

Solving Component Family Identification Problems on Manufacturing Shop Floor

Mousumi Modak, Tamal Ghosh and Pranab K Dan
Department of Industrial Engineering & Management,
West Bengal University of Technology.
BF 142, Salt lake City, Kolkata 700064, India.

Email: mousumimodak10@gmail.com, tamal.31@gmail.com, danpk.wbut@gmail.com

Abstract

This article demonstrates effective techniques for component/part family formation problem in the vicinity of Cellular Manufacturing Systems (CMS). Past investigations reported that part family formation techniques are typically grounded on production flow analysis (PFA) which largely considers operational requirements, sequences and time. Part coding analysis (PCA) is merely counted in cellular manufacturing which is assumed to be the most competent method to identify the part families. In present study different clustering techniques are quantified to develop proficient part families by utilizing Opitz part coding scheme and the techniques are tested on 5 different datasets of size (5×9) to (27×9) and the obtained results are compared with each other. The experimental results reported that the C-Linkage method is more effective in terms of the quality of the solution obtained, has outperformed SLCA and K-means techniques.

Keywords: part family formation, group technology, opitz coding, cellular manufacturing, similarity metric, clustering algorithms.

1. Introduction

Group technology (GT) in cellular manufacturing systems (CMS) is substantial in improving productivity for the manufacturing companies. Group technology (GT) might be considered as a simplified methodology which groups standardized similar entities such as parts, assemblies, process plans, tools, instructions, etc. to minimize the time and effort and to improve the overall productivity for batch type production [1]. As reported in the literature [2] a successful implementation of GT could eventually minimize the engineering costs, facilitate cellular manufacturing, quicken product development, enhance costing accuracy, simplify process planning, minimize tooling cost and simplify the overall purchasing process. A major prerequisite in implementing GT is the identification of part families [3]. A part family is a group of parts sharing homogeneous design and manufacturing attributes. Early research in this domain has been dedicated primarily on the formation of production-oriented part families in which similarities amongst the parts are principally recognized on the fact of their processing

requirements, operation time and operation sequences. Though, these methodologies are inadequate in achieving the needs of other extents of manufacturing. For example, parts with homogeneous shape, size, dimension or other design characteristics are believed to be clustered in a single family for design justification and elimination of part varieties, however parts which are clustered on the fact of homogeneous routing and the tooling needs are convenient to resolve the process planning issues.

Therefore the scope of this domain of research is believed to be expanded and examined to a wider span of part similarities. Part similarities are believed to be identified sooner than the formation of part families. Part attributes such as shape, length/diameter ratio, material type, part function, dimensions, tolerances, surface finishing, process, operations, machine tool, operation sequence, annual production quantity, fixtures needed, lot sizes have been considered as the basis for similarity utilization [4]. The complexity remains in acquiring an appropriate technique which provides an identifying competence of human being, such as identifying patterns in groups, and forming part families with the aid of intelligence [5].

This article depicts three different part family identification techniques based on Complete Linkage (C-Linkage), Single Linkage (SLCA) and K-means clustering algorithms to investigate the nature of similarities and to describe the effectiveness of the techniques in solving the problem in hand.

2. Literature Survey

Two different approaches are traced in past literature in order to form part families, first is production flow analysis (PFA) which deals with processing requirements of parts, operational sequences and operational time of the parts on the machines [6]. Second approach is the part coding analysis (PCA) which utilizes predefined coding schemes to facilitate the process using several attributes of parts such as geometrical shapes, materials, design features and functional requirements etc. [7].

PCA is exposed in this study as an essential and effective tool for successful implementation of GT concept. A code may be numbers (numerical) or alphabets (alphabetical) or a hybridization of numbers and alphabets (alphanumeric) which are allotted to the parts to process the information [8]. Parts are categorized based on significant attributes such as dimensions, type of material, tolerance, operations required, basic shapes, surface finishing etc. In this approach, each part is assigned a code which is a string of numerical digits that store information about the part. Study of coding systems which depicts either hierarchical structure (monocode), or chain structure (polycode) or hybrid mode structure mixed with monocode and polycode [9].

Several coding systems have been developed, e.g. Opitz [10], MICLASS [11], DCLASS [12] and FORCOD [13]. Han and Ham [14] have claimed that part families could be established more realistically by practicing the PCA due to the advantage of using the manufacturing and design attributes concurrently. Offodile [15] reported a similarity metric

based on the numeric codes for any pair of parts which could be utilized to an appropriate clustering method such as agglomerative clustering algorithm to form efficient part families.

Clustering analysis is practiced in Cellular Manufacturing System (CMS) as a competent methodology to facilitate the machine/part grouping problems. Various machine/part grouping techniques are developed to solve manufacturing cell formation problems since last forty years, these include similarity coefficient methods, clustering analysis, array based techniques, graph partitioning methods etc. The similarity coefficient approach was first suggested by McAuley [16]. The basis of similarity coefficient methods is to calculate the similarity between each pair of machines and then to group the machines into cells based on their similarity measurements. Few studies have been proposed to measure dissimilarity coefficients instead of similarity coefficient for machine-part grouping problems [17]. Most of the similarity coefficient methods utilized machine-part mapping chart. Few of them are Single linkage clustering algorithm [16], Average linkage clustering algorithm [18] etc.

Array based methods consider the rows and columns of the machine-part incidence matrix as binary patterns and reconfigure them to obtain a block diagonal cluster formation. The rank order clustering algorithm is the most familiar array-based technique for cell formation [19]. Substantial alterations and enhancements over rank order clustering algorithm have been described by King and Nakornchai [20] and Chandrasekharan and Rajagopalan [21]. The direct clustering analysis has been stated by Chan and Milner [22], and bond energy analysis is performed by McCornick et al. [23].

Graph Theoretic Approach depicts the machines as vertices and the similarity between machines as the weights on the arcs to form machine cells [24]. Chandrasekharan and Rajagopalan [21] proposed an ideal seed nonhierarchical clustering algorithm for cellular manufacturing. Graph searching algorithms was developed to select a crucial machine or part according to a pre-fixed criterion [25]. A non-heuristic network method was stated by Vohra et al. [26] to construct manufacturing cells with minimum inter-cell moves. Srinivasan [27] implemented a method using minimum spanning tree (MST) for the machine-part cell formation problem.

During past few decades soft computing techniques are exhaustively practiced by researchers in the vicinity of CMS. Lee-Post [28] proposed that GT coding system (DCLASS) could be efficiently used with simple GA method to cluster part families which is well suited for part design and process planning in production. A hybrid methodology based on Boltzmann function from simulated annealing and mutation operator from GA was proposed by Wu et al. [29] to optimize the initial cluster obtained from similarity coefficient method (SCM) and rank order clustering (ROC). Arkat et al. [30] developed a sequential model based on SA for large-scale problems and compared their method with GA. Ateme-Nguema and Dao [31] investigated an ACO based TS heuristic for cellular system design problem (CSDP) and the methodology proved to be much quicker than traditional methods when considering operational sequence, time and cost. These Authors further proposed quantized Hopfield network for CFP to find optimal or near-optimal solution and TS was employed to improve the performance and the quality of solution of the network [32]. Durán et al. [33] reported a modified PSO with proportional likelihood instead of using velocity vector on CF problems

where the objectives are the minimization of cell load variation and inter cellular parts movement and reported the stability of the method with low variability. A similar study was also performed by Anvari et al. [34] where a hybrid particle swarm optimization technique for CFP was reported. The initial solutions generated either randomly or using a diversification generation method and the technique also utilized mutation operator embedded in velocity update equation to avoid reaching local optimal solutions. Thereafter with due consideration, a wide variety of machine/part matrices were effectively solved by this approach.

A detailed study on metaheuristic based approaches in CMS could be obtained from recent review works [35, 36]. These studies report that PCA has merely been adopted to form part families using the stated methodologies. Therefore this article would explore an unexplored area based on PCA in CMS by employing different clustering techniques.

3. Problem Definition

Opitz classification and coding system is used in this article which was developed by Opitz [10] at Aachen Technology University in West Germany. The basic code comprises of nine digits that can be extended by additional four digits. The general interpretations of the nine digits are as indicated in Fig. 1. The first 5 digits are called the form code and designate the design or the general form of the part and hence aid in design retrieval. Later, 4 more digits were added to the coding scheme in order to enhance the manufacturing information of the specific work part. These last four digits are also called supplementary code. All four digits are integers, and respectively represent: Dimensions, Material, Original shape of raw stock, and Accuracy of the work part. The extra four digits, A, B, C, and D (not shown in Fig. 1) called the secondary code, are practiced by the particular organization to include those features that are specific to their organization.

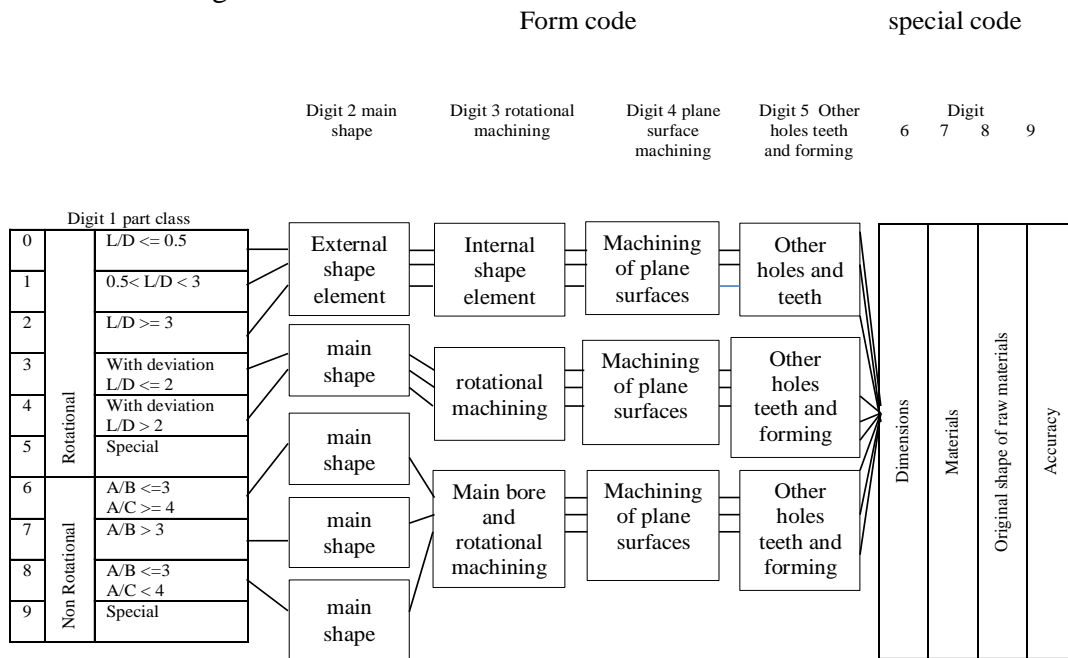


Fig. 1: Opitz part Coding system

The interpretation of first 9 digits are,

- Digit 1:** General shape of workpiece, otherwise called 'part-class'. This is further subdivided into rotational and non-rotational classes and further divided by size (length/diameter or length/width ratio.)
- Digit 2:** External shapes and relevant form. Features are recognized as stepped, conical, straight contours. Threads and grooves are also important.
- Digit 3:** Internal shapes. Features are solid, bored, straight or bored in stepped diameter. Threads and grooves are integral part.
- Digit 4:** Surface plane machining, such as internal or external curved surfaces, slots, splines.
- Digit 5:** Auxiliary holes and gear teeth.
- Digit 6:** Diameter or length of workpiece.
- Digit 7:** Material Used.
- Digit 8:** Shape of raw materials, such as round bar, sheet metal, casting, tubing etc.
- Digit 9:** Workpiece accuracy.

All the 9 digits are interpreted numerically (0-9). Examples of a mild steel forged round bar is further demonstrated in Fig. 2. The Opitz codes of the round rod is 11103 2302 [8]. The attributes are denoted as a1-a9 for the round bar,

- a1=1 (Rotational parts, $0.5 < L/D < 3$.)
- a2=1 (External shape element, stepped to one end.)
- a3=1 (Internal Shape element, smooth or stepped to one end.)
- a4=0 (No surface machining.)

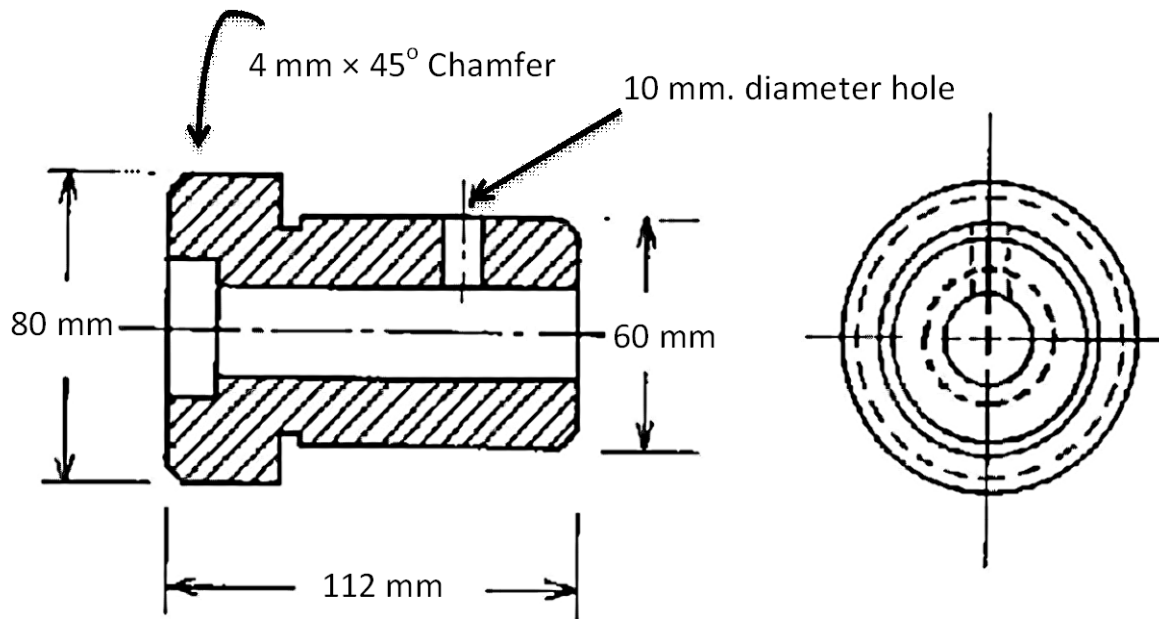


Fig. 2: Mild Steel Forged Round Rod

- a5=5 (Auxiliary holes, radial.)
- a6=2 (50 mm. < diameter <=100 mm.)

a7=3 (material is mild steel.)
 a8=0 (Internal form: Round bar.)
 a9=2 (Accuracy in coding digit.)

The part family formation problem stated in this research can be formulated using a part-attribute incidence matrix $B=[b_{ij}]$, of size $m \times n$, where m is the number of parts and n is the number of attributes of that part. b_{ij} represents the coding value (0-9) of j^{th} attribute of i^{th} part. A (10×9) example problem based on Opitz coding system is shown in Fig. 3.

	a1	a2	a3	a4	a5	a6	a7	a8	a9
p1	0	0	1	0	0	9	1	3	6
p2	0	0	1	0	1	2	6	5	6
p3	0	0	2	0	0	6	2	1	7
p4	0	0	2	3	0	4	1	6	9
p5	0	0	3	0	0	8	0	3	5
p6	0	1	1	0	0	7	7	8	9
p7	0	1	1	0	2	4	4	8	1
p8	0	1	2	3	0	0	9	5	9
p9	0	2	1	0	4	8	4	4	2
p10	0	2	1	3	0	4	4	8	2

Fig. 3: Problem #2 (10×9) dataset

According to Opitz coding scheme every part of Fig. 3 shows different attributes. First 5 columns are form codes and rest of the 4 columns represent supplementary codes. The solution to the problem is to form the families of parts in such a way that the sum of similarities among each pair of parts in a same family would be maximized. Therefore clustering methods are used in this article which groups the parts into families.

4. Research Methodology

Dissimilarity coefficient based techniques are massively practiced in formation of manufacturing cells [37]. In this article a dissimilarity/distance measure method based on part family identification technique is utilized [15]. It is presented as,

$$d_{ij} = \frac{\sum_{k=1}^K d_{ijk}}{K} \tag{1}$$

Where

$$d_{ijk} = \frac{|b_{ik} - b_{jk}|}{R_k} \tag{2}$$

Where

d_{ijk} is dissimilarity measured between part i and part j on attribute k ,

K is total number of attributes considered,

b_{ik} is part coding for part i on attribute k ,

b_{jk} is part coding for part j on attribute k ,

R_k is range of possible part codings for all parts on attribute k .

Therefore sum of similarities is derived using the formula,

$$S_{tot} = \sum_{n=1}^N S_n \quad (3)$$

Where

$$S_n = \frac{\sum_{i \in n, j \in n} S_{ij}}{0.001 + C_2^{P_n}} \quad (4)$$

$$S_{ij} = \frac{\sum_{k=1}^K [(1 - d^k)_{ijk}]}{K} \quad (5)$$

where S_{ij} = similarity measure between part i and part j

$C_2^{P_n}$ = Number of pair-wise combinations formed in part family n , and P_n is the number of parts in family n (in the denominator a small value of 0.001 is added to avoid the division by zero rule). Complete Linkage Clustering Algorithm (C-Linkage), Single Linkage Clustering Algorithm (SLCA) and K-means algorithm are adopted in this study as the solution methodologies.

This abovementioned dissimilarity metric technique is utilized in this study to calculate the dissimilarities/distances between pair of parts presented as rows of part-attribute mapping matrix as given in Fig. 3. Therefore the dissimilarity matrix generated for (10×9) part-attribute mapping matrix using equation (1) is presented in Fig. 4.

4.1. Linkage Clustering Techniques

C-Linkage and SLCA are theoretically and mathematically simple algorithm practiced in hierarchical clustering analysis of data [38]. It delivers informative descriptions and visualization of possible data clustering structures. When there exists hierarchical relationship in data this approach can be more competent. Complete linkage also known as furthest neighbor method which uses the distance between two clusters. The distance between cluster i and another cluster j is defined as:

$$C_{ij} = \max(d_{ij}), i \in (1, \dots, n_r), j \in (1, \dots, n_s) \quad (6)$$

Single linkage also known as nearest neighbor method which uses the distance between two clusters. The distance between cluster i and another cluster j is defined as:

$$E_{ij} = \min(d_{ij}), i \in (1, \dots, n_r), j \in (1, \dots, n_s) \quad (7)$$

	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10
p1	0	0.185185	0.098765	0.185185	0.061728	0.209877	0.259259	0.333333	0.185185	0.271605
p2		0	0.185185	0.197531	0.222222	0.17284	0.17284	0.17284	0.222222	0.209877
p3			0	0.160494	0.111111	0.209877	0.259259	0.283951	0.234568	0.271605
p4				0	0.197531	0.197531	0.246914	0.17284	0.320988	0.185185
p5					0	0.246914	0.271605	0.345679	0.197531	0.283951
p6						0	0.197531	0.197531	0.246914	0.209877
p7							0	0.320988	0.148148	0.08642
p8								0	0.37037	0.259259
p9									0	0.185185
p10										0

Fig. 4: Dissimilarity matrix obtained from (10×9) part-attribute mapping matrix of Fig. 3

Using equation (6) and (7), an intermediate matrix could be obtained which is a $(p-1) \times 3$ matrix, where p is the number of parts in the original dataset. Columns of the matrix contain cluster indices linked in pairs to form a binary tree. The leaf nodes are numbered from 1 to p . Leaf nodes are the singleton clusters from which all higher clusters are built. Further the dendrogram could be obtained from the matrix which indicates a tree of potential solutions. An example dendrogram structure obtained for the parts using C-Linkage method is shown in Fig. 5, which shows the clear cluster information of part families. Similar dendrogram could be produced using SLCA method (not shown in the article). From Figure 6 it can be stated that 3 part families are obtained using the hierarchical notations. The part families are, Family 1 {7,9,10}, Family 2 {1,3,4,5}, Family 3 {2,6,8}. Similarly using the SLCA approach 7 part families could be achieved as, Family 1 {22}, Family 2 {23}, Family 3 {25}, Family 4 {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,26}, Family 5 {24}, Family 6 {21}, Family 7 {27}.

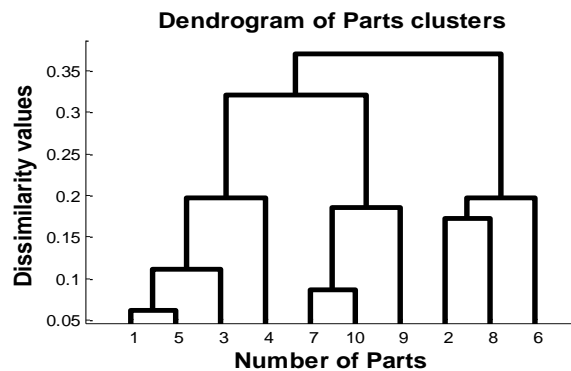


Fig. 5: Dendrogram of part clusters of example problem (10×9)

This pseudocode would explain both the hierarchical method SLCA and C-Linkage:

Input: part-attribute incidence matrix A

1. Procedure dissimilarity()

1.1. Compute dissimilarity values between pair of machines using equation (1)

1.2. Compute the dissimilarity matrix of the parts

1.3. End

2. Procedure Cluster()

2.1. loop

2.2. Compute the smallest Euclidian distance between two clusters for SLCA method

2.3. Compute the furthest Euclidian distance between two clusters for C-Linkage method

2.4. Construct matrices of size $(m-1) \times 3$ to form the hierarchical tree structure for both SLCA and C-Linkage method

2.5. Construct dendrograms for SLCA and C-Linkage using matrices obtained from 2.2 and 2.3

2.6. loop

2.7. create part families for the minimum level of dissimilarity coefficient

2.8. Compute the sum of similarities using equation (5) for SLCA and C-Linkage method

2.9. End

Output: Part family Configurations for SLCA and C-Linkage method

4.2. The K-means clustering algorithms

K-means clustering is an algorithm to classify objects based on attributes into K number of groups [39]. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. Thus the purpose of K-mean clustering is to classify the data exploiting the Euclidean distance metric between the data-points. The basic steps of k-means clustering are simple. Number of clusters K is fixed and the centroid of these clusters is assumed randomly or the first K objects in sequence could also be chosen as the initial centroids. Thereafter the K-means algorithm would follow three steps, (i) Iterate until no object is left to be clustered, (ii) Determine the centroid coordinate, (iii) Determine the Euclidean distance of each object to the centroids, (iv) Group the object based on minimum distance. The pseudocode of K-means is given as,

Input: part-attribute incidence matrix, number of clusters 'K'

Procedure k-means()

1.1. Loop

1.2. First K parts are marked as K centroids

1.3. flag = false

1.5. loop

1.6. For each cluster calculate minimum distance d_k between each part and cluster centroids ($k=1,2,3,\dots,K$)

1.7. Find minimum of d_k , marked as 'min'

1.8. assign the part to k^{th} cluster

1.9. flag = true

1.10. stop until all parts are assigned to K clusters

1.11. stop until maximum number of iterations reached

5. Results And Discussion

The proposed techniques are tested on 5 different problem datasets of size (5×9) to (27×9). The largest dataset has been obtained from Haworth [40] using Opitz coding system. Remaining 4 problems are designed using the aforesaid coding system. Problem datasets are provided in Fig. 3 and Fig. 6 to 9. The algorithms are coded in Matlab 7.0 environment and executed on PIV laptop computer. The obtained results are compared and shown in Table 1.

	a1	a2	a3	a4	a5	a6	a7	a8	a9
p1	0	0	1	0	0	9	1	3	6
p2	0	0	1	0	1	2	6	5	6
p3	0	0	2	0	0	6	2	1	7
p4	0	0	2	3	0	4	1	6	9
p5	0	0	3	0	0	8	0	3	5

Fig. 6: Problem #1 (5×9) dataset

	a1	a2	a3	a4	a5	a6	a7	a8	a9
p1	0	0	1	0	0	9	1	3	6
p2	0	0	1	0	1	2	6	5	6
p3	0	0	2	0	0	6	2	1	7
p4	0	0	2	3	0	4	1	6	9
p5	0	0	3	0	0	8	0	3	5
p6	0	1	1	0	0	7	7	8	9
p7	0	1	1	0	2	4	4	8	1
p8	0	1	2	3	0	0	9	5	9
p9	0	2	1	0	4	8	4	4	2
p10	0	2	1	3	0	4	4	8	2
p11	0	2	2	0	4	6	8	8	8
p12	0	2	3	0	4	7	5	6	7
p13	1	0	0	0	0	9	2	8	1
p14	1	0	1	0	0	7	6	6	0
p15	1	1	0	0	0	1	8	3	8

Fig. 7: Problem #3 (15×9) dataset

	a1	a2	a3	a4	a5	a6	a7	a8	a9
p1	0	0	1	0	0	9	1	3	6
p2	0	0	1	0	1	2	6	5	6
p3	0	0	2	0	0	6	2	1	7
p4	0	0	2	3	0	4	1	6	9
p5	0	0	3	0	0	8	0	3	5
p6	0	1	1	0	0	7	7	8	9
p7	0	1	1	0	2	4	4	8	1
p8	0	1	2	3	0	0	9	5	9
p9	0	2	1	0	4	8	4	4	2
p10	0	2	1	3	0	4	4	8	2
p11	0	2	2	0	4	6	8	8	8
p12	0	2	3	0	4	7	5	6	7
p13	1	0	0	0	0	9	2	8	1
p14	1	0	1	0	0	7	6	6	0
p15	1	1	0	0	0	1	8	3	8
p16	1	1	1	0	0	4	0	2	1
p17	1	2	0	3	0	9	6	3	2
p18	2	0	0	0	0	9	3	5	6
p19	2	3	0	0	0	4	8	7	2
p20	2	5	0	0	0	8	5	3	4

Fig. 8: Problem #4 (20×9) dataset

	a1	a2	a3	a4	a5	a6	a7	a8	a9
0	0	1	0	0	9	1	3	6	
0	0	1	0	1	2	6	5	6	
0	0	2	0	0	6	2	1	7	
0	0	2	3	0	4	1	6	9	
0	0	3	0	0	8	0	3	5	
0	1	1	0	0	7	7	8	9	
0	1	1	0	2	4	4	8	1	
0	1	2	3	0	0	9	5	9	
0	2	1	0	4	8	4	4	2	
0	2	1	3	0	4	4	8	2	
0	2	2	0	4	6	8	8	8	
0	2	3	0	4	7	5	6	7	
1	0	0	0	0	9	2	8	1	
1	0	1	0	0	7	6	6	0	
1	1	0	0	0	1	8	3	8	
1	1	1	0	0	4	0	2	1	
1	2	0	3	0	9	6	3	2	
2	0	0	0	0	9	3	5	6	
2	3	0	0	0	4	8	7	2	
2	5	0	0	0	8	5	3	4	
7	0	0	0	3	0	7	8	0	
7	0	0	3	3	3	4	5	9	
7	0	0	5	3	8	3	3	5	
7	0	0	6	3	0	1	7	4	
2	0	8	0	1	1	1	5	5	
1	5	1	0	0	2	6	4	3	
6	5	4	4	3	6	0	7	0	

Fig. 9: Problem #5 (27×9) dataset

Table 1 demonstrates that each of the proposed methodologies are competent to attain good solutions and effective in constructing the families of parts. Despite the fact that the solutions obtained are not identical therefore the sum of similarities are not identical for the test datasets. It depicts that C-Linkage technique outperformed the SLCA and K-means methods for the problems # 2, 3 and 4. However for the smallest dataset #1 (5×9) K-means has produced nearly 50% improved result over SLCA and C-Linkage technique. Thereafter for the dataset #5 (27×9) nearly 10% improvement has been shown by K-means over the other two technique. For continuous data K-means technique is superior over the linkage clustering methods [38]. Due to that fact K-means method is heavily adopted as an efficient clustering technique in literature. Thus according to the past study K-means should incorporate improved solutions for all the problem datasets. However data based on classification and coding systems are categorical in nature. Therefore normal distance metrics are not advised to be utilized to solve such problems [23]. Therefore, a different dissimilarity coefficient measure is utilized in this article [15]. Due to that reason linkage methods, practiced in this article, are modified substantially. Because of that modifications both the linkage methods attain improved solutions for problem datasets 2, 3 and 4 over the K-means technique. For all the test problems C-linkage has reported an overall improvement of 60% and outpaced other two techniques. Hence it could be stated that although C-Linkage is moderately better than the other methods in terms of the goodness of solutions obtained, but K-means is also capable to produce substantially good results. In terms of computational time all the proposed methods are equally good and took minimum CPU time (less than 10 seconds) for all the datasets tested. Table 2 reports the percentage of perfection achieved while developing the part groups. Further Fig. 10 establishes the superiority shown by C-Linkage over other two methods.

Table 1. Comparison of performance shown by SLCA, C-Linkage and K-Means

#	Dataset size	Part families obtained			Maximum similarities		
		SLCA	C-Linkage	K-Means	SLCA	C-Linkage	K-Means
1	5×9	Family 1 {2}, Family 2 {1,3,4,5}	Family 1 {2}, Family 2 {1,3,4,5}	Family 1 {1,3,5}, Family 2 {2,4}	0.8641	0.8641	1.7108
2	10×9	Family 1 {1,3,4,5}, Family 2 {2, 7, 8, 9, 10}, Family 3 {6}	Family 1 {7,9,10}, Family 2 {1,3,4,5}, Family 3 {2,6,8}	Family 1 {1, 5, 9}, Family 2 {2, 6, 7, 8, 10}, Family 3 {3,4}	1.6492	2.5425	2.4902
3	15×9	Family 1 {4}, Family 2 {1,3,5}, Family 3 {6, 7, 9, 10, 11, 12, 13, 14}, Family 4 {2, 8, 15}	Family 1 {6, 11, 12}, Family 2 {2, 8, 15}, Family 3 {1, 3, 4, 5}, Family 4 {7, 9, 10, 12, 13, 14}	Family 1 {7, 9, 10, 13, 14}, Family 2 {2, 6, 8, 11, 12}, Family 3 {1, 3, 5}, Family 4 {4}	2.5707	3.4338	2.5658
4	20×9	Family 1 {6, 9, 11, 12}, Family 2 {1, 3, 5, 7, 10, 13, 14, 17, 18, 19, 20}, Family 3 {4}, Family 4 {16}, Family 5 {2, 8, 15}	Family 1 {7,10,14,19}, Family 2 {9, 13, 17, 18, 20}, Family 3 {6, 11, 12}, Family 4 {2, 8, 15}, Family 5 {1, 3, 4, 5, 16}	Family 1 {9, 13, 14, 17, 18, 20}, Family 2 {2, 6, 7, 8, 10, 11, 12, 15, 19}, Family 3 {1, 3, 5}, Family 4 {4}, Family 5 {16}	2.4907	4.2510	2.5323
5	27×9	Family 1 {22}, Family 2 {23}, Family 3 {25}, Family 4 {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 26}, Family 5 {24}, Family 6 {21}, Family 7 {27}	Family 1 {6,11,12}, Family 2 {2,8,15}, Family 3 {21}, Family 4 {22,23,24}, Family 5 {27}, Family 6 {1, 3, 4, 5, 16, 25}, Family 7 {7, 9, 10, 13, 14, 17, 18, 19, 20, 26}	Family 1 {9, 17, 18, 20, 23}, Family 2 {2, 8, 15, 26}, Family 3 {1, 3, 5}, Family 4 {4, 22, 24, 25}, Family 5 {16}, Family 6 {6, 11, 12}, Family 7 {7, 10, 13, 14, 19, 21, 27}	0.7744	4.1631	4.9185

Table 2. Comparison among perfection of the obtained results by the algorithms

Problem Dataset	No. of part families formed(N)	$\sum S_{r_i}$ Sum of similarities,			$\sum \frac{S_{r_i}}{N} \times 10$ Perfection percentage,		
		SLCA	C-Linkage	K-means	SLCA	C-Linkage	K-means
1	2	0.8641	0.8641	1.7108	43.205	43.205	85.54
2	3	1.6492	2.5425	2.4902	54.97	84.75	83.07
3	4	2.5707	3.4338	2.5658	64.27	85.84	64.14
4	5	2.4907	4.2510	2.5323	49.81	85.02	50.65
5	7	0.7744	4.1631	4.9185	11.06	59.47	70.26

6. Conclusions

Three different clustering techniques, SLCA, C-Linkage and K-Means algorithms are employed in this research to form part families. Since part coding and classification techniques are merely adopted in group technology problems, therefore the objective of this study is to utilize the heavily practiced coding system namely Opitz part coding. 5 test datasets ranging from 5×9 to 27×9 are tested using the aforementioned techniques. Due to the NP-complete

nature of the reported problems these methods are equally efficient to produce optimal solutions. The proposed methods are compared among each other successfully. The objective function utilized in this study is to maximize the sum of similarities among parts in all the part families formed. From the beginning of this research work the number of part families to be formed is considered as constant. As shown in Table 1 C-Linkage algorithm has outperformed SLCA and K-means in terms of solution quality (sum of similarities value) which is further depicted in Fig. 10 in terms of the percentage of perfections achieved while forming the part families. This study has assumed identical weightage for each and every attribute, however in formation of part families some attributes could be more significant than the other attributes. Therefore future work could be done by considering fractional weightage for each of the attributes. This work could also be extended by considering operational time and sequence of each part to develop more effective and robust part families.

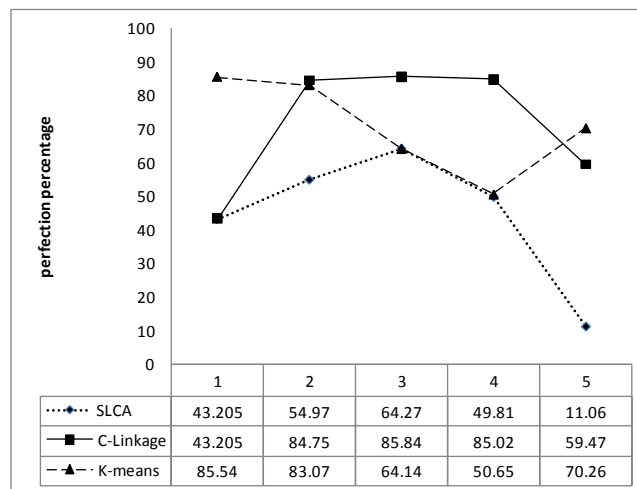


Fig. 10: Comparison among SLCA, C-linkage, K-means algorithms in terms of solution quality

References

- [1] Burbidge, J.L., "Production flow Analysis", Production Engineers, Vol. 42, 1963, pp. 742-752.
- [2] Guerrero, L.C., Lozano, S., Canca, D., Guerrero, F., Larrañeta, J. and Onieva, L., "Cell formation using sequence information and neural networks", Proceedings of the 10th International Conference Flexible Automation and Intelligent Manufacturing, 2000, pp. 566-575.
- [3] Kaparathi, S. and Suresh, N., "A neural network system for shape-based classification and coding of rotational parts", International Journal of Production Research, Vol. 29, 1991, pp. 1771-1784.
- [4] Groover, M.P. and Zimmers E.W. Jr., "CAD/CAM: Computer-Aided Design and Manufacturing", Prentice-Hall, New Jersey, USA, 1984.
- [5] Moon, Y.B., 'Establishment of a neurocomputing model for part family/machine group identification', Journal of Intelligent Manufacturing, Vol. 3, 1992, pp. 173-182.
- [6] Burbidge, J.L., "Production Flow Analysis For Planning Group Technology", Oxford University Press, USA, 1996.
- [7] Mitrofanov, S.P., "Scientific Principles of Group Technology", Part I, Boston: National Lending Library of Science and Technology (Originally published in 1959 as Russian text), 1966.
- [8] Ham, I., Hitomi, K. and Yoshida, T., "Group Technology: Applications to Production Management", Kluwer-Nijhoff Press, Boston, USA, 1985.

- [9] Singh, N. and Rajamani, D., “Cellular Manufacturing Systems: Design, Planning and Control”, London: Chapman and Hall Press, 1996.
- [10] Opitz, H., “A Classification System to Design Workpieces”, Pergamon Press, UK, 1970.
- [11] TNO., “An Introduction to MICLASS”, Organization for Industrial Research, MA, USA, 1975.
- [12] Gallagher, C.C. and Knight, W.A., “Group Technology Production Methods in Manufacture”, Chichester, Ellis Horwood, 1985.
- [13] Jung, J., Ahluwalia, R.S., “FORCOD: a coding and classification system for formed parts”, *Journal of Manufacturing Systems*, Vol. 10, 1992, pp. 223-232.
- [14] Han, C. and Ham, I., “Multi-objective cluster analysis for part family formation”, *Journal of Manufacturing Systems*, Vol. 5, 1986, pp. 223-229.
- [15] Offodile, O.F., “Application of similarity coefficient method to parts coding and classification analysis in group technology”, *Journal of Manufacturing Systems*, Vol. 10, 1992, pp. 442-448.
- [16] McAuley, J., “Machine grouping for efficient production, *Production Engineer*”, Vol. 51, 1972, pp. 53-57.
- [17] Prabhakaran, G., Janakiraman, T.N. and Sachithanandam, M., “Manufacturing data based combined dissimilarity coefficient for machine cell formation”, *International Journal of Advanced Manufacturing Technology*, Vol. 19, 2002, pp. 889–897.
- [18] Seifoddini, H. and Wolfe, P.M., “Application of the similarity coefficient method in group technology”, *IIE Transactions*, Vol. 18 no. 3, 1986, pp. 271-277.
- [19] King, J.R., “Machine-component grouping in production flow analysis: an approach using a rank order-clustering algorithm”, *International Journal of Production Research*, Vol. 18, 1980, pp. 213-232.
- [20] King, J.R. and Nakornchai, V., “Machine-component group formation in group technology: review and extension”, *International Journal of Production Research*, Vol. 20, 1982, pp. 117-133.
- [21] Chandrasekharan, M.P. and Rajagopalan, R., “An ideal seed non-hierarchical clustering algorithm for cellular manufacturing”, *International Journal of Production Research*, Vol. 24 no. 2, 1986, pp. 451-464.
- [22] Chan, H.M. and Milner, D.A., “Direct Clustering algorithm for group formation in Cellular Manufacturing”, *Journal of Manufacturing Systems*, Vol. 1, 1982, pp. 65-75.
- [23] McCormick, W.T., Schweitzer, J.P.J. and White, T.W., “Problem decomposition and data reorganization by a clustering technique”, *Operations Research*, Vol. 20 no. 5, 1972, pp. 993-1009.
- [24] Rajagopalan, R. and Batra, J.L. “Design of Cellular production system –A graph theoretic approach”, *International Journal of Production Research*, Vol. 13 no. 6, 1975, pp. 567-579.
- [25] Ballakur, A. and Steudel, H. J., “A with-in cell utilization based heuristic for designing cellular manufacturing systems”, *International Journal of Production Research*, Vol. 25 no. 5, 1987, pp. 639-655.
- [26] Srinivasan G., “A clustering algorithm for machine cell formation in group technology using minimum spanning trees”, *International Journal of Production Research*, Vol. 32, 1994, pp. 2149–58.
- [27] Vohra, T., Cheng, D. and Chang, J.L., “A network model for cell formation in CMS, *International Journal of Production Research*”, Vol. 28 no. 11, 1990, pp. 2075-2084.
- [28] Lee-Post, A., “Part family identification using a simple genetic algorithm”, *International Journal of Production Research*, Vol. 38, 2000, pp. 793-810.
- [29] Wu, T.H., Chung, S.H. and Chang, C.C., “Hybrid simulated annealing algorithm with mutation operator to the cell formation problem with alternative process routings”, *Expert Systems with Applications*, Vol. 36, 2009, pp. 3652–3661.
- [30] Arkat, J., Saidi, M. and Abbasi, B., “Applying simulated annealing to cellular manufacturing system design”, *International Journal of Advanced Manufacturing Technology*, Vol. 32, 2007, pp. 531-536.
- [31] Ateme-Nguema, B.H. and Dao, T.M., “Optimization of cellular manufacturing systems design using the hybrid approach based on the ant colony and tabu search techniques”, *Proceedings of the IEEE IEEM, 2007*, pp. 668-673.
- [32] Ateme-Nguema, B.H. and Dao, T.M., “Quantized Hopfield networks and tabu search for manufacturing cell formation problems”, *International Journal of Production Economics*, Vol. 121, 2009, pp. 88-98.
- [33] Anvari, M., Mehrabad, M.S. and Barzinpour, F., “Machine–part cell formation using a hybrid particle swarm optimization”, *International Journal of Advanced Manufacturing Technology*, Vol. 47, 2010, pp. 745-754.
- [34] Durán, O., Rodriguez, N. and Consalter, L.A., “Collaborative particle swarm optimization with a data mining technique for manufacturing cell design”, *Expert Systems with Applications*, Vol. 37, 2010, pp. 1563–1567.
- [35] Papaioannou, G. and Wilson, J. M., “The evolution of cell formation problem methodologies based on recent studies (1997–2008): Review and directions for future research”, *European Journal of Operational Research*,

Vol. 206 no. 3, 2010, pp. 509-521.

- [36] Ghosh, T., Sengupta, S., Chattopadhyay, M. and Dan, P.K., “Meta-heuristics in cellular manufacturing: A state-of-the-art review”, International Journal of Industrial Engineering Computations, Vol. 2 no. 1, 2011, pp. 87-122.
- [37] Yin, Y. and Yasuda, K., “Similarity coefficient methods applied to the cell formation problem: a comparative investigation”, Computers & Industrial Engineering, Vol. 48, 2005, pp. 471–489.
- [38] Anderberg, M.R., “Cluster Analysis for Applications”, Academic Press Inc., New York, NY, 1973.
- [39] Hartigan, J.A. and Wong, M.A., “Algorithm AS 136: A K-Means Clustering Algorithm”, Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 28 no. 1, 1979, pp. 100-108.
- [40] Haworth, E.A., “Group technology-using the opitz system”, Production Engineer , Vol. 47 no. 1, 1968, pp. 25-35.

APPENDIX

A numerical example of SLCA approach

The following information is given.

(i) Information about parts (relevant part attributes as explained in this article with 5 form codes of Opitz coding system):

Part1 1 2 3 4 5

Part2 0 2 4 3 5

Part3 2 1 3 4 0

Part4 3 5 1 4 7

Part5 8 4 3 5 7

(ii) Information about significant part attributes when,

Attribute 1 is numeric, $R_1=9$; Attribute 2 is numeric, $R_2=9$

Attribute 3 is numeric, $R_3=9$; Attribute 4 is numeric, $R_4=9$

Attribute 5 is numeric, $R_5=9$.

Step 1: Calculated pairwise similarity coefficients (S_{ij}) from equation (5).

$$S_{12} = 0.9333$$

$$S_{13} = 0.8444 \quad S_{23} = 0.7778$$

$$S_{14} = 0.8000 \quad S_{24} = 0.7333 \quad S_{34} = 0.6889$$

$$S_{15} = 0.7333 \quad S_{25} = 0.6667 \quad S_{35} = 0.6222 \quad S_{45} = 0.8000$$

S_{12} is determined as follows:-

$$S_{121} = 1 - \frac{|b_{11} - b_{21}|}{R_1} = 1 - \frac{|1 - 0|}{9} = 0.888$$

$$S_{122} = 1 - \frac{|b_{12} - b_{22}|}{R_2} = 1 - \frac{|2 - 2|}{9} = 1.000$$

$$S_{123} = 1 - \frac{|b_{13} - b_{23}|}{R_3} = 1 - \frac{|3 - 4|}{9} = 0.888$$

$$S_{124} = 1 - \frac{|b_{14} - b_{24}|}{R_4} = 1 - \frac{|4 - 3|}{9} = 0.888$$

$$S_{125} = \frac{1 - \frac{|b_{15} - b_{25}|}{R_5}}{5} = \frac{1 - \frac{|5 - 5|}{9}}{5} = 1.000$$

$$\text{Hence } S_{12} = \frac{(0.888 + 1.000 + 0.888 + 0.888 + 1.000)}{5} = 0.9328$$

All other S_{ij} are determined in the same method.

Step 2: Calculate pairwise distances (d_{ij}) using equation (1)

$$d_{12} = 0.0667$$

$$d_{13} = 0.1556 \quad d_{23} = 0.2222$$

$$d_{14} = 0.2000 \quad d_{24} = 0.2667 \quad d_{34} = 0.3111$$

$$d_{15} = 0.2667 \quad d_{25} = 0.3333 \quad d_{35} = 0.3778 \quad d_{45} = 0.2000$$

Step 3: Form initial part family.

Because d_{12} is the smallest distance, therefore PF1 is formed with parts 1 and 2 as its members.

Step 4: Calculate minimum distance E_{ij}

$$E_{(1,2)3} = 0.1556$$

$$E_{(1,2)4} = 0.2000 \quad E_{34} = d_{34} = 0.3111$$

$$E_{(1,2)5} = 0.2667 \quad E_{35} = d_{35} = 0.3778 \quad E_{45} = d_{45} = 0.2000$$

$E_{(1,2)3}$ is obtained as,

$$E_{(1,2)3} = \min(d_{13}, d_{23}) = \min(0.1556, 0.2222) = 0.1556$$

Other $E_{(1,2)j}$ are calculated in the similar manner. Thereafter $E_{(1,2)3}$ has the lowest value, hence part 3 is introduced in family PF1.

Step 5: Repeat step 4 for part 4 and 5.

$$E_{(1,2,3)4} = 0.2000$$

$$E_{(1,2,3)5} = 0.2667 \quad E_{45} = d_{45} = 0.2000$$

$E_{(1,2,3)4}$ is obtained as,

$$E_{(1,2,3)4} = \min(d_{(1,2)4}, d_{34}) = \min(0.2000, 0.3111) = 0.2000$$

E_{45} has the lowest value therefore part 1, 2 and 3 are grouped in family PF1, and part 4 and 5 are grouped in the 2nd family PF2.

Step 6: Thus all the parts are grouped. PF1 contains {parts 1, 2, 3} and PF2 contains {parts 4, 5}. Stop.

Sum of similarities for PF1 is achieved as,

$$S_1 = \frac{S_{12} + S_{13} + S_{23}}{0.001 + C_2^3} = \frac{0.9333 + 0.8444 + 0.7778}{3.001} = 0.8518$$

Sum of similarities for PF2 is achieved as,

$$S_2 = \frac{S_{45}}{0.001 + C_2^2} = \frac{0.8000}{1.001} = 0.8000$$

Therefore sum of similarities for PF1 and PF2 is achieved as,

$$S=S_1+S_2= 0.8518+0.8000=1.6518$$

C-Linkage method follows the same method of SLCA. The only difference is instead of calculating minimum distances in step 3 to 5, C-Linkage calculates the maximum distances.

A numerical example of K-means approach

The following information is given.

(i) Information about parts: (illustrated only with two attributes for four parts)

Part1 1 1

Part2 2 1

Part3 4 3

Part4 5 4

1. Initial value of centroids : Suppose we use part1 and part2 as the first centroids. Let c_1 and c_2 denote the coordinate of the centroids, then $c_1 = (1,1)$ and $c_2 = (2,1)$.

2. Parts-Centroids distance : we calculate the distance between cluster centroid to each part. Let us use Euclidean distance, then we have distance matrix at iteration 0 is,

$$D^0 = \begin{bmatrix} 0 & 1 & 3.61 & 5 \\ 1 & 0 & 2.83 & 4.24 \end{bmatrix}$$

Each column in the distance matrix symbolizes the part. The first row of the distance matrix corresponds to the distance of each part to the first centroid and the second row is the distance of each part to the second centroid. For example, distance from part3 = (4, 3) to the first centroid c_1 is

$$\sqrt{(4-1)^2 + (3-1)^2} = 3.61$$

And to the second centroid c_2 is,

$$\sqrt{(4-2)^2 + (3-1)^2} = 2.83$$

3. Objects clustering : We assign each part based on the minimum distance. Thus, part1 is assigned to group 1, part2 to group 2, part3 to group 2 and part4 to group 2.

4. Iteration-1, determine centroids : Knowing the members of each group, now we compute the new centroid of each group based on these new memberships. Group 1 only has one member thus the centroid remains in $c_1 = (1,1)$. Group 2 now has three members, thus the centroid is

the average coordinate among the three members: $c_2 = \left(\frac{2+4+5}{3}, \frac{1+3+4}{3} \right) = \left(\frac{11}{3}, \frac{8}{3} \right)$

5. Iteration-1, Parts-Centroids distances : The next step is to compute the distance of all objects to the new centroids. Similar to step 2, we have distance matrix at iteration 1 is,

$$D^1 = \begin{bmatrix} 0 & 1 & 3.61 & 5 \\ 3.14 & 2.36 & 0.47 & 1.89 \end{bmatrix}$$

6. Iteration-1, Parts-clustering: Similar to step 3, we assign each part based on the minimum distance. Based on the new distance matrix, we move the part2 to Group 1 while all the other parts remain.

7. Iteration 2, determine centroids: Now we repeat step 4 to calculate the new centroids coordinate based on the clustering of previous iteration. Group1 and group 2 both has two members, thus the new centroids are

$$c_1 = \left(\frac{1+2}{2}, \frac{1+1}{2} \right) = \left(\frac{3}{2}, 1 \right) \text{ and } c_2 = \left(\frac{4+5}{2}, \frac{3+4}{2} \right) = \left(\frac{9}{2}, \frac{7}{2} \right)$$

8. 8. Iteration-2, Parts-Centroids distances : Repeat step 2 again, we have new distance matrix at iteration 2 as,

$$D^2 = \begin{bmatrix} 0.5 & 0.5 & 3.2 & 4.61 \\ 4.3 & 3.54 & 0.71 & 0.71 \end{bmatrix}$$

9. Iteration-2, Parts clustering: Again, we assign each object based on the minimum distance. Therefore group 1 contains part1 and part2 and group 2 contains part3 and part 4.

10. We obtain, comparison of the grouping of step 6 and step 9 reveals that the parts does not move group anymore. Thus, the computation of the k-mean clustering has reached its stability and no more iteration is needed. We get the final grouping as the results PF1 {parts 1 and 2} and PF2 {parts 3 and 4}.