

Research Article

Comparison of Hierarchical Clustering Methods in Machine Cell Formation: A Case study

Mayuresh Kutte^{†*} and Keshav Valase[†]

[†]Mechanical Engineering Department, Walchand Institute of Technology, P. B. NO. 634, Solapur, Maharashtra, India

Accepted 07 April 2015, Available online 10 April 2015, Vol.5, No.2 (April 2015)

Abstract

It has always been a vital issue for any manufacturing organization to optimize the shop layouts, with the inclination of industries towards cellular layout. Cell design is a critical aspect to handle as most of the future planning depends on this parameter. Much of the research is carried out on an arbitrary data leading to disparity with real time situations. In this paper, an attempt is made to compare various hierarchical clustering methods for a batch type setup.

Keywords: Similarity co-efficient, Hierarchical clustering, Cell formation algorithm, Performance measures, Group efficiency, Group efficacy.

1. Introduction

Group technology (GT) is an approach to the organization of the work in which the organizational units are relatively independent groups, each responsible for production of a given family of products (Burbidge, 1975). GT has attracted a lot of attention because of its positive impacts in the batch-type production. Cellular manufacturing (CM) is one of the applications of GT principles to manufacturing. In the design of a CM system, similar parts are grouped into families and associated machines into groups so that one or more part families can be processed within a single machine group. The process of determining part families and machine groups is referred to as the cell formation (CF) problem.

Since the evolution of Group technology (by Mitrofonov, 1966), N number of theories and models have been researched to form part families and machine groups but, there is very less research available on robustness of these methods. Most of the research has been carried out on the predefined data sets, which give idealistic solutions and may not be suitable for real time problems. For practicing manufacturing managers, it would be nice to know as to which one of the available algorithms would be most effective and efficient for their specific situation (Prafulla Joglekar *et al.*, 2001).

In this work, the data is collected from an ancillary unit of a leading pump manufacturing industry. This unit is of type of typical conventional machine shop equipped with lathe, drilling and CNC machines

arranged in functional layout. Here, we first calculate similarity co-efficient between machines and carryout clustering analysis using SLC, ALC and CLC and compare their performances to get the most efficient and effective solution.

2. Problem Definition

With the existing layout company is facing problem of delivering the products in time. Due to functional layout the cost of and time spent on material handling is high and also most of the machines are bottleneck. The owner wishes to change the layout to overcome these problems by adopting cellular manufacturing.

Under this work the targets to achieve are as follows

- Analyze the data using three Hierarchical clustering methods.
- To compare hierarchical methods on common performance scale.
- To design cells using the best solution obtained

3. Case study

The company under survey is a machine shop and supplies machined parts to the parent company. The shop is equipped with 22 machines which consists of lathe, drill and CNC. The variation of products machined in the shop is 470.

In accordance of data collection, process plans and process routing for each part were collected and studied. For the ease of data handling, parts with similar process routings were grouped together and designated with series ranging from (A to N) equals to 14 series. From the data collected machine component incidence matrix (MCIM) has been prepared, as shown in Table 1.

*Corresponding author: mayureshkutte@gmail.com

Table 1 Initial machine component incidence matrix

	Series A	Series B	Series C	Series D	Series E	Series F	Series G	Series H	Series I	Series J	Series K	Series L	Series M	Series N
M1	1				1	1	1			1		1	1	
M2			1					1	1					1
M3	1						1	1	1			1		
M4									1					
M5	1								1					
M6			1		1	1							1	
M7	1	1	1	1	1	1	1	1	1			1	1	
M8	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M9	1	1	1	1	1	1	1	1	1					
M10	1	1	1	1	1	1	1		1	1	1	1	1	
M11									1					
M12	1	1	1	1		1								
M13														
M14								1	1					
M15									1					
M16							1							
M17										1	1			
M18														
M19							1							
M20														
M21			1			1								
M22	1	1		1		1								

Table 2 Similarity co-efficient matrix

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22
M1	1	0	0.5	0	0.22	0.54	0.7	0.636	0.5	0.636	0	0.33	0	0	0	0.25	0.22	0	0.25	0	0.22	0.545
M2		1	0.44	0.4	0.33	0.25	0.4	0.44	0.461	0.25	0.4	0.22	0	0.666	0.4	0	0	0	0	0	0.333	0
M3			1	0.33	0.33	0	0.625	0.526	0.5	0.47	0.33	0.22	0	0.571	0.33	0.33	0	0	0.33	0	0	0.22
M4				1	0.666	0	0.16	0.2	0.2	0	0.33	0	0	0.66	1	0	0	0	0	0	0	0
M5					1	0	0.307	0.25	0.3636	0.2857	0.666	0.2857	0	0.5	0.66	0	0	0	0	0	0	0.33
M6						1	0.533	0.44	0.4615	0.5	0	0.44	0	0	0	0	0	0	0	0	0.66	0.25
M7							1	0.88	0.9	0.869	0.16	0.625	0	0.3076	0.166	0.16	0	0	0.16	0	0.3076	0.533
M8								1	0.7826	0.923	0.133	0.5263	0	0.25	0.133	0.133	0.25	0	0.133	0	0.25	0.44
M9									1	0.7619	0.2	0.7142	0	0.3636	0.2	0.2	0	0	0.2	0	0.22	0.615
M10										1	0.1538	0.588	0	0.1428	0.1428	0.1428	0.2857	0	0.1538	0	0.2857	0.5
M11											1	0	0	0.66	1	0	0	0	0	0	0	0
M12												1	0	0	0	0	0	0	0	0	0.5714	0.88
M13													1	0	0	0	0	0	0	0	0	0
M14														1	0.66	0	0	0	0	0	0	0
M15															1	0	0	0	0	0	0	0
M16																1	0	0	1	0	0	0
M17																	1	0	0	0	0	0
M18																		1	0	0	0	0
M19																			1	0	0	0
M20																				1	0	0
M21																					1	0.33
M22																						1

4. Similarity co-efficient and SC matrix

A similarity coefficient is calculated for each pair of machines to determine how “alike” the two machines are in terms of the number of parts which visit both machines and the number of parts that visit each machine (Yong Yin). A 2 x 2 table is a convenient way to show the different alternatives, as illustrated in Fig. 1. Sorenson defined similarity co-efficient as in (1).

		Machine j	
		1	0
machine i	1	a	b
	0	c	d

Fig. 1 Legends of Similarity co-efficient

Legends:

- a = Part visits both machine
- b = Parts visit only machine i
- c = Parts visit only machine j
- d = Parts visit neither machines

$$SC_{ij} = \frac{2a}{2a+b+c} \tag{1}$$

Sample calculation:

$$SC_{12} = \frac{2(0)}{2(0)+7+4} = 0$$

Similarly, calculating SC between each pair of machine we get SC matrix as shown in Table 2.

5. Clustering analysis

5.1 Single Linkage Clustering (SLC)

It is the method that considers the distance between two clusters to be equal to the shortest distance from any member of one cluster to any member of the other cluster. If the data consist of similarities, the similarity between a pair of clusters is considered to be equal to the greatest similarity from any member of one cluster to any member of the other cluster (Lior Rokach and Oded Maimon).

Algorithm

Step 1: Find a pair that is most similar, i.e., they are the closest and have highest similarity co-efficient. In our case, SC = 1

Step 2: Merge clusters: Merge the pair of column (i and j) and row (k and l) with highest SC, always merge

column towards left and rows to above. Eliminate column j and row l after merging. In our problem merge M4-M11-M15 and eliminate M11, M15 columns and rows

Step 3: Dendogram: It is the graphical representation of merging pair. Show the merged pair on dendogram at respective level as shown in Fig. 2.



Fig 2. Dendogram

Step 4: Update co-efficient: The similarity between merged pair and the others will be maximum of the pair of similarities in each case. Here, Similarity between M4 and M1 is 0, M11 and M1 is 0 and M15 and M1 is 0, hence we choose maximum similarity i.e., ‘0’ to quantify similarity between (M4-M11-M15) and M1.

Another example, similarity between M4 and M2 is 0.4, M11 and M2 is 0.4 and M15 and M2 is 0.4, hence we choose ‘0.4’

Step 5: Similarly, repeat above steps till all machines are clustered in one cell or similarity between cell and machines become zero.

5.2 Average Linkage Clustering (ALC)

Methods that consider the distance between two clusters to be equal to the average distance from any member of one cluster to any member of the other cluster (King B., 1967). If the data consist of similarities, the similarity between a pair of clusters is considered to be equal to be average of similarity from any member of one cluster to any member of the other cluster.

Algorithm

Step 1: Find a pair that is most similar, i.e., they are the closest and have highest similarity co-efficient. In our problem, SC = 1

Step 2: Merge clusters: Merge the pair of column (I and j) and row (k and l) with highest SC, always merge column towards left and rows to above. Eliminate column j and row l after merging. In our problem, merge M4-M11-M15 and eliminate M11, M15 column and rows

Step 3: Dendogram: Show the merged pair on dendogram at respective level.

Step 4: Update co-efficient: The similarity between merged pair and the others will be average of the pair of similarities in each case. Here, Similarity between M4 and M1 is 0, M11 and M1 is 0 and M15 and M1 is 0, hence the average, 0+0+0/3 i.e., ‘0’ to quantify similarity between (M4-M11-M15) and M1.

Another example, similarity between M4 and M2 is 0.4, M11 and M2 is 0.4 and M15 and M2 is 0.4, hence average $(0.4+0.4+0.4)/3=0.4$

Step 5: Similarly, repeat above steps till all machines are clustered or similarity between cell/ machine and machines become zero.

5.3 Complete Linkage Clustering (CLC)

Methods that consider the distance between two clusters to be equal to the longest distance from any member of one cluster to any member of the other cluster (Ward, J.H., 1963). If the data consist of similarities, the similarity between a pair of clusters is considered to be equal to the least similarity from any member of one cluster to any member of the other cluster.

Algorithm

Step 1: Find a pair that is most similar, i.e., they are the closest and have highest similarity co-efficient. In our case, SC = 1

Step 2: Merge clusters: Merge the pair of column (*I* and *j*) and row (*k* and *l*) with highest SC, always merge column towards left and rows to above. Eliminate column *j* and row *l* after merging. In our problem merge M4-M11-M15 and eliminate M11, M15 column and rows

Step 3: Dendogram: Show the merged pair on dendogram at respective level.

Step 4: Update co-efficient: The similarity between merged pair and the others will be *minimum* of the pair of similarities in each case. Here, Similarity between M4 and M1 is 0, M11 and M1 is 0 and M15 and M1 is 0, hence we choose maximum similarity i.e., '0' to quantify similarity between (M4-M11-M15) and M1.

Another example, similarity between M4 and M2 is 0.4, M11 and M2 is 0.4 and M15 and M2 is 0.4, hence we choose '0.4'

Step 5: Similarly, repeat above steps till all machines are clustered in one cell or similarity between cell and machines become zero.

6. Performance measures

6.1 Number of Exceptional Elements

EE are off- diagonal positive elements. These source of inter-cell movements of parts. One objective of cell formation is to reduce the total cost of material handling. Therefore, EE is the most simple and intuitive measure for evaluating the cell formation solution (Fionn Murtagh and Pedro Contreras).

6.2 Number of voids

Voids indicate that all parts assigned to the cell do not require a machine assigned to a cell. Leads to large,

inefficient cells and can potentially contribute to low utilization.

6.3 Group Efficiency (η)

GE is an aggregate measure that takes both the number of exceptional elements and the machine utilization into consideration. A convex combination of both terms is considered to reveal the relative importance of each term (V. Satheeshkumar, et al., 2014). As a rule as higher the GE the better the clustering results. One drawback of grouping efficiency is the low discriminating capability (i.e. the ability to distinguish good quality grouping from bad).

Grouping Efficiency,

$$\eta = w\eta_1 + (1-w)\eta_2 \quad (2)$$

Where,

$$\eta_1 = \frac{o-e}{o-e+v} \quad (3)$$

$$\eta_2 = \frac{MP-o-v}{MP-o-v+e} \quad (4)$$

M = number of machines

P = number of parts

o = number of operations (1s) in the machine-part matrix {*aik*}

e = number of exceptional elements in solution

v = number of voids in the solution

w = 0.5 (recommended)

6.4 Group Efficacy (τ)

Group efficacy overcomes the problem of selecting weight factor and the limiting range of GE. Group efficacy has the requisite properties like non-negativity, 0-1 ranges and is not affected by the size of the MCIM i.e., the number of parts or machine is not considered.

Grouping Efficacy:

$$\tau = \frac{o-e}{o+v} \quad (5)$$

o = number of operations (1s) in the machine-part matrix {*aik*}

v = number of voids in the solution

7. Results and discussion

Fig 7.1 shows the dendogram for SLC and as can be seen from figure, we have obtained a single cluster. Table 7.1 represents the final machine component incidence matrix for Single Linkage Clustering with no exceptional element as we have only one cluster and voids equal to 179. Fig 7.2 shows dendogram for ALC. Table 7.2 shows final machine component matrix for Average Linkage clustering. Though the dendogram for

Table 3Final incidence matrix for Single linkage clustering

	Series E	Series F	Series B	Series D	Series C	Series I	Series H	Series N	Series L	Series A	Series G	Series J	Series K	Series M
M1	1	1				1				1		1		1
M7	1	1	1	1	1	1	1		1		1			1
M9	1	1	1	1	1	1	1	1	1		1			
M8	1	1	1	1	1	1	1		1	1	1	1	1	1
M10	1	1	1	1	1	1			1		1	1	1	1
M12	1	1	1	1	1				1					
M22	1	1	1	1	1				1					
M6	1	1												1
M21	1	1	1	1										
M2						1	1	1		1				
M14						1								
M11						1								
M15						1								
M5						1				1				
M4						1								
M3							1		1	1	1			
M16											1			
M19											1			
M17												1	1	
M13														
M18														
M20														

Table 4Final incidence matrix for Average linkage clustering

	Series E	Series F	Series B	Series D	Series C	Series I	Series H	Series N	Series L	Series A	Series G	Series J	Series K	Series M
M1	1	1				1				1		1		1
M7	1	1	1	1	1	1	1		1		1			1
M9	1	1	1	1	1	1	1	1	1		1			
M8	1	1	1	1	1	1	1		1	1	1	1	1	1
M10	1	1	1	1	1	1			1		1	1	1	1
M3							1		1	1	1			
M2						1	1	1		1				
M14						1								
M4						1								
M11						1								
M15						1								
M5						1				1				
M6	1	1												1
M21	1	1	1	1										
M12	1	1	1	1	1				1					
M22	1	1	1	1	1				1					
M16											1			
M19											1			
M17												1	1	
M13														
M18														
M20														

Table 5 Final incidence matrix for Complete linkage clustering

	Series E	Series F	Series B	Series D	Series C	Series I	Series H	Series N	Series L	Series A	Series G	Series J	Series K	Series M
M6	1	1												1
M21	1	1	1	1										
M12	1	1	1	1	1				1					
M22	1	1	1	1	1				1					
M2						1	1	1		1				
M4						1								
M11						1								
M15						1								
M14						1	1							
M3							1		1	1	1			
M16											1			
M19											1			
M17												1	1	
M1	1	1				1				1		1		1
M7	1	1	1	1	1	1	1		1		1			1
M9	1	1	1	1	1	1	1	1	1		1			
M8	1	1	1	1	1	1	1		1	1	1	1	1	1
M10	1	1	1	1	1	1			1		1	1	1	1
M5						1				1				
M13														
M18														
M20														

ALC and SLC vary in natures we get similar matrix for ALC and SLC. Figure 7.3 represents the dendrogram of CLC. The Table 7.3 is the final machine component incidence matrix for Complete Linkage Clustering with 87 operation elements, 3 exceptional elements and 61 voids in cells.

The results obtained under the study are astonishing. Since we have obtained similar machine component incidence matrix for SLC and ALC the competition between these methods remain only with CLC. The group efficiency obtained for CLC was 78.04 %,taking weight factor as 0.5 (Rajagopalan, 1975) and the group efficacy was 56.75%.

As the results clearly infer, the choice of hierarchical clustering method depend upon the problem. Here the main point be noted is, it is said that ALC gives the better solution over SLC and CLC as this method compromises with the extremes, taking the averages(Yong Yin). Average Linkage Clustering method was not able to give satisfactory clusters for this problem.

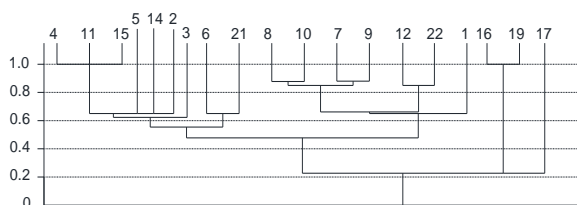


Fig. 3 Final Dendrogram for Single linkage clustering

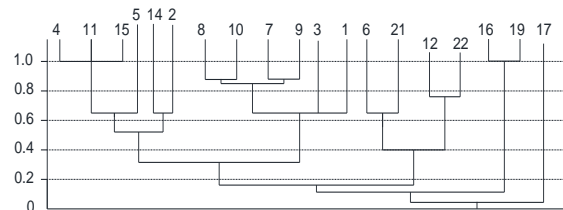


Fig. 4 Final Dendrogram for Average linkage clustering

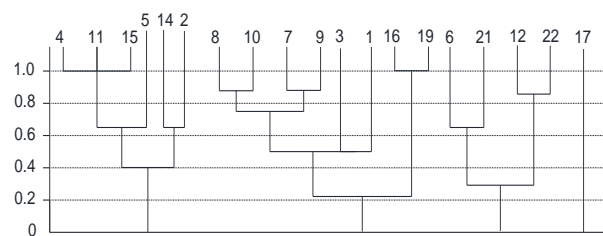


Fig. 5: Final Dendrogram for Complete linkage clustering

Since Complete linkage clustering gave better group Efficiency and group efficacy over Single linkage clustering and Average linkage clustering, the cells were designed with the solution obtained from CLC. The Table 6 shows the part families and machine cells formed. The graphical comparison of the performance measures is given in the figures below:

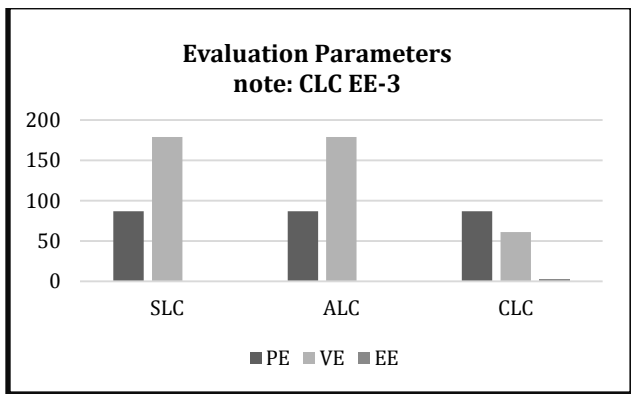


Fig. 6 Evaluation parameters

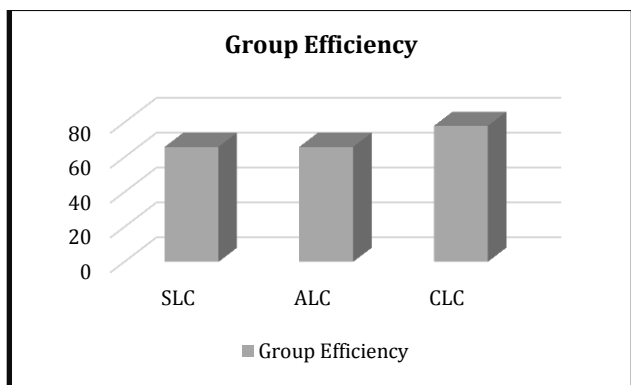


Fig. 7. Group efficiency

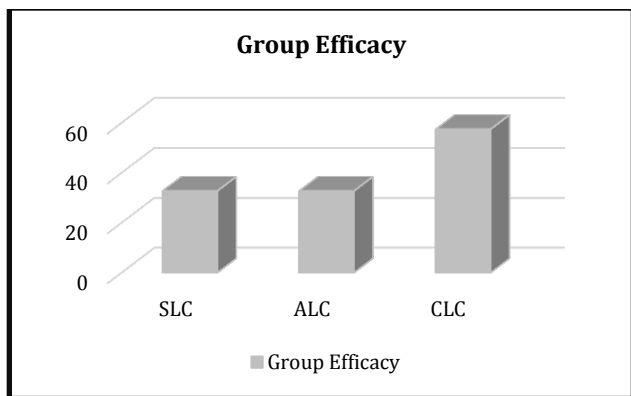


Fig. 8 Group Efficacy

Table 6 Cells formed

Cell	Machines	Parts
Cell 1	M6, M21, M12, M22	Series (E, F, B, D, C)
Cell 2	M2, M4, M11, M15, M14, M3	Series (I, H, N, L, A)
Cell 3	M16, M19, M17	Series (G, J, K, M)
Cell 4	M1, M7, M9, M8, M10, M5	All Series
Reserved Capacity	M13, M18, M20	Special orders

Conclusion

Under this work, the cellular manufacturing system was proposed to the case industry. Exceptional Elements, voids, Group efficiency and Group efficacy were used as performance scales for clustering

methods namely Single Linkage Clustering (SLC), Average Linkage Clustering (ALC) and Complete Linkage Clustering (CLC). The clusters formed by CLC were efficient and effective for this problem with **78.04 %** group efficiency and **56.75 %** group efficacy. The ALC could have given better performance but, as clusters formation gave only two cells out of which one is reserved cell we have to go with CLC based cell design. We can also conclude that, the selection of clustering method for cell design depends upon the type of problem.

References

John L. Burbidge, (1975), Introduction to Group Technology, J. Rezaeian, N. Javadin, Design of incremental cellular manufacturing system based on heuristic methods, *Trends in computers*, vol. 2, 539-544

Prafulla Joglekar, Q. B. Chung and Madid Tavana, (2001), Note on comparative evaluation of nine well-known algorithms for solving the cell formation problem in group technology, *Journal. Of Applied Mathematics & Decision science*, 5(3), 253-268.

Nikola Suzic, et al., (2012) Customisating product through application of group technology: A case study of furniture shop, *Journal of Mechanical Engineering*, 58(12), 724-731.

Antonio Batocchino, Shahruxh Adilrani, (1996) Design of manufacturing cell: Strategy, software and performance measurement, *Gestao & Producao*, 3(3), 248-261.

A. Gunasekaran, et al., (2001) Experiences of small to medium size enterprise in the design and implementation of manufacturing cell, *International journal of Cellular Integrated Manufacturing*, 14(2), 212-223

Ming Zhou, Ronald G. Askin, (1998) Formation of general GT cell: An operation based approach, *Journal of computers and industrial. Engineering*, 34(1), 147-157.

A. Gunasekaran, et al., (1994) An investigation into application of group technology in advanced manufacturing system, *International. Journal of Cellular Integrated Manufacturing*, 7(4), 215-228.

Shu Ming Ng., (1996) On the characterization and measure of machine cells in group technology", *Journal of Operations Research*, 44(5), 735-744.

P K Arora, A. Haleem, M. K. Singh, (2013) Recent development of cellular manufacturing systems, *Sadhana*, 38(3), 421-428

Lior Rokach, Oded Maimon, Clustering methods

Sanaa Ali Hamza, Erry Yulian T. Adesta, (2013) Similarity co-efficient measures applied to integrate feasibility assessment and the design of cellular manufacturing system, *Australian Journal Of Basics and Applied Science*, 7(6), 257-266.

Yong Yin, Application similarity co-efficient method to cellular manufacturing.

Rajagopalan, R., Batra, J.L., 1975. Design of cellular production system: a graph theoretic approach. *International Journal of Production Research* 13, 567-579.

Fionn Murtagh, Pedro Contreras, Methods of hierarchical clustering

King, B. (1967) Step-wise Clustering Procedures, *Journal of American Statistical Association* 69, 86-101.

Ward, J. H. (1963) Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58, 236-244.

Jose´ Fernando Goncalves, Mauricio G.C. Resende, (2007) An evolutionary algorithm for manufacturing cell formation *Journal of computers and industrial Engineering* 47, 247-273

V. Satheshkumar, et al., (2014) Evaluation of Cell Formation Algorithms and Implementation of MOD-SLC Algorithm as An effective Cellular Manufacturing System in a Manufacturing Industry, *International Journal of Current Engineering and Technology*, Special issue 2, 183-190.