IOP Conf. Series: Materials Science and Engineering **400** (2018) 042001 doi:10.1088/1757-899X/400/4/042001

Instance-based comparative assessment with application in manufacturing

C Afteni, G Frumuşanu and A Epureanu
"Dunărea de Jos" University, Manufacturing Engineering Department,
Domnească Street 111, 800201 – Galaţi, Romania

E-mail: gabriel.frumusanu@ugal.ro

Abstract. The management of MTO-based production involves multiple decisions needing to be taken. They concern all stages of manufacturing activity – order acceptance, products design, processes planning and jobs scheduling. Such decision consists in selecting, at a given moment, the most suitable alternative of the potential ones. This means that comparative assessment of potential alternatives is required. The usual method underpinning the comparative assessment supposes the direct, separate evaluation of each alternative, before decision making. In this paper we suggest a different approach in performing the comparative assessment, based on alternatives rankings, considered as the most relevant information about them. More specific, this means that rankings are assigned to potential alternatives, by referring them to the cases of already performed manufacturing activities, recorded as past instances database, after ranking criteria such as cost, time span, consumed energy etc. Thus, the selection decision results by comparing potential alternatives rankings. For finding the ranking of a given alternative, a solution to assess the difference between instances is needed, at first. Then, after iteratively defining instances neighborhoods from database and modeling them by multiple nonlinear regression, its ranking is determined. Here, we propose an expression for the distance-function together with an algorithm for actually finding the ranking of the analyzed alternative. A numerical simulation for the instance-based comparative assessment, with the help of an instances artificial database is also presented.

1. Introduction

The management of make-to-order (MTO) production involves multiple decisions needing to be taken. They concern all stages of manufacturing activity – order acceptance, products design, processes planning, machine programming, and operations scheduling. Such a decision consists in selecting, at a given moment, the most suitable alternative among the potential ones. The selection is made after a given criterion, in connection with a feature of potential alternatives whose value is not necessarily known. This selection requires comparative assessment, which aims to establish a relation of order over potential alternatives set. The usual method underpinning the comparative assessment supposes the direct, separate evaluation of feature value for each alternative, before decision-making.

A very common example of available alternatives evaluation concerns the manufacturing cost estimation, which is necessary in quotation process. There are many researches addressing the manufacturing cost estimation problem. The result of these researches consists in a large number of approaches and derived methods developed in this purpose. They can be grouped in two categories: quantitative and qualitative estimation methods [1-3].

Quantitative estimation methods work by analyzing in detail the product, its features and their needed manufacturing processes. Among these methods, there are parametric methods, which use a relation of cost calculus that takes into account the values of main product features [4], statistical methods [5], based on finding the relationship between product features and costs from historical data and empirical tests. There are also analytical methods, which determine product cost as sum of its components costs, found after identifying the relevant processes needed for their manufacturing [6] and semi-analytical methods, according to which, in the first stage, the analogical approach is used to search for analogies between the shapes to be machined [7].

Qualitative estimation methods draw on the experience accumulated in previous manufacturing activities. Case-based systems store the information related to features and costs for previously manufactured products. When cost estimation for a new product is required, a similar previous case is searched, and after finding it, a function is applied for adapting the previous case to needed estimation circumstances. According to [8], these case-based systems work fast, and, as they only consider one part of the known parameters, can offer different alternatives for the unknown parameters based on previous cases. Examples of case-based reasoning can be found in [9, 10].

The most important requirements needing to be satisfied by estimation methods used in decision-making process are promptness, easiness and accuracy. In our opinion, it is hard to say that one of above mentioned methods meet all these attributes. The quantitative estimation methods show good accuracy, but their application is based on processing a high amount of information, with specific character, so when decision object changes, all changes. Thereby, they cannot be too fast and their algorithms are often complicated. Moreover, they are rigid, being used in the same manner no matter the complexity of decision to be taken. The disadvantages of qualitative estimation methods include the difficulty of using general criteria to index the cases and the need for a certain number of base cases, a certain amount of similarity and an adaptation function, so their accuracy may be questionable.

In this paper, we suggest a different approach in performing the comparative assessment, based on assigning rankings to potential alternatives, by referring them to past cases recorded in an instances database. The selection decision further results by comparison of potential alternatives rankings. This method better responds to the above-mentioned requirements, because *the nature, the accuracy and the amount of processed information are adapted to the problem to be solved*, which enables it to work faster and to deliver results at the appropriate level of accuracy. At the same time the method is easy to apply and flexible, being not specific to decision criterion (e.g. cost, time span, consumed energy etc.).

The following section defines the concept of instance-based comparative assessment based on rankings. The third section presents an algorithm for case ranking assignment. The fourth section deals with a numerical simulation of method application, performed on an artificial database, while the last one is dedicated to conclusions.

2. Comparative assessment problem

Basic, the comparative assessment means to establish rankings for two or more alternatives to proceed, after a given criterion. When criterion value is known for all potential alternatives or its evaluation is simple, problem solution is trivial. The issue is much more complicated if criterion value is harder to be found (not to speak about the situation of multi-criteria based decisions).

The enounce of comparative assessment problem addressed here is “being given a set of potential alternatives and a criterion, alternatives rankings are required”. Because criterion values are not known, instead of it, they are used values of some variables in causal relation with this criterion. There is also available a set formed by other values of these variables together with the corresponding values of the related criterion.

Problem solution has been developed starting from the following key-ideas:

- The most relevant information about a process can be reached by recording the information concerning its past deployments in the form of an instances database, here including, obviously, both conditions and results

IOP Conf. Series: Materials Science and Engineering **400** (2018) 042001 doi:10.1088/1757-899X/400/4/042001

- There is no need to know the precise value of criterion for two (or more) potential alternatives in order to compare them after this criterion. Instead of this, the comparison is easier to make by finding neighborhoods of already performed cases (with known results) to which each potential alternative belongs
- The most efficient comparative assessment can be reached by adapting the amount of processed information to required comparison level. Sometimes many potential alternatives can be rejected after relative superficial analysis, while few (often only last two) potential alternatives might ask a more serious effort to decide which the best is.

Problem solution consists in replacing the direct comparison of potential alternatives by successive comparisons between each of them and the cases from instances database, on the base of the variables with known values. The result of such comparison is the ranking of the potential alternative, this being found without directly evaluating the value of the criterion on which the comparison is based.

The key-issue of comparative assessment is the definition of the nearness between two cases, which should be assessed starting from the values of cases known descriptors. The form of nearness function should be actually determined by modeling a set of cases from instances database, appropriately chosen. The ranking of a certain potential case results by analyzing the nearness between it and the ones from instances dataset (sorted after the targeted criterion). Thus, potential case neighborhoods can be delimited and iteratively narrowed, until the resulted ranking becomes precise enough to distinguish between it and the other(s) potential cases in competition.

3. Algorithm for case ranking assignment

The algorithm after which the appropriate ranking is assigned to a given alternative (further referred as *current case*) by comparing it to the cases recorded in *Instances database* is illustrated in figure 1.

In order to facilitate the explanations concerning algorithm functioning, we make the hypothesis that a case is defined by *result* (the scalar variable T), and by three *causes* (the scalar variables x, y and z). In these conditions the instances database means a recorded set of n lines:

$$\left\{ (x_k, y_k, z_k, T_k) \middle| \; k = 1 \ldots n \right\}, \tag{1}$$

where result values T_k, obtained for given values of cause-variables x_k, y_k and z_k, are known (e.g. by measurement). The instances database is here considered as an ordinate set, its lines being sorted after the ascending value of T_k, hence to each line its current number from this list may be associated as ranking. The current case, for which we search the rank, is defined through its cause-variables values x, y and z, while its result T is unknown. The values of variables recorded in database are separately scaled on columns, hence $x_k, y_k, z_k, T_k \in [0,1], \forall k = 1 \ldots n$. The values of x, y and z are also scaled, together with their corresponding column from database.

The algorithm works on the base of two procedures, especially conceived in this purpose: *Neighborhood delimitation* and *Nearness modeling*, which are further presented.

3.1. Neighborhood delimitation

From the very beginning, we need to state that by *similar cases* we mean cases having close values of cause-variables, which is expected to have also close values of result-variable. A certain number of similar cases can be grouped around a *pivot-case* (x_v, y_v, z_v, T_v), hence forming a neighborhood of it. Neighborhood delimitation action targets to find neighborhood profile, defined by the specific combination of *windows* $x_v \pm \delta_x$, $y_v \pm \delta_y$, $z_v \pm \delta_z$, and $T_v \pm \delta_T$ (see figure 2), which include all coordinates sets (x_k, y_k, z_k, T_k) of neighborhood cases.

IOP Conf. Series: Materials Science and Engineering **400** (2018) 042001 doi:10.1088/1757-899X/400/4/042001

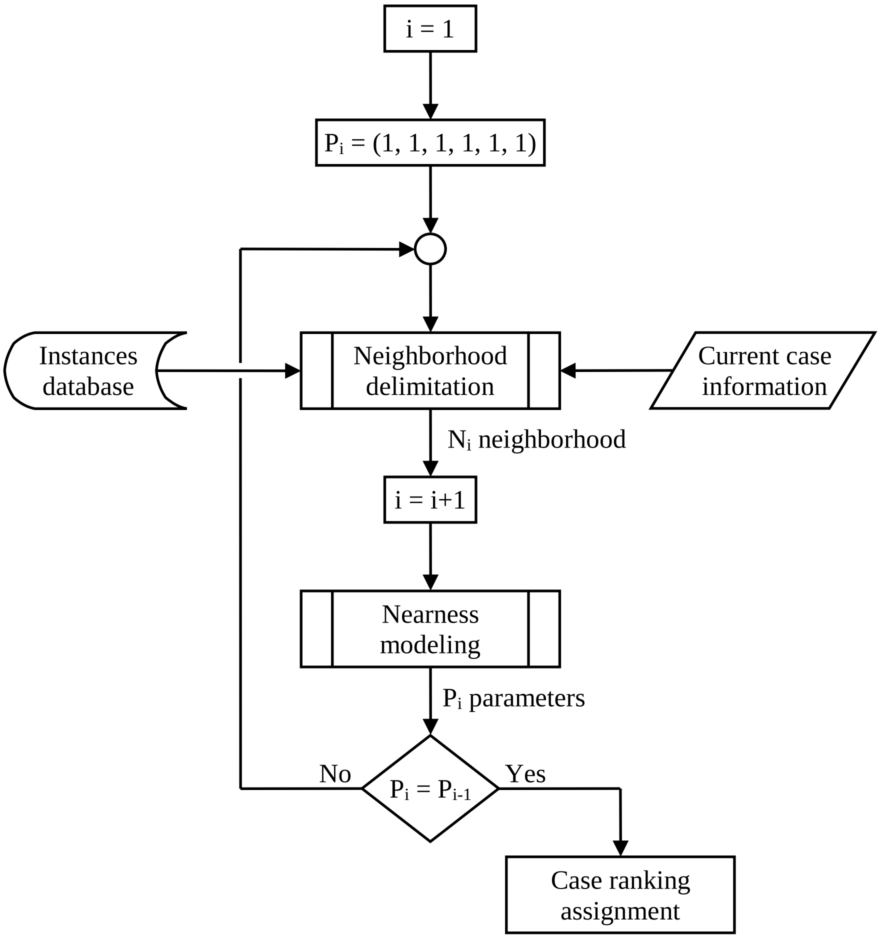

Figure 1. Ranking assignment algorithm.

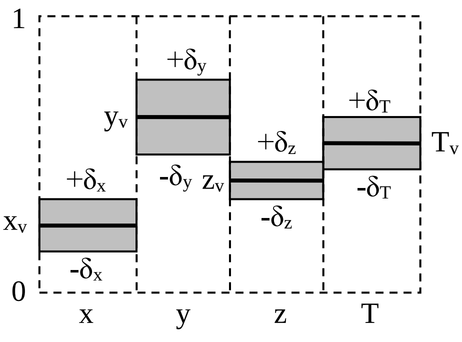

Figure 2. Neighborhood profile.

The values of δ_x, δ_y and δ_z are considered as interdependent, satisfying the relation:

$$A \cdot \delta_x^{\ \alpha} = B \cdot \delta_y^{\ \beta} = C \cdot \delta_z^{\ \gamma} = \varepsilon \,. \tag{2}$$

In relation (2), ε means the *nearness degree* of cases from neighborhood, hereby ε = zero corresponds to identical (coincident) cases, while α, β, γ exponents and A, B, C coefficients mean parameters of nearness function d, characterizing its form, which is considered to be specific for each distinct set of cases, belonging to a given neighborhood. The expression of nearness function, when calculated between current case and database generic case k is:

$$d_k = \operatorname{sgn}(x - x_k) \cdot |x - x_k|^{\alpha} \cdot A + \operatorname{sgn}(y - y_k) \cdot |y - y_k|^{\beta} \cdot B + \operatorname{sgn}(z - z_k) \cdot |z - z_k|^{\gamma} \cdot C\,. \tag{3}$$

In relation (3) *sgn* is the well-known *signum* function (from Algebra). The six parameters can be grouped as coordinates of vector $P = P(\alpha, \beta, \gamma, A, B, C)$, their values resulting by modeling of the considered set of cases.

The target of *Neighborhood delimitation* procedure is to select from *Instances database* the set of cases corresponding to a given profile of neighborhood, resulted for a given value of nearness degree, ε and a given pivot-case (x_v, y_v, z_v, T_v). This can be done by calculating, at first, from (2), the values of δ_x, δ_y and δ_z:

$$\delta_x = \left(\varepsilon / |A|\right)^{1/\alpha},\ \delta_y = \left(\varepsilon / |B|\right)^{1/\beta},\ \delta_z = \left(\varepsilon / |C|\right)^{1/\gamma}, \tag{4}$$

followed by extracting from Instances database the cases whose cause-variables satisfy:

$$x_k \in [x_v - \delta_x, x_v + \delta_x],\ y_k \in [y_v - \delta_y, y_v + \delta_y],\ z_k \in [z_v - \delta_z, z_v + \delta_z]. \tag{5}$$

The value of ε is a parameter of algorithm application, depending on both database dimension and structure, and also on the specific of the comparative assessment problem intended to be solved by case ranking assignment.

When running for the first time this procedure, for a given current case, because any neighborhood is not available, hereby a specific form of nearness function could not be yet identified, cases selection from database (for finding the first version of current case neighborhood) is made by using the nearness function resulted by implicitly setting all six values of P vector to 1.

At i^{th} iteration of procedure, the selection aiming N_i neighborhood delimitation is made by using the nearness function resulted by replacing in (3) the values from current form P_i of parameters vector, as they resulted after nearness modeling on N_{i-1} neighborhood, at previous iteration.

3.2. Nearness modeling

After delimiting the current neighborhood N_i of the current case, the nearness between included cases is modeled in order to find a more precise expression of nearness function, resulted for a new set of its parameters values, $P_i(\alpha, \beta, \gamma, A, B, C)$. In this purpose, at first, the case from N_i that is closest to current case is chosen as pivot. Then, the coordinates differences $\Delta x_j = x_j - x_v$, $\Delta y_j = y_j - y_v$, $\Delta z_j = z_j - z_v$ and $\Delta T_j = T_j - T_v$ are calculated for each of the n_i cases (x_j, y_j, z_j, T_j) from N_i. Finally, the relation between ΔT, on one side and Δx, Δy and Δz, on the other side, is modeled by *nonlinear multiple regression*, model expression being chosen in connection with nearness function (3):

$$\Delta T = b_4 \cdot \operatorname{sgn} \Delta x \cdot |\Delta x|^{b_1} + b_5 \cdot \operatorname{sgn} \Delta y \cdot |\Delta y|^{b_2} + b_6 \cdot \operatorname{sgn} \Delta z \cdot |\Delta z|^{b3}. \tag{6}$$

The values resulted from modeling for model parameters vector $B = (b_1, b_2, b_3, b_4, b_5, b_6)$ are then transferred to the new form of nearness parameters vector $P_i(\alpha, \beta, \gamma, A, B, C)$, following to be further used for delimitating a new version of current case neighborhood.

Procedures 3.1 and 3.2 are successively run until two consecutive forms of nearness parameters vector P_{i-1} and P_i result identical (see figure 1). At that moment, it may be concluded that the results in applying the rank assignment algorithm are stable. As consequence, the current case result may be calculated by using the last identified form of (6) and current case rank is established after the resulted value of T, by inserting it in the *Instances database* at the suitable position.

4. Numerical simulation

The solution of comparative assessment problem developed in this paper and the algorithm for case ranking assignment have been tested by running a numerical simulation, on an artificial instances database, generated in this purpose. The database has four columns (first three for x, y, and z cause-variables and the last for result-variable, T) and n = 150 lines, see relation (1). The values for each cause-variable were considered as a non-uniform division of [0, 1] interval, separately randomized. The values of result-variable were calculated with relation:

$$T(x,y,z) = 2\cdot x^3 + 3\cdot y^2 + 2.5\cdot z + 0.5\,. \tag{7}$$

4.1. Case ranking assignment

We supposed the current case (x_1 = 0.6, y_1 = 0.2, z_1 = 0.7), needing to be ranked relative to the instances database from above. At first, the pivot (x_{v1} = 0.58889, y_{v1} = 0.18333, z_{v1} = 0.72859, T_{v1} = 0.35724) has been chosen from instances database. Then, the algorithm for case ranking assignment has been iteratively run, the results being presented in Tables 1 and 2. The modeling by nonlinear multiple regression has been performed in MatLab (*Optimization tools* package).

Table 1. Results of applying the ranking assignment algorithm (parameters).

Iteration	A	B	C	α	β	γ	ε	δ_x	δ_y	δ_z	RMSE
1	1	1	1	1	1	1	0.2	0.2	0.2	0.2	0.0155
2	0.62104	0.09293	0.19422	1.1493	0.60595	0.60142	0.0067	0.1440	0.5827	0.1703	0.0179
3	2.0026	0.71333	0.1981	1.7058	1.6861	0.63114	0.0087	0.1590	0.2871	0.2715	0.0166
4	1.5403	0.4389	0.18048	1.5344	1.3975	0.54105	0.0087	0.1536	0.3141	0.2595	0.0166
5	1.5403	0.4389	0.18048	1.5344	1.3975	0.54105	0.0087	0.1536	0.3141	0.2595	0.0166

Table 2. Results of applying the ranking assignment algorithm (neighborhoods).

	Instances database lines											
N_1	11	19	20	35	36	38	64	72	105	109	123	141
N_2	11	19	20	35	36	38	64	89	105	124	125	141
N_3	11	19	20	23	35	36	38	64	105	109	125	141
N_4	11	19	20	23	35	36	38	64	105	109	125	141
N_5	11	19	20	23	35	36	38	64	105	109	125	141

The values for ε parameter have been selected at each iteration such as the current case neighborhood includes the same number of cases (here, 12 cases). The quality of modeling the cases neighborhood by nonlinear multiple regression is revealed by calculating the *Root Mean Square Error* (*RMSE*) parameter, well-known in Statistics. As one can easily notice, the algorithm stabilizes rapidly, after only four iterations – the fifth iteration gives same results as the previous one. As consequence, relation (6) can be used (in the form resulted after last modeling by multiple nonlinear regression) for calculating $\Delta T_1 = T_1 - T_{v1}$. The obtained value is ΔT_1 = -0.02339, hereby T_1 = 0.33385 and considered case ranking is R_1 = 47.

4.2. Comparative assessment

In order to sample the comparative assessment based on case ranking, let us consider two different current cases: first is the one addressed in previous section, while second is (x_2 = 0.25, y_2 = 0.15, z_2 = 0.45). One has to choose from the two cases the one having the smallest value of T result.

The algorithm for case ranking assignment is applied once again, for second potential case, to which the pivot (x_{v2} = 0.26296, y_{v2} = 0.13333, z_{v2} = 0.44271, T_{v2} = 0.16722) is associated from the same instances database. This time, the algorithm stabilizes after only three iterations and delivers the following results: A = 6.4576, B = 0.57096, C = 0.33746, α = 3.644, β = 1.563, γ = 0.89287, with *RMSE* = 0.00742. In the same manner as above, we find ΔT_2 = 0.00511, T_2 = 0.17233 and case ranking is R_2 = 17. In conditions of the addressed problem, we have $R_2 < R_1$ hence second case must be taken.

Despite the things may look very simple, there is still a problem needing to be clarified: what happens when the difference between the values of *T* result is too small. For example, let us suppose one need to choose from the potential case considered at 4.1 and other two potential cases: (x_3 = 0.15, y_3 = 0.25, z_3 = 0.8) and (x_4 = 0.75, y_4 = 0.15, z_4 = 0.55). By applying again the algorithm for case ranking assignment, we find T_3 = 0.33147 (*RMSE* = 0.0127) and R_3 = 46, respective T_4 = 0.34991 (*RMSE* = 0.0136) and R_4 = 48, which must be compared to T_1 = 0.33385 (*RMSE* = 0.0166), R_1 = 47. Because the differences between the values T_1, T_3 and T_4 are of the same range as *RMSE* (0.01...0.02), this means the resulted rankings are uncertain and the three cases might be considered equivalent.

4.3. Conditions of algorithm application
The presented numerical simulation addresses a favorable situation: the result *T* depends on no less and no more than the three causes *x*, *y* and *z*. In practical problems, sometimes it is hard to find exactly the cause-parameters needed for properly modeling the result-variable. In connection to this, there are two possibilities: *i)* one or more cause-variables taken into account for modeling have, in fact, no significant influence onto result-variable, and/or *ii)* one or more cause-variables impacting the result are ignored. The question is how can one realize, when applying the algorithm for case ranking assignment that he is in one of the two unwanted situations from above? Both situations have been simulated successively.

4.3.1. Cause-variable with no influence
We supposed that *y* variable has no influence onto *T* result, hence in relation (7) the coefficient of the term including *y* was annulled before generating the artificial database. Still the algorithm was applied in the form presented in section 3, which supposes that *T* = *T*(*x*, *y*, *z*), in order to assign the ranking to potential case (x_1 = 0.6, y_1 = 0.2, z_1 = 0.7). The same pivot as at subsection 4.1 has been chosen from instances database, namely (x_{v1} = 0.58889, y_{v1} = 0.18333, z_{v1} = 0.72859, T_{v1} = 0.35724). At the first iteration, the modeling by multiple nonlinear regression applied to N_1 neighborhood has found out of range values for the parameters concerning the second term from (6) and consequently from (3): *B* = -64684 and β = 82249, while at the second iteration the modeling of N_2 failed to deliver a result.

As the second term is the one involving *y* variable, we can state that out of range values resulted from modeling for the parameters concerning a term from (6) show that the variable from that term might have not significant impact onto result-variable.

4.3.2. Ignored cause-variable
This time we generated the instances database by adding a fourth cause-variable, *w*, whose values were generated in the same manner as for *x*, *y* and *z*, and by calculating *T* result value after considering a supplementary term in (7):

$$T(x, y, z, w) = 2 \cdot x^3 + 3 \cdot y^2 + 2.5 \cdot z + 1.5 \cdot w^{1.5} + 0.5\,. \tag{8}$$

The algorithm for case ranking assignment has also been applied in the form presented in section 3, which supposes that *T* = *T*(*x*, *y*, *z*). The same potential case and pivot as above have been considered. At the first iteration, for performed modeling, we obtained *RMSE* = 0.0697, which is a value several times higher than the normal ones (around 0.01). This shows that modeling result is bad, a possible cause being the absence of at least one cause-variable from formula (6).

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

- The linear interpolation, which is currently used in modeling diverse problems (inclusive in manufacturing domain), often does not deliver satisfactory results in the case when modeled function depend on several variables
- Within manufacturing domain, when selecting the best alternative to proceed from a set of potential ones, the difference between two cases can not be assessed according to an universal,

unique criterion (such as Euclidean distance, or, more general, Minkowski distance). Instead of this, an adequate criterion, depending on the specific of modeled causal relation, should be used. This paper proposes a parametric criterion, whose expression is adjusted by customization to the problem needing solution.

- There is an inverse proportionality between the difference among the analyzed cases and the assessment-required accuracy: the smaller is the difference, the higher must be the accuracy. The method introduced in this paper, namely the instance-based comparative assessment, enables to make the difference between the analyzed cases with a minimum of both initial information and computational effort. The suggested assessment is an iterative process, which stops when the imposed conditions concerning accuracy are satisfied.
- The maximum accuracy of the presented method, for a given number of cause-variables, is shown by the root mean square error (deviation) in the process of modeling the cases neighborhood. If this limit is reached and the difference between two or more cases still can not be made, then the solution is to declare these cases equivalent or to enlarge the number of cause-variables taken into account for describing the result-variable.

References

[1] Garcia-Crespo A, Ruiz-Mescua B, Lopez-Quadrado JL and Gonzalez-Carrasco I 2009 A review of conventional and knowledge-based systems for machining price quotation *J Intel Manuf* **22(6)** 823-841

[2] Niazi A, Dai J, Balabani S and Seneviratne L 2006 Product cost estimation: technique classification and methodology review *J Manuf Sci Eng* **128(2)** 563-575

[3] Ghelase D, Daschevici L, Marinescu V and Epureanu A 2011 Method for control of the make-to-order manufacturing system on the base of earning power assessment *Int J Adv Manuf Technol* **65(9-12)** 1439-1458

[4] Kirchain R and Field F R 2001 Process-based cost modelling: Understanding the economics of technical decisions *Encyclopedia of Materials Science and Engineering* **2** 1718-1727

[5] Layer A, Ten Brinke E, Van Houten F, Kals H and Haasis S 2002 Recent and future trends in cost estimation *International Journal of Computer Integrated Manufacturing* **15** 499-510

[6] Ruffo M, Tuck C, & Hague R 2006 Cost estimation for rapid manufacturing – laser sintering production for low to medium volumes *Proceedings of the Institution of Mechanical Engineers, Part B (Journal of Engineering Manufacture)* **220** 1417-1427

[7] Bouaziz Z, Ben Younes J, Zghal A 2006 Cost estimation system of dies manufacturing based on the complex machining features *Int J Adv Manuf Technol* **28** 262-271

[8] Duverlie P and Castelain J M 1999 Cost estimation during design step: Parametric method versus case based reasoning method *Int J Adv Manuf Technol* **15(12)** 895-906

[9] Chan F 2005 Application of a hybrid case-based reasoning approach in electroplating industry Expert systems with applications **29** 121-130

[10] Chougule R G and Ravi B 2006 Casting cost estimation in an integrated product and process design environment *Int Journal of Computer Integrated Manufacturing* **19** 676-688