International Journal of Production Research
Vol. 33, No. 9, pp2493-2509, 1995

Component grouping for GT applications-a fuzzy clustering approach with validity measure

N. N. Z. GINDY†*, T. M. RATCHEV† and K. CASE‡

†Department of Manufacturing Engineering and Operations Management, University of Nottingham, University Park, Nottingham NG7 2RD, UK.

‡Department of Manufacturing Engineering, Loughborough University of Technology, Loughborough, Leicester LE11 3TU, UK.

The variety of the currently available component grouping methodologies and algorithms provide a good theoretical basis for implementing GT principles in cellular manufacturing environments. However, the practical application of the grouping approaches can be further enhanced through extensions to the widely used grouping algorithms and the development of criteria for partitioning components into an 'optimum' number of groups. Extensions to the fuzzy clustering algorithm and a definition of a new validity measure are proposed in this paper. These are aimed at improving the practical applicability of the fuzzy clustering approach for family formation in cellular manufacturing environments. Component partitioning is based upon assessing the compactness of components within a group and overlapping between the component groups. The developed grouping methodology is experimentally demonstrated using an industrial case study and several well known component grouping examples from the published literature.

1. Introduction

Group Technology (GT) is a manufacturing philosophy which seeks to exploit similarity between components in order to achieve increased productivity and efficiency in the planning and operation of manufacturing systems. One of the most widespread applications of GT is in the development of cellular manufacturing. Here, similar components are grouped into part families and each part family is designated for production using a group of processing equipment organized as a cell.

In the past three decades a number of methodologies have been proposed for part family formation in GT applications (for a recent review refer to Singh 1993). Most conventional grouping methodologies assume well defined boundaries between groups and therefore assign each component to one component family. Such crisp models often fail to fully reflect the complex nature of component data, where boundaries between groups are fuzzy, and where a more nuanced description of the affinity of components to different groups is required (Xu and Wang 1989).

According to conventional grouping methodologies, if *n* components and *m* machines are grouped into *c* groups, the results can be described using a binary matrix of the form:

$$A = \begin{array}{c} \\ G_1 \\ G_2 \\ \dots \\ G_c \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & \dots & C_n \end{array} \\ \left\{ \begin{array}{cccc} 0 & 0 & \dots & 1 \\ 1 & 0 & \dots & 0 \\ & & \dots & \\ 0 & 1 & \dots & 0 \end{array} \right\} \end{array}$$

where

$$\alpha_{ij} = \begin{cases} 1, \text{ if the } j^{\text{th}} \text{ part belongs to the } i^{\text{th}} \text{ family;} \\ 0, \text{ if the } j^{\text{th}} \text{ part does not belong to the } i^{\text{th}} \text{ family} \end{cases}$$

and

$$(1) \quad \alpha_{ij} = 0 \vee 1 \quad \text{for } i = 1, 2, \dots, c; \ j = 1, 2, \dots, n$$

$$(2) \quad \sum_{i=1}^{c} \alpha_{ij} = 1 \quad \text{for } j = 1, 2, \dots, n$$

$$(3) \quad \sum_{j=1}^{n} \alpha_{ij} > 0 \quad \text{for } i = 1, 2, \dots, c$$

As a result component C_j belonging to a group G_i (α_{ij} = 1) does not belong to any other group ($\alpha_{kj} = 0, \forall k \neq i, k = 1, 2, \dots c$).

Fuzzy clustering has been advocated as an appropriate methodology for part family formation in cases where no clear division between component groups can be achieved and hence crisp logic of family formation does not seem appropriate (Wang and Li 1991). In such cases fuzzy logic reflects more realistically the industrial environment and provides a convenient mathematical platform for problem solving as well as facilitating tasks like production planning and scheduling (Singh and Mahanty 1991, Chu and Hayya 1991, Singh 1993).

According to the fuzzy logic (Kaufmann 1975), components may belong to different groups with various probabilities (fuzzy membership) reflecting the similarity between the component and the component groups. Component membership, therefore, is not restricted to a binary value of 0 or 1. Instead it is defined in the whole interval [0, 1] and can be represented by a matrix of the form:

$$A = \begin{array}{c} \\ G_1 \\ G_2 \\ \dots \\ G_c \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & \dots & C_n \end{array} \\ \left\{ \begin{array}{cccc} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ & & \dots & \\ \alpha_{c1} & \alpha_{c2} & \dots & \alpha_{cn} \end{array} \right\} \end{array}$$

where component membership is defined by:

$$(1) \quad 0 \leq \alpha_{ij} \leq 1 \quad \text{for } i = 1, 2, \dots, c; \ j = 1, 2, \dots, n$$

$$(2) \quad \sum_{i=1}^{c} \alpha_{ij} = 1 \quad \text{for } j = 1, 2, \dots, n$$

(3) $\sum_{j=1}^{n} \alpha_{ij} > 0$ for $i = 1,2,\ldots,c$

Grouping algorithms are traditionally based on assuming a well-defined (preferred) routing for each component. This reflects the preferred set of machine tools to be used for its processing. However, the increased capabilities of modem machine tools and the advances that have been made in developing generative process planning systems capable of generating alternative routes for component processing, have led to a significant increase in the number of available alternatives for component processing. Consequently the partition of components into crisp groups is becoming more difficult and increasingly inappropriate. If processing alternatives are to be considered, then fuzzy clustering can provide a more realistic environment for decision making by capturing the 'fuzziness' of the component routings based on the available alternatives.

Two basic issues limit the practical usefulness of fuzzy clustering techniques for part family formation in cellular manufacturing applications: (1) the unrealistic nature of the distribution of components in the groups formed using fuzzy clustering algorithms, and (2) the lack of a representative cluster validity measure for partitioning components into meaningful groups-a problem known as cluster validation (Davies and Bouldin 1979).

In most of the reported techniques (Chu and Hayya 1991, Xu and Wang 1989) the number of groups is a pre-defined input to the grouping algorithm. This may be appropriate in cases where limited machining resources exist for component processing, grouping is used for clustering new components around already existing centroids or *a priori* knowledge on the character of the component set strongly suggests a finite number of groups. This approach, however, is not valid in the general component grouping case, where such pre-knowledge cannot be assumed.

The main target of component grouping is to achieve a stable partitioning of a set of components into a number of meaningful groups to serve a specific manufacturing application. For example, in a cellular manufacturing environment, production is normally organized for processing families of components in several cells where the machines in each cell are located in close proximity. As we are talking about concurrent formation of component groups and groups of processing equipment, the grouping requirements can be formulated as:

(1) Distinctive division between component groups such that the number of repeated machines between cells (overlapping of cell capability) is minimized;
(2) Maximizing the number of components in each group which require the full set of machining resources available in the cell, i.e. maximum compactness of component clusters.

The approach adopted in this work is based on extensions to the widely used divisive fuzzy C-means (FCM) algorithm (Bezdek 1980). A new cluster validation measure based on cluster compactness and machine repetition is proposed for group partitioning. Several of the validity measures used in cluster analysis are evaluated and the applicability of the proposed fuzzy grouping algorithm is tested using industrial data and data from several well known component grouping methodologies from the published literature.

As a prelude to the proposed solution methodology detailed in § 3, some of the issues involved in partitioning components into part families are overviewed in § 2, using an industrial case study. The grouping methodology has been validated for a large sample (approximately 10000) of components randomly selected from the database of a major engineering company. The validation was based on comparing the results produced using the new methodology with the results of other clustering techniques by assessing the quality of grouping in terms of compactness and overlapping of the component groups. To simplify the presentation of results, the approach adopted is illustrated using a representative sample of 120 components. The routing information for the sample component set is shown in Table 1.

Components	Machines	Components	Machines	Components	Machines
1	5	41	27 33	81	8 15 29
2	9 40 41	42	27	82	40 41
3	9 23	43	15 32 45	83	9 43
4	13 34	44	17 33	84	8 15 31
5	13	45	26 27	85	38 40 41
6	13 27	46	17 27	86	8 15 32
7	13 26	47	25 27	87	28 38 40
8	6 11 14	48	26 27 33	88	17 25
9	30 33	49	17 20 27	89	6 16 37
10	4 15	50	20 27	90	1 7 16
11	22 29 31 45	51	5 41	91	1 16
12	48	52	25 26 30	92	4 8 31
13	4	53	5 12	93	4 31
14	12	54	8	94	39
15	3	55	8 15	95	17 18 25
16	2 27	56	32 45	96	3 15 32
17	2 13	57	16 42	97	28 36 40
18	2	58	17 20 33	98	9
19	2 13 27	59	34	99	1 6 24
20	3 19	60	42	100	6 16
21	33	61	35	101	1 6 14
22	15 32	62	38	102	1 6 23
23	7 16	63	36	103	1 6 7
24	21 40	64	11	104	I 16 43
25	57	65	25	105	6 16 35 43
26	6 7 23	66	5 40	106	20 26
27	6 23 43	67	9 41	107	17 20 26
28	41 12 44	68	17 25 27	108	17 18 33
29	26	69	27 30	109	32
30	25 26	70	16 21	110	7 37
31	4 32	71	9 40	111	7 43
32	4 32 45	72	9 21 16	112	34 35
33	39 45	73	10	113	9 16 23
34	20	74	28	114	9 16
35	7 6 16	75	38 40	115	16
36	8 19	76	40	116	4 12
37	9 21 16	77	41	117	4 29
38	17	78	41 43	118	14
39	17 26	79	8 15 38	119	1 6 16 43
40	26 30	80	28 40	120	37

Table 1. Component data set.

2. Fuzzy clustering approach

2.1. *Clustering algorithm*

Components are defined as vectors in a Euclidean vector space $\mathcal{Q}$ by the sets of machines

$$C_k = \{c_{1k}, c_{2k,\dots} c_{nk}\} \tag{1}$$

where c_{ik} (i = 1, 2, … , n) indicates the relationship between component k and machine i:

$$c_{ik} = \begin{cases} 1, & \text{component } k \text{ requires machine i,} \\ 0, & \text{component } k \text{ does not require machine } i \end{cases} \tag{2}$$

Component grouping is formally defined as a mapping of the component set into a set of component clusters $Gj \subset \mathcal{Q}$ = 1, 2, … , m), where each component is associated with at least one cluster.

The FCM algorithm includes an iterative repetition of the following three steps to reach a stable partition of a component set into groups:

(1) Initialization of the membership function μ_{ij} of component C_i to group G_j such that:

$$\sum_{j=1}^{m} \mu_{ij} = 1 \tag{3}$$

(2) Computation of the fuzzy grouping centroids G_i for i = 1, 2, … , m defined as weighted sums of all data points in the set

$$G_{i=} \frac{\sum_{j=1}^{n}(\mu_{ij})^{f} C_j}{\sum_{j=1}^{n}(\mu_{ij})^{f}} \tag{4}$$

where n is the total number of components

(3) Updating of the fuzzy memberships μ_{ij} using:

$$\mu_{ij} = \frac{\left(\frac{1}{e^2(C_{j,}\ G_i)}\right)^{1/(f-1)}}{\sum_{i=1}^{m}\left(\frac{1}{e^2(C_j, G_i)}\right)^{1/(f-1)}} \tag{5}$$

where $f > 1$ is the fuzziness index (Bezdek 1980, 1987, Cannon *et al.* 1986) and e^2 (C_i, G_j) is the Euclidean metric norm.

During the iterative process of group formation, using an FCM algorithm, it was observed that the new components tended to gravitate towards the groups which already had the largest number of components (see Fig. 1). This leads to a very uneven and unrealistic distribution of components in the formed groups which differ significantly from the practically expected results.

In the main, this is a consequence of the way by which grouping centroids, used to calculate component membership, are defined. Each machine in the group vector is defined by the relative weight sum of the components requiring it (equation (4)). Unless a machine is required by a large number of components, the resulting coordinate is very small. The calculation of the membership is based on the distance between the component and the group centroid and since some machines have small coordinates their

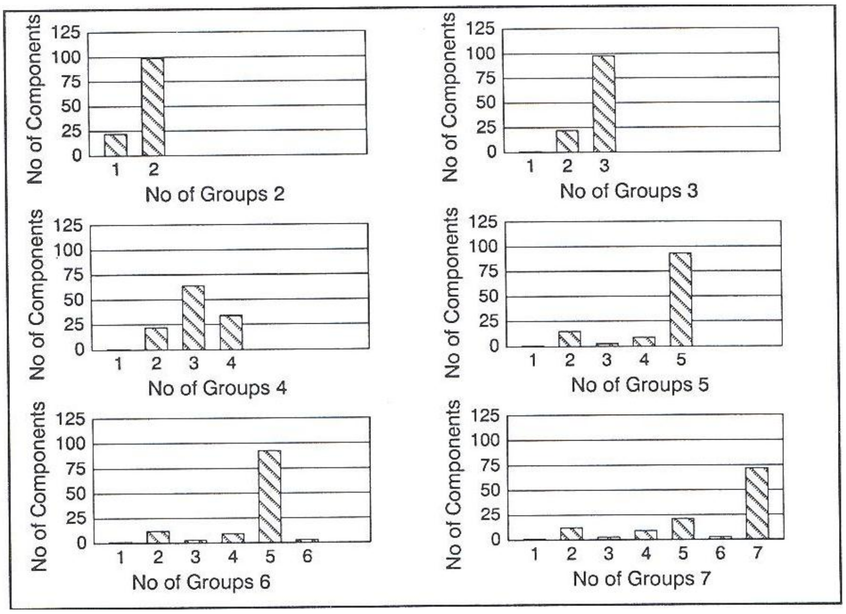

Figure 1. Component distribution – results of the FCM algorithm.

resulting influence on the distance is consequently minimal. As a result, new machines tend to be added to the more powerful component groups.

2.2. *Selection of validity measure*

Probably the single most important feature of a successful cellular manufacturing environment is the organization of production into several cells of dissimilar production equipment that are product-based rather than process-based (Warren and Moodie 1993). It is therefore important that the validity measure used for partitioning the component set into groups should lead to formation of machining cells with maximum diversity, i.e. minimum overlapping in terms of the repeated machines between cells.

The assessment of cluster tendency in terms of both compactness of component groups and differentiation between the clustering centroids has a fundamental impact on the quality of the grouping process (Jain 1988, Selim and Izmail 1986). Several cluster validation measures have been proposed for partitioning data sets in a wide variety of grouping applications varying from astronomy to colour television defect analysis.

An unsupervised fuzzy clustering approach for fuzzy classification without a priority assumption of the number of clusters in the data set was used by Davies and Bouldin (1979). The proposed cluster separation measure assesses the average dispersion and distance between clusters measured by the Minkowski metric. One of the problems associated with this method is that, in some cases, it leads to extra local minima and therefore fails to generate the correct number of groups (Oath and Geva 1989).

A partition coefficient for measuring the amount of overlapping between component clusters was proposed by Bezdek (1975). It is defined by the sum of the variations of all groups using component membership function μ_{ij}:

$$F = \sum_{i=1}^{n}\sum_{j=1}^{m}(\mu_{ij})^2 \tag{6}$$

Optimum partitioning of components into groups is indicated by the minimum value of the partition coefficient F.

The main disadvantages of validity measures which rely solely on cluster compactness for group partitioning are their tendency to monotonically decrease with the increase of the number of component groups and their insensitivity in differentiation between the grouping centroids.

A cluster validity function which overcomes most of these difficulties has been proposed by Xie and Bebi (1991). The validity measure S= π/s, defined as the ratio of the overall compactness π and cluster separation s, takes into account both average compactness and separation of each fuzzy partition.

The overall compactness of the partition is defined as:

$$\pi = \frac{1}{m}\sum_{j=1}^{m}\pi_j \tag{7}$$

where $\pi_j = (\sigma_j/n_j)$ is average deviation of group G_i and m is the total number of groups.

The deviation of each group is recorded as:

$$\sigma_j = \sum_{i=1}^{n}(d_{1j})^2 + (d_{2j})^2 + \cdots + (d_{nj})^2 \tag{8}$$

where:

$$d_{ij} = \mu_{ij}e(C_i, G_j),$$

is the weighted Euclidean distance between component C_i and group G_j.

The cluster separation s, indicating the level of overlapping between clusters, is defined as the minimum distance between any two clusters of the partition:

$$S = \min_{ij}(b_{ij})^2 \tag{9}$$

where $b_{ij} = e(G_i, G_j)$ is the Euclidean distance between group centroids G_i and G_j.

Minimum S indicates 'optimum' partitioning of components into groups, thus significantly simplifying the calculation procedure (Xie and Beni 1991).

The validity measure S has a monotonic decreasing tendency as m reaches a value close to the number of components n. This, however, is not a serious handicap since in practice the feasible number of clusters m is much smaller than the number of components n. Heuristic methods can be used to determine a 'stop-value' beyond which grouping is considered infeasible. One of the possible approaches is to define the optimum value for S for m = 2, 3, ... , n- 1 and select the starting point of monotonic decreasing tendency as the maximum number of groups to be considered. Optimum value of m can be found by solving:

$$\min_{2\le m\le m^o}\left\{\min_{\Omega m} S_m\right\} \tag{10}$$

where Ω_m is the optimum partition for each number of groups m and m^0 is the stop-value.

The full grouping results based on an FCM grouping algorithm and the validity measures F and S are shown in Fig. 2. As can be seen from the figure, a major difficulty when using partition coefficient F (equation (6)) is its unreliability due to its monotonic decreasing tendency with the increase in number of component groups.

The compactness measure π (equation (7)) is dependent on the number of machines in

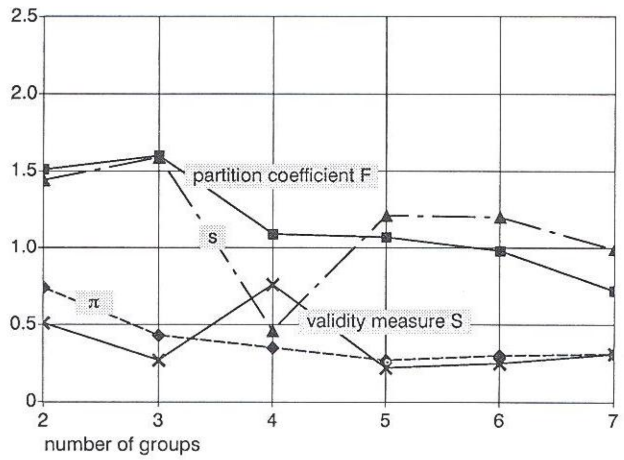

Figure 2. Variation of the validity measures F and S.

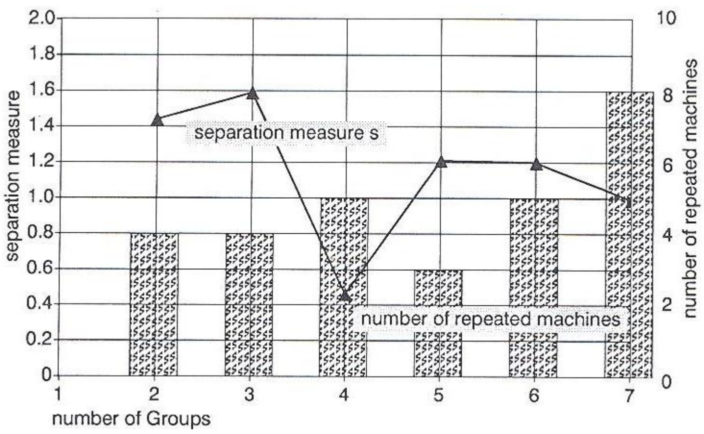

Figure 3. Variation of the separation measure s and the number of repeated machines.

the group. It is therefore infeasible for it to be used as a determining factor for judging the quality of component partitions, in cases when component grouping is done using alternative machines.

Among the two elements (s and π) making up the validity measure S, it is the separation measure s that is formulated to measure inter-group overlapping. However, when the number of duplicated machines in the groups is considered (see Fig. 3), it can be seen that maximum separation (minimum overlapping) is achieved when the component set is divided into five groups, while using s alone (equation (9)) for determining 'optimum' group formation will lead to partitioning the component set into three groups. This suggests that the methodology for calculating group separation can be improved in order to better reflect inter-group overlapping in terms of the repeated machines between components groups-a crucial factor in determining the success of family formation for cellular manufacturing applications.

The deficiencies of the FCM algorithm and available validity measures undermine the practical usefulness of the fuzzy clustering approach for component grouping. The extensions to the FCM algorithm and the definition of a new validity measure outlined below are aimed at improving the applicability of the fuzzy clustering methodology for family formation tasks in cellular manufacturing.

3. Methodology

Three extensions to the FCM algorithm are proposed here: (1) an algorithm for selecting the initial grouping centroids; (2) to base the definition of a group centroid only on components belonging to it; (3) to improve the representation of the centroid vectors by introducing a machine membership function in their calculation.

3.1. *Formation of initial part families*

In general, during component group formation, *a priori* knowledge of the approximate locations of the initial grouping centroids cannot be assumed. Clustering, for each number of groups, starts with an initial selection of clustering prototypes and an iterative algorithm results in convergence of cluster centroids to local minima. The quality of the grouping results is, therefore, dependent on the proper selection of the initial clustering prototype.

In this work, the initial clustering prototypes are defined as the components which require the largest number of machines and are maximally dissimilar (a concept similar to the 'host and guest' approach suggested by Purchek 1985). The large number of machines required by the clustering prototypes makes them natural points of concentration for components with similar processing requirements, while their mutual exclusion guarantees that most components have clear affinity to the nearest prototype.

The centroids of the initial split of components are defined as:

$$G_1^0 \equiv C_k \text{ and } G_2^0 \equiv C_{p,}$$
$$if\ \exists\left(C_{k,}C_p\right) \in \mathbb{C} \text{ for which } d\left(C_k, C_p\right) = \max_{ij} d\left(C_i, C_j\right) \tag{11}$$

where $d(C_i, C_j)$ is a component distance based on the Euclidean norm (Kaufmann 1975) between component vectors C_i and C_j representing the number of exclusive machines in the component vectors.

The next grouping centroid is then defined by finding the most distant component to the already selected clustering prototypes:

$$G_{k+1}^0 \equiv C_p,$$
$$\text{if } \exists C_p \text{ for which } \mathrm{d}\left(C_p, G_k\right) = \max_{ij}\left(\min_{ij} d\left(C_i, G_j\right)\right) \tag{12}$$

Equations (11) and (12) ensure that each cluster prototype is a data point less likely to be clustered to any of the existing grouping prototypes and therefore more likely to be a natural concentration point.

3.2. *Definition of machine membership function*

The FCM algorithm is based on the perception that each grouping centroid is defined by the membership of all components in the set. This is a valid point in fuzzy clustering analysis, but it does not reflect the practical requirements of component grouping where technological success of each group is dependent only on the machines selected for processing the components in that group. Moreover, basing the definition of clustering centroids on the membership of all components in the set does not offer a clear representation of the group boundaries in terms of required machines.

The definition of the clustering centres, in this work, is limited only to the components having the same cluster label and excludes the influence of the components already clustered to other groups. The goal here is to achieve a concurrent formation of components and machine groups. Machine selection is carried out at each iterative grouping step and group partitions are continuously assessed on the basis of the actual machines associated with each group.

The affinity of machines to different grouping centroids (machining cells) is addressed through the use of a machine membership function g_{ij} in defining the coordinates of the centroid vectors:

$$g_{ij} \in [0,1], \quad 1 \le i \le m, \qquad 1 \le j \le q$$
$$\sum_{i=1}^{m} g_{ij} = 1 \qquad 1 \le i \le m \tag{13}$$

where q is the total number of machines.

To satisfy equation (13), group centroids are defined by dividing the weighted sum of the component coordinates in each cluster by the total weighted sum of the corresponding component coordinates for all clusters denoting the same machine:

$$g_{ik} = \frac{\gamma_{ik}}{\sum_{i=1}^{m} \gamma_{ik}} \tag{14}$$

where λ_{ik} is the weighted sum of the component coordinates representing machine k and belonging to group i:

$$\gamma_{ik} = \frac{\sum_{j=1}^{n_i} (\mu_{ji})^f}{\sum_{j=1}^{n_i} \mu_{ji}{}^f}$$

According to equation (14) the allocation to different groups is related to the number of times machines are repeated in the component vectors of each cluster as well as the overall demand for a specific machine across the cluster borders. If a machine is in demand by only one cluster then its membership function g_{ij} is equal to 1 for that cluster and 0 for all others. The machines required by several clusters will have membership function in the range $0 < g_{ij} <$ I and will be equal to 0 for all other clusters.

The modified definition of group centroids restricts the membership of the machines only to the groups in which they are required, compared to the fuzzy clustering (Chu and Hayya 1991) where all components and machines appear in the centroids.

The machine membership function (equation (14)) makes the definition of the cluster centroids 'harder' and, therefore, reflects more realistically the presence (inclusion) of machines in the grouping centroids. This change gives a clearer definition of the cell content and makes decision making and group assessment closely related to the output requirements of the grouping task.

3.3. *Definition of new validity measure*

The extensions to the fuzzy grouping algorithm outlined above are aimed at improving the quality of component clustering. However, the issue of selecting a partition measure that takes into account the number of repeated machines between component groups, still needs to be addressed.

The problem lies in that a separation measure based on minimum distance to represent group overlapping is considered unsatisfactory in representing machine repetition between component groups. The difference is due to the influence of the exclusive machines in the clusters. If the number of exclusive machines is high, the cluster separation distance increases without reflecting machine overlapping. Thus two highly overlapped clusters, if they are large enough, will lead to a larger distance, hence better separation, than two small, but fully exclusive clusters.

In this work, a measure for machine repetition r is included as one of the elements defining the new validity measure R to be used for partitioning the component set into groups. r is based on machine membership function (equation (14)) which has a value in the range of $0 < g_{ij} < 1$ for all repeated machines and is defined as:

$$r = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{q_i}(G_i)^2 = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{q_i}\sum_{j=1}^{q_i}\left(g_{ij}\right)^2 \tag{15}$$

where gi_j is membership function of machine M_j to group G_i and q_i is the number of machines in the group centroid G_i (number of the coordinates $g_{ij} > 0$).

The repetition measure r is defined in the interval [1/m, 1], where m is the total number of groups in the partition. If there is no overlapping then all machines have group membership of 1 or 0. The partition is crisp and the machine repetition measure has a maximum value $r_{\max} = 1$. If all machines are equally distributed between the groups, then repetition measure tends to a minimum, $r_{\min} = 1/m$.

The validity measure R is defined as the ratio of the average compactness λ to the machine repetition measure r:

$$R = \frac{\lambda}{r} = \frac{\sum_{j=1}^{m}\lambda_j}{\sum_{i=1}^{m}\frac{1}{q_i}(G_i)^2} \tag{16}$$

To make the compactness measure independent from the number of machines in the group, the average group compactness λ_j is defined as:

$$\lambda_{j=} \frac{1}{q_j}\frac{1}{n_j}\sum_{i=1}^{n_j}\mu_{ij}\, e^2\left(C_i, G_j\right) \tag{17}$$

Using equation (17) it is possible for the validity measure R to be used for comparison between group partitions based on sets of alternative machines. 'Optimum' component partitions can be found by minimizing the validity measure R, a combination of minimum component deviation λ within groups (maximum compactness) and minimum total number of repeated machines (maximum value of repetition measure r).

The validity measure R is defined in a similar way to that used by Xie and Beni (1991) in defining their validity measure S, the difference being in the formula used to measure group overlapping. The existence and uniqueness of R can therefore be proved via its relationship to the validity measure S. Using this relationship it can be further transformed to a well established hard-partition validity measure (Dunn 1974).

The solution methodology for part family formation using the extended algorithm and the compactness and repetition validity measure R is summarized as follows:

Step 1. Initialize the number of groups $m = 2$, $R_{min} = \infty$, $m^* = 1$.
Step 2. Initialize the grouping centres.
Step 3. Compute fuzzy membership functions μ_{ij}
Step 4. Compute machine membership functions g_{ij} and grouping centroids G_i

Step 5. Re-cluster the components by repeating steps 3-4 after assigning each component.
Step 6. Compute cluster validiation measure *R*.
Step 7. Define optimum partition Ω_m by repeating steps 2-6 for each number of groups *m* until the value of validity function *R* is no longer decreasing.
Step 8. If $R < R_{rmin}$, then $R_{min} = R$, $m^* = m$.
Step 9. Repeat steps 2-8 by increasing the number of groups $m = m + 1$ until $m = m^{o}$ (stop value) then select the partition with $m = m^*$ groups.

4. Results and discussion

The results of applying the extended algorithm for grouping the same set of 120 components used in this case study are shown in Figs 4 and 5. Figure 4 shows that the new definition of the cluster centroids leads to a better distribution of the components among the groups. The machine membership function included in the definition of the grouping centroids influences component membership and therefore limits the undesirable tendency of components to be clustered to the more powerful groups observed previously.

Figure 5 shows that, accordingly to R_{min}, the 'optimum' combination (in terms of compactness and machine repetition) occurs when the component set is partitioned into 4 groups. This leads to a repetition of two machines (see Fig. 6) compared with three repeated machines and partitioning into five groups obtained when *S* is used as the validity measure.

Figure 6 shows the number of repeated machines at various group partitions when the validity measure *R* is used for partitioning the component set. The total number of repeated machines is lower at most partition levels and its variation consistent with the character of

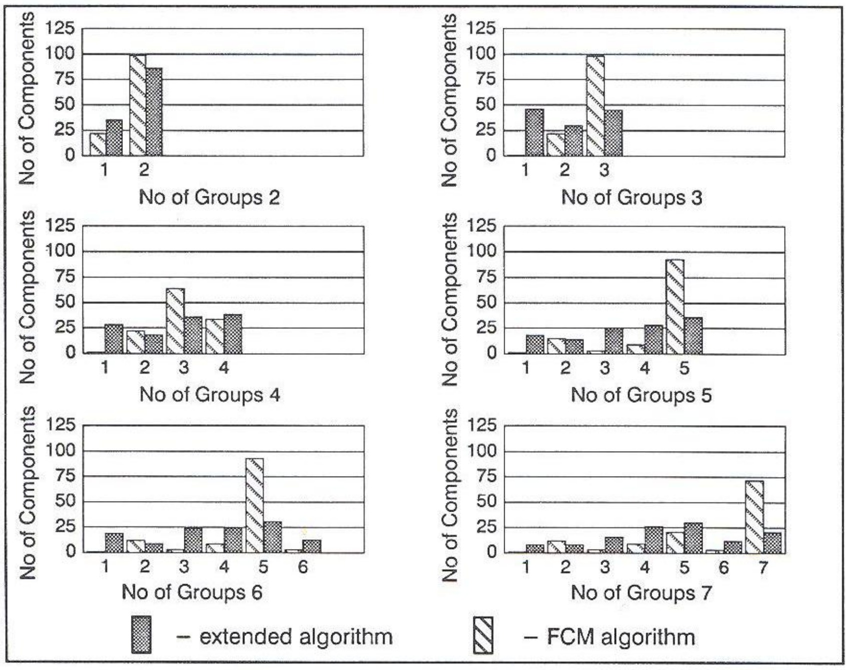

Figure 4. Component distribution – results of the extended fuzzy algorithm

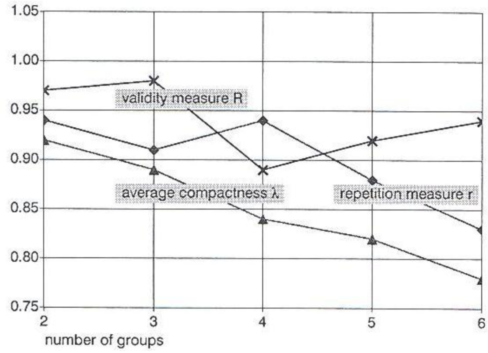

Figure 5. Grouping results based upon the extended fuzzy algorithm – data for 120 components.

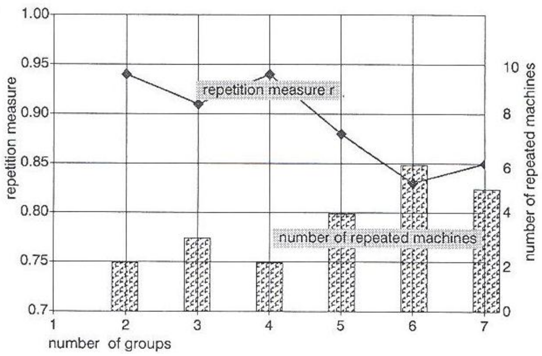

Figure 6. Variation of the repetition measure r and the number of repeated machines.

the repetition measure r, compared with the results obtained when group separation is measured using s as shown in Fig. 3.

Two well known grouping examples from the published literature were selected to further test the proposed methodology: the case of 9 machines and 9 components described in Gongaware and Ham (1981) and Chu and Hayya (1991), and the set of 41 components represented by 30 machines described in Kumar and Vannelli (1987) and Vannelli and Hall (1993).

The results of the first example (Gongaware and Ham 1981, Chu and Hayya 1991) obtained using the validity measure R are shown in Fig. 7. Minimum R suggests 'optimum' partitioning of the components set into three groups. It can be seen that the contents of the resulting groups match closely the results of Gongaware and Ham (1981) and Chu and Hayya (1991). The difference here is that the number of groups is not pre-selected but deduced based on the minimum value of the validity measure R ($R_{min} = 0{\cdot}3$) as shown in Fig. 8. The machines required by each group are represented by their membership function and if attached solely to the clusters where membership is maximum, the results are identical to those obtained by hard clustering for $m = 3$ as the pre-defined number of groups (Gongaware and Ham 1981).

The results of the second example are shown in Fig. 9 and the variation of the validity measure R at various partition levels are shown in Fig. 10. According to R_{min},

Partition into 2 groups Validity Measure R = 0.32	
group 1	
components	3 4 5 7 8
machines	3(1)* 7(1) 8(1) 5(0.77) 1(0.22) 4(0.66)
group 2	
components	1 2 6 9
machines	2(1) 6(1) 9(1) 1(0.78) 5(0.23) 4(0.34)
Partition into 3 groups Validity Measure R = 0.30	
group 1	
components	3 7 8
machines	3(1) 4(0.50) 7(1) 8(0.62) 5(0.34)
group 2	
components	2 6 9
machines	1(0.42) 2(0.72) 4(0.35) 6(1) 9(1)
group 3	
components	1 4 5
machines	1(0.58) 2(0.28) 5(0.66) 4(0.15) 8(0.38)
Partition into 4 groups Validity Measure R = 0.52	
group 1	
components	3 7 8
machines	3(1) 4(0.34) 7(1) 8(0.41) 5(0.24)
group 2	
components	2 6 9
machines	1(0.39) 2(0.64) 4(0.26) 6(1) 9(1)
group 3	
components	1 5
machines	1(0.61) 2(0.36) 5(0.38) 8(0.18)
group 4	
components	4
machines	4(0.40) 5(0.38) 8(0.41)

* Machine membership is shown in brackets

Figure 7. Grouping results – Example 1: 9 components and 9 machines (Gongaware and Ham 1981).

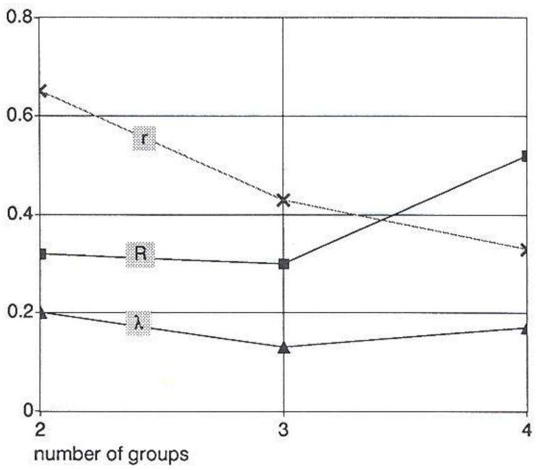

Figure 8. Variation of the validity measure *R*: Example 1.

Partition into 3 groups Validity Measure R = 0.71	
group 1	
components	1 3 6 8 9 13 14 15 21 22 28 29 30 35
machines	8(1)* 19(1) 20(1) 29(1) 30(1) 9(1) 16(1) 27(1) 28(1) 18(0.16) 6(0.58) 4(0.47)
group 2	
components	2 10 11 12 18 20 23 26 31 32 33 39 40 41
machines	10(1) 23(1) 1(1) 11(1) 21(1) 22(1) 2(1) 3(1) 12(0.80) 14(0.30) 6(0.42)
group 3	
components	4 5 7 16 17 19 24 25 27 34 36 37 38
machines	13(1) 24(1) 25(1) 5(1) 17(1) 7(1) 26(1) 15(1) 12(0.20) 14(0.70) 4(0.53) 18(0.84)

Partition into 4 groups Validity Measure R = 0.63	
group 1	
components	1 3 8 9 13 14 15 21 22 29 30 35
machines	8(1) 19(1) 20(1) 29(1) 30(1) 9(1) 27(1) 28(1) 18(0.12) 4(0.20)
group 2	
components	10 12 18 23 31 32 33 39 40 41
machines	1(1) 11(1) 21(1) 22(1) 3(1) 2(0.86) 10(0.68) 12(0.55) 23(0.84) 14(0.11)
group 3	
components	4 5 6 16 17 26 27 34 36 37
machines	25(1) 5(1) 6(1) 7(1) 26(1) 15(1) 14 (0.88) 16(0.41) 17(0.78) 18(0.87)
group 4	
components	2 7 11 19 20 24 25 28 38
machines	13(1) 24(1) 10(0.32) 23(0.16) 4(0.80) 17(0.22) 2(0.14) 12(0.45) 16(0.59)

Partition into 5 groups Validity Measure R = 0.67	
group 1	
components	1 3 8 9 13 14 21 22 29 30 35
machines	8(1) 19(1) 20(1) 29(1) 30(1) 9(1) 28(1) 27(0.57) 18(0.12) 4 (0.11)
group 2	
components	2 11 20 23 28 39 41
machines	10(0.68) 23(0.55) 2(0.44) 12(0.49) 3(0.39) 22(0.35) 4(0.17) 16(0.42) 1(0.44) 11(0.66)
group 3	
components	4 5 7 15 16 17 27 34 36 37
machines	14(0.51) 25(1) 5(1) 4(0.26) 17(1) 27(0.42) 7(1) 18(0.88) 26(1) 15(1)
group 4	
components	12 19 24 25 38 40
machines	13(1) 24(1) 2(0.24) 3(0.36) 10(0.32) 12(0.51) 21(0.43) 22(0.32) 23(0.29) 4(0.46)
group 5	
components	6 10 18 26 31 32 33
machines	6(1) 16(0.58) 1(0.56) 11(0.34) 21(0.56) 22(0.33) 2(0.32) 14(0.49) 3(0.25) 23(0.15)

* Machine membership is shown in brackets

Figure 9. Grouping results – Example 2: 41 components and 30 machines (Kumar and Vanelli 1987).

'optimum' partitioning is achieved at four component groups (see Fig. 9). This compares favourably with a pre-defined number of 3 groups suggested by Kumar and Vannelli (1987), possible partitioning into 4 groups (one of the variants suggested by Vannelli and Hall 1993) and the selection of 5 groups proposed by Kaparti and Suresh (1992). The distribution of components among the groups is similar to the results obtained by Vanelli and Hall (1993) using the eigenvector methodology.

5. Conclusions

A new component grouping methodology for cells formation is presented. It is an extended version of the fuzzy C-means clustering algorithm for component grouping with cluster validation procedure for selection of 'optimum' component partitions.

The validity measure *R,* proposed in this paper, is aimed at component grouping for

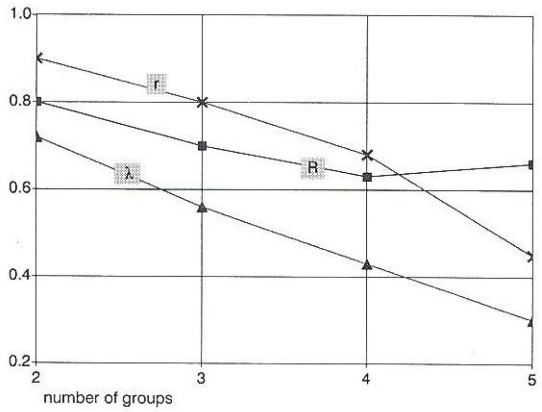

Figure 10. Variation of the valididity measue R – Example 2.

cellular manufacturing applications where maximum diversity between manufacturing cells is considered of prime importance. The validity measure R has proved very useful in 'optimizing' component partitioning by forming component groups with the maximum compactness of the components within groups and of machining cells with a minimum number of repeated machines.

Extensions are also proposed for improving the performance of fuzzy grouping algorithms in part family formation through a procedure for initializing the grouping prototypes and the introduction of a machine-group membership function. The definition of the grouping centroids and machine membership function has provided an appropriate basis for the simultaneous definition of component groups and machining cells through continuous evaluation of component partitions. Moreover, the proposed machine membership function allows post-grouping decision-making in cases where restricted machining resources are required by several groups.

The validity measure has been experimentally assessed using industrial data and compared with similar validity measures used in fuzzy clustering analysis. The results show that the fuzzy clustering approach and the validity measure R, proposed in this work, provide a realistic solution methodology useful for part family formation in cellular manufacturing applications.

6. Acknowledgments

The research reported here is supported by a grant provided by ACME Directorate of SERC; their financial contribution is gratefully acknowledged.

References

BEZDEK, J., 1975, Numerical taxonomy with fuzzy sets. *Journal of Mathematical Biology,* **l,** 57-71.

BEZDEK, J., 1980, A convergence theorem for the fuzzy ISODATA clustering algorithms. *IEEE Transactions of Pattern Analysis and Machine Intelligence,* 2 (1), 1-8.

BEZDEK, J., 1987, *Analysis of Fuzzy Information,* 3, Applications in Engineering and Science (Florida: CRC Press).

CANNON, R., JITENDRA, V. D., and BEZDEK, J., 1986, Efficient implementation of the fuzzy C-means clustering algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 8 (2), 248-255.

CHU, C. H., and HAYYA, J. C., 1991, A fuzzy clustering approach to manufacturing cell formation. *International Journal of Production Research,* 29 (7), 1475-1487.

DAVIES, D. L., and BOULDIN, D. W., 1979, A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 1(2), 224-227.

DUNN, J. C., 1974, Well separated clusters and optimal fuzzy partitions. *Journal of Cybernetics,* 4, 95-104.

GONGAWARE, T. A., and HAM, I., 1991, Cluster analysis applications for group technology manufacturing systems. *9th North American Manufacturing Research Conference (NAMARC) Proceedings.* Society of Manufacturing Engineers, 503-508.

GATH, I., and GEVA, A. B., 1989, Unsupervised optimal fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 11 (7),773-781.

JAIN, A. K., 1988, *Handbook of Pattern Recognition and Image Processing* (New York: Academic Press).

KAPARTHI, A., and SURESH, N.C., 1992, Machine-component cell formation in group technology; a neural network approach. *International Journal of Production Research,* 30 (6), 1353-1367.

KAUFMANN, A., 1975, *Introduction to the Theory of Fuzzy Subsets* (New York: Academic Press).

KUMAR, K. R., and VANNELLI, A., 1987, Strategic subcontracting for efficient disaggregated manufacturing. *International Journal of Production Research,* 25, 1715-1728.

PURCHEK, G., 1985, Machine component group formation: an heuristic method for flexible production cells and flexible manufacturing systems. *International Journal of Production Research,* 23 (5), 911-943.

SELIM, S. Z., and IZMAIL, M.A., 1986, Efficient implementation of the fuzzy C-means clustering algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 8 (2), 284-288.

SINGH, N., 1993, Design of cellular manufacturing systems-an invited review, *European Journal of Operational Research,* 69 (3), 284-291.

SINGH, N., and MAHANTY, B. K., 1991, A fuzzy approach to multi-objective routing problem with applications to process in manufacturing systems. *International Journal of Production Research,* 29 (6), 1161-1170.

VANNELLI, A., and HALL, R. G., 1993, An eigenvector solution methodology for finding part-machine families. *International Journal of Production Research,* 31 (2), 325-349.

WANG, H.-P., and LI, J.-K., 1991, *Computer Aided Process Planning* (New York: Elsevier).

WARREN, G., and MOODIE, C., 1993, Cellular manufacturing. Report TAP930104, Purdue University.

XIE, X. L., and BENI, G., 1991, A Validity Measure for Fuzzy Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 13 (8), 841-847.

XU, H., WANG, H.-P., 1989, Part family formation for GT applications based on fuzzy mathematics. *International Journal of Production Research,* 27 (9), 1637-1651.