BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES
Vol. 57, No. 3, 2009

Machine-part grouping and cluster analysis: similarities, distances and grouping criteria

J.W. OWSIŃSKI*

Systems Research Institute, Polish Academy of Sciences, 6 Newelska St., 01-447 Warszawa, Poland

Abstract. The paper considers the machine-part grouping problem, as equivalent to partitioning the set of machines and operations into subsets, corresponding to block diagonalisation with constraints. The attempts to solve the problem with clustering methods are outlined. The difficulties encountered are presented, related to (i) ambiguity of formulations; (ii) selection of criteria; and (iii) lack of effective algorithms. These are illustrated in more detail with a limited survey of similarity and distance definitions, and of criteria used, constituting the main body of the paper. The return is proposed to the basic paradigm of cluster analysis, as providing simple and fast algorithms, which, even if not yielding optimal solutions, can be controlled in a simple manner, and their solutions improved.

Key words: group technology, cell formation, machine-part grouping, cluster analysis, agglomerative clustering, similarity definitions, distance definitions.

1. Introduction

Among the formal problems in so-called "flexible manufacturing" a persistent element is constituted by "optimum division of production space" according to machines and operations. This issue is referred to as "group technology cell formation", "machine-part grouping", "cellular manufacturing", "part family & machine cell formation" etc. If, namely, machines used, indexed $i \in I = \{1, \ldots, n\}$, can execute operations from the sets $K_i = \{k_{1i}, \ldots\}$ of cardinality m_i, then, given the financial and physical constraints (buildings, transport) and costs, the problem arises of dividing the set of machines, I, into subsets and assigning to these subsets operations $k \in K$, so as to optimise a production organisation quality criterion. This very general criterion, if maximised, usually reflects "density" of operations within the machine groups established and "sparseness" of operations outside these groups. Additional aspects, complicating the formulation and solution of the general problem are timing, costs, sequencing of operations, possibility of multiplication and costs of machines, as well as various limitations on groups. In what follows, we shall be referring to the general problem outlined as MPG, for machine-part grouping.

Since the beginning of the 1970s attempts were undertaken to solve this problem with the clustering methods. A number of respective studies appeared, in which (i) the quality criterion is formulated in a variety of manners (the differences resulting from both the ambiguity of the problem and consideration of additional aspects); (ii) various solution methods are proposed, starting with the classical algorithms of clustering, through their newer variants, including the use of various metaheuristics, up to the methods only indirectly referring to cluster analysis (e.g. uniquely by the form of the quality criterion).

This paper contains a short survey of the MPG models and the proposed clustering methods, as well as those referring to the notions proper for cluster analysis. The existing difficulties, technical and substantial, are characterised, and certain proposals are forwarded, meant to help in further work in the field. The main body of the paper is constituted by (limited) surveys of similarity or distance definitions used and the criteria proposed and optimised.

2. The domain, fundamental notions and issues

Assume we deal with a set of "machines", numbered by index i, $i \in I = \{1, \ldots, n\}$, and a set of "operations" ("part processes"), numbered by index $k \in K = \{1, \ldots, m\}$. Each „machine" can perform operations from a set $K_i \subseteq K$, with $\cup_i K_i = K$, and, in general (luckily), $K_i \cap K_j \neq \phi$, $\forall i,j \in I$. Assume also that the set of "operations" K exhausts a certain "technological process" (or a set of "technological processes"), having as output a definite product or a group of products.

In quite a natural manner, then, a problem arises of "(optimum) production organisation": to group together machines engaged in similar sets of operations, and to separate machines engaged in different ones. Such production organisation would lead to minimisation of the (unnecessary) „transport" activities, transitions between machines, and other "slack" operations. Let us add that it is unimportant whether the groups of machines obtained ("production cells") are actually spatially isolated or only appropriately spatially organised.

Numbers i and k can be treated, respectively, as indices of rows and columns of a matrix, containing problem data. This is the "incidence matrix", $A = \{a_{ik}\}$, whose elements inform, whether machine i performs operation k ($a_{ik} = 1$), or not ($a_{ik} = 0$). It is easily seen that this formulation of

*e-mail: owsinski@ibspan.waw.pl

the MPG problem can be interpreted as the problem of such ("optimal") permutation of columns and rows of matrix A as to obtain the block-diagonal structure, formed out of the elements $a_{ik} = 1$. The blocks, composed of 1's, situated along the main diagonal of the matrix, would correspond to the machine groups obtained.

Assume we obtained through such a permutation a matrix A^*, resulting from our effort to build possibly homogeneous blocks of 1's along the main diagonal of the matrix, leaving possibly all 0's outside of the blocks. Assume further that the number of such explicit blocks, indexed q, is p^*, so that $q = 1, \ldots, p^*$. We can now classify all entries of A^* in the following manner: (i) 1's inside blocks; (ii) 0's outside blocks; (iii) 1's outside blocks, and (iv) 0's inside blocks. In the light of the previous remarks concerning "optimum organisation" one could formulate in very general terms the following objective function, depending on A^*, and reflecting cost, to be minimised:

$$\text{Total cost } (A^*) = \Sigma_q C(q) + k^{\text{iii}}\Sigma c(ik : \text{1's outside blocks}) + k^{\text{iv}}\Sigma c \ (ik\text{: 0's inside blocks}),$$

where block-proper costs $C(q)$ are composed of block setup cost (which would depend on the magnitude of the block, but less than linearly) and a function of operation costs, c_{ik}, for *ik* belonging to the block. Thus, generally, the less blocks, the better from the point of view of the first component of the cost function, but, of course, the other two components are then bigger. The coefficients k^{iii} and k^{iv} reflect the difference of costs, related to the necessity of performing "isolated" operations, involving, in particular, losses related to internal transport, and to the unused capacity of a block. This formulation, even though only illustrative, serves to emphasise the essential "structural" features of the problem considered.

This illustration introduces, as well, some important aspects of the problem, namely: (i) one should not expect obtaining, in general, the "ideal" solutions, in which there are no 1's outside of blocks and 0's inside them; (ii) the thus formulated problem lacks uniqueness (how to evaluate the potentially different structures A^* obtained?); and, crucial for our purpose here: (iii) the pairs of rows (machines), as well as pairs of columns (operations), situated in the ("well-behaved") output matrices side by side – within the blocks or outside of them – are, as a rule, "similar", while the pairs of rows and columns, situated far away, are clearly different.

Thus, permutation of rows and columns, leading to the block-diagonal structure, appears to yield the desired properties of the solution and feature simplicity, but its implementation requires resolving of a number of basic issues:

(i) **The quality criterion and the ideal structure**. It is natural to evaluate solution A^* with (say, without loss of generality, non-negative and increasing) function $Q(A^*, A)$ of the number of 0's in blocks and 1's outside of them. The "*ideal structure*" would, then have $Q(A^*, A) = 0$. This case is, though, particular, as ("statistically") exceptional, and specialised methods can be applied to obtain it, if we know it exists. Hence, we assume that for the best practicable solutions, $Q(A^*, A) > 0$.

(ii) **Uniqueness of formulation.** This issue is reflected in diversity of forms of $Q(A^*, A)$. Let us note at least the following sources of this ambiguity: (a) do 0's in blocks and 1's outside of them weigh the same? (b) what should be the reference for the numbers of 0's in blocks and 1's outside of them?

(iii) **Similarity of machines and parts.** It is the observation that block-diagonal structure is formed by groups of machines and/or parts mutually "similar" that constitutes the basis of using the paradigm of cluster analysis. Yet, from this intuitive observation to effective algorithms, optimising definite functions $Q(A^*, A)$, it is indeed far.

There is also a very important issue of realistic modelling of aspects, for which the model presented is only a rough approximation. Let us only mention some these aspects: (a) possibility of multiplying the machines: if solution can be improved by using machine i in more than one block (which ought to have a reflection in terms of cost and can be accounted for by applying function $Q(A^*, A)$ from a definite class); (b) use of matrix A containing natural or real numbers, reflecting not just the fact of using machine i in operation k, but also, say, cost or time; (c) constraints, concerning parameters of the blocks, such as number of machines in a block, or number of operations, etc.; (d) requirements (taking also the form of constraints), concerning sequencing of operations (so that permutations of columns k are limited). Note that aspect (b) can be accounted for through appropriate distance definition. Aspects (a) and (c), though, require much more complex approaches.

On the top of these there are, of course, numerous other considerations, of quite specific and detailed nature, like, for instance, the way the parts and subassemblies are moved around production space (transport means, routes, distances, etc.). All this has an impact on the potential concrete formulation of the problem in terms of both the quality criterion and the similarity of machines / parts. Hence, the overall structure of the domain may be perceived as composed of three layers, like in Fig. 1.

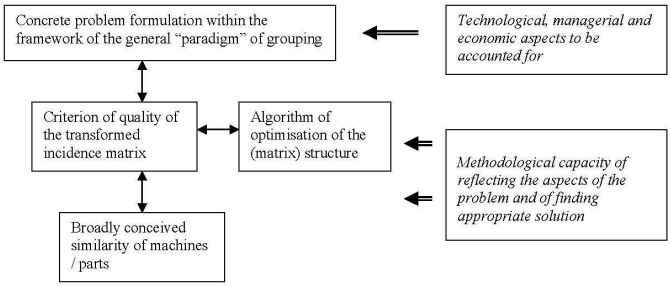

Fig. 1. The overall perception of the problem domain and the elements of modelling and solution approaches

As indicated already, we assume that the general form of the problem is adequately represented by the drive towards block-diagonalisation of the incidence matrix (which is, indeed, a sweeping assumption), and focus on specific issues, associated with two other layers, that is – the one of criteria

and algorithms, with emphasis on clustering, and of similarities between machines / parts. (We shall be using indistinctly the terms "distance" and "similarity" in view of the unique mutual relation between the two, given the strict monotonicity of this mutual relation).

Let us also add that MPG, being a problem in organisation of production space, is closely related to the facility layout problem as it is often formulated and the methods used are often the same or very similar.

3. Attempts to apply clustering

3.1. The concept and its realisation. The fact that the block-diagonal structure corresponds in a way to the structure of the appropriately defined similarities / distances caused already at the beginning of the 1970s an interest in cluster analysis as a potential solution methodology. It appeared that the paradigm of cluster analysis: for a set of given objects divide it into subsets so as to have objects belonging to the same subsets possibly similar and objects belonging to different clusters – possibly dissimilar, fits well the problem solved (similarity and dissimilarity being, of course, expressed through distances). The proposal for such a way of proceeding was forwarded, largely independently, by Burbridge [1], McAuley [2], and Carrie [3]. This started an ample literature on the subject. The studies and the publications were especially abundant in the 1980s. In the middle of the 1990s the publications appeared summing up the work done, achievements and failures (e.g. Cheng, Kumar and Motwani [4]; Crama and Oosten [5]; Sarker and Mondal [6]; Sarker and Khan [7]; Kulkarni and Kiang [8]).

Initial attempts concerned the use of the classical, general clustering algorithms. Yet, they did not fulfil the expectations. Then, specialised methods were developed, mostly special cases of the general algorithms. The consecutive stage consisted in formulation of mathematical programming tasks, modelling possibly precisely the considered aspects of the problem, see, e.g., papers by Kusiak, Vannelli and Kumar and [12], by Askin et al. [13], Veeramani and Mani [14]. Solving of the problems obtained with classical methods of optimisation, though, turned out to be too difficult. The next stage, still underway today, consisted in application of meta-heuristics: tabu search, simulated annealing [15, 16]; genetic algorithms [17–19]; simulated neural networks [8, 20–22], or [23] (fuzzy neural networks); or ant colony (see [24] for the facility layout problem). Regarding the methods used to solve the facility layout problem, mentioned before, it is worth noting that quite a similar reasoning occurred, see, e.g. [25, 26] and [27].

Within the same "paradigm" of solving the MPG problem, hybrid algorithms have been devised (e.g. [28]), as well as quite special clustering procedures (e.g. [29], or [30]). Likewise, specialised heuristics have been designed and applied ([31], or [32], the latter dealing with an incremental cell formation scheme). The continuing use of the fuzzy-set-based clustering is represented, in particular, by [33–35], as well as [36]. There are also, of course, attempts to use graph theoretical approaches, like [37]. In this context let us mention that a whole stream of MPG-oriented methods refer to the p-median problem (e.g. [38, 39]), as we shall also see further on.

3.2. The essential elements of the approach. The publications mentioned, and especially the surveys, contain characterisation of both the approaches associated with the classical methods of cluster analysis, and the consecutive methodological attempts. Here, we should mention apart the papers by Shafer and Rogers, [40–41], devoted to a large extent to the review of the distance / similarity definitions used (altogether 23 such definitions). Some of them are, again, well known from the literature on clustering (Jaccard, Russell and Rao, Dice, Rogers and Tanimoto, etc.), while other were developed specially for MPG. Since the papers by Shafer and Rogers several new proposals for distance definitions appeared (see, e.g. [42, 43], or [44, 45]). Attention should be paid to the paper by Dimopoulos and Mort [46], where an evolutionary algorithm is proposed to evolve distance for classical agglomerative clustering procedures. Within the block layout context, [25] proposes simultaneous use of various distance definitions, in view of different conditions applying to, for instance, various transport means. A more general perspective is outlined in [47], where similarity models and their applications are considered in a general context.

For many MPG methods, especially related directly to clustering, distance definition constitutes, in fact, one of three essential components, indicated in Fig. 1: (A) definition of distance (or similarity, or "proximity"); (B) quality criterion $Q(A^*, A)$; and (C) the method of search for A^*. The components B and C appear in all methods, and are treated in them in a variety of manners. In particular, in methods based on mathematical programming, component B is complemented with constraints, describing the problem.

Yet, a particular problem, appearing in case of application of clustering methods is to consider both "dimensions" of the problem, i.e. machines i and parts-operations k. The basic paradigm of cluster analysis, namely, refers to only one "dimension", i.e. numbering of objects along only one index. This causes the necessity of complementing the respective algorithms with procedures concerning the other "dimension". Alas, such manipulations do not provide unambiguous effects and their computational burden is comparable with that of the initial problem.

This is also largely why in later studies direct application of clustering algorithms was abandoned, first of all to the advantage of mathematical programming and respective methods. Yet, clustering continues to be applied, or at least referred to, in view of the limitations to all the other methodologies (examples are [29, 30, 36, 48]),

At the end of this section let us mention that the entire development of the domain resulted from the difficulties encountered on the way. Also in the most recent period, with the use of the newest methods of optimisation, neural networks and parallel computing, it is possible at most to somewhat improve computational efficiency of solving MPG problems.

The fundamental issues, related to problem formulation and solution methods, remain unchanged. Therefore the studies of summarising and comparative nature search for new definitions of distance and similarity, new criterion functions and optimisation methods.

We shall now proceed to the short surveys of similarity / distance definitions (A), first, and quality criteria (B) of the transformed incidence matrix, thereafter. It is obvious that the two are intimately interrelated, but, despite this, there are virtually no methods, in which this association would really be effectively exploited. This is due to the fact that, like in Fig. 1, there must be an algorithm to optimise, and the algorithms used and developed refer (directly) either to A or to B, but virtually never to both.

4. Similarity and distance definitions

We have already indicated that the problem at hand has two essential dimensions – machines and parts (processes). Yet, it is also true for the vast majority of approaches, whether explicitly using the clustering paradigm, or not, that they overlook this duality. There are two essential aspects to the duality, one merely software-technical and related to reconstruction of the block-diagonal structure from clusters (realisation of permutation), but the other one much deeper, associated with the previously mentioned relation to quality criteria. We shall not consider these here, but simply concentrate on the definitions and their consequences. In this section we shall largely follow Saiful Islam and Sarker, [42], to then develop over their valuable presentation. Thus, we shall be basically considering the distances between machines (machine descriptions), notwithstanding potential applicability to distances between part (part routes or processes), and the (possibly different) consequences thereof.

4.1. The basic properties and definitions. We shall start with examples of machine descriptions in order to refer to them in providing similarity definitions, their interpretations and envisaged consequences. The examples show, in a way, situation for $n = 4$, $m = 8$, as an excerpt from a binary incidence matrix (Table 1). We shall use notation s_{ij} and d_{ij} for, respectively, similarities and distances between machines i and j.

The basic structure of the fundamental set of similarity definitions relies on the following notions:

- $\boldsymbol{a}(i,j)$ – number of 1's on the same positions in i and j (number of parts processed by both machines),
- $\boldsymbol{b}(i,j)$ – number of 1's on definite positions only in i (number of parts processed **only** by machine i of the two),
- $\boldsymbol{c}(i,j)$ – number of 1's on definite positions only in j (number of parts processed **only** by machine j of the two),
- $\boldsymbol{d}(i,j)$ – number of 0's on the same positions in i and j (number of parts not processed by either machine).

The values of these four basic indicators for the example of Table 1 are shown in Table 2.

Table 1
An academic example to illustrate similarity definitions, a_{ik} values for $n = 4$ and $m = 8$

i=	k = 1	2	3	4	5	6	7	8
1	**0**	**0**	**1**	**1**	**1**	**0**	**0**	**1**
2	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**
3	**1**	**1**	**0**	**0**	**0**	**1**	**1**	**1**
4	**1**	**0**	**1**	**0**	**0**	**1**	**1**	**0**

Table 2
Values of the indicators $\boldsymbol{a}(i,j)$, $\boldsymbol{b}(i,j)$, $\boldsymbol{c}(i,j)$ and $\boldsymbol{d}(i,j)$ for the example of Table 1

i,j:	$\boldsymbol{a}(i,j)$	$\boldsymbol{b}(i,j)$	$\boldsymbol{c}(i,j)$	$\boldsymbol{d}(i,j)$	$\boldsymbol{a}(i,j) + \boldsymbol{b}(i,j) + \boldsymbol{c}(i,j) + \boldsymbol{d}(i,j) = m$
1,2	2	2	0	4	8
1,3	1	3	4	0	8
1,4	1	3	3	1	8
2,3	0	2	5	1	8
2,4	0	2	4	2	8
3,4	3	2	1	2	8

Let us also add at this point that the notion of similarity (and, implicitly, of distance) is used in the studies reported in quite a "relaxed" manner, meaning that certain (usual) formal requirements (like, e.g., non-negativity) are not necessarily kept to. Given this caveat, Table 3 presents the set of similarity coefficients that are basic for the domain. To make the content of the table clear note yet that it is most often required that the similarity between the machines satisfy the following properties (in further notation we omit indices i and j whenever they are not necessary):

- P1: no mismatch, s_{ij} tends to 1 for $\boldsymbol{b}$ and $\boldsymbol{c}$ close to 0;
- P2: minimum match, s_{ij} tends to 0 (or -1^*) for $\boldsymbol{a}$ and $\boldsymbol{d}$ = 0 (or close to 0);
- P3: no match, $s_{ij} = 0$ (or -1^*) for $\boldsymbol{a}$ = 0;
- P4: complete match: $s_{ij} = 1$ for $\boldsymbol{a}$ = m;
- P5: maximum match: s_{ij} tends to 1 for $\boldsymbol{a}$ + $\boldsymbol{d}$ tending to m, $\boldsymbol{a} \neq 0$ and the higher the $\boldsymbol{a}$, the higher the s_{ij}.

* in reference to previous remark: the so-called non-Jaccardian similarities (see Table 3) do not satisfy the condition of non-negativity – their values range between -1 and $+1$.

The definitions, quoted as illustration in Table 3, can be classified into three groups (i) six "Jaccardian" coefficients, being simple relative measures, displaying varying emphasis on different components of respective vectors (i.e. $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, $\boldsymbol{d}$); (ii) two "non-Jaccardian" coefficients, in which "penalisation" occurs for the mismatches between the machines, represented by $\boldsymbol{b}$ and $\boldsymbol{c}$; and (iii) two coefficients, explicitly developed to satisfy the formal requirements like P1 through P5, and yet other ones.

Following the previous review of literature let us note that Table 3 presents just an excerpt from the exhaustive list of definitions forwarded in the literature, of which there are not only many more, but, as well, devices have been developed to evolve the "most appropriate" parametric definitions within an approach to MPG. All these definitions satisfy (to a different degree) various (sets of) requirements, formal as well as pragmatic, and produce different values of similarity for the same machine descriptions.

Table 3
Machine similarity measures used in MPG

Name of measure	Definition	Satisfaction of properties or their conditions				
		P1	P2	P3	P4	P5
Jaccard	$a/(a+b+c)$	Yes	n.a.	Yes	Yes	n.a.
Russell & Rao	$a/(a+b+c+d)$	$d=0$	n.a.	Yes	Yes	n.a.
simple matching	$(a+d)/(a+b+c+d)$	Yes	Yes	$d=0$	Yes	Yes
Rogers & Tanimoto	$(a+d)/(a+2(b+c)+d)$	Yes	Yes	$d=0$	Yes	Yes
Sorenson	$2a/(2a+b+c)$	Yes	n.a.	Yes	Yes	n.a
Sneath & Sokal	$2(a+d)/(2(a+d)+b+c)$	Yes	Yes	$d=0$	Yes	Yes
Yule	$(ad - bc)/(ad + bc)$	Yes	Yes	Yes	n.a.	Yes
Hamann	$((a+d)-(b+c))/((a+d)+(b+c))$	Yes	Yes	$d=0$	Yes	Yes
Ochiai	$a/[(a+b)(a+c)]^{1/2}$	Yes	n.a.	Yes	Yes	n.a.
Baroni-Urbani & Buser	$(a+(ad)^{1/2})/(a+b+c+(ad)^{1/2}))$	Yes	Yes	Yes	Yes	Yes
Saiful-Islam & Sarker	$(a+(ad)^{1/2})/(a+b+c+d+(ad)^{1/2}))$	No	Yes	Yes	Yes	Yes

Table 4
Values of similarity coefficients for the example of Tables 1 and 2

Name of measure	Definition*	Values of similarity for machines i, j					
		1,2	1,3	1,4	2,3	2,4	3,4
Jaccard	$a/(a+b+c)$	2/4 = 0.50	1/8 = 0.125	1/7 = 0.14	0	0	3/6 = 0.50
Russell & Rao	a/m	2/8 = 0.25	1/8 = 0.125	1/8 = 0.125	0	0	3/8 = 0.375
simple matching	$(a+d)/m$	6/8 = 0.75	1/8 = 0.125	2/8 = 0.25	1/8 = 0.125	2/8 = 0.25	5/8 = 0.625
Rogers & Tanimoto	$(a+d)/(m+b+c)$	6/10 = 0.60	1/15 = 0.07	2/14 = 0.14	1/15 = 0.07	2/14 = 0.14	5/11 = 0.45
Sorenson	$2a/(2a+b+c)$	4/6 = 0.67	2/9 = 0.22	2/8 = 0.25	0	0	6/9 = 0.67
Sneath & Sokal	$2(a+d)/(m+a+d)$	12/14 = 0.86	2/9 = 0.22	4/10 = 0.40	2/9 = 0.22	4/10 = 0.40	10/13 = 0.77
Yule	$(ad - bc)/(ad + bc)$	8/8 = 1.00	− 12/12 = − 1.00	− 8/10 = − 0.80	− 10/10 = − 1.00	− 8/8 = − 1.00	4/8 = 0.50
Hamann	$((a+d) - (b+c))/m$	4/8 = 0.50	− 6/8 = − 0.75	− 4/8 = − 0.50	− 6/8 = − 0.75	− 4/8 = − 0.50	2/8 = 0.25
Ochiai	$a/[(a+b)(a+c)]^{1/2}$	2/2.83 = 0.71	1/4.47 = 0.22	1/4 = 0.25	0	0	3/4.47 = 0.67
Baroni-Urbani & Buser	$(a+(ad)^{1/2})/(m - d + (ad)^{1/2}))$	4.83/6.83 = 0.71	1/8 = 0.125	2/8 = 0.25	0	0	5.45/8.45 = 0.64
Saiful-Islam & Sarker	$(a+(ad)^{1/2})/(m+(ad)^{1/2}))$	4.83/10.83 = 0.45	1/8 = 0.125	2/9 = 0.22	0	0	5.45/10.45 = 0.52

* here, identity $a+b+c+d=m$ is used for simplicity

Now, Table 4 shows the values of the similarity coefficients for the pairs of machines from Tables 1 and 2.

The values, appearing in Table 4, are telling not just because of the differences between various definitions (why should one bother forwarding new definitions, if they were only to reproduce values known for other ones?), but in view of the "reversals", essential for the grouping, or clustering procedures. Thus, while it is obvious that in the example treated the pairs of machines (1,2) and (3,4) are candidates for groups, the degree of similarity in these two groups differs significantly depending upon the definition:

pair (3,4) is more coherent than (1,2) according to 2 measures,

pairs (3,4) and (1,2) are equally coherent according to 2 measures,

pair (3,4) is less coherent than (1,2) according to 7 measures,

should we apply a voting procedure?

So, Tables 3 and 4 illustrate the fundamental aspect that we wish to emphasise here, and to which we shall yet be returning: on the one hand, there is relative facility of forwarding and using various distance and similarity measures, ensuring high degree of flexibility with respect to the potential different concrete problem formulations, but, on the other hand, the choice is not guided by a formal procedure or strict conditions that would relate the problem formulations to the (properties of the) similarity coefficients.

This aspect is also very well visible within the next level of consideration, namely that of quality criteria.

5. Evaluation of solution quality

5.1. Initial notions. Let us introduce the following notations (matrices A and A^* are of dimensions $n \times m$):

G – number of 1's in the matrix,
T – number of 0's in the matrix, i.e. $G+T=nm$,
E – number of 1's outside of the block-diagonal structure A^*,
F – number of 1's in the block-diagonal structure, i.e. $E+F=G$,

V – number of 0's in the block-diagonal structure,
W – number of 0's outside of the block-diagonal structure, i.e $V + W = T$,

hence:
$E + F + V + W = G + T = nm$, and:
$E + W$ = number of elements outside of the block-diagonal structure A^*,
$F + V$ = number of elements in the block-diagonal structure.

Note that, for a concrete problem, G and T are given "parameters" of the problem. In particular, $G/(G + T) \in (0, 1)$ is the density of the matrix. This information is very "superficial", and we would like to know more on the configuration of 1's and 0's in matrix A, and not only their shares. Yet, this particular kind of information is contained in the solution we look for.

The paper by Nair and Narendran [49], contains quite a detailed analysis of the quality indicators, based primarily on the quantities introduced above.

5.2. The quality criteria applied. The simplest indicator, broadly used in literature, is the "share of exceptions" ("exceptional operations"), denoted *SE*, $SE = E/G = E/(E + F)$. It is the share of 1's outside of the block-diagonal structure in the total number of 1's, which is being minimised in the set of admissible matrices A^*, that is – the divisions of G into E and F. Yet, the most known forms of the indicators are "grouping efficiency", *GI*, and "grouping efficacy", *GE*.

The first one is defined as $GI = rG(1) + (1 - r)g(0)$, $r \in [0, 1]$ being the weight coefficient, $G(1) = F/(V + F)$ (share of 1's in the block-diagonal structure, BDS), and $g(0) = W/(E + W)$ (share of 0's outside of BDS). This indicator is, of course, maximised.

Grouping efficacy, on the other hand, is defined as $GE = (1 - SE)/(1 + SV)$, where *SE* is the "share of exceptions", and *SV*, analogously, is the "share of empty operations", i.e. $SV = V/G = V/(E + F)$, that is – the ratio of 0's in BDS to the total number of 1's. By simple transformations we obtain the expression for *GE*, i.e. $GE = (G - E)/(G + V) = F/(G + V) = F/(F + V + E)$.

Another class of BDS quality criteria relates not so much to the entire matrix, or to all the operations, G, as to the obtained "block-diagonal field", BDS $(F + V)$. And so, the counterpart of the "grouping efficiency", related to the "block-diagonal field", is $ge = (1 - E/(F + V))/(1 + V/(F + V)) = (F + V - E)/(F + 2V)$. This indicator was yet further modified in the successive studies to the definitely much more complex parametric form (parameter r), that is:

$$ger = (1 - (rV + (1 - r)E)/(F + V))/$$
$$(1 + (rV + (1 - r)E)/(F + V)) =$$
$$= (F + (1 - r)(V - E))/(F + V(1 + r) + (1 - r)E).$$

This indicator was then transformed to other forms of the BDS quality criteria. In particular, the form

$$gerq = (1 - (rV + (1 - r)(E - Q))/$$
$$(F + V))/(1 + (rV + (1 - r)(E - Q))/(F + V)),$$

was proposed, with yet another variable, defined as follows:

$$Q = 0 \quad \text{for} \quad E \leq F + V,$$
$$Q = E - (F + V) \quad \text{for} \quad E > F + V.$$

This indicator was analysed in detail in [49]. Sarker and Khan [7], consider yet, mainly following other authors, a broader range of criteria, referring to the here introduced notions.

The first of those is the "weighted grouping efficiency", *Ger*, defined as $Ger = (r(G-E))/(r(G+V-E)+(1-r)E)$, with weight r. It can be easily established that $= rF/(r(F + V) + (1 - r)E)$. Another indicator, quite simple, on the other hand, is $SI = 1-E/G$, i.e. $SI = 1-E/(E+F) = F/(E+F)$, that is – the counterpart of the minimized indicator *SE* on the "positive" (maximized) side. The next simple indicator was called "grouping measure", and was defined as $GM = F/(F + V) - E/G$.

Sarker himself proposed the "doubly weighted grouping efficiency measure",

$$GI2 = ((r_1F + (1 - r_1)V)/(F + V))$$
$$((r_2F + (1 - r_2)E)/(F + E)),$$

referring to similar prerequisites as the preceding indicators.

The subsequent indicators have somewhat different prerequisites, such as, for instance, grouping of 1's possibly close to the main diagonal of the matrix A^*, or maximising the similarities of rows and columns. We shall quote here the indicator, corresponding to the former prerequisite, the so-called "clustering measure", *GC*, $GC = (\Sigma_{ik}(\delta_h^2(a_{ik}) + \delta_v^2(a_{ik}))^{1/2})/\Sigma_{ik}a_{ik}$, where $\delta_h(.)$ and $\delta_v(.)$ are functions of distance along, respectively, rows and columns, of the elements a_{ik} of matrix A^* from the diagonal, defined as:

$$\delta_h(a_{ik}) = i - k(n - 1)/(m - 1) - (m - n)/(m - 1)$$
$$\text{for } a_{ik} \neq 0, \text{ and } \delta_h(a_{ik}) = 0 \text{ for } a_{ij} = 0,$$

and, analogously,

$$\delta_v(a_{ik}) = k - i(m - 1)/(n - 1) + (m - n)/(n - 1)$$
$$\text{for } a_{ij} \neq 0, \text{ and } \delta_v(a_{ik}) = 0 \text{ for } a_{ik} = 0.$$

For the sake of completeness, we shall quote yet in this short survey of the quality indicators of the BDS a criterion which gained significant popularity, even though its sense is not directly related to the problem here considered. This is, namely the "bond energy measure", see, e.g., [9], considered in [7] in a normalised version, that is,

$$Gep = (\Sigma_{ik}a_{ik}a_{i,k+1} + \Sigma_{ik}a_{ik}a_{i+1,k})/\Sigma_{ik}a_{ik},$$

where we consider the elements a_{ik} of matrix A^* within the appropriately defined summation limits or assume values of a_{ik} along the borders of the matrix.

A separate group, that we only mention here, is constituted by the indicators based on quality measures defined for the quality of partitioning into groups along rows and columns, usually in the form of a pair of indicators, for rows and for columns.

And so, similarly as for the definitions of distance, the literature of the subject contains numerous indicators of quality

of the BDS, some of them quite complicated and far from intuition. A typical example of such a quite complex indicator is provided in [17]. It refers directly to the division into groups, entailing division into blocks, and accounts for three components: (i) linked with measure of similarity between blocks (though only with respect to one dimension); (ii) linked with the similarity inside blocks (actually: the number of 1's inside blocks), and (iii) the number of blocks.

This quality function is characteristic for the situation here considered in which, not being able of formulating a unique quality criterion, we look for solutions, which satisfy several criteria, considered simultaneously within the framework of one function, or the objective function and constraints. Let us note that the problem considered is indeed very similar to identification of "notions" or "rules" in data matrices, especially if these matrices are binary, or strongly discrete (small number of values of a_{ik}). This is one of the fundamental problems in data analysis (see Bock [50]). No wonder, then, that the objective function analysed in [17] in the context of application of genetic algorithms in solving the problem considered, is analogous to the objective function proposed by Stańczak [51], in the search for rules in a discrete matrix.

5.3. The reasons for and the directions of search. The diversity of formulations of the quality indicators of the BDS stems from the wish of making them match certain intuitive, or more formal, representations of the role of such an indicator. And so, in design of successive forms here quoted, the following prerequisites were taken into consideration:

- non-negativity of the indicator value and/or taking of values from the interval [0,1];
- independence of the problem dimensions, expressed through n and m (e.g. normalisation);
- independence of the "density" of matrix A, expressed through $G/(G+T)$;
- possibly equilibrated weight of the 1's outside of the BDS (E) and 0's inside it (V); yet, with respect to this aspect there is a frequent postulate of regulating the weights of E and V; it is necessary to know a priori the principles of this regulation, as well as the range of the respective coefficient (this applies, in particular, to "grouping efficiency", *GI*, in which, in cases of large and sparse matrices, in order to balance out the influence of values composing E and V, very small values of r should be selected).

Notwithstanding these "detailed" postulates, hard to satisfy simultaneously, there are also two more general ones, applying by no means only to the here considered problem:

- the simplicity and intuitive appeal of the quality indicator as the model of the problem,
- facility of (carrying out, designing the procedure of) optimising with the help of the indicator.

Let us add that since in the majority of methods explicit or implicit distance or similarity measures are used, as shortly commented upon in Sec. 4, it should have been hoped that a correspondence exist between the BDS quality indicators and these distance/similarity definitions, ensuring both better results of the respective procedure, and its more effective working. Alas, owing, on the one hand, to the variety of formulations, supposed to express the same aspect of the concrete problem considered, and lack of formal relations between the quality criteria and similarity definitions, the choices with this respect are made on the basis of "common sense", with little possibility of a priori assessment of the consequences of choices made.

In effect, the whole domain considered is devoted to the attempts of fulfilling the above stipulations. Therefore the multiplicity of empirical, theoretical and comparative studies. In the face of the still existing difficulties, computational and interpretative, as witnessed by the ample literature, it is proposed here to return to the basic paradigm of cluster analysis, from which the studies in this domain started.

In order to illustrate the issues, associated with the development or selection of the methods, we quote the data from Viswanathan [43], and Saiful Islam and Sarker [42], concerning the results of some chosen methods for quite a group of concrete tasks.

Let us explain that the OG method appearing in Table 5 is based on an integer model, reducing, actually to the problem of cluster analysis, while the heuristics compared is a specialised variety of the classical clustering algorithms. The paper [42] is largely devoted to demonstration that the use of an adequate distance definition may significantly improve the results achieved in optimisation.

Table 5
Comparison of selected methods for examples of the MPG problems

Dimension $n \times m$	Grouping efficiency, *GI* [%], methods:				CPU time [sec.]*	
	p-median[1]	p-median modified[2]	OG model[3]	heuristic[4]	OG model[3]	heuristic[4]
5×7	85.62	85.62	85.62	83.00	0.54	0.28
5×7	78.57	78.57	79.61	77.96	0.56	0.27
8×12	85.53	85.53	85.53	85.53	0.69	0.29
8×20	71.72	71.88	72.76	71.88	2.78	0.29
10×20	100.0	100.00	100.00	100.0	0.63	0.29
11×22	87.82	87.82	87.82	88.58	1.21	0.30
14×24	58.63	82.16	82.34	82.34	1.43	0.30
16×43	59.01	81.80	80.04	80.04	3.24	0.37
24×40	100.0	100.00	100.00	100.0	1.49	0.37
24×40	56.77	95.20	95.20	95.20	2.06	0.35
5×6	–	–	90.00	90.00	0.52	0.27
20×35	–	–	88.38	88.38	1.52	0.29
40×100	–	–	95.10	94.33	5.24	0.62
16×30	–	–	76.00	77.50	1.57	0.32
7×11	–	–	88.00	88.00	0.51	0.26
10×25	–	–	83.22	82.35	1.67	0.30
15×30	–	–	64.50	63.60	3.21	0.31
8×10	–	–	96.00	96.00	0.61	0.28
25×55	–	–	100.00	100.0	2.93	0.41
30×60	–	–	100.00	100.0	2.86	0.43
40×60	–	–	100.00	100.0	3.09	0.48
40×70	–	–	99.54	99.54	4.50	0.56

First ten examples after Ref. 43, the subsequent ones after Ref. 42

* IBM-3090-600-E; [1][12]; [2][43]; [3,4][42].

Table 5 shows clearly how limited (if any) are the gains from the use of explicit optimisation, achieved at the cost of increased computational burden (by almost one order of magnitude). Besides, some of these methods fail to a much bigger degree than the simple heuristics. Particularly surprising are the poor results of the p-median method for examples 7, 8 and 10.

6. Application of clustering

6.1. General justification. It is not without reason that the model of clustering problem was at the foundations of the studies from the domain. At least, on the verbal level this model is a very appropriate representation: "to separate complex objects possibly internally coherent, and at the same time possibly differing among themselves".

Besides, the methods of cluster analysis feature a number of characteristics that allow for considering them as appropriate with respect to both a more detailed modelling and to solving the problem considered. These characteristics are:

- flexibility and facility of accounting for various assumptions related to the properties of the problem (e.g., various distance or proximity definitions, various clustering algorithms, etc.);
- simplicity of both the general manner of reasoning and a vast majority of concrete algorithms, allowing for an intuitive understanding of the procedure and the sources of its results;
- possibility of developing and using computationally effective algorithms (at the order of at most O(n^3) or O(n^3m^3), or better);
- possibility of dealing away with the stage of "learning" of "learning sets";
- possibility of applying (choosing) various assumptions and algorithms (see above), yet without the necessity of designing algorithms for the needs of concrete applications (a frequent case with metaheuristics); we put apart, of course, the stage of modelling, where the need of securing adequate fit is obvious.

6.2. The principle of work. We shall comment on the functioning of a clustering general algorithm from the popular group of progressive merger algorithms for the MPG problem. These algorithms start from the situation, in which all objects (here: machines and parts/operations) are considered as separate entities (the initial matrix A). Then, the objects are aggregated, columns or rows, which are most similar. In case of binary or strongly discrete objects this often means aggregating identical objects. Such objects (or their respective elements) ought, indeed, be placed in matrix A^* together in the same blocks (or together outside of blocks).

Let us consider in a bit more detail a certain academic example, to which we shall yet refer. Matrix A of dimensions $n \times m = 8 \times 9$ is as shown below:

i/k	1	2	3	4	5	6	7	8	9
1	1	1	0	1	0	0	0	1	0
2	0	0	1	1	0	0	1	0	1
3	0	0	1	0	0	0	1	1	1
4	0	0	1	1	0	0	0	1	1
5	1	0	0	0	1	1	1	0	0
6	1	1	0	1	0	0	0	1	0
7	1	0	0	1	1	1	1	0	0
8	1	1	1	0	0	0	0	1	0

The density of this matrix is equal G/nm = 33/72 $\cong$ 46%, proper more for an academic example, since the real-life incidence matrices usually feature much lower densities, at 10% or lower.

The first operation in agglomerative clustering is to calculate distances (or proximities). We can define distances either between machines i or between parts k. Assume we use the apparently simplest in this situation distance definition, i.e. $d_{ii'} = \Sigma_k |a_{ik} - a_{i'k}|$, and, analogously, $d_{kk'} = \Sigma_i |a_{ik} - a_{ik'}|$ (in this way we put apart, at least for this exercise, if not in more general terms, the considerations concerning the "best" similarity or distance measure, as illustrated in Sec. 4). Then, we obtain the triangular distance matrices as shown below:

Distances between machines, $d_{ii'}$

$d_{ii'}$	1	2	3	4	5	6	7	8
1	0	6	6	4	6	**0**	5	2
2		0	2	2	6	6	5	6
3			0	2	6	6	**7**	4
4				0	**8**	4	**7**	4
5					0	6	**1**	6
6						0	5	2
7							0	**7**
8								0

Distance between parts, $d_{kk'}$

$d_{kk'}$	1	2	3	4	5	6	7	8	9
1	0	2	**7**	4	3	3	5	4	**8**
2		0	5	4	5	5	**7**	2	6
3			0	5	6	6	4	3	**1**
4				0	5	5	5	4	4
5					0	**0**	2	**7**	5
6						0	2	**7**	5
7							0	**7**	3
8								0	4
9									0

In both matrices the extreme, and close to extreme, values are shown in bold. The progressive merger procedures aggregate first the objects closest to each other, to then aggregate the closest groups (clusters). The solution proper should be reconstructed from the thus arising hierarchy. Without specifying the course of the procedure for the above example we can, on the basis of the basic paradigm, estimate the possible result. It could be composed of three blocks, formed by the following subsets of machines and parts:

Block I: $i = 5, 7$ ($d_{57} = 1$), $k = 5, 6, 7$ (average distance = 1.33)

Block II: $i = 1, 6, 8$ (average distance = 1.33), $k = 1, 2, 8$ (average distance = 2.67)

Block III: $i = 2, 3, 4$ (average distance = 2), $k = 3, 4, 9$ (average distance = 3.33).

We could thus obtain the matrix A^* of the form:

i/k	5	6	7	1	2	8	4	3	9
5	**1**	**1**	**1**	**1**	0	0	0	0	0
7	**1**	**1**	**1**	**1**	0	0	1	0	0
8	0	0	0	**1**	**1**	**1**	0	1	0
6	0	0	0	**1**	**1**	**1**	1	0	0
1	0	0	0	**1**	**1**	**1**	1	0	0
2	0	0	1	0	0	0	**1**	**1**	**1**
3	0	0	1	0	0	1	**0**	**1**	**1**
4	0	0	0	0	0	1	**1**	**1**	**1**

with clearly visible blocks. To obtain such a matrix A^* with clustering algorithms we should answer two already mentioned basic questions: (**i**) how to conduct aggregation simultaneously according to two dimensions of the matrix (or: how to coordinate the aggregations, if they are not simultaneous?); (**ii**) how to choose a solution in the hierarchy, produced by the agglomerative algorithms? Other issues, mentioned before, especially in the context of indicators of quality of the block-diagonal structure, are mainly solved by selecting (**iii**) distances $d_{ii'}$ and $d_{kk'}$, and (**iv**) the method of calculating distances between clusters. Finally, since clustering algorithms can have complexity of O(n^2m^2), or even less, it is possible to (**v**) verify other constraints on the BDS during the functioning of the procedure (acceptance or rejection of aggregations, resulting from the algorithm).

6.3. Relation to the indicators of the quality of block-diagonal structure: the ideal structure. It can be easily noticed that in the case of existence of the ideal structure the respective indicators take extreme values (most often = 0), which, in the terminology of cluster analysis, is equivalent to identification of groups (clusters), corresponding to blocks, in which distances between columns and rows are zero. Identification of such structure (under milder conditions) is one of the basic requirements on clustering methods. This condition is fulfilled by, in particular, the classical agglomerative schemes, but not necessarily so by the procedures from the K-means group.

Such a structure arises in the agglomerative schemes by joining the rows (columns) having zero distances. At a certain instant no more such aggregations can be performed and the rows (columns) would have to be joined having non-zero distances.

Note that such a course of the initial aggregation phase is not exceptional. In many practical problems subsets of identical objects (in terms of descriptions used) are encountered and the first iterations of the algorithms consist in identification of these subsets. In particular, of course, there may be no such steps at all. It is important what kind of structure we do obtain after joining the identical objects. Here of importance are these aspects of the quality indicators, which refer to the magnitudes of blocks (groups), or their number, putting preference on large blocks and/or their small number. Hence, if steps of joining the "identical" objects end up with creation of a small number of small blocks, one can hardly speak of a correct solution, the ambiguity of the very notion of solution put apart.

Let us also note that in case of ideal structure the fact that blocks are not disjoint is of no importance. And so, in matrix A^* from the last example, two first blocks have column no. 1 in common. This does not change the fact that distances inside of the blocks between (in this case) either rows or columns are zero, and similarly extreme values are attained by the indicators of quality of the block-diagonal structure.

6.4. Relation to the indicators of the quality of block-diagonal structure: steps of the procedure. Irrespective of whether on a given step of the agglomerative procedure the objects (groups) are joined, whose distances are zero, or not, it can be posed that if joining occurs for the closest objects, then it is equivalent to such a change in the quality indicator of the BDS, which is the best from the point of view of the preceding structure.

Such a local property can be demonstrated for a broad class of problems, quality indicators of the BDS and agglomerative clustering algorithms. This correspondence is not just a question of mathematical properties of the functions used. It also has a substantial meaning, as we have already indicated that the distance function, used in clustering ought to (and indeed, can) correspond to the desired properties of the structure from the point of view of the respective indicator (therefrom the ample studies like those of Shafer and Rogers [40, 41], or of Saiful Islam and Sarker [42], devoted to analysis of various distance definitions and attempts of designing specialized distanced, endowed with such properties).

Even though we do not dispose (now) of the proof that an appropriately designed agglomerative scheme leads to an optimal solution, or its sufficiently good approximation, the very simplicity of the algorithms allows for seeking solutions satisfying corresponding conditions, including those of quality of the BDS.

6.5. Satisfying the constraints. We have already mentioned various constraints that happen to be formulated for the solution of the flexible manufacturing problem. Certainly, if we use to solve this problem the clustering algorithms, the very block-diagonal form must be treated as a constraint, if it is not obtained directly from the procedure.

The clustering algorithms have polynomial complexity, most often O(n^3) or O(n^4) (in our case – more like *nm*), although some simplified algorithms, especially those from the "data mining" domain, register better results (like O($n\log n$)). Taking into account the usually encountered dimensions of the flexible manufacturing problems, much smaller than for the "real-life" data analysis problems, we gain a significant margin, both for potential improvement of solution quality and for their selection with respect to constraints.

Thus, at every step of the procedure fulfillment of constraints would be checked for the objects joined or the structure obtained. Considering the manner of proceeding in many clustering algorithms, this is a natural procedure. In a wide class of these algorithms, namely, the subsets are formed of the objects the closest to the currently considered one (k nearest neighbours). Such subsets are formed in a natural way during the review of the objects. Selecting among them does not, therefore, significantly weigh on the computational complexity.

We do not mean here, of course, for this approach, verification of fulfillment and selection of objects for fulfillment of very complicated constraints, which occur in the literature and in reality of flexible manufacturing. Such tasks, though, are being usually solved with highly specialised methods, and the solutions obtained can hardly be assessed for the "degree of optimality" (e.g. the question of multimodality of the objective function) or for their sensitivity (increased probability of violating constraints under the change of problem conditions).

In a vast majority of the clustering methods, used to solve the problem considered, the block-diagonal structure is obtained after the proper algorithm terminates its work. Note that this is also the moment, when some constraints can still be checked and their satisfaction attempted, e.g. through "backward movement" in the hierarchy established through the working of the algorithm.

6.6. Once more on ambiguity. Let us return to ambiguity, illustrated with the indicators of quality of the BDS. This issue ought to be considered in the context of computational cost of obtaining a solution and organisation of its use. If, namely, the economic and organisational aspects do not lead to a unique ("standard") problem formulation, then a doubt arises whether it is worthwhile to try to establish the methods of solving accurately the concrete problems or rather to ensure the possibility of effective solving a broad class of problems with the possibility of applying heuristics, leading to solutions with desired properties (values of the quality indicators or fulfillment of constraints).

In this context, we should, of course, assure fulfillment by the approach selected (e.g. based on agglomerative clustering procedures) of certain formal requirements, related to the quality of solutions, even if it were not a precise representation of the indicators of quality of the BDS. This is needed in view of both the potential establishment of correspondence between the problem of clustering and the one of flexible manufacturing, and of the necessity of having an internal criterion of quality of the methods applied. The approach proposed by this author ([52, 53]) satisfies these conditions.

7. Conclusions

Thus, the problem of flexible organization of production space, analysed for some thirty years already with the purpose of finding effective solution methods, remains the object of methodological attempts, originating from various domains. This is caused, first, by the ambiguity of formulation of the problem, in conjunction with the multiplicity of the forms, and, second, the actual lack of effective methods for solving the essential forms of the problem. Since the very beginning application of clustering algorithms was attempted. These attempts, however, were abandoned, in the hope that other methods, allowing for a more precise modelling of various aspects of the problem, shall also yield more accurate solutions. Yet, these efforts proved also to a large extent vain, except for very specialized forms of the problem, or, in any case a partial success was paid for by a high cost. The latter concerns, in particular, such aspects as simplicity, intuitive interpretation, possibility of application to various forms of the problem, flexibility in the sense of facility of change under varying conditions, or lack of necessity of providing learning examples. The clustering algorithms, even if in general do not provide precise solutions to concrete forms of the problem, are characterized by just these aspects. The arguments, brought forward in the paper, suggest the return to application of clustering algorithms, fulfilling appropriate formal requirements, possibly complemented with heuristics, allowing for a more precise solution of the concrete forms of the problem.

At the same time, explicit application of clustering might allow for an appropriate analysis of the essential problems, indicated in this paper, crucial for the success in effective solving of MPG, such as:

- correspondence between similarity definitions and grouping criteria and problem formulations,
- numerical effectiveness of the algorithms accounting for more intricate constraints,
- two-way clustering and quality assessment (machines and parts).

Indeed, these are the challenges to the domain.

REFERENCES

[1] J.L. Burbridge, "Production flow analysis", *Production Engineering* 4, 139–152 (1971).
[2] J. McAuley, "Machine grouping for efficient production", *The Production Engineer* 51, 53–57 (1972).
[3] A.S. Carrie, "Numerical taxonomy applied to group technology and plant layout", *IJPR* 11, 399–416 (1973).
[4] Ch.H. Cheng, A. Kumar, and J. Motwani, "A comparative examination of selected cellular manufacturing clustering algorithms", *Int. J. Operations & Prod. Mgmt*. 15 (12), 86–97 (1995).
[5] Y. Crama and M. Oosten, "Models for machine-part grouping in cellular manufacturing", *IJPR* 34 (6), 1693–1713 (1996).
[6] B.R. Sarker and S. Mondal, "Grouping efficiency measures in cellular manufacturing: a survey and critical review", *IJPR* 37 (2), 285–314 (1999).
[7] B.R. Sarker and M. Khan, "A comparison of existing grouping efficiency measures and a new weighted grouping efficiency measure", *IIE Trans.* 1, 32 (2000).
[8] U.R. Kulkarni and M.Y. Kiang, "Dynamic grouping of parts in flexible manufacturing systems – a self-organizing neural networks approach", *EJOR* 84, 192–212 (1995).

[9] A. Kusiak, A. Vannelli and K.R. Kumar, "Clustering analysis: models and algorithms", *Control and Cybernetics* 15 (2), 139–154 (1986).

[10] K.R. Kumar, A. Kusiak, and A. Vannelli, "Grouping parts and components in flexible manufacturing systems", *EJOR* 24, 387–397 (1986).

[11] A. Vannelli and K.R. Kumar, "A method for finding minimal bottleneck cells for grouping part-machine families", *IJPR* 24, 387–400 (1986).

[12] A. Kusiak, "The generalized group technology concept", *IJPR* 25, 561–569 (1987).

[13] R.G. Askin, S.H. Cresswell, J.B. Goldberg, and A.J. Vakharia, "A Hamiltonian path approach to reordering the part machine matrix for cellular manufacturing", *IJPR* 29 (6), 1081–1100 (1991).

[14] D. Veeramani and K. Mani, "A polynomial-time algorithm for optimal clustering in a special class of $\{0–1\}$ matrices", *IJPR* 34 (9), 2587–2611 (1996).

[15] F.F. Boctor, "A linear formulation of the machine-part cell formation problem", *IJPR* 29 (3), 343–356 (1991).

[16] F.F. Boctor, "The minimum-cost machine-part cell formation problem", *IJPR* 34 (4), 1045–1063 (1996).

[17] A. Stawowy, Z. Mazur, and J. Obrzud, "Application of a genetic algorithm in product grouping"; in *Optimization in Combinatorial Problems*, vol. 1, pp. 185–193, eds A. Straszak, J. Joszczuk-Januszewska, and Z. Nahorski, Gdynia, 1996, (in Polish).

[18] D.F. Rogers and Sh.S. Kulkarni, "Optimal bivariate clustering and a genetic algorithm with an application in cellular manufacturing", *EJOR* 160 (2), 423–444 (2005).

[19] K. Yasuda, L. Hu, and Y. Yin, "A grouping genetic algorithm for the multi-objective cell formation problem", *IJPR* 43 (4), 829–853 (2005).

[20] A. Kusiak and Y.K. Chung, "GT/ART: using neural network to form machine cells", *Manufacturing Review* 4 (4), 293–301 (1991).

[21] S. Kaparth and N.C. Suresh, "Machine-component cell formation in group technology: MACE", *IJPR* 30 (6), 1353–1367 (1992).

[22] I. Arizono, M. Kato, A. Yamamoto, and H. Ohta, "A new stochastic neural network model and its application to grouping parts and tools in flexible manufacturing systems", *IJPR* 33 (6), 1535–1548 (1995).

[23] R.J. Kuo, Y.T. Su, C.Y. Chiu, K.-Y. Chen, and F.C. Tien, "Part family formation through fuzzy ART2 neural network", *Decision Support Systems* 42 (1), 89–103 (2006).

[24] R. Shankar, M. Solimanpur, and P. Vrat, "Ant colony optimization algorithm to the inter-cell layout problem in cellular manufacturing", *EJOR* 157 (3), 592–606 (2004).

[25] A.E. Smith, G. Ozdemir, and B.A. Norman, "Incorporating heterogeneous distance metrics within block layout design", *IJPR* 41 (5), 1045–1056 (2003).

[26] I. Castillo and B. Peters, "An extended distance-based facility layout problem", *IJPR* 41 (11), 2451–2479 (2003).

[27] T. Dunker, G. Radons, and E. Westkämper, "A coevolutionary algorithm for a facility layout problem", *IJPR* 41 (15), 3479–3500 (2003).

[28] D. Lei and Z. Wu, "Tabu search approach based on a similarity coefficient for cell formation in generalized group technology", *IJPR* 43 (19), 4035–4047 (2005).

[29] M.-Sh. Yang, W.-L. Hung, and F.-Ch. Cheng, "Mixed-variable fuzzy clustering approach to part family and machine cell formation for group technology applications", *IJPE* 103 (1), 185–198 (2006).

[30] F. Alhourani and H. Seifoddini, "Machine cell formation for production management in cellular manufacturing systems", *IJPR* 45 (4), 913–934 (2007).

[31] C.O. Kim, J.-G. Baek, and J.-K. Baek, "A two-phase heuristic algorithm for cell formation problems considering alternative part routes and machine sequences", *IJPR* 42 (18), 3911–3927 (2004).

[32] I. Mahdavi, J. Rezaeian, K. Shanker, and Z.R. Amiri, "A set partitioning based heuristic procedure for incremental cell formation with routing flexibility", *IJPR* 44 (24), 5343–5361 (2006).

[33] C.H. Chu and J.C. Hayya, "A fuzzy clustering approach to manufacturing cell formation", *IJPR* 29 (7), 1475–1487 (1991).

[34] N.N.Z. Gindy, T.M. Ratchev, and K. Case, "Component grouping for group technology applications – a fuzzy clustering approach with validity measure", *IJPR* 33 (9), 2493–2509 (1995).

[35] S. Susanto, R.D. Kennedy and J.W.H. Price, "A new fuzzy-c-means and assignment-technique-based cell formation algorithm to perform part-type-clusters and machine-type-clusters separately", *Production Planning and Control* 10 (4), 375–388 (1999).

[36] J. Li, C.-H. Chu, Y. Wang, and W. Yan, "An improved fuzzy clustering method for cellular manufacturing", *IJPR* 45 (5), 1049–1062 (2007).

[37] H. Wang, B. Alidaee, F. Glover, and G. Kochenberger, "Solving group technology problems via clique partitioning", *IJFMS* 18 (2), 77–97 (2006).

[38] Y. Won and K.R. Currie, "Efficient p-median mathematical programming approaches to machine-part grouping in group technology manufacturing", *Engineering Optimization* 36 (5), 555–573 (2004).

[39] Y. Won and K.-Ch. Lee, "Modified p-median approach for efficient group technology cell formation", *Computers & Industrial Engineering* 46 (3), 495–510 (2004).

[40] S.M. Shafer and D.F. Rogers, "Similarity and distance measures for cellular manufacturing. Part I: A survey", *IJPR* 31 (5), 1133–1142 (1993).

[41] S.M. Shafer and D.F. Rogers, "Similarity and distance measures for cellular manufacturing. Part II: An extension and comparison", *IJPR* 31 (6), 1315–1326 (1993).

[42] K.M. Saiful Islam and B.R. Sarker, "A similarity coefficient measure and machine-parts grouping in cellular manufacturing systems", *IJPR* 38 (3), 699–720 (2000).

[43] S. Viswanathan, "A new approach for solving the P-median problem in group technology", *IJPR* 34 (10), 2691–2700 (1996).

[44] K. Yasuda and Y. Yin, "A dissimilarity measure for solving the cell formation problem in cellular manufacturing", *Computers & Industrial Engineering* 39, 1–17 (2001).

[45] Y. Yin and K. Yasuda, "Manufacturing cells design in consideration of various production factors", *IJPR* 40, 885–906 (2002).

[46] C. Dimopoulos and N. Mort, "Evolving knowledge for the solution of clustering problems in cellular manufacturing", *IJPR* 42 (19), 4119–4133 (2004).

[47] G.K. Adil and J.B. Ghosh, "Maximum diversity / similarity models with extension to part grouping", *ITOR* 12 (3), 311–324 (2005).

[48] M.S. Akturk and H.O. Balkose, "Part-machine grouping using a multi-objective cluster analysis", *IJPR* 34 (8), 2299–2315 (1996).

[49] G.J.K. Nair and T.T. Narendran, "Grouping index: a new quantitative criterion for goodness of block-diagonal forms in group technology", *IJPR* 34 (10), 2767–2782 (1996).

[50] H.H. Bock, "Classification and clustering: problems for the future", *New Approaches in Classification and Data Analysis* 1, 3–24 (1994).

[51] J. Stańczak, *Seminar at the Systems Research Institute*, IBS, Warsaw, 2002.

[52] J.W. Owsiński, "On a quasi-objective global clustering method", in *Data Analysis and Informatics*, pp. 293–305, ed. E. Diday, North Holland, Amsterdam, 1984.

[53] J.W. Owsiński, "On a new naturally indexed quick clustering method with global objective function", *Appl. Stoch. Models & Data Analysis* 6, 157–171 (1991).