Management and Production Engineering Review
Volume 10 • Number 4 • December 2019 • pp. 90–100
DOI: 10.24425/mper.2019.131449

GENERALIZED UTILIZATION-BASED SIMILARITY COEFFICIENT FOR MACHINE-PART GROUPING PROBLEM IN CELLULAR MANUFACTURING

Tamal Ghosh

Norwegian University of Science and Technology, Department of Manufacturing and Civil Engineering, Norway

Corresponding author:
Tamal Ghosh
Norwegian University of Science and Technology
Department of Manufacturing and Civil Engineering
Teknologivegen 22, 2815 Gjøvik, Norway
phone: +47 909 12 002
e-mail: tamal.ghosh@ntnu.no

Received: 11 November 2018
Accepted: 29 November 2019

Abstract

This article intends to justify the gap in the research of similarity coefficient driven approaches and cell formation problems (CFP) based on ratio data in cellular manufacturing systems (CMS). The actual implication of ratio data was vaguely addressed in past literature, which has been corrected recently. This research considered that newly projected CFP based on ration data. This study further revealed the lack of interest of researchers in investigation for an appropriate and improved similarity coefficient primarily for CFP based on ratio data. For that matter a novel similarity coefficient named as Generalized Utilization-based Similarity Coefficient (GUSC) is introduced, which scientifically handles ratio data. Thereafter a two-stage cell formation technique is adopted. First, the proposed GUSC based method is employed to obtained efficient machine cells. Second, a novel part allocating heuristic is proposed to obtain effective part families. This proposed approach is successfully verified on the test problems and compared with algorithms based on another similarity coefficient and a recent metaheuristic. The proposed method is shown to obtain 66.67% improved solutions.

Keywords

Cellular manufacturing, machine utilization percentage, ratio data, similarity coefficient.

Introduction

Group Technology (GT) is a contemporary manufacturing philosophy that develops part families exploiting the resemblances of parts based on the geometric shapes, features or manufacturing requirements and allocates them to the suitable machine groups/cells. GT exploits the benefits of flow production such as decreased throughput times, reduced work in progress, curtailed tool requirements, enhanced product quality and improved control of operations. CMS is an application of GT, which presents a hybrid system of jobshop (production variety) and flowshop (production volume). The design process of CMS is initiated with an effective solution to the machine-part grouping problem that would attain the competent machine cells to further process the appropriate part families in optimized production condition [1]. Designing effective cells in CMS is the most crucial and basic task in CMS research. There have been numerous approaches proposed in CMS literature, which effectively solve the CFPs in CMS. Among these, production flow analysis (PFA) based method is mostly explored [2], which deals with processing requirements of parts, operational sequences and operational time of the parts on the machines. Many review and survey articles are published based on the cell formation techniques in CMS [3]. Among which array-based methods [4, 5], clustering methods [6, 7], graph theory driven methods [8, 9], mathematical programming methods [10, 11] and similarity coefficient driven approaches [12, 13] are prominent. Recently soft-computing based optimization techniques are being adopted for CFPs to find near-optimal solutions efficiently. References [14] and [15] have covered nearly the entire horizon of soft-computing based methodologies being applied for machine-part grouping problems.

Among all these various techniques, similarity coefficient (SC) based approaches are acknowledged to be more adaptable, computationally inexpensive, prompt and exact while compared to the other methods. References [16] pointed out these SC based approaches are more sensitive while including manufacturing information in the CFP. The research in engineering computation is becoming critical due to the complexities in real world applications. Hence the computationally inexpensive or data-driven models are being evolved recently [17] and SC based approaches are ideal in such cases. Therefore, the main research questions are being set as, how could SC be used in newer CFP based on ratio data? Could this new SC be compared with latest cell formation techniques? To answer these, an attempt is made in this paper to construct a novel utilization-based SC to solve the CFP based on ratio data and compared the results with latest cell formation techniques.

Similarity coefficient

The SC based scientific measures are being adopted in many disciplines of scientific research such as physical science, biological science, medical science, engineering science etc. [18]. The SC describes numerical score of similarity between a pair of objects (machines in CMS). It is required to adopt a suitable clustering method while employing SC to obtain data groups/clusters. SC approaches are substantially popular in the domain of CMS [19]. The step by step procedure of SC based methods can be competently employed in CMS [20], which is as follows,

Input:

- Machine part Incidence Matrix (MPIM).
- Rows and columns of MPIM represent machines and parts respectively. The MPIM matrix is denoted by $U_{q\times p}$ with q machines and p parts. An element of $U_{q\times p}$ is denoted by u_{ij} such that,

$$u_{ij} = \begin{cases} \text{non zero}, & \text{if part } j \text{ goes machine } i, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Step 1:

Employ an appropriate SC and compute the numerical score of similarity between each pair of machines to obtain a symmetric similarity matrix. For an example, the machine-machine similarity matrix is denoted by $S_{q\times q}$ and its elements are s_{ij} expressed as,

$$s_{ij} = \begin{cases} 1, & \text{if } i = j, \\ \text{similarity value}, & \text{otherwise,} \end{cases} \tag{2}$$

and

$$s_{ij} = s_{ji}. \tag{3}$$

Equation (2) states that the diagonal elements of matrix $S_{q\times q}$ has maximum similarity (similarity between same machines for each row) and the score is 1. Equation (3) portrays that the similarity between machines i and j is same as the similarity between machines j and i, therefore the matrix is symmetric in nature.

Step 2:

Once the similarity matrix $S_{q\times q}$ is obtained, a suitable clustering approach is required to apply on the similarity scores available through matrix $S_{q\times q}$. This procedure would find out clustering scores based on some threshold values specified for the algorithm and finally obtains a tree structure known as dendrogram. This shows the hierarchical structure of each pair of machines. This dendrogram visually depicts the machine clusters to be selected depending upon the number of cells.

Output:

Machine-Cell assignment matrix $MC_{q\times c}$ where each elements mc_{ij} is presented as,

$$mc_{ij} = \begin{cases} 1, & \text{if machine } i \text{ is assigned to cell } j, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Cell formation problems are combinatorial optimization (CO) problems due to their intrinsic complexities. Due to this fact many methodologies are available in past literature as stated in section #1. Among these approaches SC based techniques are the primitive techniques since the basic objective of SC and GT are indifferent. Both exploit similarities between objects (machines) and group them systematically [21]. Unlike other methodologies such as mathematical programming approaches or soft-computing techniques, the SC based approaches are more flexible in terms of choice of SC and clustering algorithms. Therefore, selection of a particular SC doesn't influence the selected clustering algorithms since the step 1 and step 2 of the procedure mentioned previously are independent to each other [18].

One of the earliest papers of 1972 successfully initiated the use of SC with an adoption of single linkage clustering algorithm (SLCA) to form efficient cells in CMS [12]. Thereafter many researchers successfully practiced SC based approaches in CMS [22–34].

SC based approaches can be further classified as, (1) problem specific approaches (solely designed for CFPs) [25, 30, 31] and (2) generic approaches (applicable for any data classification problems) [22, 34]. References [18] suggested that the generic approaches are more inclined in finding the similarities between pair of machines rather than the appropriateness while the problem specific SC would act inversely. This fact implies that the problem specific approaches are more suitable when important production factors such as part volume, unit processing time, alternative routing, operation sequence, machine utilization etc. are considered in the design. However, the aim of this discussion is to restrict this section within the scope of SC oriented approaches to the CFP based on ratio data. References [25] first proposed a novel SC known as Cell Bond Strength (CBS) based on production volume and unit processing time, which is shown to be more effective than Jaccard's index driven SC. References [26] has suggested a problem model other than the 0-1 binary CFP, which can incorporate part volume or unit processing time. This work exploited seven existing similarity measures such as McAuley's coefficient [12], modified multiplicative weight, modified Hamann's index, modified Baroni-Urbani's measure etc. along with four hierarchical clustering techniques efficiently. A primitive and generic matching measure is used as the grouping measure along with product moment correlation coefficient and inter-cell moves [7]. In another study a new similarity coefficient is demonstrated, which successfully incorporates alternative part routings, machine capacity, part demand and processing times in an effective manner and conducted an optimal clustering analysis henceforth [35]. A production volume driven similarity coefficient is introduced by ref. [36], which modified Jaccard's index based McAuley's coefficient. This approach competently minimized the inter-cell and intra-cell moves and obtained an improved scheduling process. Ref. [34] proposed an e-learning tool, which utilizes the similarity between machines and parts, which is calculated using part volume and processing time. This minimizes inter-cell and intra-cell moves. It can be concluded that the CBS coefficient by [25] is the only SC, which solely deals with ratio data. This fact is also confirmed by Yin's major review work on SC [18]. The reason behind this limited use of SC based approaches could be found in a recent study [37]. SC based approaches either seems substantially simple or it is absolutely complicated to be developed. Exact SC would need much of the information from the problem in hand and a real effect of SC on performance metrics would also be an interesting area to be covered.

Problem definition

Ratio level data or *processing time* is being used synonymously in CMS since past few decades. However, the accurate explanation of *ratio level data* was rarely portrayed in past. In the earlier research works in CMS many researchers discussed part volumes, processing time, machining sequences, machine capacities as production factors to the CFPs, however these factors are not quantified reasonably in their studies [7, 25, 26]. Many other interrelated issues such as cell utilization, exceptional utilization, and machine-machine similarity measures are overlooked. References [38] first described *ratio level data* in the mathematical form. As proposed, the incidence matrix is generated using ratio of *total processing time* and *available machine hours*, which is practically termed as *machine utilization percentage* by the manufacturing personnel. This is stated as,

$$u_{ij} = \frac{(t_{ij} \times n_j)}{MH_i}, \tag{5}$$

where

$$u_{ij} = \begin{cases} \text{zero}, & \text{if part } j \text{ does not go in machine } i, \\ \text{non zero}, & \text{if part } j \text{ goes in machine } i, \end{cases} \tag{6}$$

t_{ij} – unit processing time (hour/unit) of part j on machine i; $1 \le i \le q$ and $1 \le j \le p$, n_j – production volume of part j, MH_i – available machine hours of machine i, $U = [u_{ij}]$ is $(p \times q)$ – machine-component incidence matrix where u_{ij} – percentage utilization of machine i induced by part j.

Equation (5) produces an MPIM, that is U, which is recognized as processing time or ratio data in past literature.

Since then the ratio data based CFPs are being used often in many articles [31, 39–43]. References [39] specified that the ratio data can be synonymously used as workload data in CMS and this phenomenon converts the binary incidence matrix into real valued U matrix. All the '1's of the binary incidence matrix are changed to fractional values, which would be called as the workload data or ratio data. This proposition is reinforced by the statement of [42, p. 637], *'The real valued matrix is produced by assigning random numbers in the range of 0.5 to 1 as uniformly distributed values by replacing the ones in the incidence matrix and zeros to remain in its same positions.'*. This procedure produces the real valued matrix U at random in unrestricted manner, which is unscientific realistically.

In real-world practice the elements (u_{ij}) of U point to ratio values. These are obtained using (Hours ÷ Hours) expression, which is unit-less. Therefore, these cannot be termed as *processing time/operational time* (absolute values). Some researcher mentioned this as *capacity percentage* [44], which is partially correct as capacity of a machine is also known as *available machining hours*. However, *capacity percentage* is a vague term since it is the *utilization* of machine expressed in percentage value. u_{ij} demonstrates the fraction of *available machine hours* of i-th machine needed to process the demanded quantity of j-th part. This is, termed as *percentage utilization of machine* in real factory shop-floor, more reasonable and appropriate terminology than the *processing/operational time*.

To correctly present the real valued matrix U, a constraint is prescribed with Eq. (5).

Constraint:

Eqn. (7) depicts that the sum of *percentage utilization* of all parts over i^{th} machine is required to be less than or equal to 1. This is because the total utilization of any machine would never exceed its total available machine hours, which is 100%.

$$\sum_{j=1}^{p} u_{ij} \leq 1. \tag{7}$$

In experience closely all the authors of the published articles, who considered ratio data in CMS, ignored the above indicated constraint of Eq. (7), which is a crucial proposition while designing the test datasets. Even after proposing the correct mathematical formulation of ratio data, [38] also overlooked constraint of Eq. (7) while generating their test problems. Thus, practicing the test problems from the past literature would not be a correct selection for the ratio data driven CFP. To serve the purpose a novel step by step technique for ratio data matrix generation is described as,

Input:
Number of machines q and parts p

Routine:
Create random real valued matrix of size $q \times p$
Case 1: if $q \leq 10$
Limit the density of zeroes in the range of 40% to 50%
Limit sum total of each row ≤ 1
Case 2: if $q \leq 20$
Limit the number of zeroes in the range of 60% to 70%
Limit sum of each row ≤ 1
Case 3: if $q > 20$
Limit the number of zeroes in the range of 80% to 90%
Limit sum of each row ≤ 1

Output:
$q \times p$ real valued incidence matrix

This proposed technique carefully impose a control mechanism for the density of zeros in the incidence matrix while satisfying the constraint of Eq. (3). A thorough inspection throughout all the real valued test problems of past literature reveals that the density of zeroes is restricted in the range of 40% to 50% in small size problems ($q \leq 10$), 60% to 70% in medium test problems ($q \leq 20$) and 80% to 90% in large test problems ($q > 20$). In future students/researchers can promptly obtain the test problem of any size to use it in their study/research.

Research methodology

This section introduces a new Generalized Utilization-based Similarity Coefficient (GUSC) for CFP based on ration data. Thereafter a two-stage technique is introduced, which efficiently solves the CFP. In the first stage the GUSC and the single linkage clustering algorithm (SLCA) is used to obtain machine cells first. In the next stage an appropriate part allocating heuristic is applied to assign the part families to the newly developed cells.

Generalized Utilization-based Similarity Coefficient (GUSC)

The CBS similarity coefficient is defined by the sum of the ratios of the total utilization percentage on machine M1 and M2 owing to the shared parts on machines M1 and M2 [25]. It is expressed as,

$$\mathrm{CBS}[\ \mathrm{M1,M2}\ ] = \frac{a+c}{a+b+c} + \frac{d+e}{d+e+f}, \tag{8}$$

a, b, c, d, e and f are the percentage utilization of parts P1, P2, P3 and P4 on machines M1 and M2 shown in Table 1.

Table 1
MPIM example.

	P1	P2	P3	P4
M1	a	b	c	
M2	d		e	f

CBS is a basic SC, which works well with ratio data and the solution generated using CBS is not the *best solution* to the problem but is considerably *good solution*. A good design of SC for utilization driven CFP must incorporate the following ideas,

1) total utilization percentage of shared parts on both machines M1 and M2,
2) total utilization percentage of shared parts on machine M1 only,
3) total utilization percentage on of shared parts on machine M2 only,

4) total number of operations of shared parts on none of the machines M1 and M2.

The formulation of the CBS did not consider all the stated facts while computing similarity scores between a pair of machines. Therefore, a more generalized concept of a similarity coefficient is required for utilization driven CFP. Henceforth a new similarity coefficient named as the Generalized Utilization-based Similarity Coefficient (GUSC) is proposed and it is expressed as,

$$s_{ij} = \left[\frac{\sum\limits_{k=1}^{p} u_{ik}x_{ijk}}{\sum\limits_{k=1}^{p} u_{ik}} + \frac{\sum\limits_{k=1}^{p} u_{jk}x_{ijk}}{\sum\limits_{k=1}^{p} u_{jk}}\right] \times \left[\frac{1}{1+\sum\limits_{k=1}^{p} y_{ijk}}\right], \tag{9}$$

S_{ij} – similarity between machine i and machine j, u_{ik} – utilization of machine i induced by part k, p – total number of parts, q – total number of machines, x_{ijk} – 1 if part k visits both machines i and j; 0 otherwise, y_{ijk} – 1 if part k visits neither of the machines i and j; 0 otherwise.

The major objectives of the proposed GUSC are to minimize total exceptional utilization (TEU) which is the total utilization percentage induced by exceptional elements (TEU) and maximize total cell utilization (TCU) which is the sum total of the individual in-cell utilization percentage. This will indirectly control the number of voids and exceptional elements.

Linkage clustering techniques

Single Linkage Clustering Algorithm (SLCA) is ideally a simple hierarchical clustering algorithm, which can be employed in conjunction with some appropriate similarity or distance measure for the clustering analysis of data [45]. This technique produces explanatory narrations and graphically present the structure of obtained data clusters. This method would be more suitable if the hierarchical correlation exists in data. SLCA is recognized as an adjacent neighbor technique, which exploits the distance between two clusters. The distance between cluster i and another cluster j is defined as:

$$E_{ij} = \min(d_{ij}), \quad i \in (1, ..., n_r) j \in (1, ..., n_s). \tag{10}$$

Equation (10) hierarchically generates a $(m-1) \times 3$ intermediate matrix, where m is the number of machines in the CFP. Columns of this intermediate matrix depict indices of machine clusters, which are hierarchically pairwise connected to obtain a binary tree. The leaf nodes of the tree are assigned machine numbers from 1 to m. The higher clusters are obtained from the leaf nodes and visualized further. This visual tree is termed as dendrogram that indicates the potential clustering results.

Part allocating heuristic

Next step is to assign the appropriate part families to the machine cells obtained from the previous step. For that matter a part allocating heuristic is developed based on the part allocation factor for part j in cell k. The formula is stated as,

$$\mathrm{PAH}_{jk} = \frac{\left(\sum\limits_{i=1}^{q_k} u_{ij}\right)^2 \sum\limits_{i=1}^{q_k} a_{ij}}{\sum\limits_{i=1}^{q_k}\sum\limits_{j=1}^{p} u_{ij} \sum\limits_{i=1}^{q} u_{ij} \sum\limits_{i=1}^{q} a_{ij}}, \tag{11}$$

q_k – the number of machines in cell k, q – total number of machines in plant, p – total number of parts in plant, u_{ij} – utilization on machine i induced by part j (non-zero or zero ratio level data), a_{ij} – if part j is being processed by machine j (0-1 binary data), PAH_{jk} – association score of part j in cell k.

Equation (11) is the product of three fractional figures, percentage of total utilization of part j in cell k, percentage of total utilization of cell k required to process part j and percentage of total number of operations for part j being processed in cell k. The design of part allocating heuristic carefully incorporates two major objectives,

- minimization of total exceptional utilization (TEU) by reducing number of exceptional utilizations,
- maximization of total cell utilization (TCU) by reducing number of voids.

The cell formation algorithm

The two-stage cell formation algorithm is demonstrated as,

Stage 1:

Input:

Machine-part incidence matrix U

Output:

Machine-cell assignment string MCA

Machine Cell Assignment Procedure:

Step 1.	**Procedure** *similarity()*
1.1	**Set** $[m, n] = \mathrm{size}(U)$
1.2.	**Initialize** $m \times m$ empty matrix S
1.3.	**Initialize** number of cells c
1.4.	**For** $i = 1$ to m
1.4.1.	**For** $j = i + 1$ to m
1.4.1.1.	**Compute** S_{ij} for pair of machines (i, j) using Eq. (9)
1.4.1.2.	**Set** $S_{ij} = S_{ji}$
1.4.2.	**End**

1.5. **End**
1.6. **For** $i = 1$ to m
1.6.1. **For** $j = 1$ to m
1.6.1.1.1. **if** $i == j$
1.6.1.1.2. **Set** $Sij = 1$
1.6.2. **End**
1.7. **End**
1.8. **Return** S
Step 2. **End**
Step 3. **Procedure** *Cluster()*
Step 4. **Initialize** an one-dimensional empty array MCA of size m
Step 5. Initialize an empty matrix MC of size $m \times c$
5.1. **Set** all the machines as leaf nodes or singleton clusters
5.2. **Compute** the smallest Euclidian distance between two clusters
5.3. **Construct** a matrix $INTM$ of size $(m-1) \times 3$ to from the hierarchical tree structure
5.4. **Construct** dendrograms using matrix $INTM$
5.5. **Select** appropriate machine cells for the maximum level of similarity
5.6. **For** $i = 1$ to m
5.6.1. **For** $k = 1$ to c
5.6.1.1. **Set** $MC_{ik} = 1$ or *0* depending upon the machine-cell assignment
5.6.2. **End**
5.7. **End**
5.8. **For** $i = 1$ to m
5.8.1. **For** $k = 1$ to c
5.8.1.1. **if** $MC_{ik} == 1$
5.8.1.2. **Set** $MCA(i) = k$
5.8.2. **End** loop
5.9. **End** loop
5.10. **Return** MCA

Stage 2:
Input:
MCA and U
Output:
PCA

Part Assignment Heuristic:

Step 1. **Initialize** an empty matrix MAT of size $q \times c$
Step 2. **Set** $[q, p] = \text{size}(U)$
Step 3. **Set** $c = \max(MCA)$
Step 4. **Initialize** empty matrix PAH of size $p \times c$
Step 5. **Initialize** an one-dimensional empty array PCA of size p
Step 6. **For** $i = 1$ to q
 a. **Set** $MAT(i, MCA(i)) = 1$
Step 7. **For** $j = 1$ to p
 a. **For** $k = 1$ to c
  i. **Calculate** PAH_{jk}, elements of PAH using *Eq. (11)*
 b. **End** loop
Step 8. **End** loop
Step 9. **For** $j = 1$ to p
 a. **For** $k = 1$ to c
  i. **if** $PAH_{jk} == \max(j\text{-th row of } PAH)$
  ii. **Set** $PCA(j) = k$
 b. **End** loop
Step 10. **End** loop
Step 11. **Return** PCA

A numerical example

- To illustrate the proposed two stage approach, a small test problem (5×10) is considered in Table 2. The first step is to obtain similarity matrix of size (5×5) from this incidence matrix using GUSC. The similarity matrix S is a symmetric matrix. Therefore, only the upper half is shown in Table 3.
- In next step SLCA algorithm is applied to obtain the dendrogram and machine-cell assignment vector MCA. The dendrogram is shown in Fig. 1, where five machines are visible along horizontal axis as singleton clusters or leaf nodes.

Table 2
(5×10) MPIM U.

	m1	m2	m3	m4	m5
p1	0.149741	0.165117	0.197491	0	0.133516
p2	0.01799	0	0.034073	0	0
p3	0.037942	0	0	0	0
p4	0	0	0.174167	0	0.203849
p5	0.175803	0.197782	0.000131	0	0
p6	0.19799	0.105536	0.119881	0	0
p7	0.099619	0.18017	0	0	0.18466
p8	0	0	0	0	0.15649
p9	0.132189	0	0.222688	0.982303	0.192257
p10	0	0.185663	0.145023	0.016983	0.030215

Table 3
Symmetric matrix S (5×5) obtained using GUSC.

	m1	m2	m3	m4	m5
m1	1	0.51519	0.736592	0.381982	1.036835
m2		1	0.433907	0.047908	0.341034
m3			1	0.35289	0.72445
m4				1	0.249384
m5					1

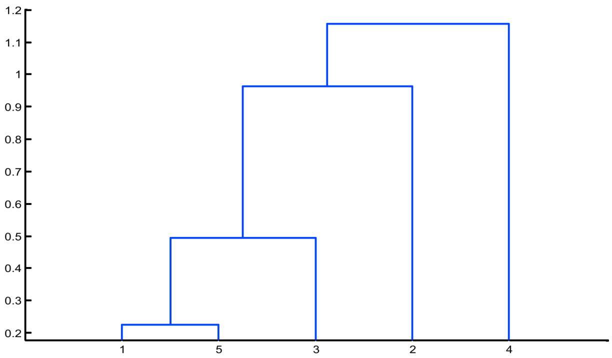

Fig. 1. Dendrogram obtained for (5×5) similarity matrix.

Thereafter the hierarchical links are also visible, and two cells are identified as cell 1 {machine 1, 2, 3, 5} and cell 2 {machine 4}. The MCA vector is presented as,

$$\text{MCA} = \left[\begin{array}{ccccc} 1 & 1 & 1 & 2 & 1 \end{array}\right].$$

In the next stage, the MCA vector is provided as input to the part allocating heuristic algorithm. After applying part allocating heuristic method, a part-cell mapping matrix is obtained and presented in Table 4, which depicts closeness score of each part in each of the cells. The higher is the closeness score, the better is the mapping of that part to the corresponding cell. Table 4 shows that part 1 to 8 and part 10 are assigned to cell 1 and part 9 is assigned to cell 2. The part-cell assignment vector PCA is obtained from this result and presented as,

$$\text{PCA} = [\begin{array}{cccccccccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 1 \end{array}].$$

Table 4
Part-cell mapping matrix based on closeness score.

	cell 1	cell 2
p1	**0.187752**	0
p2	**0.015135**	0
p3	**0.01103**	0
p4	**0.109889**	0
p5	**0.108639**	0
p6	**0.123084**	0
p7	**0.135015**	0
p8	**0.045491**	0
p9	0.042674	**0.157837**
p10	**0.075149**	0.000191

Thereafter the solution matrix with block diagonal cellular structure (grey colored) is obtained from MCA and PCA and depicted in Table 5.

Table 5
Solution matrix for (5×10) test problem.

	m4	m1	m2	m3	m5
p9	0.982303	0.132189	0	0.222688	0.192257
p1	0	0.149741	0.165117	0.197491	0.133516
p2	0	0.01799	0	0.034073	0
p3	0	0.037942	0	0	0
p4	0	0	0	0.174167	0.203849
p5	0	0.175803	0.197782	0.000131	0
p6	0	0.19799	0.105536	0.119881	0
p7	0	0.099619	0.18017	0	0.18466
p8	0	0	0	0	0.15649
p10	0.016983	0	0.185663	0.145023	0.030215

Computational experiments

In this study a novel SC based cell formation approach is proposed considering percentage utilization of machines instead of the binary (0-1) or processing time. The proposed model minimizes TEU and maximizes TCU while improving the score of performance measure, which is developed by [46]. This is known as Utilization-based Grouping Efficiency (UGE), which effectively counts all the overseen disputes in all the previous performance measures for ratio data driven CFP. UGE is defined as,

$$\text{UGE} = \frac{\left(\sum_{k=1}^{c}\left[U_{\text{cell}}^{k}\left(1-\frac{V_k}{E_k}\right)\right]\right)\left(1-\frac{U_{ee}}{\sum_{k=1}^{c} U_{\text{cell}}^{k}}\right)}{U_{\text{plant}}}, \tag{12}$$

where

$$U_{\text{cell}}^{k} = \left\{\sum_{i=1}^{mic}\sum_{j=1}^{pic} u_{ij}\right\}^{k}, \tag{13}$$

$$U_{ee} = \sum_{i=1}^{moc}\sum_{j=1}^{poc} u_{ij}, \tag{14}$$

$$U_{\text{plant}} = \sum_{i=1}^{mtp}\sum_{j=1}^{ptp} u_{ij}, \tag{15}$$

c – number of cells, m – number of parts, p – number of machines, k – index of cell $\{k = 1, 2, ..., c\}$, i – index of machines $\{i = 1, 2, ..., m\}$, j – index of parts $\{i = 1, 2, ..., p\}$, U_{cell}^{k} – total utilization of k-th cell, U_{plant} – total utilization of plant, U_{ee} – total utilization outside the block diagonal cell structure, u_{ij} – utilization of machine i induced by part j; $1 \leq i \leq q$ and $1 \leq j \leq n$, V_k – total number of voids in cell k

$\{k = 1, 2, ..., c\}$, E_k – total number of elements in cell k $\{k = 1, 2, ..., c\}$, mic – number of machines in cell, pic – number of parts in cell, moc – number of machines outside of cells, poc – number of parts outside of cells, mtp – total number of machines in plant, ptp – total number of parts in plant.

UGE yields 100% efficiency score in the absence of exceptional elements or voids, which is referred as a perfect solution for CFPs.

Result and discussion

The proposed SC based approach is coded in MATLAB on a 2.4 GHz Intel i3 computer. The proposed algorithm is tested on a set of 15 test problems generated using the algorithm demonstrated in section #3. The obtained solutions are compared with the solutions obtained by CBS SC method and a recently developed metaheuristic method known as Non-dominated Sorting Buffalo Optimization (NSBUF II) [47]. Table 6 presents the comparison among the results of all three methods. It can be observed that the sizes of test problems are varying between very small (5×10) to very large (45×120). GUSC outperforms the CBS based approach in most instances by attaining 66.67% better UGE scores. However, GUSC produces results at per with the latest NSBUF II. For data #14, GUSC even outpace NSBUF II. While CBS attains best solutions for 3 test problems (9×15, 12×12, 22×35) out of 15. Only for the 7×11 test problem both the SC techniques obtains same solution as NSBUF II. This indicates that CBS is also a good measure and comparable with newer similarity indices. However, GUSC could be established as the most suitable SC for the ratio data-based CFP. Table 6 also portrays the TCU and TEU scores obtained by all the approaches. It is observed that the corresponding TCU and TEU values of an improved UGE score are higher and lower respectively. This fact clearly indicates that the UGE considers the utilization inside the cells and outside the cells. Another insight of utilization-based CFP is that, even if the number of voids is in the higher side but maximizing TCU value is more crucial while obtaining a good solution. GUSC is not only better as a SC, but also it can contest with latest metaheuristic algorithm.

This fact clearly demonstrates the effectiveness of GUSC based technique, which is computationally inexpensive, prompt and improved one for CFP based on ratio data.

Figure 2 shows the pictorial view of the results obtained by GUSC, CBS, and NSBUF II. The superiority of GUSC over CBS is clearly displayed in this graphical representation, however it is substantially close to the NSBUF II. In this study computational time is not considered as a performance criterion because SC based approaches produce solutions quickly even for the large size data. Therefore, time consumption is negligible. Moreover, the prime focus of this research is to provide effective and prompt solutions for CFP to the shop-floor personnel with enhanced TCU and TEU scores. For the inquisitive readers, the CPU time is provided for the largest test problem (45×120) solved using GUSC based approach, which is 0.4452 seconds.

Table 6
Comparison of results obtained using GUSC, CBS, NSBUF II.

No.	Size	No. of cell	GUSC			CBS			NSBUF II		
			UGE	**TCU**	**TEU**	**UGE**	**TCU**	**TEU**	**UGE**	**TCU**	**TEU**
1	5×10	2	**52.932**	3.8752	0.564	36.656	3.6203	0.819	**52.932**	3.8752	0.564
2	7×11	2	48.202	5.496	0.796	48.202	5.496	0.7966	48.202	5.496	0.7966
3	9×9	2	**50.322**	6.206	0.828	47.818	5.8975	1.1372	**50.322**	6.206	0.828
4	9×15	2	44.998	7.4956	0.64	**45.099**	6.7077	1.4279	**45.099**	6.7077	1.4279
5	10×10	3	**38.054**	7.0204	1.064	35.299	6.4671	1.618	**38.054**	7.0204	1.064
6	10×25	3	**37.492**	7.6846	1.100	36.028	7.454	1.3314	**37.492**	7.6846	1.100
7	12×12	3	21.918	6.2079	2.379	**28.839**	6.6607	1.9272	**28.839**	6.6607	1.9272
8	15×15	3	**36.923**	10.670	2.097	34.936	9.6504	3.1176	**36.923**	10.670	2.097
9	20×20	4	**25.069**	12.835	4.13	9.716	10.045	6.9205	**25.069**	12.835	4.13
10	20×35	4	24.503	14.326	3.635	19.556	13.242	4.7194	**26.912**	16.627	2.833
11	22×35	4	22.923	16.304	1.644	**24.678**	16.074	1.8751	**24.678**	16.074	1.8751
12	24×40	4	**19.737**	14.986	2.917	17.197	14.908	2.9959	**19.737**	14.986	2.917
13	30×48	5	16.473	20.720	4.779	15.669	20.701	4.7983	**20.016**	21.226	4.171
14	35×48	5	**15.539**	16.220	2.693	15.208	15.153	3.7607	15.143	15.011	3.983
15	45×120	6	12.089	34.602	7.918	8.5653	28.680	13.840	**14.767**	36.382	4.897

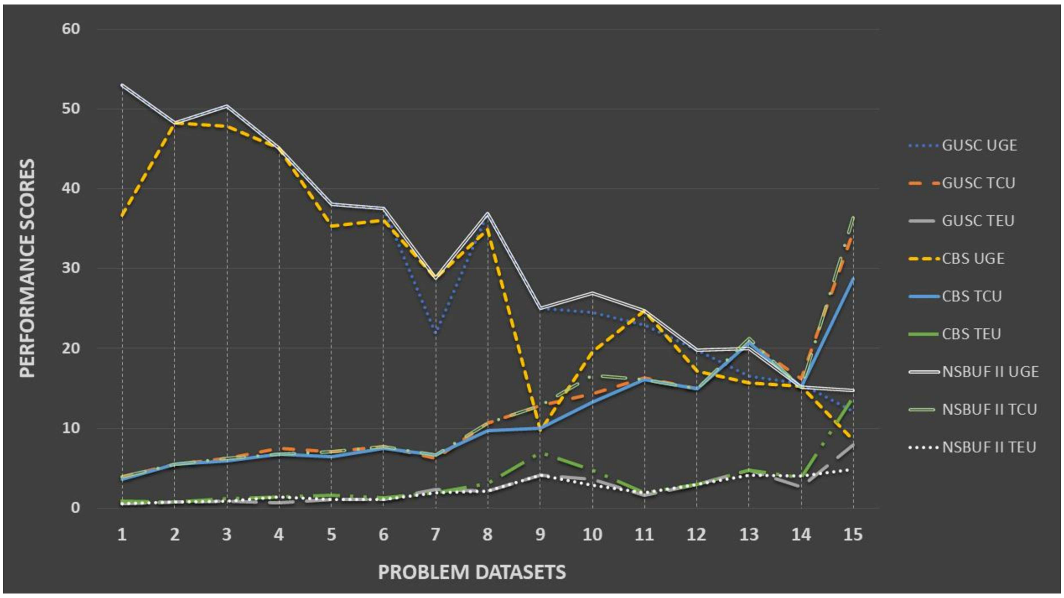

Fig. 2. Pictorial view of performance comparison among GUSC, CBS, NSBUF II in terms of UGE, TCU and TEU.

Managerial implications

This study has several perceptions, which are advantageous for the operations/production managers while taking decision on the shop-floor. Trading off between number of exceptional elements and voids are the primitive decisive factor for the managers while obtaining effective cells and minimizing the inter-cell and intra-cell movement costs. However, the prime objectives of utilization-based CFP are to increase the TCU and reduce the TEU. These in turn reduce the overall production time and costs, which further reduce inventories and market response times. Therefore, a prompt and realistic solution obtained from the proposed model can make the decisions of the managers easier and quick, which further possibly influences some important decisions such as the need of subcontracting of exceptional parts or machine duplications etc. The clustering analysis can also help to select the number of cells, which is not pre-defined in this approach. Thus, this approach shows more flexibility for the decision makers. Moreover, this approach is less complex mathematically, which is useful for the layman practicing in industry.

Conclusions

This article identifies the gap in the research of SC based approaches and ratio data driven machine-part grouping problems. Firstly, it is identified that very few articles practically introduced SC based approaches for CFP based on ratio data. Among those, only ref. [24] introduced a SC based method solely applicable to the ratio data CFP, which is useful. Thereafter this SC has never been improved, extended or modified. A novel data generation technique is adopted to generate the realistic test problems. Further, a new similarity coefficient namely GUSC is introduced in this study. This GUSC is successfully applied on 15 test problems generated beforehand and the applied algorithm obtains the ideal machine cells further. A novel part allocation heuristic is adopted subsequently to produce the part families. The proposed SC based approach is shown to outperform CBS by attaining 66.67% improved solutions and perform at per with a latest metaheuristic algorithm known as NSBUF II. The proposed approach can promptly generate the near-optimal solutions that could be useful for the shop-floor personnel in future. This work can be extended in future for other production factors such as product sequence, lot sizes etc. This GUSC could also be combined with latest metaheuristic algorithm to obtain more improved cells in CMS.

This work is supported by the SFI Manufacturing (Project No. 237900) and funded by the Research Council of Norway (RCN).

References

[1] Wemmerlöv U., Hyer N.L., *Research issues in cellular manufacturing*, Int. J. Prod. Res., 25, 3, 413–431, 1987.

[2] Burbidge J.L., *AIDA and group technology*, Int. J. Prod. Res., 11, 4, 315–324, 1973.

[3] Selim H.M., Askin R.G., Vakharia A.J., *Cell formation in group technology: review, evaluation and directions for future research*, Comput. Ind. Eng., 34, 1, 3–20, 1998.

[4] King J.R., *Machine-component grouping in production flow analysis: an approach using a rank order-clustering algorithm*, Int. J. Prod. Res., 18, 2, 213–232, 1980.

[5] Chandrasekaran M.P., Rajagopalan R., *MODROC: an extension of rank order clustering for group technology*, Int. J. Prod. Res., 24, 5, 1221–1233, 1986.

[6] Waghodekar P.H., Sahu S., *Machine-component cell formation in group technology: MACE*, Int. J. Prod. Res., 22, 6, 937–948, 1984.

[7] Mosier C.T., Taube L., *Weighted similarity measure heuristics for the group technology machine clustering problem*, Omega, 13, 6, 577–583, 1985.

[8] Ballakur A., Steudel H.J., *A with-in cell utilization based heuristic for designing cellular manufacturing systems*, Int. J. Prod. Res., 25, 5, 639–655, 1987.

[9] Srinivasan G., *A clustering algorithm for machine cell formation in group technology using minimum spanning trees*, Int. J. Prod. Res., 32, 9, 2149–2158, 1994.

[10] Kusiak A., *The generalized group technology concept*, Int. J. Prod. Res., 25, 4, 561–569, 1987.

[11] Rabbani M., Samavati M. Ziaee M.S., Rafiei H. *Reconfigurable dynamic cellular manufacturing system: a new bi-objective mathematical model*, RAIRO – Oper. Res., 48, 1, 75–102, 2014.

[12] McAuley J., *Machine grouping for efficient production*, Prod. Eng., 51, 2, 53–57, 1972.

[13] Seifoddini H., Wolfe P.M., *Application of the similarity coefficient method in group technology*, IIE Trans., 183, 3, 271–277, 1986.

[14] Papaioannou G., Wilson J.M., *The evolution of cell formation problem methodologies based on recent studies (1997–2008): Review and directions for future research*, Eur. J. Oper. Res., 206, 3, 509–521, 2010.

[15] Chattopadhyay M., Sengupta S., Ghosh T., Dan P.K., Mazumdar, S., *Neuro-genetic impact on cell formation methods of Cellular Manufacturing System design: A quantitative review and analysis*, Comput. Ind. Eng., 64, 1, 256–272, 2013.

[16] Seifoddini H., *Single linkage versus average linkage clustering in machine cells formation applications*, Comput. Ind. Eng., 16, 3, 419–426, 1989.

[17] Ghosh T., Martinsen K., Dan P.K., *Data-Driven Beetle Antennae Search Algorithm for Electrical Power Modeling of a Combined Cycle Power Plant*, [in:] Le T. H., Le H., Pham D.T. [Eds], Optimization of Complex Systems: Theory, Models, Algorithms and Applications, WCGO 2019, Advances in Inteligent Systems and Computing, Springer, 991, 906–915, 2019.

[18] Yin Y., *Application Similarity Coefficient Method to Cellular Manufacturing*, In Manufacturing the Future, V. Kordic, A. Lazinica, M. Merdan [Eds], ISBN: 3-86611-198-3, InTech, 2006.

[19] Sarker B.R., *The resemblance coefficients in group technology: a survey and comparative study of relational metrics*, Comput. Ind. Eng., 30, 1, 103–116, 1996.

[20] Romesburg H.C., *Cluster analysis for researchers. Lifetime Learning Publications*, Wadsworth Inc., Belmont, 1984.

[21] Yin Y., Yasuda K., *Similarity coefficient methods applied to the cell formation problem: a comparative investigation*, Comput. Ind. Eng., 48, 3, 471–489, 2005.

[22] Carrie A.S., *Numerical taxonomy applied to group technology and plant layout*, Int. J. Prod. Res., 11, 4, 399–416, 1973.

[23] Rajagopalan R., Batra J.L., *Design of cellular production system: a graph theoretic approach*, Int. J. Prod. Res., 13, 6, 567–579, 1975.

[24] Chandrasekharan M.P., Rajagopalan R., *ZODIAC: an algorithm for concurrent formation of part families and machine cells*, Int. J. Prod. Res., 25, 6, 451–464, 1987.

[25] Steudel H.J., Ballakur A., *A dynamic programming based heuristic for machine grouping in manufacturing cell formation*, Comput. Ind. Eng., 12, 3, 215–222, 1987.

[26] Mosier C.T., *An experiment investigating the application of clustering procedures and similarity coefficients to the GT machine cell formation problem*, Int. J. Prod. Res., 27, 10, 1811–1835, 1989.

[27] Srinivasan G., Narendran T.T., *GRAFICS – a nonhierarchical clustering algorithm for group technology*, Int. J. Prod. Res., 29, 3, 463–478, 1991.

[28] Kusiak A., Boe W.J., Cheng, C., *Designing cellular manufacturing systems: branch-and-bound and A* approaches*, IIE Trans., 25, 4, 46–56, 1993.

[29] Vakharia A.J., Wemmerlöv U., *A comparative investigation of hierarchical clustering techniques and dissimilarity measures applied to the cell formation problem*, J. Oper. Manag., 13, 2, 117–138, 1995.

[30] Nair G.J.K., Narendran T.T., CASE: *A clustering algorithm for cell formation with sequence data*, Int. J. Prod. Res., 36, 1, 157–179, 1998.

[31] Nair G.J.K., Narendran T.T., ACCORD: *A bicriterion algorithm for cell formation using ordinal and ratio-level data*, Int. J. Prod. Res., 37, 3, 539–556, 1999.

[32] Sarker B.R., Xu Y., *Designing multi-product lines: job routing in cellular manufacturing systems*, IIE Trans., 32, 3, 219–235, 2000.

[33] Dimopoulos C., Mort N., *A hierarchical clustering methodology based on genetic programming for the solution of simple cell-formation problems*, Int. J. Prod. Res., 39, 1, 1–19, 2001.

[34] Ilić O.R., *An e-Learning tool considering similarity measures for manufacturing cell formation*, J. Intell. Manuf., 25, 3, 617–628, 2014.

[35] Gupta T., *Design of manufacturing cells for flexible environment considering alternative routing*, Int. J. Prod. Res., 31, 6, 1259–1273. 1993.

[36] Seifoddini H., Djassemi M., *Merits of the production volume based similarity coefficient in machine cell formation*, J. Manuf. Syst., 14, 1, 35–44, 1995.

[37] Pachayappan M., Panneerselvam R., *A Comparative Investigation of Similarity Coefficients Applied to the Cell Formation Problem using Hybrid Clustering Algorithms*, Mater. Today: Proc., 5, 5, 12285–12302, 2018.

[38] Venugopal V., Narendran T.T., *Cell formation in manufacturing systems through simulated annealing: An experimental evaluation*, Eur. J. Oper. Res., 63, 3, 409–422, 1992.

[39] George A.P., Rajendran C., Ghosh, S., *An analytical-iterative clustering algorithm for cell formation in cellular manufacturing systems with ordinal-level and ratio-level data*, Int. J. Adv Manuf. Technol., 22, 1–2, 125–133, 2003.

[40] Zolfaghari S., Liang M., *A new genetic algorithm for the machine/part grouping problem involving processing times and lot sizes*, Comput. Ind. Eng., 45, 4, 713–731, 2003.

[41] Ponnambalam S.G., Pandian R.S., Mahapatra S.S., Saravansankar S., *Cell formation with workload data in cellular manufacturing system using genetic algorithm*, IEEE International Conference on Industrial Engineering and Engineering Management, Singapore, 2007.

[42] Mahapatra S.S., Pandian R.S., *Genetic cell formation using ratio level data in cellular manufacturing systems*, Int. J. Adv. Manuf. Technol., 38, 5–6, 630–40, 2008.

[43] Sengupta S., Ghosh T., Dan P.K., FAKMCT: *Fuzzy ART K-Means Clustering Technique: a hybrid neural network approach to cellular manufacturing systems*, Int. J. Comput. Integr. Manuf., 24, 10, 927–938, 2011.

[44] Wu N., *A concurrent approach to cell formation and assignment of identical machines in group technology*, Int. J. Prod. Res., 36, 8, 2099–2114, 1998.

[45] Anderberg M.R., *Cluster Analysis for Applications*, Academic Press Inc., New York, 1973.

[46] Ghosh T., Doloi B., Dan P.K., *Utilization-based grouping efficiency and multi-criteria decision approach in designing of manufacturing cells*, J. Eng. Manuf., 231, 3, 505–522, 2016.

[47] Ghosh T., Martinsen K., Dan P.K., *Development and correlation analysis of non-dominated sorting buffalo optimization NSBUF II using Taguchi's design coupled gray relational analysis and ANN*, Appl. Soft Comput., 85, 105809, 2019.