MATEC Web of Conferences **368**, 01001 (2022) https://doi.org/10.1051/matecconf/202236801001
NEWTECH 2022

Study on the Application of the Holistic Optimization Method of the Manufacturing Process in the Case of a Reduced Instances Database

Cezarina Afteni[1*[0000-0003-2296-4751]] Mitica Afteni[1,2[0000-0001-8292-1863]] and Gabriel-Radu Frumusanu[1[0000-0003-3038-8758]]

[1] ”Dunarea de Jos” University of Galati, Domneasca str. 111, 800201, Galati, Romania
[2] Rulmenti S.A., Republicii str. 320, 731108, Barlad, Romania
*Cezarina.Afteni@ugal.ro

Abstract. The optimal management of the manufacturing processes is achieved through a set of optimal decisions, which must be made for choosing the best way to follow, every time we find ourselves in a point from which several potential manufacturing paths start. A dedicated method, namely the Holistic Optimization Method has been already developed in this purpose, and validated in a number of studies based on artificial and real instances databases. In the current papers that approach the optimal management of the manufacturing processes, in order to estimate the consequences of a decision, are used known methods, such as: NN modeling, big data analysis, statistics, etc. In all these cases, the database size plays an essential role in terms of estimation quality. The present study aims to prove the feasibility of applying the Holistic Optimization Method when the decision-maker does not dispose of a consistent database. This can be a significant advantage relative to the other methods. The study is performed using an artificially generated instances database in the case of a turning process, and the results obtained are promising.

Keywords: Decision Making, Holistic Optimization Method, Instances Database, Comparative Assessment, Turning Process.

1 Introduction

Today, in the global economy, manufacturing plays a key role. To succeed in the highly competitive global production environment, a company must be able to deliver products that customers request at the requested time. The companies also need to meet the challenges of the transition from mass production to mass customization, as well as low product costs with shorter product life cycles.

The activities or actions that take place at each stage of the manufacturing process are specific to a particular product. The essence of manufacturing management is to lead these activities/actions in order to efficinetly and effectively use the necessary resouces as: raw material, time, energy, personnel for product manufacturing. The production activity management means organizing it by making decisions in accordance with the defined scope. The manufacturing process management means the control of this process. For example, in the case of cutting process, the compensation of the deviations occured is made by changing the reference with a value equal to the difference between the target value and the predicted deviation of a certain parameter/set of parameters. The set-up of the references values is performed when the production system is programmed, so setting them is also a decisions-making act.

The need for evaluation is obvious in both management options (process/activity) and is required whenever: *i)* an analysis ”what if” is performed to adopt an alternative to proceed in manufacturing process case, and *ii)* the characteristics of the task that have the greatest impact on the effect (objectives) must be determined in order to effectively control the manufacturing process.

This evaluation requires the existence of a model, therefore, finding the model of the considered process is a matter of general interest. However, such a model can be complicated by involving many variables, so finding it becomes a difficult job. Additionally, the model applicability is limited – when the premises on which the model was determined modify the model may become useless or, at least – inaccurate.

Model construction involves two stages: *i)* establishment of the model structure, which means, first of all, the selection of the cause-variables by which the effect-variable can be evaluated, and *ii)* model formalization (through the concrete relation linking the effect-variable to the cause-variables) – for example, starting from a parametric model, the parametric values are adjusted until the model properly expresses, in a quantitative way, the causal link.

A large number of techniques for performing the second stage are available in dedicated literature, [1]-[12]. Papers addressing the first stage can also be highlighted, based on different features selection techniques, e.g. [13].

The optimal management of the manufacturing process and, implicitly, the making of optimal decisions can be done by applying a new optimization method, developed exclusively for application in the case of this type of process, namely the holistic optimization, [14]. Holistic optimization generally involves using as process model the history of its operation under similar conditions, in the form of a database.

This paper target is to demonstrate the ability of the holistic optimization method, in general, and of the causal link identification algorithm (as an essential stage of the method application), in particular, to provide convincing results when using a previous cases database of relative small size.

The paper is structured as follows: the next section presents the holistic optimization method and the stages of its application, together with the specific actions. The third section describes the methodology for simulating the application of the causal identification algorithm for a small database. The fourth section is dedicated to results and discussions related to the topic of this study. The last section gives the conclusions.

2 Holistic Optimization Method - HOM

In holistic optimization, the optimization request format it is not predefined. In fact, the desiderate formalization is part of the optimization problem solving.

In manufacturing, the managerial policy imposes the desideratum concerning the process. This can be different for different products. Moreover, the desideratum can change over time even for the same product. At the same time, the desideratum reaching can be evaluated according to various criteria, specific objective functions (effect-variables) can be assigned for each criterion, and for evaluating such a function different sets of arguments (independent cause-variables) can be used. For this reason, the presented method requires this stage for identifying the potential goals, criteria, functions and arguments, among which the most suitable ones will be selected, according to method algorithm presented below, [15].

The optimal decisions about the manufacturing process should be maken based on process models. A process model generally means the relationship between a considered effect-variable and a set of job descriptors (cause-variables). Usually, due to the complexity of the problems, such a model is neither unique nor precisely defined; thus, more or less descriptors may be considered, in different combinations, for the for the same effect-variable.

The proposed HOM consists in succesive performing the folowing loop of actions, (see Fig. 1).

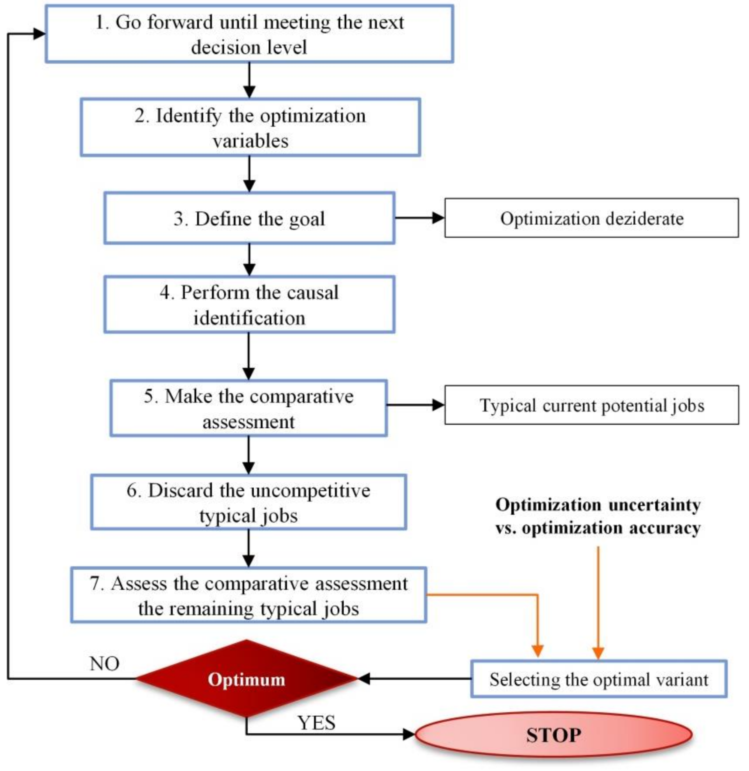

Fig. 1. Flowchart of the HOM application.

In what concerns the choice of the most suitable arguments (job effect-variables), this can be done by instance based causal identification of the manufacturing system, [14], while the comparative assessment between two or more typical jobs can be realized after the values of their effect-variables, according to the method presented in [16].

The causal identification algorithm can be used to find the most suitable structures for the model of a certain manufacturing process. It aims to identify the sets of variables with potential application in manufacturing process modeling.

The main objective in the development of the algorithm is to allow the selection of the most influential, easy to measure and with as few as possible variables, such as the resulting model has the lowest complexity, according to the required level of estimation accuracy.

The method uses the past instances related to the manufacturing system, registered as a database, to reveal the causal link between the variables that characterize the process ongoing on the considered manufacturing system.

The finality of algorithm application is the elaboration of the causal links tree, which can be considered a Decision Support System, DSS, [17]. The causal identification algorithm works on the base of the existing information, by processing a database associated with the manufacturing process (Instances-based learning, IBL, [18]) and involves to go through several succesive stages. The specific actions from these stages are: *i)* process identification, *ii)* data concatening, *iii)* instances comparing, *iv)* variables assessing and *v)* causal models identification, (see Fig. 2).

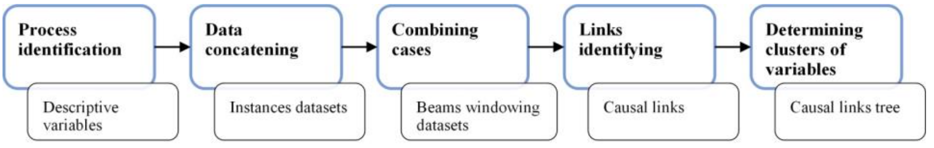

Fig. 2. Causal identification algorithm.

The specific actions to be performed at each stage are presented below.

- *Process identification*

The first step of the algorithm involves analyzing the targeted manufacturing process, in order to choose the variables that characterize its achievement. Thus, the set of variables (both cause-variables and effect-variables) with potential in process modeling are defined.

- *Data concatening*

The purpose of this stage is to generate the database of previous cases, regarding the considered manufacturing process. More cases refer to same type of activity if they can be characterized by the same cause-variables and effect-variables.

Three actions are necessary in order to do data concatening, namely *clustering*, *updating* and *homogenization*, [14].

- *Instances comparing*

The core idea in finding causal models is to look for relations between the variations of cause- and effect-variables (denote by p_i respective q_i), instead to see each instance as event that illustrates the causal relation between these variables, [14]. Variables variations can be revealed by instances comparing. The comparison of k^{th} and l^{th} instances from a certain dataset means to calculate the differences $\delta p_i(k,l)$ and $\delta q_j(k,l)$ between their corresponding variables:

$$\delta p_i(k,l) = \left|p_{ik} - p_{il}\right|,\ i = 1\ldots n_p \text{ and } \delta q_j(k,l) = \left|q_{jk} - q_{jl}\right|,\ j = 1\ldots n_q \qquad (1)$$

In relation (1) n_p and n_q represent the number of cause- respective effect-variables. The result of such a comparison will be further referred as *beam*(k, l). It consists in the reunion of the vectors $\delta p_i(k,l)$ and $\delta q_j(k,l)$. Hereby, the beam(k, l) includes *beam components* (more specific, n_p *cause-components* and n_q *effect-components*).

The instances, as well as the beams resulted by their comparing have identical dimension and similar structure. Thus, from instances as (p_1, p_2, …, p_{i}, p_{np}, q_1, q_2, …, q_{j}, …, q_{nq}) result beams of the same structure, (δp_1, δp_2, …, δp_{i}, δp_{np}, δq_1, δq_2, …, δq_{j}, …, δq_{nq}). For this reason, it will be hereinafter made a natural correspondence between the cause-variable p_i and the cause-component δp_i, as well as between the effect-variable q_j and the effect-component δq_j. Obviously, each instance from the n composing the dataset can be compared to all other n–1. The ensemble of beams resulting after making all possible comparisons represents the *beams dataset*.

Because the beam(k, l) and beam(l, k) are identical (with k, l = 1...n), only one of them is recorded. Hereby, the beams dataset has $N = C_n^2$ lines.

- *Variables assessing*

The purpose of this step is to find and evaluate the dependency relationships between variables with the potential to describe the causal relationship between the cause-variables and the effect-variables.

The variables assesing consists in the succesive application of two procedures:

– *The procedure for dimensionality reduction*, for eliminating the cause-variables with dependence on other cause-variables;
– *The procedure for assessing the modeling potential* of each remaining cause-variable.

The results of applying the first procedure is the cause-variables maximal cluster. Starting of this, based on the values of the specific characteristics that characterize the cause-variables in terms of their modeling capacity, a sub-cluster of the maximal cluster can be generated (simply called clusters), [14].

- *Causal models identification*

The use of modeling potential characteristics defined in the previous step for the cause-variables can be extended to the case of the variable clusters, after the necessary adaptations have been made. Let us consider the case of a causal model whose maximal cluster has n_{mc} cause-variables, p_1, p_2, ... p_{nmc}. This should have, at least in principle, the highest potential of modeling the effect-variable q. However, they might be encountered situations when the values for one or more of cluster variables are not available, or, as well, it might be useless a complicated model, involving all variables from the maximal cluster. In both cases, the solution is to use a causal model defined by fewer cause-variables. This can be realized by successively and repetitively applying a couple of algorithms [19], namely:

– The algorithm for generating smaller clusters, and
– The algorithm for assessing the modeling potential of a cluster.

- *Causal links tree*

Finally, the causal links tree is elaborated. The causal models tree is the representation of causal models concerning the same effect-variable. The representation shows the value of a criterion assessing the modeling potential for each variables cluster.

3 Methodology for simulating the application of the causal identification algorithm for a reduced-size database

Within this chapter, the applicability of the causal links identification algorithm among variables that describe the turning process of a cylindrical part, using a smaller artificially generated instances database is being assesed.

The study was performed by comparing the results of the algorithm application in two cases: a database with 150 instances, and another with 50 instances. The causal identification in the case of the database with 150 lines has already been done and the results presented in the paper [14]. For the application of the algorithm in the addressed case (database with 50 lines) the steps presented in Fig. 2 were followed.

The following set of 11 cause-variables was considered: turned part lenght *L* [mm] and diameter *D* [mm], required level of part accuracy *A* [mm], machinability of part material *M* [mm], rigidity *R* [mm], cutting speed *v* [m/min], feed *s* [mm/rot], cutting depth *t* [mm], main cutting force *F* [daN], power absorbed by lathe *P* [kW], removed chips volume V [cm^3] and 3 effect-variables: machining cost *C* [EURO], machining timespan *TS* [min] and consumed energy *E* [kWh]. The values for the first 2 variables were chosen in the range of variation [30, 300] and [20, 200], the next 3 variables take conventional values in the range 1 to 10. Starting from here, the values for the other 6 cause-variables were calculated with:

$$t = \frac{5.1 \times R - 0.1 \times A}{10} \text{ [mm]}, \tag{2}$$

$$s = \frac{4.4 - 0.4 \cdot A}{10} \text{ [mm/rot]}, \tag{3}$$

$$v = \frac{C_v}{s^{0.3} \cdot t^{0.2} \cdot T^m} \left(\frac{10}{M} \cdot x_v + \frac{R}{10} \cdot y_v \right) \text{ [m/min]}, \tag{4}$$

$$F = C_F \cdot s^{0.8} \cdot t \left(x_F + \frac{M}{10} \cdot y_F \right) \text{ [daN]}, \tag{5}$$

$$P = \frac{F \cdot v}{6000} \cdot \frac{1}{\eta} \text{ [kW]}, \tag{6}$$

$$V = \frac{\pi \cdot D \cdot L \cdot t}{10^3} \text{ [cm}^3\text{]}. \tag{7}$$

In relations (4) and (5) C_v, x_v and y_v, respective C_F, x_F and y_F means constants to which the values are given. Based on these, values were calculated for *C*, *TS* and *E*:

$$C = \frac{V}{v \cdot s \cdot t}\left[\left(1 + k + \frac{\tau_{sr}}{T}\right)c_\tau + \frac{\tau_{sr} \cdot c_\tau + c_s}{T} + \frac{P \cdot c_e}{60}\right] \text{ [Euro]}, \tag{8}$$

$$TS = \frac{V}{v \cdot s \cdot t}\left(1 + k + \frac{\tau_{sr}}{T}\right) \text{ [min]}, \tag{9}$$

$$E = \frac{P \cdot V}{v \cdot s \cdot t} \cdot \frac{1}{60} \text{ [kWh]}. \tag{10}$$

In the relations from above, η means the energy efficiency of the lathe, k – the ratio between the auxiliary time and the machining time, τ_{sr} – the time for worn tool changing [min], T – tool durability [min], c_τ – the wage specific cost [Euro/min], c_s – the tool expenditure [Euro], c_e – the energy price [Euro/kWh].

Using the above formulas, the database of 150 cases was artificially generated at first, [14]. Here, 50 of them were randomly selected. The comparative study of the causal identification algorithm application in both cases was performed to reveal the effect of using a smaller database on the algorithm performance.

4 Results and discussions

The steps of the causal identification algorithm were followed (see Fig. 2). The values in each of the 14 columns (assigned to the 11 cause-variables, and 3 effect-variables) were generated and scaled separately in the range [0, 1]. In the case of the database with 50 lines a combination of $N' = C_{50}^2 = 1225$ lines (beams) results.

In the case of the database with 50 lines for the causal link identification stage, the reference threshold was set at $h_{ref} = h_5 = 0.3277$. The values obtained for Δ_i' using the same MatLab application that was used for the entire database, are shown in Table 1. As it can be seen, $\Delta_{min} = 0.2066$ corresponds to the variable *V*, therefore it can be eliminated. At step 2, the action from previous step is repeated for the remaining ten cause-variables and another one is discarded, namely *P*, and so on. After step 5, $\Delta_{min} = 0.4535 > h_5$, so the seven cause-variables remaining until here can be considered relative independent and the maximal cluster is [*L*, *D*, *A*, *M*, *R*, *v*, *F*], the same maximal cluster as when using the entire database.

Table 1. Images dimensions Δ_i' values.

Cause-variables	Successive steps of dimensionality reduction				
	Step 1	Step 2	Step 3	Step 4	Step 5
L	0.8037	0.8444	0.8444	0.8444	0.8444
D	0.7167	0.9056	0.9056	0.9056	0.9056
A	0.3192	0.3192	0.3192	0.7197	0.7197
M	0.7800	0.7800	0.9189	0.9189	0.9189
R	0.3278	0.3278	0.3278	0.3278	0.6001
v	0.4535	0.4535	0.4535	0.4535	0.4535
s	0.3192	0.3192	0.3192	-	-
t	0.3240	0.3240	0.3240	0.3240	-
F	0.5849	0.5849	0.5849	0.5849	0.5849
P	0.2661	0.2661	-	-	-
V	0.2066	-	-	-	-

One can notice that the actually independent cause-variables (the first five from Table 1) retrieve themselves all in the maximal cluster, which confirms what it was known from the very beginning (when the artificial instances database has been built) and proves the reliability of the proposed method. Another important remark is that only 7/11 cause-variables remained for modeling the effect-variables, which means a significant ease of the modeling problem.

Table 2 shows the results obtained after the causal links identification stage, where the lines containing sets other than those resulting from the data set with 150 cases [14] are shaded in gray. The same clusters were evaluated by implementing the specific algorithm based on both data sets.

The modeling potential of a cause-variables belonging to a given cluster is assessed with one of the criteria: *i) the modeling power MP*, which shows how much the cause-variable variation is found in the effect-variable variation and *ii) the modeling capacity MC*, meaning the measure on which the cause-variable is able to describe the effect-variable. The resulting values for MP_c and MC_c are shown in Fig. 3.

Table 2. The values of a_c, b_c and RMSE.

Clusters variables	a_c	b_c	RMSE
[L, D, A, M, R, v, F]	0.2995	0.0167	0.0049
[L, D, M, R, v, F]	0.2717	0.0243	0.0029
[L, D, A, R, v, F]	0.2596	0.0291	0.0035
[L, D, A, M, v, F]	0.3072	0.0159	0.0068
[L, D, R, v, F]	0.2576	0.0299	0.0030
[L, D, M, v, F]	0.2841	0.0232	0.0053
[L, D, A, v, F]	0.2801	0.0251	0.0038
[L, D, A, M, v]	0.2818	0.0237	0.0050
[L, D, v, F]	0.2969	0.0217	0.0042
[L, D, R, F]	0.2332	0.0310	0.0019
[L, D, M, v]	0.2543	0.0334	0.0052
[L, D, A, F]	0.2436	0.0301	0.0040
[L, D, A, v]	0.2304	0.0400	0.0012

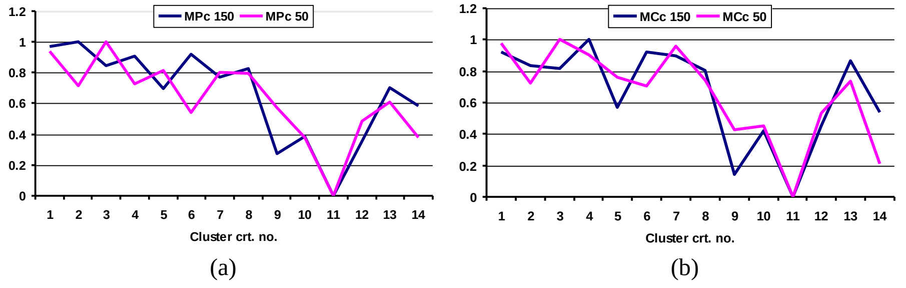

Fig. 3. Comparison between the values of the criteria obtained for the data sets with 150 and 50 cases respectively: (a) Modeling power (MP_c) and (b) Modeling capability (MC_c).

Following the algorithm application of the 50 cases dataset, the following observations can be made:

- The same maximum cluster results after the application of the dimensionality reduction algorithm,
- Most sets of variables (about 2/3) have the same composition in both cases,
- The values of the criteria for assessing the ability to model clusters are different in some cases from those obtained from the extended database, but the monotony of the poles of the lines in Fig. 3 is the same, as are the clusters with extreme behavior.

Despite the low number of cases, one can conclude that the HOM method works with satisfactory results even when the information on the manufacturing process to be modeled is not (very) consistent.

5 Conclusions

At the end of the research presented in this paper, the following conclusions can be drawn:

- The results obtained in implementing the HOM in the addressed case are showing reliability, efficiency in application and a high potential for solving diverse practical problems in manufacturing field optimization.
- The causal identification algorithm also works with satisfactory results when using a database with a smaller number of cases (50 versus 150).
- In both cases studied (50 and 150 data respectively), following the causal identification algorithm application, the same maximum cluster is obtained [L, D, A, M, R, v, F].
- HOM is proving to be a viable alternative to causal modeling for NN modeling methods, which have the disadvantage that their operation is problematic when a small amount of information is available.
- The HOM application accuracy improves on itsown with each new case added to the database, as its size increases.

References

1. Assidjo, E., Yao, B., Kisselmina, K., Amané, D.: Modeling of an industrial drying process by artificial neural networks. In: Brazilian Journal of Chemical Engineering, 25(3), 515-522 (2008).
2. Tafazzoli, E., Saif, M.: Application of combined support vector machines in process fault diagnosis. In: Proceedings of the American Control Conference, pp. 3429-3433, Publisher: IEEE, St. Louis, MO, USA (2009).
3. Deja, M., Siemiatkowski, M.: Machining process sequencing and machine assignment in generative feature-based CAPP for mill-turn parts. In: Journal of Manufacturing Systems 48, 49-62 (2018).
4. Rehman, N.: Data Mining Techniques Methods Algorithms and Tools. In: International Journal of Computer Science and Mobile Computing 6(7), 227-231 (2017).
5. Denno, P., Dickerson, C., Harding, J.A.: Dynamic Production System Identification for Smart Manufacturing Systems. In: Journal of Manufacturing Systems 48, 1-11 (2018).
6. Corne, R., Nath, C., Mansori, M., Kurfess, T.: Study of spindle power data with neural network for predicting real-time tool wear/breakage during inconel drilling. In: Journal of Manufacturing Systems 43, 287-295 (2017).
7. Kotsiantis, S.B.: Supervised Machine Learning: A Review of Classification Techniques. In: Informatica 31, 249-268 (2007).
8. Su, W., Bo, M.: Ant Colony Optimization for Manufacturing Resource Scheduling Problem. In: Wang, K., Kovacs, G.L., Wozny, M., Fang, M. (eds) Knowledge Enterprise: Intelligent Strategies in Product Design, Manufacturing, and Management. IFIP International Federation for Information Processing 207, 863-868, Publisher: Springer, Boston, MA (2006).
9. Song, Y., Huang, J., Zhou, D., Zha, H., Giles, C.L.: IKNN: Informative K-Nearest Neighbor Pattern Classification. In: Kok, J.N., Koronacki, J., Lopez de Mantaras, R., Matwin, S., Mladenič, D., Skowron, A. (eds) Knowledge Discovery in Databases: PKDD 2007. Lecture Notes in Computer Science, 4702, 248-264, Publisher: Springer, Berlin, Heidelberg (2007).
10. Bi, Z.M., Wang, L.: Optimization of machining processes from the perspective of energy consumption: A case study. In: Journal of Manufacturing Systems, 31(4), 420-428 (2012).
11. Rogalewicz, M., Piłacińska, M., Kujawińska, A.: Selection of Data Mining Method for Multidimensional Evaluation of the Manufacturing Process State. In: Management & Production Engineering Review (MPER), vol. 3, no. 2, 27-35 (2012).
12. Sika, R., Ignaszak, Z.: Data acquisition in modeling using neural networks and decision trees. In: Archives of Foundry Engineering 11(2), 113-121 (2011).
13. Feature selection, https://en.wikipedia.org/wiki/Main_Page, last accessed 2022/05/10.
14. Frumusanu, G.R., Afteni, C., Epureanu, A.: Data-driven causal modelling of the manufacturing system. In: Transactions of Famena 45(1), 43-62 (2021).
15. Afteni, C., Frumusanu, G.R., Epureanu, A.: Method for Holistic Optimization of the Manufacturing Process. In: International Journal of Modeling and Optimization 9(5), 265-270 (2019).
16. Afteni, C., Frumusanu, G.R., Epureanu, A.: Instance-based comparative assessment with application in manufacturing. In: IOP Conference Series: Materials Science and Engineering 400(042001) (2018).
17. Decision support system, https://en.wikipedia.org/wiki/Main_Page, last accessed 2022/05/10.
18. Instance-based learning, https://en.wikipedia.org/wiki/Main_Page, last accessed 2022/05/10.
19. Afteni, C.: Holistic optimization of manufacturing process. PhD Thesis, 'Dunarea de Jos' University of Galati, Series I 4: Industrial Engineering, no. 70 (2020).