Part-machine grouping using a multi-objective cluster analysis

Article *in* International Journal of Production Research · August 1996

DOI: 10.1080/00207549608905026

2 authors, including:

Selim Akturk
Bilkent University

INT. J. PROD. RES., 1996, VOL. 34, NO. 8, 2299–2315

Part-machine grouping using a multi-objective cluster analysis

M. S. AKTURK† and H. O. BALKOSE†

In the existing literature, the part-family formation problem is handled either by the coding systems or the cluster analysis. In this study, we propose a new method that will consider both design and manufacturing attributes and operation sequences simultaneously, in conjunction with the related performance measures such as the machine investment, the amount of workload deviations within and between the cells, and the number of skippings. Finally, the proposed method is compared with the similarity coefficient method under different experimental settings and its robustness is checked against the varying system parameters.

1. Introduction

There are many studies related to the part-family and machine-cell formation (PFMCF) problems in the context of cellular manufacturing (CM) systems. In the literature, these studies can be categorized into two major groups, which are the classification and coding (CC) systems and the clustering methods. Coding systems like OPITZ and MICLASS are used to group parts according to their design attributes, geometric features and machining requirements as discussed in Opitz (1972) and Houtzeel (1975), respectively. The clustering methods can be further grouped into the matrix formulation and mathematical programming techniques. The matrix formulation techniques use the binary machine-part incidence matrix as an input, and process on it by rearranging rows and columns according to some measures until visible clusters are formed, such as the rank order clustering method proposed by King (1980). Askin and Subramanian (1987) developed a three-stage cluster analysis heuristic to minimize the variable production, setup, production cycle inventory, work-in-process inventory, fixed machine and material handling costs, where the initial part-families were formed by using the machining requirements only. Seifoddini and Wolfe (1986) utilized the similarity coefficient method (SCM) to compute the pairwise similarity coefficients between the parts and collected parts into groups using a preset threshold value. An overview of similarity and distance measures for solving the cell formation problem can be found in Miltenburg and Zhang (1991), and Shafer and Rogers (1993). The p-median model by Kusiak and Chow (1988) is a good representative of the mathematical programming approaches. Offodile *et al.* (1994) provide a comprehensive review of the CM literature and present an extensive bibliography of the PFMCF problems by citing more than 100 group technology (GT) related works.

GT is a manufacturing philosophy that exploits similarities in product design and production process both to achieve efficiencies in design rationalization and variety reduction, and to reduce setup times, lead time and work-in-process inventory in the shop floor. GT is one of the most important aspects in the design of CM systems. Generally, the basic objective in designing a CM system is the identification of part

Revision received March 1995.

† Department of Industrial Engineering, Bilkent University, 06533 Bilkent, Ankara, Turkey.

families and machine cells based on the similarity of characteristics, representing the basis for the design and implementation process of advanced manufacturing systems such as just-in-time (JIT), flexible manufacturing system (FMS) and computer integrated manufacturing (CIM). It is well known that JIT includes a simplified production line and standardized products. GT can be used to form a family of parts and machines which would lead to standardized products and a simplified production line. A GT cell will provide logistic support for employing pull/kanban material flow in JIT manufacturing systems. In FMS, GT cells form the core of the whole system. Krajewski *et al.* (1987) developed a simulation model to compare a kanban system's performance with that of the reorder point system, and shown that the improving operating conditions were more important than the changing the features of the material planning and control system to achieve improved system performance. A review by Gunasekaran *et al.* (1994) reveal that there is a need to develop GT models and techniques, which will focus on the application to such advanced manufacturing systems. This needs to consider various operational and technological constraints of the manufacturing systems as well as suitable criteria while forming part families and GT cells.

The CC system forms an important vehicle for implementing the GT concept in manufacturing systems. Moreover, a CC/GT database can be an important part of a computer-integrated design and manufacturing system. Hyer and Wemmerlöv (1989) reported the findings of a survey of 53 U.S. users of GT. In the majority of the cases, firms used CC systems as tools in applying GT. Welke and Overbeeke (1988) argue that CM is the best technique available to achieve JIT manufacturing philosophies and report the experiences at Deere & Co., where they group parts into part families by geometric similarities or part likeness. Obviously, a well designed CC system, including both design and manufacturing attributes, will be more comprehensive and facilitate the formation of part families and machine cells more effectively than any approach that uses only machining requirements as an input. Furthermore, a CC system can provide other benefits such that it permits quick retrieval of designs and process plans, minimizes design and route sheet duplication, promotes standardization and improves cost estimation.

The existing models often deal with the part-families and they do not make any effort on the spatial arrangement of machines inside cells, even though the cell layout can have a significant impact on the cell loading problem, the amount of work-in-process inventories, throughput, and throughput variability. Greene and Sadowski (1982) discussed the modified flow shop idea as being more appropriate to take advantage of CM benefits. In a modified flow shop, each part will move in only one direction, i.e. it is not necessarily processed by all of the machines, but the backtracking is not allowed. The parts with routeings in reverse direction with the cell layout are called the backtracking parts. Modified flow has the advantages of less flow distance, easier material handling and less complicated scheduling-control problems. Another advantage is the applicability of kanban systems, since the cell configuration must be flow-based in a kanban system to reduce transport and waiting times, and to highlight interdependencies between the work centres. Currently similarity coefficients typically rely on substitute measures, such as common machines required in processing, to evaluate the similarity between parts. Similarity coefficients that employ more direct measures, such as similarities between parts in terms of design and manufacturing attributes and operation sequences, may yield better results. The operation sequences not only specify the type of machine tools needed but also impact the flow of the material. Moreover, the application of GT cells in advanced manufacturing systems such as JIT, FMS and CIM will certainly enhance the design and operational efficiency of such systems.

There are a few studies in the literature that propose similarity coefficients based on part routeings. Tam (1990) used the Levenshtein's distance measure to represent any part routeing as a string of characters, and to find the similarities between two routeings based on the operations sequences. Choobineh (1988) also presented a similarity measure based on the operation sequences and proposed an integer programming model. The objective function of the model minimized the sum of the production costs and the acquisition and maintenance costs of the machines. Vakharia and Wemmerlöv (1990) proposed a four-stage cell formation method by considering the demand and available productive time data with the operation sequences. They also emphasized the importance of the within-cell layout problem to achieve a flow-line cell by stating its advantages over the job shop cells. Harhalakis *et al.* (1988) developed a heuristic method of grouping machines in order to minimize the within-cell material movements under the constraint that the specified number of machines per cell was not exceeded. Most of the above methods, however, ignored the interactions between the part-families and the machine-cell formation process, which might affect the structure and the efficiency of the cells as discussed above. Furthermore, the configuration and the performance measures of the cell must be considered throughout the design process in order to provide information for the selection of the best part-family and machine-cell formation and spatial arrangements of machines in each cell.

The remainder of this paper is organized as follows. In the following section, a multi-objective mathematical programming formulation is presented which can be very helpful in determining the part-families and machine-cells simultaneously in conjunction with the cell layout. We discuss the proposed multi-objective cluster analysis heuristic in § 3. A full factorial design is developed in § 4 to evaluate the effects of several system parameters. Finally some concluding remarks are provided in § 5.

2. Problem statement

We propose a new approach that will consider both the design and manufacturing attributes and the operation sequences, since two parts might be similar in terms of the design attributes but not in machining requirements. The design attributes are important in the part design, manufacturing, and purchasing phases, whereas the operation sequences are important factors in determining the spatial arrangement of machines inside the cells. Furthermore, in the existing literature, the part-family formation process is usually made independent of the machine-cell formation process. In other words, there is no consideration of how the machines will be arranged within the cell, how the materials move in a cell, how much machine investment is required to process the parts assigned to that cell, and the amount of workload deviations within and between the cells.

Due to a disaggregated cell layout, CM systems are vulnerable to imbalanced workloads, which cause underutilization of certain machines in cells. Furthermore, the delays in front of the highly utilized machines will cause a high level of work-in-process inventories, which will be detrimental to the success of JIT systems. A recent survey by Wemmerlöv and Hyer (1989) has shown that volume/capacity imbalance problems occurred where either too much or too little work was loaded onto the cells.This was the most frequent problem experienced by the firms. Although detailed assessment of machine loading is not possible without operational information regarding batch size, setup time and scheduling rules, it is still valuable to balance aggregate workloads between and within cells in the cell formation stage. Hence, two additional objective functions relevant to cell utilization and load balancing are incorporated into the multi-objective cluster analysis.

As stated by several authors, the design of a CM system should not only be based on one objective; rather it should be a decision based on several objectives which are usually conflicting and thus need to be prioritized. In addition to the maximizing similarities in terms of the design and manufacturing attributes and operation sequences, the following objectives will be considered to maximize the effectiveness of a CM system:

(1) Machine investment: The investment in the machines plays an important role in establishing the cells. If there is no limitation on the amount of machine investment, then it is certain that, forming machine cells in such a way, neither backtracking parts nor intercell movements will take place. At the other extreme, if the investment capacity is too limited, backtracking parts and intercell movements would be unavoidable, then we have to deal with the bottleneck machines. Another observation is that the increase in the amount of machine investment might cause a decrease in the utilization of the machines, therefore decision makers should be able to specify the upper and lower limits on the machine utilization levels.
(2) Balanced workloads within cells: The mean squared deviation, defined as the average of the squared differences between each individual machine utilization and the cell utilization, is the measure of load imbalances within a cell.
(3) Balanced workloads between cells: It would make sense if we think of cells as separate plants processing various parts of a variety of products, which will be sent to the final assembly area to assemble the final products. Then the arrival times of these parts to the assembly area should be synchronized to prevent high level inventories. Another drawback of late arrivals is the increase in the flow time and decrease in the throughput rate. For this reason, in our model there is a measure of variability of total workloads between the cells.
(4) Number of skippings: This is used as a surrogate objective to measure the material handling activities in a flow-line manufacturing cell.

Examining these objectives individually, one might observe that some of them are inter-conflicting. Specifically, the increase in machine investment could cause a reduction in machine utilization levels and the number of skippings, and the workload between cells is less deviated. Therefore, our objective is to find a compromise solution for all six objectives as shown below. Vakharia and Wemmerlöv (1990) also developed a four-stage cell formation algorithm based on the operation sequences to establish a flow-line cell under the restriction that the capital available to make additional investments in equipment is limited and the utilization of cell equipment must be above an acceptable level. The proposed approach is more comprehensive than the previous studies since we can link the part-family formation to the cell formation process by combining design and manufacturing attributes with the operation sequences to form a modified shop structure. Furthermore, the characteristics and various cost aspects that are directly relevant to JIT, FMS and CIM are considered by selecting several objectives while designing GT cells.

2.1. *The mathematical programming formulation*

A multi-objective mathematical programming model for the PFMCF problem is formulated below for the purpose of defining the scope of our study. For a given set of objectives, the solution of this model will specify the number of cells, the type and number of machines required to meet the demand and other constraints along with the spatial arrangement of machines in each cell and the assignment of the part families to the cells. The following notation is used in the mathematical formulation, in which subscripts i, c and d refer to parts, whereas k and l refer to machine types, and j refers to cells.

Decision variables

x_{ij} : 0–1 binary variable which is equal to 1 if part i is assigned to cell j
λ_j : 0–1 binary variable which is equal to 1 if cell j is opened
E_{kj} : 0–1binary variable which is equal to 1 if machine type k exists in cell j
γ_{klj} : 0–1 binary variable which is equal to 1 if machine type k is placed before machine type l (not necessarily immediately) in cell j
N_{kj} : number of type k machines assigned to cell j
m_{kj} : number in the sequence of machine type k in cell j
L_{kj} : the expected load of machine type k in cell j
AL_j : average load in cell j
TL_j : total load in cell j
EL : average load of cells
S_i : total number of skippings for part i

Parameters

p : number of cells
n : number of parts
M : set of all machine types
DA_{cd} : dissimilarity between parts c and d based on design attributes
OS_{cd} : dissimilarity between parts c and d based on operation sequences
Y_{ikl} : 0–1 binary indicator which is equal to 1 if part i should be processed by machine type k immediately before machine type l
B_{cd} : 0–1 binary indicator which is equal to 1 if parts c and d should not be in the same cell due to their reverse operation sequences
α_{kj} : a lower limit on the utilization of machine type k in cell j
β_{kj} : an upper limit on the utlization of machine type k in cell j
P_i : the average demand rate for part i
t_{ik} : per unit processing time of part i on machine type k
A_k : available capacity for machine type k
C_k : fixed cost of machine type k
CS_j : an upper limit on the number of machines assigned to cell j
U : a very large constant

$$\min \quad f_1 = \sum_{c=1}^{n}\sum_{d=1}^{n}\sum_{j=1}^{p} DA_{cd} \cdot x_{cj} x_{dj} \tag{1}$$

$$\min \quad f_2 = \sum_{c=1}^{n}\sum_{d=1}^{n}\sum_{j=1}^{p} OS_{cd}\, x_{cj} x_{dj} \tag{2}$$

$$\min \quad f_3 = \sum_{j=1}^{p} \sum_{k \in M} N_{kj} C_k + E_{kj} \tag{3}$$

$$\min \quad f_4 = \sum_{j=1}^{p} \sum_{k \in M} \frac{\left[(L_{kj}/N_{kj}) - AL_j \right]^2}{\sum_{k \in M} E_{kj}} \tag{4}$$

$$\min \quad f_5 = \sum_{j=1}^{p} \frac{(TL_j - EL)^2}{p} \tag{5}$$

$$\min \quad f_6 = \sum_{i=1}^{n} S_i \tag{6}$$

subject to

$$\sum_{j=1}^{p} x_{ij} = 1 \quad \forall i \tag{7}$$

$$\sum_{j=1}^{p} \lambda_j \le p \tag{8}$$

$$x_{ij} \le \lambda_j \quad \forall i, j \tag{9}$$

$$B_{cd}(x_{cj} + x_{dj}) \le 1 \quad \forall c, d, j \tag{10}$$

$$x_{ij} Y_{ikl}(m_{lj} - m_{kj}) \ge 1 \quad \forall i, j, k, l \tag{11}$$

$$(m_{lj} - m_{kj}) + U(1 - \gamma_{klj}) \ge 1 \quad \forall k, l, j \tag{12}$$

$$(m_{kj} - m_{lj}) + U \gamma_{klj} \ge 1 \quad \forall k, l, j \tag{13}$$

$$L_{kj} - \sum_{i=1}^{n} t_{ik} P_i x_{ij} = 0 \quad \forall k, j \tag{14}$$

$$L_{kj} - N_{kj} A_k \le 0 \quad \forall k, j \tag{15}$$

$$\alpha_{kj} \le (L_{kj}/N_{kj} \le \beta_{kj} \; \forall k, j \tag{16}$$

$$N_{kj} - E_{kj} U \le 0 \quad \forall k, j \tag{17}$$

$$\sum_{k \in M} \sum_{l \in M} x_{ij} Y_{ikl}(m_{lj} - m_{kj} - 1) - S_i = 0 \quad \forall i, j \tag{18}$$

$$TL_j - \sum_{k \in M} L_{kj} = 0 \quad \forall j \tag{19}$$

$$AL_j - \frac{\sum_{k \in M} (L_{kj}/N_{kj})}{\sum_{k \in M} E_{kj}} = 0 \quad \forall j \tag{20}$$

$$EL - \left(\frac{1}{p}\right)\sum_{j=1}^{p} TL_j = 0 \tag{21}$$

$$\sum_{k\in M} N_{kj} \le CS_j \,\forall j \tag{22}$$

$$x_{ij},\ \lambda E_{kj},\ \gamma_{klj} = 0, 1 \text{ and } m_{kj},\ S_i,\ N_{kj} \text{ are integer } \forall\, i,k,l,j \tag{23}$$

In this formulation, the first objective function, f_1 minimizes the dissimilarities based on the design and manufacturing attributes, while the second objective, f_2, minimizes the dissimilarities based on the operation sequences. The parameters DA_{cd} and OS_{cd} will incur if both parts c and d are assigned to the same cell j as explained below, and we do not allow intercell movements. The third objective function, f_3, computes the total machine investment cost. The criterion of the minimal number of machine types required for forming part families for a given number of cells is clearly preferable in any industrial application. The fourth objective, f_4, presents the sum of workload variabilities in each cell, while the term in the denominator is simply the number of machine types in each cell. The objective function f_5 computes the variability between the cells, where p is equal to the number of cells. The last objective, f_6, computes the total number of skippings for all the parts.

Furthermore, we define a set of constraints to specify the feasible region, and in particular, to integrate the within-cell layout problem and the multiple objectives in the PFMCF process. Constraint 7 ensures that a part will only be assigned to a single cell. Constraint 8 guarantees that there are at most p cells. Constraint 9 ensures that a part can be assigned to a cell if and only if that cell exists. Constraint 10 argues that two backtracking parts cannot be in the same cell due to the modified flow shop assumption. Constraint 11 is an inequality which determines the machines' sequence in a cell by utilizing the immediate precedence data to determine the machine arrangements. If both $x_{ij} = 1$ and $Y_{ikl} = 1$, which means part i is processed in cell j and it requires machine k immediately before l, hence machine l should be placed after machine k in cell j, not necessarily immediately. Constraints 12 and 13 guarantee that two machines cannot occupy the same place in a cell. The workload of a machine in a cell and the required number of machine type k in cell j, N_{kj}, are computed by constraints 14 and 15, respectively. Constraint 16 allows the decision makers to specify the upper and lower limits on the desired machine utilization in each cell. Constraint 17 determines the number of serially arranged machines in the cell which is used in equations (6) and (20), where the alternative machines are placed in parallel. Constraint 18 counts the number of skippings for each part. Constraints 19 and 20 define the total and average workloads in a cell, respectively. Constraint 21 calculates the average cell workloads. Finally, constraint 22 ensures that the size of a cell must not be larger than a specified number of machines.

The parameters DA and OS are used to calculate the dissimilarity matrices between the parts c and d based on their design attributes and operation sequences, respectively. Offodile (1991) also has addressed the problem of how to form part-families using part classification and coding analysis, where the attributes were represented by the ordinal type variables. Design and manufacturing attributes in our study include nine digits, even though other information can easily be incorporated. The code fields carry information for main shape, i.e., rotational or non-rotational, external shape, internal shape, plane surface machining, auxiliary holes, gear teeth, raw material, size ratios, and tolerance codes, respectively. The first digit is of asymmetric binary type, the last two digits are of ordinal type, and the others are nominal variables. For the given coding scheme, the pairwise dissimilarity matrix, DA_{cd}, can be computed by using the procedure developed by Kaufman and Rousseeuw (1989). Whereas we propose a new dissimilarity measure between parts c and d, OS_{cd}, based on their operation sequences considering both the processing requirements and the flow direction in a cell to accomplish a modified flow-line structure. Tam (1990) represented an operation sequence by a string of characters, and used the Levenshtein distance measure between two strings to calculate the pairwise similarity matrix. We revised the Levenshtein distance measure, since the original measure does not penalize the backtracking parts neither does award the commonality, as follows:-

$$z_{cd} = \begin{cases} H & \text{if two strings carry opposite sequences} \\ \frac{MNT_{cd}}{COM_{cd}} & \text{if}\, COM_{cd} > 0 \\ (h).MNT_{cd} & \text{if}\, COM_{cd} = 0 \end{cases} \tag{24}$$

where,

z_{cd} : dissimilarity value between strings c and d

MNT_{cd}: minimum number of insertions and deletions required to transform string c to d

COM_{cd}: number of common operations in strings c and d

H : a big constant number

h : a constant for penalizing disjoint strings, and $1 < h \ll H$

The constant number H is used to prevent two parts with backtracking sequences from being assigned to the same cell by assigning a high dissimilarity value. If two parts have no common operations, then we find a dissimilarity value by using the penalizing factor h. Finally, the OS_{cd} matrix is found by applying an eigenvalue normalization on z_{cd}, if any entity of z_{cd} is strictly greater than 1. An example of DA and OS calculations is given in the Appendix.

3. Multi-objective cluster analysis

The complexity of handling all six objectives simultaneously with nonlinear constraints and binary variables led us to develop a multi-objective cluster analysis (MCA) heuristic as discussed below.

Step 1: Combine two dissimilarity matrices, DA and OS, to obtain the combined dissimilarity matrix, CD, as follows:

$$w_1\, DA + w_2\, OS = CD \text{ where } w_1 + w_2 = 1 \text{ and } 0 \le w_1, w_2 \le 1 \tag{25}$$

Step 2: Apply the divisive clustering algorithm developed by Kaufman and Rousseeuw (1989) on CD to define the upper bound on the number of cells, NC, to be equal to the number of cells existing above a preset threshold value. The k-medoid technique is, then, applied on CD for a given number of cells, p, where $2 \le p \le NC$, to group parts into the families. The objective is to minimize the average dissimilarity of each part to other parts in the same family. A more detailed discussion on the k-medoid technique can be found in the same reference. For any family A, the average dissimilarity of a part i in that family to all other parts of A can be defined as $a(i)$, whereas for any other family C, average dissimilarity of i

to all parts of C is called $d(i, C)$. After computing $d(i, C)$ for all families $C \neq A$, we select the smallest of those as $b(i)$. The part-family for which this minimum is attained is called the neighbour of part i. This will be very helpful in defining the primary and secondary cells for each part. The part-family and machine-cell formation for a given p corresponds to a PFMCF alternative. Furthermore, the efficiency of each alternative can be evaluated by a silhouette coefficient, $s(i)$, which can be found by combining $a(i)$ and $b(i)$ into one formula, as follows:

$$s(i) = \frac{b(i) - a(i)}{\max[a(i), b(i)]} \tag{26}$$

For each part i,

$$-1 \leq s(i) \leq 1. \tag{27}$$

Average of these silhouettes gives us a measure for each alternative, and if it is close to 1, then the clustering is well structured. Among those alternatives, we eliminate those with negative average silhouette coefficients, and let ρ be the set of remaining candidate alternatives.

Step 3: After the elimination process, even we expect that no two backtracking parts will be in the same cell for each candidate alternative, this might occur depending on the relation with other parts assigned to the same cell due to the weight of the design and manufacturing attributes. If it occurs then the direction of the cell layout will be determined by the part flow with respect to the other parts. Otherwise, the spatial arrangement of the machines will be determined by using the precedence relationships in the part routeings only. The parts with routeings in the reverse direction with the above formed layout are called the *backtracking parts*. For each cell, the backtracking parts are determined after forming the initial cell layout.

Step 4: In this step, we will assign backtracking parts to one of the cells. For each backtracking part, there are as many alternatives as the number of formed cells. These alternatives might be different in terms of addition of a machine or a group of machines, then we select the one resulting in the minimum machine investment cost. Based on this selection, the modifications will be made in that particular cell according to the minimum cost alternative. In the same manner, we proceed for all backtracking parts.

Step 5: In Step 4, each part is assigned to a specific cell, and the layout is determined so we can compute the required capacity for each machine and using available capacities, determine the required number of each machine type in each cell. If the utilization of a particular machine is above its maximum or below its minimum utilization levels specified by the decision maker, check which parts strictly require that machine and if it is possible to assign those parts to another cell without extra machine investment, then shift those parts to alternative cells. Total investment for each alternative is then calculated by using the equation (3) given in the mathematical programming formulation.

Step 6: Calculate the amount of workload deviations within and between the cells, and the number of skippings using the equations (4), (5) and (6), respectively.

Step 7: Repeat steps 3 to 6 for all candidate alternatives which are determined in

Step 2. Because the four objectives have different units of measure, the following eigenvalue normalization procedure is used to have a common unit of measure for each objective.

$$N_{ij} = \frac{A_{ij}}{\sqrt{\sum_{j \in p} A_{ij}^2}} \tag{28}$$

where A_{ij} stands for the value of the ith objective in the jth alternative, and N_{ij} stands for the normalized value of the A_{ij} value. A global measure is found for each alternative by utilizing the analytical hierarchy process (AHP) developed by Saaty (1989), which gives the normalized priority vector by constructing a pairwise comparison matrix for all objectives.

As a summary, in Step 1, a dissimilarity measure is defined between pairs of parts, which takes into account the similarity of both design and manufacturing attributes and operation sequences of parts. The initial part-families are found first for a given number of cells in Step 2, then the initial layout is determined according to the part-families and the requirements on the demand, available machine capacities, and operation sequences in Step 3. In Step 4, the backtracking parts are assigned to one of the cells with the minimum machine investment. In Steps 5 and 6, the objective function values of the multiple performance measures are calculated for each candidate alternative. Finally, the AHP is used for evaluating different alternatives by unifying the multiple criteria into a global measure to select the best alternative. A step-by-step illustration of the heuristic is given in the Appendix.

4. Experimental design

An experimental design is performed for two purposes, which are comparing the proposed PFMCF approach with the SCM, and analysing the effects of the selected factors by the analysis of variance (ANOVA) test. To test the performance of our model, we generated a problem of 100 parts, and assumed that they require at most 8 machines and at least 2 machines selected randomly from 15 machine types. The processing time on each machine is uniformly distributed between 1 and 5. The available capacity for each machine is 3000 time units. The four experimental factors are listed in Table 1. In order to observe any quadratic relation between the factors and performance measures, every factor has three levels in the design. Since there are four factors and three levels, our experiment is 3^4 full-factorial design, which corresponds to eighty one treatment combinations including both the linear and quadratic interactions.

The first factor in the experimental design is the weights associated with the design and manufacturing attributes and operation sequences that are used in Step 1 as shown in equation (25). If more weight is given to the dissimilarities based on the

	Low	Medium	High
Design level (w_1)	0·2	0·5	0·8
Demand variability	U ~ (280, 320)	U ~ (150, 450)	U ~ (10, 600)
M/C cost variability	U ~ (90.110)	U ~ (50, 150)	U ~ (10, 190)
AHP level	0·3 0·3 0·3 0·1	0·462 0·231 0·231 0·077	0·65 0·147 0·147 0·056

U stands for the uniform distribution

Table 1. Summary of the factors in the experimental analysis.

design attributes, the part-families are expected to have similar parts in design, but these parts may require quite different material flow requirements. The resulting machine-cells may consist of various machines, but most of those machines will be processing only a small portion of the parts. However, giving higher weights to dissimilarities based on the operation sequences will result in part-families consisting of parts that require similar operations, then the machine investment cost might be less than the former case. If we have more machines in a cell, that means the materials movement will be increased, which will consequently increase the number of skippings. Briefly, the way we combine two dissimilarity matrices may significantly affect the proposed performance measures.

The second and third factors are the demand and the machine investment cost variabilities, respectively, whereas the processing times, the average demand rate, and the average machine investment cost are assumed as fixed parameters. The last factor is the priorities found by the AHP approach to unify different objectives. We assume that minimizing the machine investment cost is always superior to minimization of skippings, which directly affects the pairwise comparison matrix at each level.

In the computational analysis, the proposed MCA approach is compared with the SCM, corresponding to the best alternative in that setting. An important point to notice is that we run both methods under the same conditions and assumptions. We can briefly state that the only difference between the execution of two methods is the operation sequence based matrix, *OS*, which is replaced by the part-machine incidence matrix in SCM. The same dissimilarity matrix based on the design attributes, *DA*, is used in both of the methods. The results in Table 2 state that for the same conditions and assumptions, our model performed better than the SCM in most of the cases, i.e., for 57 settings out of 81. A stronger conclusion can be achieved by comparing two methods when the weight of the design attributes is at the low level, since both methods are using the same dissimilarity matrix based on design attributes, but we propose a new dissimilarity measure based on the operation sequences. Eventually, the performance of the proposed measure gets better, and it dominates the SCM method. Furthermore, to test the robustness of our method, we check the ranges of our method's best results with the SCM's best results as given in Table 3. We can conclude that the proposed approach is also more robust than the SCM.

The ANOVA tables are generated for all performance measures separately, and the significant factors, linear and quadratic terms are given at 1, 2·5 and 5%

Design level	Proposed method	SCM	Equal
High	18	8	1
Medium	15	8	4
Low	24	0	3
Total	57	16	8

Table 2. Comparison of the methods according to best results.

	Range	Average
Proposed method	(0·34–0·43)	0·38
SCM	(0·35–0·47)	0·42

Table 3. Ranges and averages for two methods.

Significance levels				
Perf. mea.	M/C Cost	Within Cell Var.	Var. Bet. Cells	Skippings
Factors	T L Q	T L Q	T L Q	T L Q
Design level	a a c	b b *	* * *	a a *
Demand var.	a a a	* * *	* * *	a a *
M/c cost var.	a a *	a a *	a a a	a a a
AHP level	a a a	a a a	a a a	a a a

T : Total L : Linear Q : Quadratic
a : 1% b : 2·5% c : 5% * : Insignificant

Table 4. Summary of ANOVA tables for the proposed method.

significance levels as summarized in Table 4 by using the Yates' algorithm, which is discussed in detail in Montgomery (1991). For machine investment, all factors are found to be significant. The design level directly affects the machine cost, since giving more weight to design attributes cause the resulting part-families to be more similar in design attributes rather than the operation sequences. Since the processing times and routeings are fixed, the machine utilizations depend highly on the demand variability. The machine investment cost is also significantly affected by the machine cost variability and the AHP level as expected.

The variability of the workloads within the cell is affected by the design level, machine cost variability, and AHP level. The total demand is slightly the same for all demand levels, since all three levels are distributed uniformly with a mean of 300. The variability of total workload between the cells is also significantly affected by the machine cost variability and the AHP level. The total number of skippings is significantly affected by all factors, since the number of skippings in a cell depends on both the number of machines in a cell and their arrangement, such as an additional machine placed in the middle of a flow-line type cell may significantly increase the number of skippings.

5. Conclusions

In this study, the PFMCF problem was studied by considering the design and manufacturing attributes and operation sequences simultaneously. After the combination of these data, four performance measures were defined, which were the machine investment, within and between cell variability, and the number of skippings, to compare the several alternatives generated by using the proposed cluster analysis procedure to minimize the average dissimilarity among the parts within each family. The proposed approach determined the number of cells, assignment of the part-families to the cells, the number of machine types, and the spatial arrangement of machines in each cell, required to meet the demand, machine capacities and other constraints, guided by six objective functions and a set of weights assessed by the decision maker. A comparison with the SCM showed that the proposed MCA approach performed better in 70% of the cases. Furthermore, it dominated the SCM in cases where the weight of the design attributes was low. This illustrated that the proposed dissimilarity measure based on operation sequences was much stronger than the measure in SCM, and also more robust, as the range of the results was smaller as shown in Table 3. The ANOVA tables indicated the importance of design level, therefore the design attributes should be taken into account in the PFMCF process.

Machine type	A	B	F	G	H	K	L	M	R	T
Investment cost	106	136	65	140	103	61	126	93	94	70

Table 5. Machine type investment costs.

		Design and Manuf. Attri.	Proc. Times	
		Attribute Number	Sequence Num.	
Part no	Operation sequence	1 2 3 4 5 6 7 8 9	1 2 3 4 5	Average demand
1	B F G	0 1 2 6 4 3 2 3 1	2 3 2	150
2	A B F L	1 2 5 1 0 3 5 5 3	2 3 4 3	226
3	F K M R	1 2 1 3 8 3 4 1 0	2 3 4 2	335
4	F G H R	0 5 2 6 8 2 2 6 2	3 2 4 2	446
5	A M T	1 0 0 3 4 3 0 6 0	2 3 4	274
6	A B L M T	0 1 1 6 3 1 1 2 2	2 3 2 1 5	171
7	F G R L T	0 4 3 0 4 3 0 2 3	3 2 3 1 2	218
8	K A R T	1 1 0 8 5 3 0 4 2	2 3 1 2	273
9	A K M R T	1 5 3 7 3 2 3 3 0	3 1 4 2 3	307
10	A K R T	0 1 0 7 5 3 1 4 2	3 2 2 1	414
11	A B R T	0 2 0 5 5 3 2 4 1	2 3 4 1	223
12	B F R L	0 1 3 5 4 2 3 3 1	3 2 1 1	378
13	M R T	1 4 3 6 4 2 5 2 2	3 2 4	328
14	A R T	1 1 1 8 4 0 0 5 2	1 3 2	280
15	F H R L	0 3 2 5 7 2 2 6 4	1 2 1 1	270
16	A B M T	1 1 0 3 2 3 2 5 3	2 3 2 3	182
17	A B F T	1 2 3 3 0 3 4 2 1	3 2 3 4	244
18	A B F G	0 0 0 0 5 2 0 0 1	5 4 2 2	152
19	F G R T	0 3 2 5 1 4 2 3 2	4 1 1 1	366
20	G R T	0 3 3 6 1 2 2 4 1	1 1 2	226

Table 6. Parts, design and manufacturing attributes, routeings and demand data.

The most of the studies in the literature consider either design attributes or machining requirements, and usually ignore the cell layout problem. So, this study brings a different perspective to the PFMCF problem. We focused on the importance of the modified flow shop idea, and found that the proposed distance measure is more effective in terms of avoiding backtracking parts, and penalizing the disjoint sequences. The backtracking parts in a cell reduce the advantages of CM systems as stated by several authors. Besides, the necessity of the proposed performance measures is expressed, and it is stated that the part-family formation ignoring the cell layout, workloads on the machines, machine utilization, and the machine investment cost might degrade the effectiveness of CM systems. Furthermore, a nonlinear mixed integer programming model is formulated by several objectives and constraints for the purpose of defining the scope of our study.

Before concluding, it is worth mentioning some future research directions. Extensions can be made to handle alternative process plans. Furthermore, an integrated approach can be developed to provide reasonable machine insertion points, so that both the machine investment and the number of skippings will be considered simultaneously. Finally, the monocode coding structure for design attributes can be studied to provide the use of it in a cluster analysis.

Appendix: A numerical example to illustrate the proposed heuristic

An example problem is generated to illustrate the proposed multi-objective cluster analysis (MCA) heuristic described in § 3. The example considers 20 parts and each part requires at most 5 and at least 3 machine types selected from 10 machine types. The investment costs for each machine type are given in Table 5. The medium levels are chosen for each experimental factor. In this study, there are 9 design and manufacturing attributes coded in a polycode structure as summarized in Table 6 along with the operation sequences, processing times on each machine and the demand forecasts for each part. Each symbol in an operation sequence corresponds to a machine type that will perform the given operation.

The steps of the proposed method are as follows:

Step 1: Calculate two dissimilarity matrices, DA and OS, using the design and manufacturing attributes and operation sequences, respectively. These matrices are given below, where the values for the first ten parts are reported due to space limitations. A combined dissimilarity measure, CD_{cd}, is found between every part of parts c and d such that $0{\cdot}2{\cdot}DA_{cd} + 0{\cdot}8{\cdot}OS_{cd} = CD_{cd}$.

$DA =$

	1	2	3	4	5	6	7	8	9	10	⋯ ⋯	20
1	0											
2	0·76	0										
3	0·73	0·6	0									
4	0·42	0·82	0·82	0								
5	0·64	0·66	0·54	0·83	0							
6	0·49	0·86	0·74	0·63	0·91	0						
7	0·52	0·72	0·77	0·77	0·6	0·69	0					
8	0·6	0·6	0·67	0·82	0·43	0·7	0·62	0				
9	0·81	0·79	0·7	0·67	0·72	0·74	0·77	0·74	0			
10	0·49	0·71	0·78	0·7	0·65	0·48	0·62	0·33	0·74	0		
⋮			⋮			⋮						
20	0·46	0·85	0·86	0·4	0·84	0·62	0·65	0·81	0·6	0·69	⋯ ⋯	0

$OS =$

	1	2	3	4	5	6	7	8	9	10	⋯	20
1	0											
2	0·0168	0										
3	0·0556	0·0668	0									
4	0·0168	0·0668	0.0222	0								
5	0·1336	0·0556	0·0556	0·1559	0							
6	0·0668	0·0111	0·088	0·2	0·007	0						
7	0·0222	0·0278	0·0278	0·0111	0·0668	0·0334	0					
8	0·1559	0·0668	0·022	0·0668	0·0278	0·0278	0·0278	0				
9	0·178	0·0779	0·0111	0·0779	0·007	0·0148	0·0334	0·4008	0			
10	0·1559	0·0668	0·0222	0·0668	0·0168	0·0278	0·0278	0·4008	0·0028	0		
⋮			⋮			⋮						
20	0·0445	0·1559	0·0556	0·0167	0·0445	0·0668	0·007	0·0167	0·0222	0·167	⋯	0

Step 2: The upper bound on the number of cells, NC, is equal to 6 for a threshold value of 0·150. The k-medoid technique is applied on CD for a given number of cells, p, where $2 \le p \le 6$, to group parts into the families. The objective is to minimize the average dissimilarity of each part to other parts in the same family. The average silhouette coefficient corresponding to 4 cells alternative has a negative value, so it is eliminated. The set of remaining candidate alternatives are {2, 3, 5, 6}.

Step 3: The spatial arrangement of machines in each cell for the remaining candidate alternatives are found by using the CRAFT algorithm for the initial part families as shown in Table 7.

Step 4: Initially, part 10 is assigned to the cell number 2, when $p = 2$, but it has reverse operations with part 8 in the same cell. Therefore it is moved to its secondary cell, i.e., the cell number 1, since it does not require any additional machine investment rather than duplicating a machine type K. There is no backtracking parts for other alternatives, which also shows the effectiveness of the proposed dissimilarity measure.

Step 5: In Step 4, each part is assigned to a specific cell and the cell layout is determined so we can compute the required number of each machine type in each cell by considering the demand requirements and capacity restrictions.

Step 6: The total machine investment, the amount of workload deviations within and between the cells, and the number of skippings are calculated for each candidate alternative as summarized in Table 8.

Step 7: After the eigenvalue normalization and the AHP process, the first alternative, i.e., $p = 2$, gives the minimum unified measure as also shown in Table 8.

In order to compare the proposed method with the SCM, the same example is repeated when $p = 2$. The only difference is that the proposed operation sequence based matrix, OS, is replaced by the 0–1 part-machine incidence matrix in SCM. The same dissimilarity matrix DA is used in both of the methods. The final solution of

# Cells	Cell number	Part family	Cell layout
2	1	1 4 7 9 12 13 15 18 19 20 10	A K M R T B FF G H R L T
	2	2 3 5 6 8 11 14 16 17	A B F L K M T A R T
3	1	1 2 6 10 11 12 17 18	A B F G A B F K R L M T
	2	3 5 8 14 16	F K A B M R T
	3	4 7 9 13 15 19 20	FF G H R L T A K M R T
5	1	1 7 12 15 18	A B F G H R L T
	2	2 3 17	A B F K M R L T
	3	4 9 13 19 20	F G H R T A K M R T
	4	5 8 14 16	K A R T B M T
	5	6 10 11	A B L M K R T
6	1	1 7 12 15	B F G H R L T
	2	2 3 17	A B F K M R L T
	3	4 9 13 19 20	F G H R T A K M R T
	4	5 8 14 16	K A R T B M T
	5	6	A B L M T
	6	10 11 18	A B F G K R T
SCM	1	1 4 7 12 15 18 19 20	A B FF G H R L R T
	2	2 3 5 6 8 9 10 11 13 14 16 17	AA B F L K MM R TT A R T

Table 7. Part families and cell layouts.

	2 cells	3 cells	5 cells	6 cells
M/C investment	2150	2910	3803	4214
Within cell deviation	602	752	627	473
Between cell deviation	38	57	142	347
Total # skippings	9894	9068	7399	7057
Unified measures	0·33	0·42	0·50	0·62

Table 8. Alternatives and objectives.

	Proposed method	SCM
M/C investment	2150	2289
With cell deviation	602	550
Between cell deviation	38	128
Total # skippings	9894	11226
Unified measures	0·60	0·77

Table 9. Comparison of the proposed method and SCM.

SCM is also reported in Table 7. A comparison of these two methods is given in Table 9, which clearly indicates the advantage of using the proposed dissimilarity measure, OS. The following example is given to explain why the proposed OS matrix is more effective than the machining requirements based SCM approaches. In OS calculations, the maximum number of the sum of insertions and deletions to transform one operation sequence to another one is given by parts 4 and 6, and $MNT_{4,6}$ is equal to 9. Therefore $Z_{4,6} = 2 \cdot 9 = 18$, since $COM_{4,6} = 0$ and $h = 2$. As a result, H is set to 36. If we look at the parts 8 and 10, their dissimilarity coefficient value, when we use SCM, is equal to 0 since they both require exactly the same machines. But, in the OS calculations, the dissimilarity coefficient, after the eigenvalue normalization, is 0·4008, because an assignment of these two parts into the same cell will require at least an additional machine type A or K. Lets now calculate the dissimilarity coefficient between parts 5 and 6. The number of common operations, $COM_{5,6} = 3$, and the required number of deletions to transform the operation sequence of part 6 to part 5 is 2, i.e., $z_{5,6} = 2/3$. After the eigenvalue normalization $OS_{5,6} = 0 \cdot 007$.

References

Askin, R. and Subramanian, S., 1987, A cost based heuristic for group technology configuration. *International Journal of Production Research*, **25** (1), 101–113.

Choobineh, F., 1988, A framework for design of cellular manufacturing systems. *International Journal of Production Research*, **26** (7), 1161–1172.

Greene, T. J. and Sadowski, R. P., 1982, Cellular manufacturing control. *Journal of Manufacturing Systems*, **2** (2), 137–144.

Gunasekaran, A., Goyal, S. K., Virtanen, I. and Yli-Olli, P., 1994, An investigation into the application of group technology in advanced manufacturing systems. *International Journal of Computer Integrated Manufacturing*, **7** (4), 215–228.

Harhalakis, G., Nagi, R. and Proth, J. M., 1988, An efficient heuristic in manufacturing cell formation for group technology applications. *International Journal of Production Research*, **28** (1), 185–198.

Houtzeel, A., 1975, An introduction to the MICLASS system. *Proceedings of Computer Aided Manufacturing International: Seminar of CAPP Applications*, **CAM-1**, 159–178.

HYER, N. L. and WEMMERLÖV, U., 1989, Group technology in the U.S. manufacturing industry: a survey of current practices. *International Journal of Production Research*, **27** (2), 1287–1304.

KAUFMAN, L. and ROUSSEEUW, P. J., 1989, *Finding Groups in Data* (New York: Wiley).

KING, J. R., 1980, Machine-component group formation in PFA: an approach using rank order clustering algorithm. *International Journal of Production Research*, **18** (2), 213–232.

KRAJEWSKI, L. J., KING, B. E., RITZMAN, L. P. and WONG, D S., 1987, Kanban, MRP, and shaping the manufacturing environment. *Management Science*, **33** (1) 39–57.

KUSIAK, A. and CHOW, W., 1988, Decomposition of manufacturing systems. *IEEE Journal of Robotics and Automation*, **4** (5) 457–471.

MILTENBURG, G. J. and ZHANG, W., 1991, A comparative evaluation of nine well-known algorithms for solving the cell formation problem in group technology. *Journal of Operations Management*, **10** (1) 44–72.

MONTGOMERY, D. C., 1991, *Design and Analysis of Experiments*, 3rd edn (New York: Wiley).

OFFODILE, O. F., 1991, Application of similarity coefficient method to parts coding and classification analysis in group technology. *Journal of Manufacturing Systems*, **10** (6), 442–448.

OFFODILE, O. F., MEHREZ, A. and GRZNAR, J., 1994, Cellular manufacturing: a taxonomic review framework. *Journal of Manufacturing Systems*, **13** (3), 196–220.

OPITZ, H., 1972, *A Classification System to Describe Workpieces, Parts I and II* (New York: Pergamon Press).

SAATY, T. L., 1989, *The Analytic Hierarchy Process*, 2nd edn (New York: Wiley).

SEIFODDINI, H. and WOLFE, P. M., 1986, Application of the similarity coefficient method in group technology, *IIE Transactions*, **18** (3) 271–277.

SHAFER, S. M. and ROGERS, D. F., 1993, Similarity and distance measures for cellular manufacturing. Part II. An extension and comparison, *International Journal of Production Research*, **31** (6), 1315–1326.

TAM, K. Y., 1990, An operation sequence based similarity coefficient for part families formation. *Journal of Manufacturing systems*, **9** (1), 55–68.

VAKHARIA, A. and WEMMERLÖV, U., 1990, Designing a cellular manufacturing system: a materials flow approach based on operations sequences. *IIE Transactions*, **22** (1), 84–97.

WEMMERLÖV, U. and HYER, N. L., 1989, Cellular manufacturing in the U.S. industry: a survey of users. *International Journal of Production Research*, **27** (9), 1511–1530.

WELKE, H. A. and OVERBEEKE, J., 1988, Cellular manufacturing: a good technique for implementing just-in-time and total quality control. *Industrial Engineering*, **20** (11), 36–41.