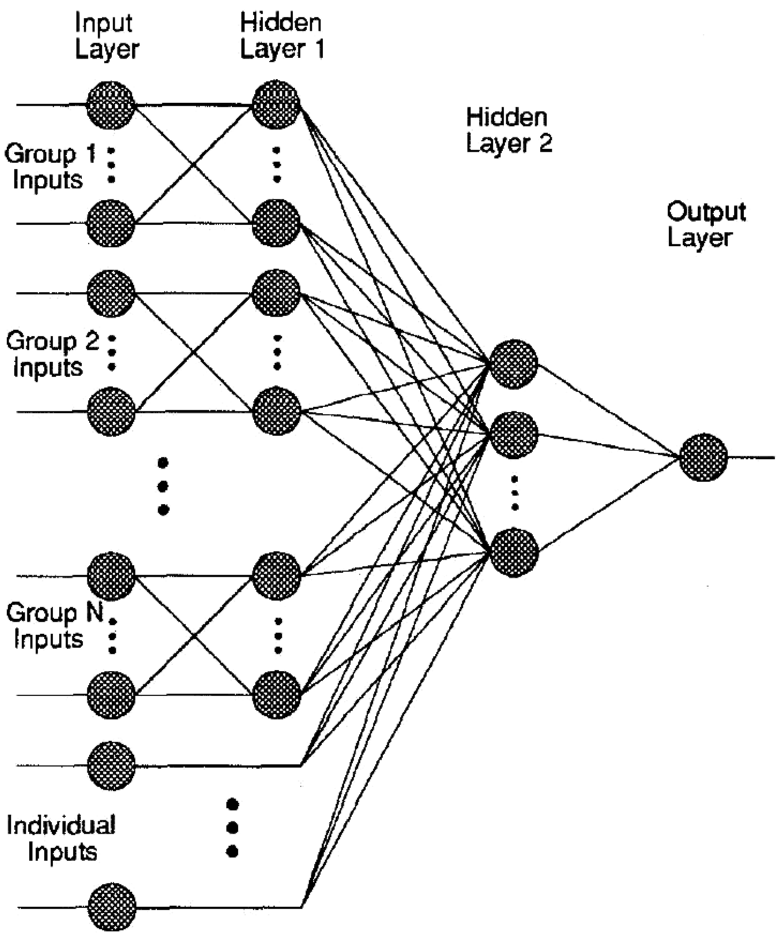


Fig. 10.3 ANN topology used in [27]

Data preprocessing was the first step of analysis. A reduction of sensors' number allowed to simplify the system and had a positive impact on computational time. Large number of data dropouts did not allow to treat this set as a time-series, so these data was considered as discrete samples. Quality records containing information about defects were also included in the data collection. Statistical analyses like spectral density estimation allowed to recognize the number of periodicities in the data. Because neural networks will firstly attempt to match the strongest feature in the data, a low-pass filter was applied in order to minimize that effect.

Network topology was designed to represent the production process. Fig. 9.3 presents the developed neural network scheme with two hidden layers. Selected inputs ware grouped into subnets base on real functional or spatial relations. Dedicated inputs reflect human factors such as shift and quality control personnel. Prepared data were divided into two groups, one for network training and the other for testing. Besides testing the defect prediction, the ANN weights were analyzed to establish most powerful factors related to an out-of-control situation. This allowed to determine the human factor as crucial for the product quality.

The results obtained in [27] showed that DM techniques may be useful as an enhancement of traditional SPC, especially in the situation when nonlinearity occurs within the data and many factors have to be considered.

The above presented examples confirm that application of computational intelligence methods in the control and fault diagnosis of production processes may positively affect both the quality of product and economic potential of a company through reduction of the defect ratio.

10.3 RELATIVE SIGNIFICANCE ANALYSIS OF PROCESS PARAMETERS

As indicated in Section 2, significances of process parameters play an important role in control and fault diagnosis of manufacturing processes. However, introducing available CI-based and statistical tools in industry is often difficult due to the lack of a deep insight into the characteristics of particular methods and algorithms. Comparative analyses of different methods made from the point of view of their performance in carrying out specific tasks are therefore very important. Those aimed at assessment of tools suitable for determination of relative significances of the process variables are relatively rare. Hence, the selection of appropriate methodology is often casual.

The significance of an input variable can be understood in different ways. One is based on the sensitivity analysis, which returns changes of the output variable due to small variations of the input variable, calculated at particular levels of the input. However, in the opinion of the authors, the practitioners in industry would be rather interested in finding potentially greatest overall effect of a process variable on the process results or equipment behavior.

For the significance analysis, a suitable input output model, based on the observations (past production data), should be built. Dependent on the output, different types of the models are suitable. For predicting the process outputs of numerical type, such as the fraction of off-spec products or their physical characteristics, the regression models are appropriate whereas for the fault diagnosis the classification models seem to be more useful as the outputs are often in the categorical form, e.g. ‘faulty’ or ‘acceptable’ for the product quality.

In the previous works [17, 18] extensive studies on the methodologies of determination of process input variables have been presented. Below some most important issues in this field will be discussed.

For the regression-type process outputs, an original procedure of determination of the input variables relative significances has been developed and evaluated in [17]. It is based on the so called "pedagogical" approach, i.e. uses a specially designed interrogation procedure to obtain the desired information. The significance factor for an input is defined as the maximum difference of the output, which can be obtained by changing the value of the analyzed input. The two extremes of the output are found by a gradient method, with the starting points found by a specially developed procedure, permitting to avoid local minima in most cases. These differences are calculated repeatedly a number of times for the other variables set at random levels. The final values of significance or interaction factors are calculated as the arithmetic averages of the differences. It is worth noticing, that the magnitude of the resulting scatter of the significance factor of a given input can be also a measure of the possible interactions with the other input variables.

For the classification tasks the other approaches are suitable and have been utilized in [18]. The first one is called 'decompositional' and is based on the analysis of the values of the model's parameters. The second approach, dedicated especially for determination of the overall significances of input variables, is based on the reduction of the prediction accuracy obtained from a model with reduced number of inputs, compared to the model built for all input variables.

It should be noticed, that some non-parametrical statistical methods such as ANOVA for and contingency tables, seem to be relative simple and natural methods for detection and measurement of the dependencies between variables represented by data sets, without necessity of making assumptions about forms of those dependencies. These tools are suggested by some commercial statistical packages for preliminary assessment of variables' significances and are also included in the presented works. All the significances are normalized by dividing them by the value obtained for the most significant variable.

For the assessments of the methodologies used for the significance analysis, two types of data sets were used: simulated (synthetic) data, with assumed hidden dependencies between inputs and output, as well as the real industrial data. The synthetic data were obtained by assuming analytical formulas of the type Y = f(X1, X2, …), from which, for random values of continuous-type input variables X1, X2, …, the continuous-type dependent variable Y was calculated. A Gaussian-type noise was then imposed on the input variables, with maximum deviation ±20%; that value was found to be characteristic of many real manufacturing processes. For the classification tasks all continuous values in most cases were converted to categorical ones, assuming the equal intervals method. The analytical formulas were assumed in the forms which correspond to the situations often observed in practice. The industrial data sets were essentially related to metal casting processes. Further details can be found in [17, 18].

The main findings related to the regression-type tasks are as follows. The proposed definitions of significance, based on specially designed interrogation procedures of CI regression models of processes, have proved to be valid and accurate for ANNs and Support

Vector Machines (SVMs), for all simulated data sets, i.e. with pre-defined hidden relationships. The non-parametric statistical method based on analysis of variance (ANOVA), sometimes used for initial analysis of the data, predict relative significances of single variables which reflect general tendencies but they remarkably underestimate the actual values. The observed tendency of the ANOVA-based significance analysis can therefore lead to unjustified elimination of a variable from a predictive-type model. Exemplary results are presented in fig. 10.4.

The developed methodology of finding the significances of process parameters appeared to be also promising for an industrial case. In fig. 10.5 some results obtained for ductile cast iron melting process are presented, where the input variables are the contents of chemical elements and the output was its tensile strength.

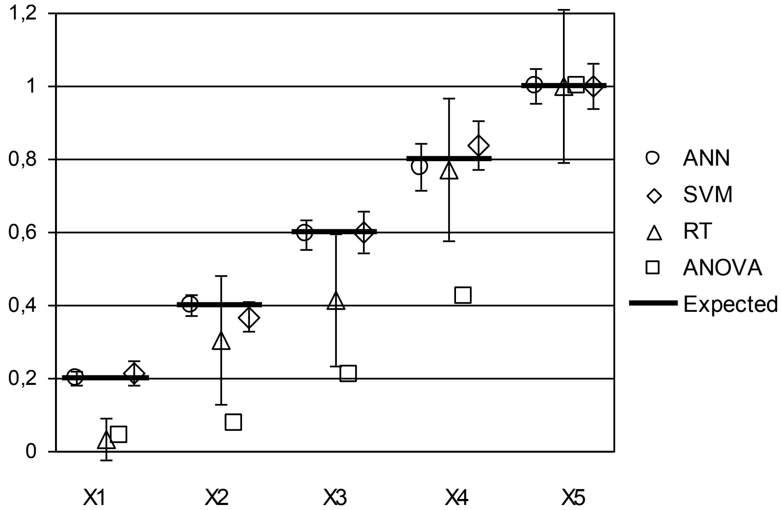


Fig. 10.4 Comparison of relative significance factors obtained from regression models using various learning systems and ANOVA for the simulated data set generated from the basic formula Y=X1+2·X2+3·X3+4·X4+5·X5;
the scatter bars are calculated as average deviations resulting from randomly set values of the other variables [17]

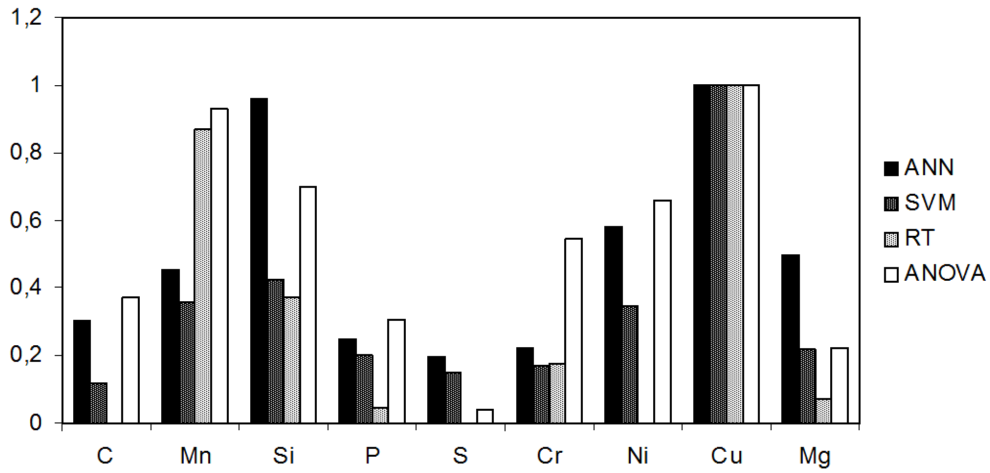


Fig. 10.5 Relative significance factors for the industrial data set (tensile strength of ductile cast iron vs its chemical composition defined by 9 elements) obtained from regression models using various learning systems and ANOVA [18]

For the classification-type tasks the results appeared to be substantially different [18]. The overall best performance in determination of relative significances of input variables revealed a simple statistical method, namely that based on contingency tables. For the simulated data, the performances of the advanced CI models, such as ANNs and SVMs as well as in several cases DTs, appeared to be worse for the simulated data sets. The generally accepted view, that any problem that can be solved with traditional modeling or statistical methods can most likely be solved more effectively with a neural network [13], is probably based on the neural models behavior in different situations from that appearing in the present study or for another expectations of their performance.

The performance of Naïve Bayes Classifier (NBC) and the classification models based on the Rough Sets Theory (RST) could be evaluated as acceptable. A summary of the classification errors is presented in fig. 10.6.

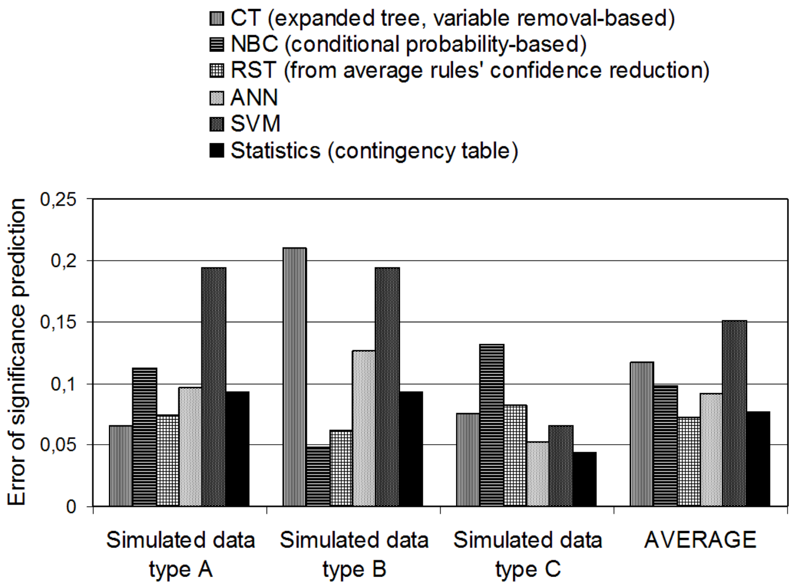


Fig. 10.6 Absolute errors of calculated significances of input variables obtained from various models for various simulated data sets containing 1000 records; simulated data of types A, B and C were generated from three basic formulas, described in [18]

For several industrial data sets, related to the melting process of ductile cast iron and feeding of grey cast iron castings, practically all the models appeared to be satisfactory and the statistical ones were also among the best ones.

10.4 PREDICTIVE CAPABILITIES OF THE TIME-SERIES ANALYSIS APPLIED FOR A FOUNDRY PROCESS

As indicated in Section 2, time-series analysis can be a useful tool in the feedback control of manufacturing processes. Time-series analysis is one of the DM methods, which deals with series of data recorded in a chronological order, usually in regular time intervals or in another sequences. There are two main purposes of that kind of analysis: the discovery of the nature of the given process or phenomenon and prediction of future values. Time-series prediction can be considered as a particular case of the regression task, where the input and output variables are the same quantity but measured at different time moments.

The analysis and prediction of time-series can be done by many different methods. Time-series models have three classical types: Auto Regressive (AR), Integrated (I) and with Moving Average (MA). The compositions of those three classes make the popular autoregressive with moving average models (ARMA) as well the autoregressive integrated with moving average (ARIMA).

An alternative is application for a time-series a generalized regression model, described in detail in [13]. The idea is to utilize a multivariate regression model (in the present work it was a linear regression) in which the input variables are values of the given quantity recorded in several consecutive moments, and the output variable is its next value (i.e. shifted by one measurement from the last input point). The regression model is built for the residual data, i.e. obtained by subtraction form the original data the following components: the means's trend, the variability amplitude trend and the periodical component. The idea of this methodology is to use a regression model for modeling finer changes than those which can be easily described by trends and periodicity.

In this section some most important findings concerning applications of the time-series analysis in predicting future values of melting process parameters are presented, based on the previous works of the current authors [19, 20].

The production data were collected in one of Polish foundries. The quality of the alloy was controlled by its chemical composition in about 0.5 hour time intervals. The general approach is based on the residual data regression modeling, described above. The computations were made using the authors' own software having a wide range of capabilities, including detection of important periodicity in data as well as linear regression modeling of the residual data.

It can be expected, that the results of modeling of the residual data should be valuable if the information content in that data is significant. The software used in the computations provides that kind of information in a verbal form, based on results of two statistical tests: the runs test (also called Wald-Wolfowitz test) and the Durbin-Watson test, both described in [19] and the literature cited there. These messages were converted to a numerical score scale ranging between 1 (residual data are only a noise) and 5 (residual data contain a significant information).

The predictions presented below concern the manganese content in a grey cast iron. Wednesdays in the period of 3 months were chosen as the working days with the largest numbers of measurements. The values recorded before 12 am were used as training data, in which the 5 consecutive measurements were inputs and the 6th one was an output in the linear regression models implemented in the software. The predictions for Mn contents were made for the first measurements after 12 am.

In fig. 10.7 the relative prediction errors are shown. The relative errors do not exceed 40% and their average value is less then 20%. For the industrial melting process it means that, with a great confidence, the operators would be able to predict weather the next value of a given chemical element content will be 'high', 'medium' or 'low'. This would enable them to adjust the current amounts of the additives with a greater accuracy than if they rely only on the current results of chemical analysis.

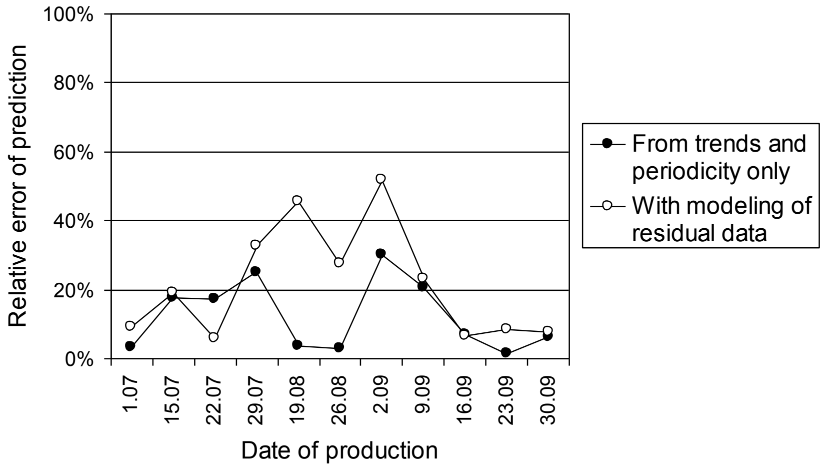


Fig. 10.7 Comparison of prediction errors for Mn contents in gray cast iron, for the first measurements recorded after 12 am, on Wednesdays, during three months

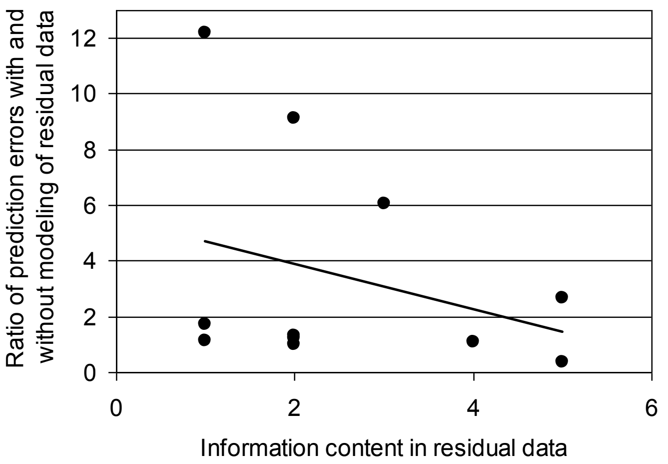


Fig. 10.8 Influence of the information contents in residual data on the effectiveness of regression modeling, data as in fig 10.7

In fig. 10.8 the ratios of the prediction errors obtained with modeling residual data to those obtained only from the both types of trends and periodicity, are presented as a function of the information contents in the residual data.

These results indicate that the linear regression modeling of the residual data may not necessary improve the prediction accuracy. For the Mn content predictions, the residual data modeling increased the prediction errors in most of the cases, compared to those based only on the both types of trends and periodicity. This can be possibly attributed to the low information contents in the residual data; the increase of the information content apparently reduces this negative effect.

Although the prediction accuracy appeared to satisfactory for the control of the considered process, a future research would be desirable in order to analyze the behavior of the regression models for residual data, including more advanced, non linear ones, e.g. artificial neural networks or regression trees.

10.5 ASSESSMENT OF CI MODELS CAPABILITIES TO IDENTIFY PROCESS DISTURBANCES ON SPC CHARTS

Application of CI models to detection and identification of the sequences of points on SPC charts, being signals of an occurrence of process disturbances, was a subject of several works. Most of them utilize ANNs to detect the process mean gradual shifts of various slopes (e.g. [2, 22, 33]) and only few of them consider different patterns of points. In an in-depth study [6] various patterns are analyzed, such as upward trend, downward trend, cyclic, decreased variation around the centerline, systematic up and down and sudden shift. In [7] the standard patterns of the points sequences are considered but the numerical results have not been reported.

In the present work a comparative study of utilization of learning systems for detection and classification of the standard patterns of points on SPC charts was made. The purpose of the numerical tests was to evaluate the capabilities of ANNs and DTs to identify patterns of points, different from the random variations. These tests should be considered as preliminary ones, leading to the application of CI models in situations different from typical ones, e.g. combined disturbances, more subtle disturbances, process-specific disturbances (when the patterns of points are not predefined but obtained from the production experience) and when the results in a stable process are not distributed normally around the actual mean.

10.5.1 Methodology

The sequences of points used in the present study included the 7 standard patterns. Some of them are defined using the notion of three zones above and below the chart centerline, typically denoted as: Zone A – the area between 2σ and 3σ above and below the center line; Zone B – the area between σ and 2σ, and Zone C – the area between the center line and σ, where σ is the standard deviation of the points from the centerline in a stable process. These patterns are defined as follows [26]:

Pattern 1: 9 points in Zone C or beyond (on one side of central line).
Pattern 2: 6 points in a row steadily increasing or decreasing.
Pattern 3: 14 points in a row alternating up and down.
Pattern 4: 2 out of 3 points in a row in Zone A or beyond.
Pattern 5: 4 out of 5 points in a row in Zone B or beyond.
Pattern 6: 15 points in a row in Zone C (above and below the center line).
Pattern 7: 8 points in a row in Zone B, A, or beyond, on either side of the center line.

The testing data were generated in the form of records including 20 points. The sequences of points, meeting one of the above requirements, where generated randomly and were placed at the end of the records. The remaining (preceding) points, were also generated randomly, according to a normal distribution. In Fig. 10.9 the examples of records are shown. For each out-of-control pattern, as well as for the stable process, the points from five records, selected by chance, are plotted. The black arrows point at the beginning of the out-of-control sequences. For comparison, the five examples of records for stable processes, denoted as Pattern 0, are also presented.

Two types of data sets were used in testing. First, the 7 sets, corresponding to the above 7 patterns, were generated, each including 500 records: 250 records of the stable process plus

250 records containing one of the out-of-control patterns.

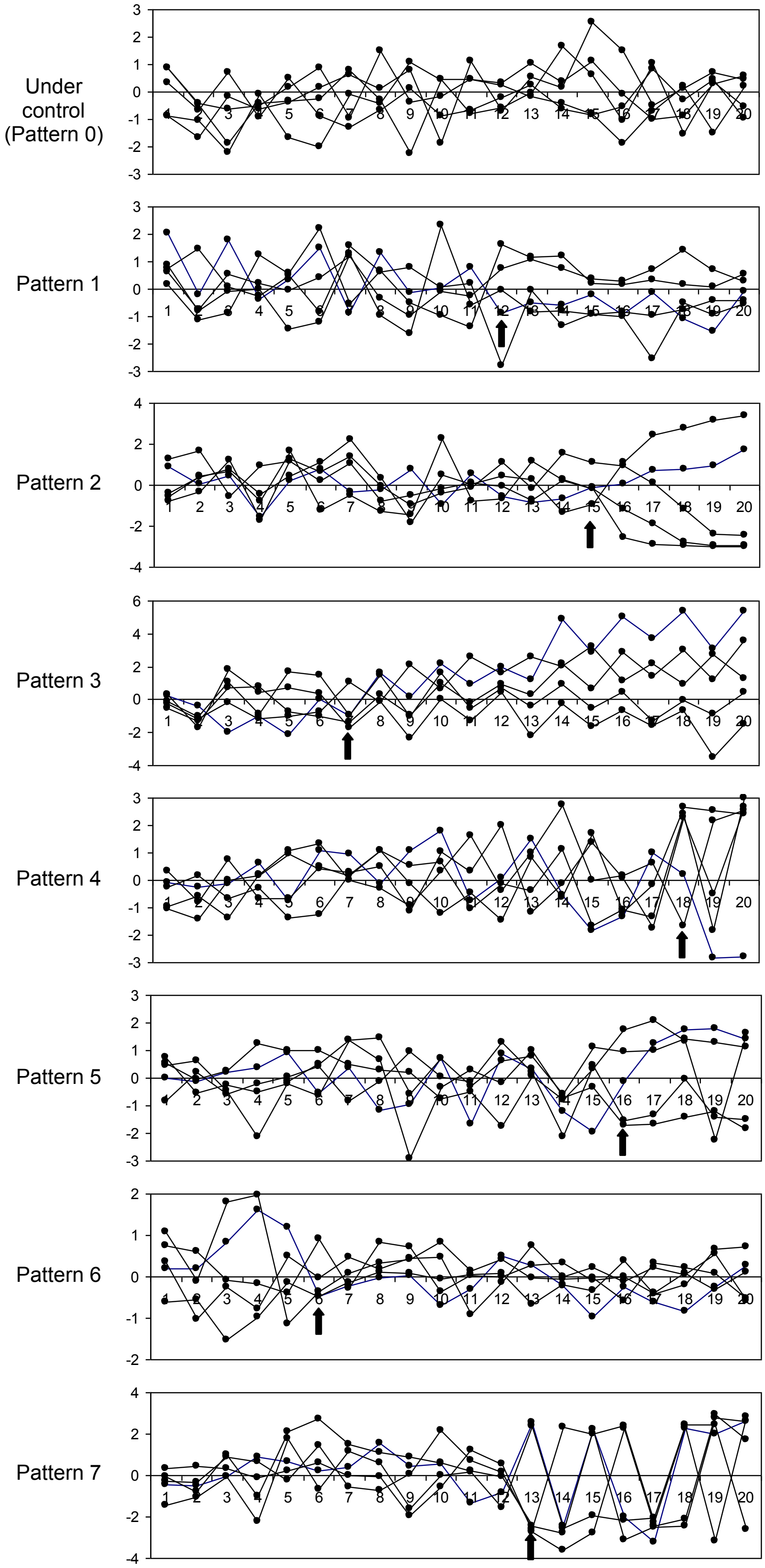


Fig. 10.9 Examples of the records used for testing the ANNs and DTs abilities to identify the out-of-control patterns

The second type of data was in the form of one 2000-records set, including about 87% of records containing all 7 out-of-control patterns and about 13% of the stable process records. The MLP classification-type ANNs were built, using Statistica v8 software, with a random choice of the number of hidden neurons (in one hidden layer) and the logistic transfer function. 10% of records were distinguished as the test subsets used for checking the stopping criterion (increase of the prediction error for these subsets). The best of 100 networks was selected in each case, basing on the lowest prediction errors obtained for the test subsets.

The classification-type DTs were built using the C&RT algorithm and Statistica v8 software. The optimum tree was obtained by application of the 10-fold cross-validation procedure, without setting the minimum number of records in nodes.

In all cases, 10% of records were not involved in the training process and were treated as new data, used for the evaluation of the classification capabilities of the models.

10.5.2 Results presentation and discussion

In fig. 10.10 the misclassification errors are shown for the training data and in fig. 10.11 for the new data. The values shown in the charts show all types of misclassifications, covering the three possible cases: (1) a stable process recognized as an out-of-control process, (2) an out-of-control process recognized as a stable process and (3) a false type of the out-of-control pattern.

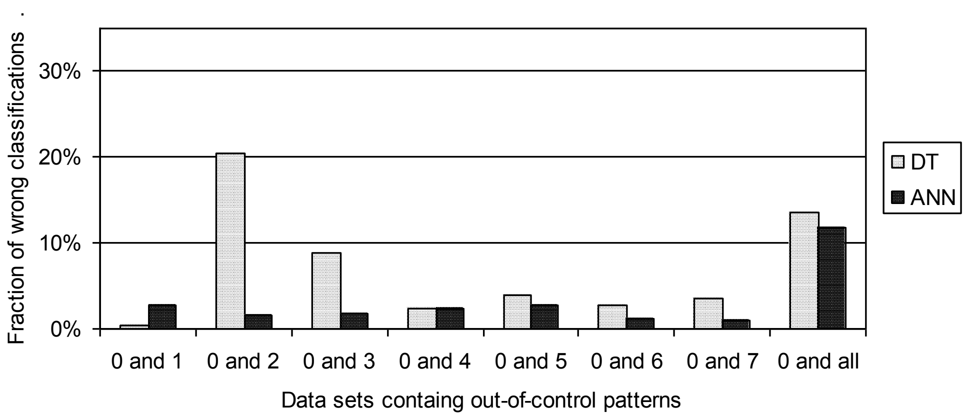


Fig. 10.10 Fractions of misclassified cases for training data

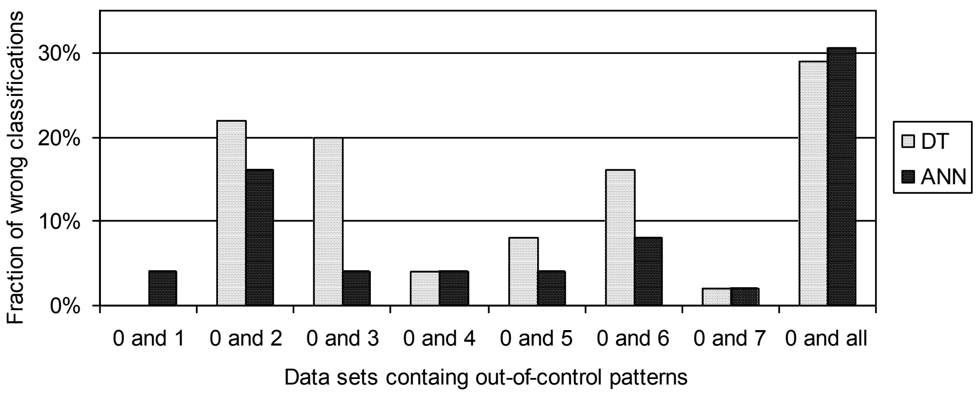


Fig. 10.11 Fractions of misclassified cases for new data

Based on the charts presented in fig. 10.10 and 10.11, the following observations can be made. In all cases the magnitude of the misclassification errors for the new data is obviously larger than those obtained for the training data. The general level of the errors obtained from DTs is comparable with those obtained from ANNs. However, in the majority of cases the misclassification fractions obtained from DTs are larger, which can be interpreted as a result of the lower flexibility or accuracy of this type models, compared to ANNs.

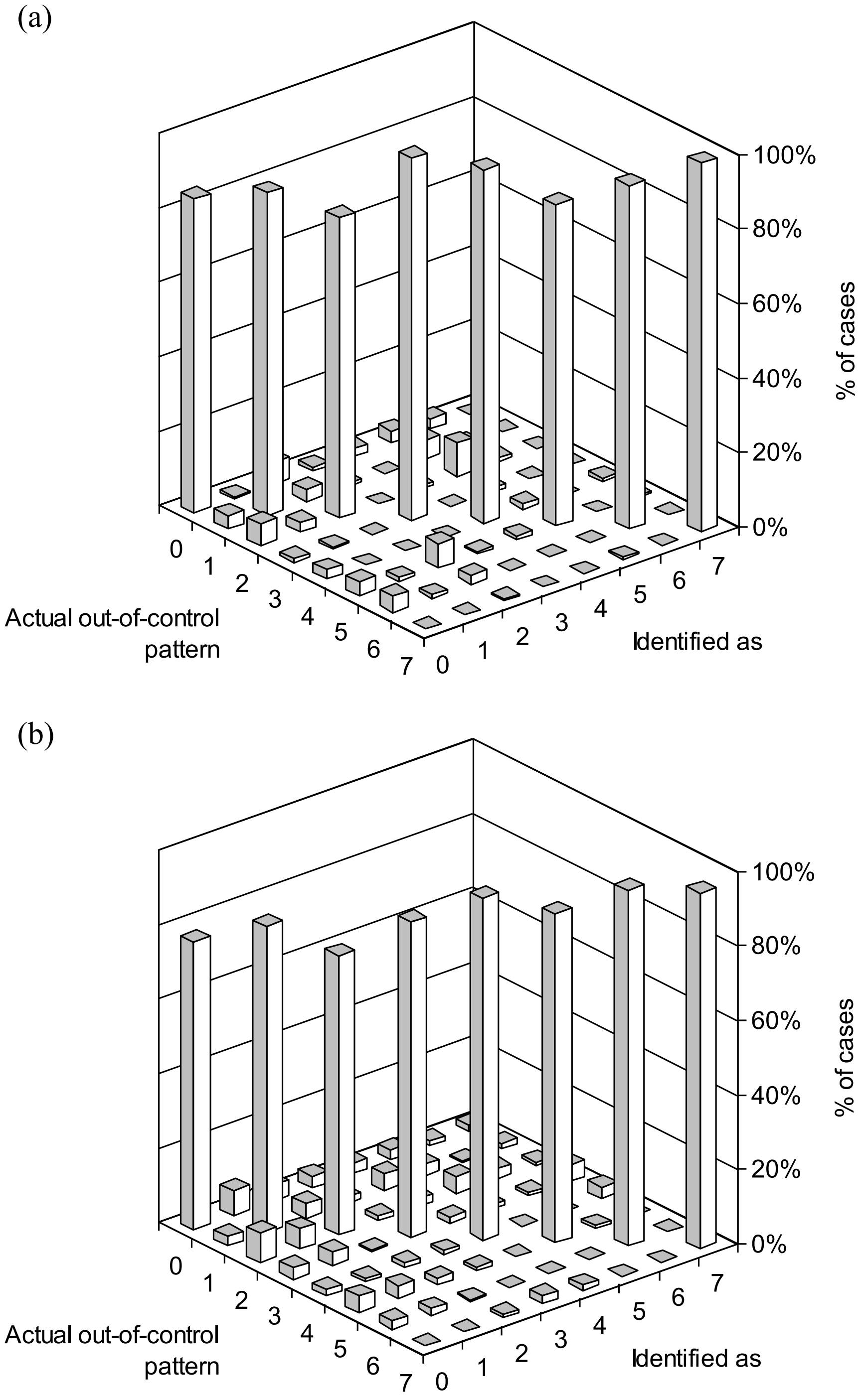
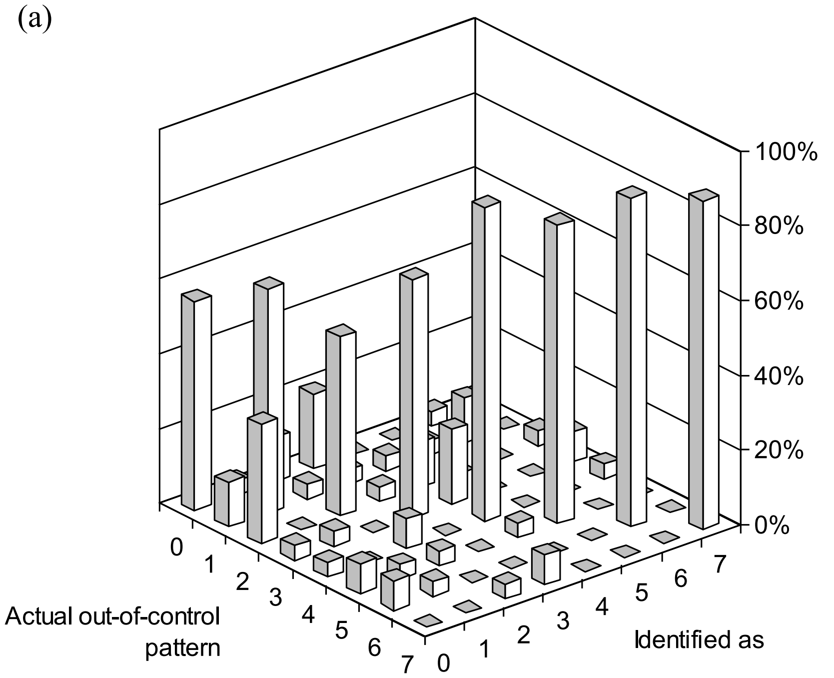


Fig. 10.12 Distribution of correct and incorrect classification fractions obtained for the training data set including all out-of-control patterns: (a) from DTs, (b) from ANNs

(a)



(b)

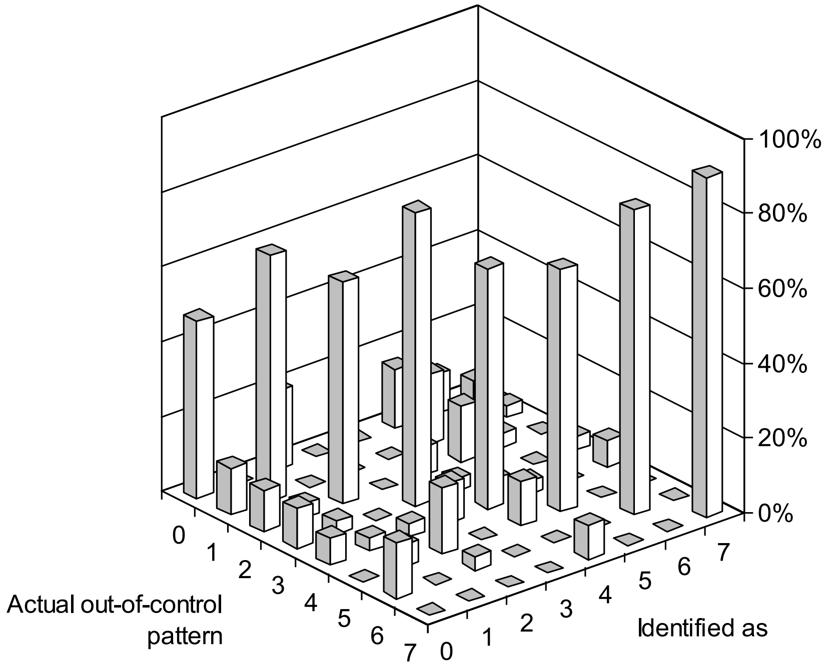


Fig. 10.13 Distribution of correct and incorrect classification fractions obtained for the new data set including all out-of-control patterns: (a) from DTs, (b) from ANNs

In particular, for the new data including single out-of-control patterns number 3, 5 and 6, the classification capability of DTs turned out to be much lower. This does not apply to the most complex data set, i.e. including all out-of-control patterns, where the misclassification fractions for both models are very similar.

In fig. 10.12 and 10.13 the misclassifications types are presented for the data set including all out-of-control patterns.

The distribution of the misclassification types is different for the both models. For example, the largest misclassification fraction for the new data was obtained from DTs in the case where the out-of-control process of the Pattern 2 was recognized as a stable process. For the ANNs, Pattern 5 turned out to be the most difficult and was often classified as Pattern 1 or Pattern 2, or recognized as a stable process.

The general evaluation of the magnitude of misclassification errors should take into account the fact that the classification task was extremely difficult. Referring to the exemplary data shown in fig. 9.9 it should be noticed that, in most cases, points having very different values have to be identified as belonging to this same class. Also, because the out-of-control patterns were of various lengths, some of the points (numbered between 7 and 18) play different roles as the input variables. Another difficulty results from the methodology assumed for the data generation. In the sequences of randomly obtained points for the stable process, some out-of-control patterns could occur accidentally. Similarly, in the records generated for obtaining an assumed pattern, some other patterns could also appear. However, the probability of these undesired instances is rather low and it should not be very significant for the results.

CONCLUSION AND FUTURE WORK

The control and fault diagnosis of manufacturing processes, both discrete and continuous, can be considerably supported by application of advanced, data-driven models, particularly obtained from learning systems and other tools making up the data mining technology. They can be successfully applied both in the Statistical Process Control and Engineering Process Control, in a form of various prediction and identification tasks.

It is a real need to present these possibilities to manufacturing industry and to develop the company-oriented systems. Since selection of the appropriate methodology is often casual, the research aimed at comparative assessments of different types of tools is also desirable.

REFERENCES

1. Chen W. H., Tseng S. S., Hsiao K. R., Liu C. C., A data mining project for solving low-yield situations of semiconductor manufacturing, Proc. IEEE International Symposium on Semiconductor Manufacturing Conference, Boston, MA, USA, May 4-6, 2004, 129-134.
2. Chiu C. C, Shao Y. E., Lee T. S., Lee K. M., Identification of process disturbance using SPC/EPC and neural networks, Journal of Intelligent Manufacturing, 14 (2003), 379-388.
3. Harding J. A., Shahbaz M., Srinivas M., Kusiak A., Data mining in manufacturing: a review, Trans. ASME, J Mfg Sci Engng 128(2006) 969-976.
4. Huang C. H. and Lin Y. N., Decision rule of assignable causes removal under an SPC-EPC integration system, International Journal of Systems Science, 33 (2002) 855-867.
5. Huang H., Wu D., Product quality improvement analysis using data mining: A case study in ultra-precision manufacturing industry, Lect Notes Comput Sci 3614LNAI (2006) 577-580.
6. Hwarng, H. B. and Hubele, N. F., Back-propagation Pattern Recognizers for X Control Charts: Methodology and Performance, Computers and Industrial Engineering, 24 (1993) 219-235.

7. Jacobs D. A., Luke S. R., Training Artificial Neural Networks For Statistical Process Control, Proceedings of the Tenth Biennial University/Government/Industry Microelectronics Symposium, Research Triangle Park, NC, USA, May 18-19, 1993, 235-239.
8. Jiang W. and Farr J. V., Integrating SPC and EPC Methods for Quality Improvement, Quality Technology & Quantitative Management, 4 (2007) 345-363.
9. Kamal A. M. M., A data mining approach for improving manufacturing processes quality control, Proc. 2nd Intern. Conf. on Next Generation Information Technology, ICNIT 2011, Hong Kong, China, 2011, p 207-212.
10. Koonce D., Fang C. H., Tsai S. C., Data mining tool for learning from manufacturing systems, Comput Ind Eng 33 (1997) 27-30.
11. Kusiak A., Data mining: manufacturing and service applications, Int. J. Prod. Res. 44 (2006) 4175-4191.
12. Lewicki P., Hill T., Multivariate Quality Control, Quality Magazine, 2007, April, 38-45. Available from: www.qualitymag-digital.com/qualitymag/200704/.
13. Masters T., Practical neural network recipes in C++, Academic Press, San Diego, 1993, 248-249.
14. Montgomery D. C., Keats J. B., Runger G. C. and Messina W. S., Integrating statistical process control and engineering process control, Journal of Quality Technology; 26 (1994) 79-87.
15. Ozturk A., Kajaligil S., Ozdemirel N. E., Manufacturing lead time estimation using data mining, European Journal of Operational Research 173 (2006) 683-700.
16. Perzyk M., Data mining in foundry production, in Research in Polish metallurgy at the beginning of XXI century, Committee of Metallurgy of the Polish Academy of Sciences, Cracow, Poland, 2006, 255-275.
17. Perzyk M., Biernacki R., Kozlowski J., Data mining in manufacturing: significance analysis of process parameters, J Eng Manuf, Proc Inst Mech Eng Part B, 222 (2008) 1503-1516.
18. Perzyk M., Kochanski A., Kozlowski J., Soroczynski A. and Biernacki R., Applications of data mining to diagnosis and control of manufacturing processes in "Knowledge – Oriented Applications in Data Mining", Kimito Funatsu (editor), ISBN 978-95-307-154-1, Publ. InTech (2011), 147-166.
19. Perzyk M., Krawiec K., Kozłowski J., Application of time-series analysis in foundry production, Archives of Foundry Engineering, 9 (2009), 109-114.
20. Perzyk M., Rodziewicz A., Application of Time-series Analysis in Control of Chemical Composition of Grey Cast Iron, Archives of Foundry Engineering, 12 (2012), 171-175.
21. Shahbaz M., Srinivas M., Harding J. A., Turner M., Product design and manufacturing process improvement using association rules, Proc. IMechE Part B: J. Engineering Manufacture, 220 (2006) 243-254.
22. Shao Y. E., Wu C. H., Ho B. Y., Liu J. F., Identifying the Change Point of a Process with the Integration of SPC Charts and Neural Networks, Proc. Second International Conference on Innovative Computing, Information and Control, 2007. ICICIC '07, Kumamoto, Japan, 400-403.

23. Shen L., Tay F. E. H., Qu L., Shen Y., Fault diagnosis using Rough Sets Theory, Computers in Industry 43 (2000) 61-72.
24. Shu-guang H. E., Li L. I., Er-shi Q. I., Study on the Continuous Quality Improvement Systems of LED Packaging based on Data Mining, Proc. International Conference on Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007, Shanghai, China, 5625-5628.
25. Stanley G. M., Guide to Fault Detection and Diagnosis, White Paper available from: http://gregstanleyandassociates.com/whitepapers/FaultDiagnosis/faultdiagnosis.htm, 2013.
26. StatSoft, Inc. (2012). Electronic Statistics Textbook. Tulsa, OK: StatSoft. WEB: http://www.statsoft.com/textbook/.
27. Stites R., Ward B. Walters R., Defect Prediction With Neural Networks, Proceedings of the Conference on Analysis of Neural Network Applications ANNA ’91, May 29-31, 1991, Fairfax, VA, USA, 199-206.
28. Tsang K. F., Lau H. C. W., Kwok S. K., Development of a data mining system for continual process quality improvement, Proc Inst Mech Eng Part B: J Eng Manuf 221 (2007) 179-193.
29. Tseng T. L., Jothishankar M. C., Wu T., Xing G., Jiang F., Applying data mining approaches for defect diagnosis in manufacturing industry, IIE Annual Conference and Exhibition, Institute of Industrial Engineers, Houston, 2004, 1441-1447.
30. Vazan P., Tanuska P., Kebisek M., Moravcik O., Data Mining Model Building as a Support for Decision Making in Production Management, Advances in Computer Science, Engineering & Applications, Proc. Second International Conference on Computer Science, Engineering and Applications (ICCSEA 2012), New Delhi, India, Vol. 1, May 25-27, 2012, 695-701.
31. Wang K., Applying data mining to manufacturing: the nature and implications, J Intell Manuf, 18 (2007) 487-495.
32. Wardell, D. G., Moskowitz, H. and Plante, R. D., Control charts in the presence of data correlation. Management Science, 38 (1992), 1084-1105.
33. Zobel C. W., Cook D. F., Nottingham Q. J., An augmented neural network classification approach to detecting mean shifts in correlated manufacturing process parameters, International Journal of Production Research, 42 (2004) 741-758.

APPLICATION OF COMPUTATIONAL INTELLIGENCE METHODS IN CONTROL AND DIAGNOSIS OF PRODUCTION PROCESSES

Abstract: *This chapter presents actual and potential applications of advanced data-driven models in control and fault diagnosis of manufacturing processes. Types of process control are discussed and the role of the computational intelligence as well as other data mining methods in them is shown. The main findings of the present authors, based on results of the previous works, are presented. They include the methodologies of determination of relative significances of process parameters and evaluation of prediction capabilities of time-series modeling. Results of a new research, aimed at assessment of capabilities of learning systems to detect out-of-control patterns of points observed in SPC charts, are presented.*

Key words: *manufacturing, process control, fault diagnosis, data mining, computational intelligence*

ZASTOSOWANIE METOD INTELIGENCJI OBLICZENIOWEJ DO STEROWANIA I DIAGNOSTYKI PROCESÓW PRODUKCYJNYCH

Streszczenie: *Niniejsze opracowanie przedstawia rzeczywiste i potencjalne zastosowania zaawansowanych modeli opartych na danych w sterowaniu i diagnostyce usterek procesów wytwarzania. Omówiono rodzaje sterowania procesem oraz pokazano rolę, jaką pełnią w nich metody inteligencji obliczeniowej i inne metody eksploracji danych. Zaprezentowano główne stwierdzenia, do jakich doszli autorzy na podstawie wyników wcześniejszych badań. Obejmują one metody określania istotności względnych parametrów procesu oraz ocenę zdolności predykcyjnych modelowania szeregów czasowych. Przedstawiono także wyniki nowych badań, mających na celu ocenę zdolności systemów uczących się do wykrywania układów punktów na kartach kontrolnych SSP, świadczących o rozregulowaniu procesu.*

Słowa kluczowe: *wytwarzanie, sterowanie procesem, diagnostyka usterek, eksploracja danych, inteligencja obliczeniowa*

prof. dr hab. inż. Marcin PERZYK, dr inż. Jacek KOZŁOWSKI
Warsaw University of Technology
Faculty of Production Engineering
Institute of Manufacturing Processes
ul. Narbutta 85, 02-524 Warszawa
tel. +4822 8499797, +48 509 093 935, e-mail: M.Perzyk@wip.pw.edu.pl,
tel. +48 512 081 953, e-mail: kozlows76@gmail.com
mgr inż. Krzysztof ZARZYCKI
Reckitt Benckiser Production (Poland) Sp. z o.o.
ul. Okunin 1, 05-100 Nowy Dwór Mazowiecki
tel. +48 663 334 552, e-mail: k.zar@wp.pl