# FORMATION OF PART FAMILY IN RECONFIGURABLE MANUFACTURING SYSTEM USING PRINCIPLE COMPONENT ANALYSIS AND K-MEANS ALGORITHM

## GUPTA, A[shutosh]; JAIN, P. K. & KUMAR, D[inesh]

*Abstract: The Reconfigurable Manufacturing Systems (RMS) is the next step in manufacturing, allowing the production of any quantity of highly customised and complex parts together with the benefits of mass production. In RMSs, parts are grouped into families, each of which requires a specific system configuration. Initially system is configured to produce the first family of parts. Once it is finished, the system is reconfigured in order to produce the second family, and so forth. The effectiveness of a RMS depends on the formation of the optimum set of part families addressing various reconfigurability issues. The aim of this work is to establish a methodology for grouping parts into families for effective working of Reconfigurable Manufacturing Systems (RMSs). The methodology carried out in three phases. In the first phase, the correlation matrix is used as similarity coefficient matrix. In the second phase, Principal Component Analysis (PCA) is applied to find the eigenvalues and eigenvectors on the correlation similarity matrix. A scatter plot analysis as a cluster analysis is applied to make parts groups while maximizing correlation between parts. In the third phase, Agglomerative Hierarchical K-means algorithm improved the parts family formation using Euclidean distance resulting in an optimum set of part families for reconfigurable manufacturing system.*
*Keywords: Reconfigurable manufacturing system, Part family formation, Principle Component Analysis, K-means algorithm, Similarity coefficient, Multivariate analysis*

## 1. INTRODUCTION

Current market trends are characterized by globalization, new product requirements, rapidly changing demands, and a continuous improvement of the existing technology for manufacturing activities. Time reduction to introduce new products to the market with high quality and low cost is a necessity for enterprise survival in this new scenario. The key factor in this highly competitive environment is the ability of the companies to launch new products to the market, with high quality, and low cost. For achieving this, the manufacturing system must have exact capacity and functionality to yield different batch sizes of different product types. Reference [1] were first , defined a RMS as "a manufacturing system designed at the outset for rapid changes in structure as well as in hardware and software components in order to quickly adjust production capacity and functionality within a part family in response to sudden changes in market or in regulatory requirements." In the same way, Refrence [2] consider a RMS as a manufacturing system configured to produce a family of products that shares some similarities. As these definitions state, the formation of part families is a central issue in RMS. The key attribute of part families is that all the components within a family require similar production systems and thus RMSs should have the exact capacity and functionality required

to manufacture a part family, allowing cost effectiveness [3]. It is suggested that grouping parts into families in RMSs has a positive effect on the introduction of new products in market [4]. Initially, RMS is configured to produce the first selected family and then it is reconfigured to effectively produce the following part family, and so forth. Thus, the RMS configuration changes over a number of part families to complete all the batches. The first issue which considered in RMS is the formation of part families. Literature presents plenty of methods to obtain families and diverse formation criteria. These methods and criteria cannot be used directly in reconfigurable manufacturing because it has its own singularities that are to be taken into consideration, which differ from other manufacturing paradigms. Reference [5] stated that formation of part family in RMS has to be based on some grouping criteria. Modularity and commonality are two important criteria for grouping considered suitable for RMS. For customized products, modularity allows the assembly of simple and functionally independent parts. These simple and functionally independent parts use standard parts which are group together to on the basis of operations similarity. Another important criterion is commonality concept which is related to the part variety and is defined as measure of how well the product uses standard parts. The standard parts require similar type of operations and grouped together to capture the inherent advantages of GT such as reduced setup times, reduced in-process inventories, improved product quality, shorter lead time, reduced tool requirements, improved productivity, and better overall control of operations.

Dedicated Manufacturing System focuses on the economic production of one specific part type only. The first manufacturing system focused on cost-effective manufacturing of several part types simultaneously was Cellular Manufacturing System (CMS). The development of CMSs has been closely linked to grouping of parts into families. In CMS, the part family formation is achieved using following techniques: descriptive procedures, mathematical programming approaches, and artificial intelligence methods.The hierarchical clustering agglomerative methods group together similar elements (products) in clusters based on their attribute similarities. The coefficients that measure similarity between two parts are calculated from the incidence matrix. After that, a dendrogram shows the similarity degree to group parts. They used similarity or dissimilarity coefficients among parts to obtain the groups. Hierarchical clustering algorithms yield a dendrogram representing the nested family of parts and similarity levels at which families change. Non-

hierarchical clustering algorithms, on the other hand, obtain a single partition of the data instead of a clustering structure.

Most of the existing Cell Formation methods suffer from one or more drawbacks. Their major common drawbacks are the inflexibility and the limited industrial application due to the non-availability of software programs supporting them. Reference [5] also focused on grouping the parts by modifying existing CMS methods instead of developing the new method in order to stand the requirements of RMS. So, new part groups formation approaches that overcome these limitations are clearly needed. This paper proposes a new approach based on new similarity coefficient method for the part family formation in RMS environment.To this effect, a new method has been proposed using correlation as a similarity coefficient. This new similarity coefficient is based on the correlation between operation sequences which are required to produce the parts. Furthermore, Principal Component Analysis (PCA) is used to cluster the parts followed by agglomerative hierarchical K-means for identifying part groups on the basis of operational sequence similarity.

Principal Component Analysis (PCA) is the best known and oldest technique in multivariate analysis [6]. Reference [7] was first to introduce it to recast linear regression analysis into a new form. PCA is frequently used for the data set with some intrinsic complexity [8, 9]. Reference [10] used PCA for concurrent part machine group formation problem in CMS. It is a quantitatively rigorous method for achieving the simplification. The method generates a new set of variables, called Principal Components (PCs). Each PC is a linear combination of the original variables. All the PCs are orthogonal to each other, so there is no redundant information. The number of PCs extracted in a PCA is equal to the number of observed variables being analyzed. However, in most analyses, only the first few components account for meaningful amount of variance, and hence those first few components are retained, interpreted, and used in subsequent analysis and rest are neglected. When the analysis is complete, the resulting components display varying degrees of correlation with the observed variables, but are completely uncorrelated with one another.

The K-means algorithm is a clustering technique. The K-means algorithm randomly selects K data points as initial cluster centroids. A centroid is an artificial point in the space which represents an average location of the particular cluster. K clusters are formed by assigning each data point to its nearest centroid. New virtual centroids are then calculated for each cluster. These processes are iterated until a predefined number of iteration is reached or the clusters did not change anymore. The major problem with the K-means algorithm is that its initial starting points are generated randomly and does not guarantee the unique clustering results [11]. Also, due to the non-hierarchical nature of the algorithm, a hierarchical relationship between the clusters is needed. This hierarchical relationship is effective to visualize and analyze the large data sets. The hierarchical technique is classified into agglomerative method and divisive method. The divisive method is the top down approach in which initially all the objects are included in a single cluster. Then, the single cluster is divided into sub-clusters until each object constitutes a cluster. An agglomerative method is the bottom up approach in which each object is assumed as a separate cluster and then they are clustered in succession until a single cluster which consists of the entire object set is formed. So, an Agglomerative Hierarchical K-means Clustering Algorithm (AHKCA) is used for the part family formation.

The outline of the paper is as follows: The proposed methodology is presented in Section 2. Afterwards, one numerical example gives illustration of the proposed methodology in phased manner in Section 3. Section 4 discusses the results. Lastly, conclusions are drawn in section 5.

## 2. METHODOLOGY

In this section, a novel method based on PCA and AHKCA is developed for part family formation. The proposed methodology consists of three phases as shown in Figure 1. The objective of the methodology is to cluster the parts into k part families based on operations sequence similarity.
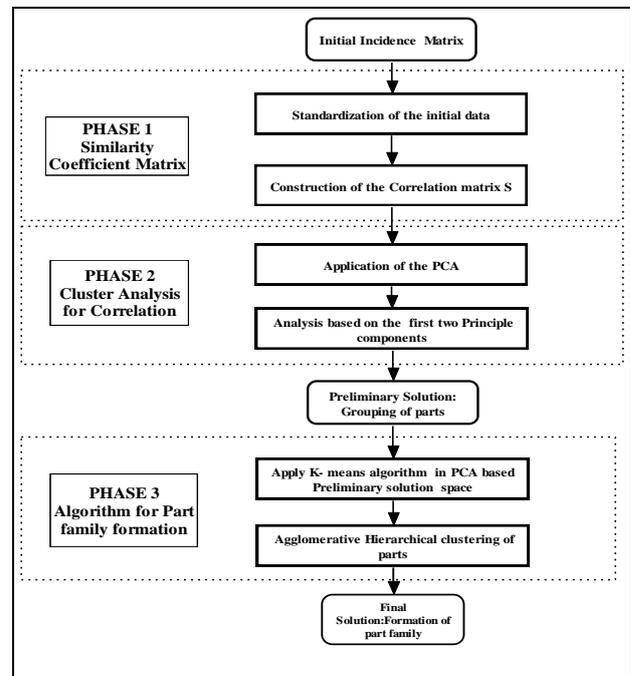


Fig.1.Proposed methodology

### 2.1 Similarity coefficient matrix

The first phase starts with building a similarity coefficient matrix. The initial part-operation incidence matrix as shown in equation (1) is a binary matrix in which rows represent the operations and columns stand for parts. This matrix looks like the transpose of the classical part-machine incidence matrix.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \quad (1)$$

Where $a_{ij} = 1$ if part j requires operation i and $a_{ij} = 0$ otherwise.

Let $P_j^A$ is a binary row vector of matrix A, such that

$$P_j^A = \begin{bmatrix} a_{1j}, a_{2j}, \ldots, a_{nj} \end{bmatrix}.$$

Initial matrix *A* is further standardized by using a suitable method of standardization [12]. In this work, the general standardization is used and applied to the initial incidence matrix. The standardization process is expressed as follows:

$$P_j^B = \frac{P_j^A - E_j}{\sigma_j} \qquad (2)$$

Where, $E_j$ is the average of row vector $P_j^A$ and $P_j^B$ is the row vector of the standardized matrix B. $E_j$ can be expressed as:

$$E_j = \frac{\sum_{k=1}^{n} a_{kj}}{n} \qquad (3)$$

Here $a_{kj}$ is the element of initial incidence matrix; *n* is the number of elements in a row vector. Similarly,

$$\sigma_j^2 = \frac{1}{n} \sum_{k=1}^{n} (a_{kj} - E_j)^2 \qquad (4)$$

To simplify the equation 4 further, Huyghens Koning theorem is applied to yield

$$\sigma_j^2 = \sqrt{E_j - (E_j)^2} \qquad (5)$$

Once the standardized matrix is formed, the proposed similarity coefficient is based on the simple correlation matrix of the standard incidence matrix. The correlation matrix S is defined as follows:

$$S = \frac{1}{n} B^T B \qquad (6)$$

S is the square matrix in which elements are given by:

$$S_{ii} = 1 \text{ and } S_{ij} = \frac{1}{n} \sum_{k=1}^{n} b_{ik} b_{jk} \qquad (7)$$

**2.2 Cluster analysis for correlation**

In the second phase of the proposed approach, the part family formation is done on the basis of operations similarity using Principle Component Analysis. Principle Component Analysis is a dimension reduction technique which attempts to model the total variance of the original data set, via new uncorrelated variables called Principal Components. PCA consists of determining a small number of principal components that recover as much variability in the data as possible. These components are linear combinations of the original variables and account for the total variance of the original data. The first principle component is a single axis in space. When each observation is projected on that axis, resultant is a new variable. The variance of this variable is maximum among all possible choices of the first axis. The second principle component is another axis in space, perpendicular to the first one. Projecting the observations on this axis generates another new variable. The variance of this new variable is again maximum among all possible choices of this second axis. The full set of principal components is as large as the original set of variables. However, the sum of the variances of the first few principal components is usually 80% or more of the total variance of the original data [13].

The first component extracted in principal component analysis accounts for a maximal amount of total variance in the observed variables. Under typical conditions, this means that the first component is correlated with at least some of the observed variables. The second component extracted is having two important characteristics. First, this component accounts for a maximal amount of variance in the data set that was not accounted for by the first component. This means that the second component is correlated with some of the observed variables that did not display strong correlations with first component. The second characteristic of the second component is that it is uncorrelated with the first component. This means that the correlation between first and second principle components is zero [14].

The remaining principle components that are extracted in the analysis display the same two characteristics mentioned above. That is, each principle component accounts for a maximal amount of variance in the observed variables that was not accounted for by the preceding components, and is uncorrelated with all of the preceding components. Principal components analysis proceeds in this fashion, with each new component accounting for progressively smaller and smaller amounts of variance.

Thus, the study of principal components is considered as putting into terms the usual developments of eigenvalues and eigenvectors for positive semi-definite matrices. The eigenvector equation where the terms $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ are real, non-negative roots of the determinant polynomial of degree P is given as:

$$\det \ (S - \lambda_i I) = 0 \ ; \ i \in <1, m> \qquad (8)$$

Let $\{F_1, F_2, \ldots, F_m\}$ be corresponding eigenvectors. When PCA was performed on the mean centered data, a model with the first and the second principal components was usually obtained. This model explains the procedure to determine the principle component in the data.

$$\text{Where,} \quad PC = \frac{\lambda_1 + \lambda_2}{\sum_{k=1}^{m} \lambda_k} = \frac{\lambda_1 + \lambda_2}{m} \qquad (9)$$

In this application of PCA, the objective is to cluster parts into families. As part-operation matrix is binary in nature, two principal components are enough to analyse correlation between elements (i.e. parts).

## 2.3 Algorithm for part family formation

The objective of the third phase is to assign parts into families after the preliminary grouping as done in the second phase of the proposed approach. An Agglomerative Hierarchical K-means Clustering Algorithm is used for this purpose. The algorithm for assigning the parts ($P_i$) into parts families is given below:
For each part k=1 to $P_i$ do

*Step 1*:Take n points (Initial seeds) in 2-dimensional plane, where n is the number of variables. (Initial seeds are taken based on score values (coordinate) of each part on first two principle components. Each seed represents a part which contains associated part operations.)

*Step 2*:Compute Euclidean distance for each part.

$$(P_k, P_i) = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2} \qquad (10)$$

Where, $x_i$ and $y_i$ are the co-ordinates of part $P_i$ on two principal components axis.

*Step 3*:Since the objective is to group parts with minimum distance, part under consideration (say $P_i$) is assigned to a family (say $P_k$ ) on the basis of least smallest distance to the part family $P_k$.

*Step 4*:Draw the dendrogram of the sequence of family formation.

The iteration continues until all parts ($P_i$) are assigned to part families.

## 3. NUMERICAL ILLUSTRATION

An example case has been taken to demonstrate the proposed methodology where seven operations (OPs) are required to manufacture the eleven parts in reconfigurable manufacturing system. Parts are labeled as $P_1$ to $P_{11}$ and operations as $OP_1$ to $OP_7$. Table 1 shows part's operational sequence requirements.

| Part Number | Required Operations Sequence |
|---|---|
| $P_1$ | $OP_1$, $OP_2$ |
| $P_2$ | $OP_2$, $OP_3$ |
| $P_3$ | $OP_1$, $OP_5$, $OP_6$ |
| $P_4$ | $OP_4$, $OP_5$, $OP_6$, $OP_8$ |
| $P_5$ | $OP_4$, $OP_5$, $OP_6$ |
| $P_6$ | $OP_2$, $OP_3$ |
| $P_7$ | $OP_1$, $OP_6$ |
| $P_8$ | $OP_6$, $OP_7$ |
| $P_9$ | $OP_3$ |
| $P_{10}$ | $OP_5$, $OP_6$, $OP_7$ |
| $P_{11}$ | $OP_1$, $OP_5$ |

Tab.1. Parts and Operations sequence data

Accordingly a Part Operation Incidence Matrix (POIM) A is constructed as shown in Table 2.

Now, applying the equations 2, 3 and 5 to the part operation incidence matrix given above, yield standardized matrix B, which represents Standard Sequence Part Operation Matrix (SSPOM) in the case. The following procedure is adopted to determine various elements of SSPOM.

| | $P_1$ | $P_2$ | $P_3$ | $P_4$ | $P_5$ | $P_6$ | $P_7$ | $P_8$ | $P_9$ | $P_{10}$ | $P_{11}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $OP_1$ | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| $OP_2$ | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| $OP_3$ | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| $OP_4$ | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| $OP_5$ | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| $OP_6$ | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 |
| $OP_7$ | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |

Tab. 2. Part Operation Incidence Matrix: POIM A

Say, for part $P_1$,

$$E_1 = 2/7 = 0.286$$

$$\sigma_1 = \sqrt{0.286 - (0.286)^2}$$
$$= 0.452$$

The member coefficient between $P_1$ and $OP_1$ (i.e. $b_{11}$) is calculated as follows:

$$b_{11} = \frac{1 - 0.286}{0.452} = 1.580$$

Similarly,

$$b_{21} = \frac{1 - 0.286}{0.452} = 1.580$$

$$and\ b_{31} = \frac{0 - 0.286}{0.452} = -0.633$$

The same procedure is repeated for other elements. Finally, SSPOM is obtained as shown in Table 3.

From the above SSPOM and on the basis of proposed similarity coefficient (as given in the eq. 7), the correlation matrix (S) is obtained as shown in Table 4.
Now cluster analysis is performed based on PCA method and by using eq. 8. The computed eigenvalues for the correlation matrix (S) and their associated variance, and cumulative variance are listed, sorted in a descending order as shown in Table 5.

The two principle components are having the maximum variance (i.e. almost 80%) and are sufficiently enough to represent the all parts operations data. Hence for the further analysis two principle components are taken.

| | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | P11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| OP1 | 1.580 | -0.633 | 1.154 | -0.633 | -0.633 | -0.633 | 1.580 | -0.409 | -0.409 | -0.633 | 1.580 |
| OP2 | 1.580 | 1.580 | -0.867 | -0.633 | -0.633 | 1.580 | -0.633 | -0.409 | -0.409 | -0.633 | -0.633 |
| OP3 | -0.633 | 1.580 | -0.867 | -0.633 | -0.633 | 1.580 | -0.633 | -0.409 | 2.449 | -0.633 | -0.633 |
| OP4 | -0.633 | -0.633 | -0.867 | 1.580 | 1.580 | -0.633 | -0.633 | -0.409 | -0.409 | 1.580 | -0.633 |
| OP5 | -0.633 | -0.633 | 1.154 | -0.633 | -0.633 | -0.633 | 1.580 | -0.409 | -0.409 | -0.633 | -0.633 |
| OP6 | -0.633 | -0.633 | 1.154 | 1.580 | 1.580 | -0.633 | -0.633 | -0.409 | -0.409 | -0.633 | 1.580 |
| OP7 | -0.633 | -0.633 | -0.867 | -0.633 | -0.633 | -0.633 | -0.633 | 2.449 | -0.409 | 1.580 | -0.633 |

Tab.3. Standard Sequence Part Operation Matrix

| | $P_1$ | $P_2$ | $P_3$ | $P_4$ | $P_5$ | $P_6$ | $P_7$ | $P_8$ | $P_9$ | $P_{10}$ | $P_{11}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $P_1$ | 1.000 | | | | | | | | | | |
| $P_2$ | 0.300 | 1.000 | | | | | | | | | |
| $P_3$ | 0.091 | -0.548 | 1.000 | | | | | | | | |
| $P_4$ | -0.200 | -0.400 | 0.091 | 1.000 | | | | | | | |
| $P_5$ | 0.400 | -0.400 | 0.091 | 0.949 | 1.000 | | | | | | |
| $P_6$ | 0.300 | 0.999 | -0.548 | -0.400 | -0.400 | 1.000 | | | | | |
| $P_7$ | 0.300 | -0.400 | 0.730 | -0.400 | -0.400 | -0.400 | 1.000 | | | | |
| $P_8$ | -0.258 | -0.258 | -0.354 | -0.258 | -0.258 | -0.258 | -0.258 | 1.000 | | | |
| $P_9$ | -0.258 | 0.645 | -0.354 | -0.258 | -0.258 | 0.645 | -0.258 | -0.167 | 1.000 | | |
| $P_{10}$ | -0.200 | -0.400 | -0.548 | 0.300 | 0.300 | -0.400 | -0.400 | 0.645 | -0.258 | 1.000 | |
| $P_{11}$ | 0.300 | -0.400 | 0.730 | 0.300 | 0.300 | -0.400 | 0.300 | -0.258 | -0.258 | -0.400 | 1.000 |

Tab.4. Correlation matris S

| No of principal components | Eigen value | % variance | Cumulative % variance |
|---|---|---|---|
| 1 | 1.52e+000 | 50.04 | 50.04 |
| 2 | 8.53e-001 | 28.11 | 78.15 |

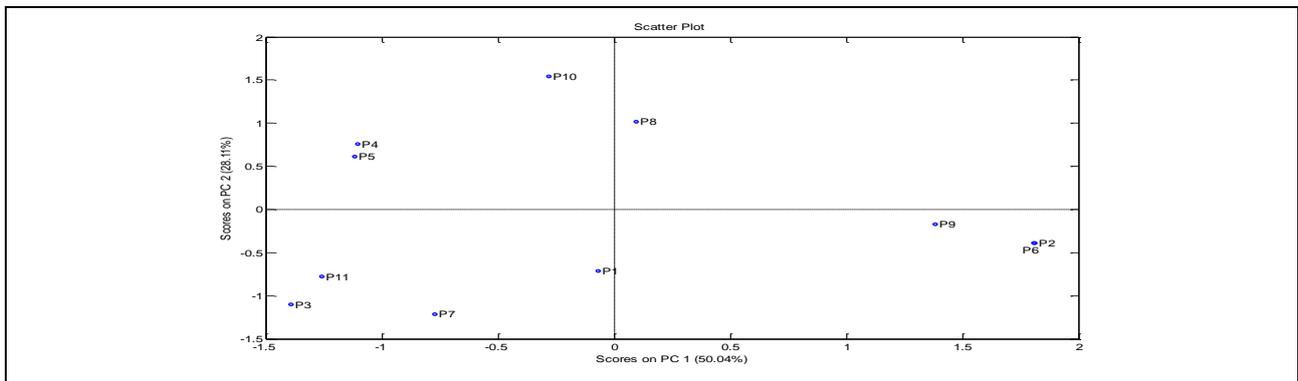Tab.5. Principle components, Eigen value and Percentage Variance



Fig. 2. Graphical representation of Scatter Plot

Further, the graphical analysis is performed by a two dimensional scatter plot where each part is represented by a dot and the two axes of the scatter plot are the two principle components. This scatter plot indicates the relationship between parts as shown in Figure 2. There is high correlation between parts which are closely placed and thereby are strongly associated with each other such as ($P_2$ and $P_6$) and ($P_4$ and $P_5$). The various correlation values are shown in Table 4. On the basis of correlation results, the following principle situations are recovered from the scatter plot:

i. Two neighboring parts having low distance measure belong to the same group such as $P_2$ and $P_6$; $P_4$ and $P_5$.

ii. Part group ($P_4$ and $P_5$) which is almost $180^0$ to the other part group ($P_6$ and $P_2$; $P_9$) are negatively correlated and thus cannot belong to the same group.

iii. Two parts which are placed almost $90^0$ to each other such as $P_9$ and $P_8$ are independent and thus cannot belong to the same group. It is clearly seen in the Figure 2.

The co-ordinate (score) of each part on first two principle components is obtained from scatter plot as shown in Table 6.

| Part Number | First Principle Component (50.04%) | Second Principle Component (28.11%) |
|---|---|---|
| $P_1$ | -0.0699 | -0.7105 |
| $P_2$ | 1.8086 | -0.3856 |
| $P_3$ | -1.3957 | -1.1034 |
| $P_4$ | -1.1100 | 0.7616 |
| $P_5$ | -1.1208 | 0.6156 |
| $P_6$ | 1.8086 | -0.3856 |
| $P_7$ | -0.7736 | -1.2148 |
| $P_8$ | 0.0917 | 1.0202 |
| $P_9$ | 1.3827 | -0.1736 |
| $P_{10}$ | -0.2835 | 1.5465 |
| $P_{11}$ | -1.2622 | -0.7806 |

Tab.6. Score (Co-ordinate) of parts

The third phase of the methodology is to cluster parts, obtained from the preliminary solution space in second phase, into family by using AHKCA mentioned in the methodology. The algorithm is initialized by taking score values (coordinates) of parts from scatter plot as a starting point. The algorithm starts with the initial solution i.e. the number of parts and then computes and stores the euclidean distance between each part using equation 10. The parts having minimum distance are grouped first and these grouped parts are removed from the subsequent iteration. This process continues until all the parts are grouped together. The algorithm finally provides output in the form of dendrogram (Figure 3).
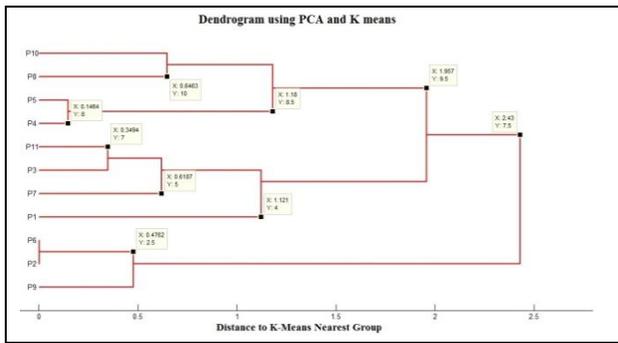
Fig.3. Hierarchical Clustering of parts

## 4. RESULT

The resulting dendrogram is based on the distances to K-means nearest group. In K-means algorithm, the measured distance is known as dissimilarity measure. Hence, the magnitude of the distance to nearest group of parts represents the dissimilarity associated with the part. The ratio of the distance to the nearest group to the maximum distance measured in the dendrogram gives the percentage of dissimilarity of the particular parts group. In Figure 3, parts $P_2$ and part $P_6$ are found at zero distance. So their percentage of dissimilarity is zero. It means that they are 100% similar in operations. The distance between parts $P_4$ and $P_5$ is 0.1464 and their associated percentage of dissimilarity is calculated to 6.0 % and level of similarity is 94%. The same procedure is repeated to calculate the other similarity level for remaining cluster of parts. The obtained results are summarized in Table 7.

| Distance between parts group | Similarity Level (%) | Formed part families | Number of families |
|---|---|---|---|
| 0.000 | 100 | $P_{2,6}$,$P_1$,$P_3$, $P_4$,$P_5$,$P_7$,$P_8$,$P_9$,$P_{10}$,$P_{11}$ | 10 |
| 0.1464 | 94.0 | $P_{2,6}$,$P_1$,$P_3$, $P_{4,5}$,$P_7$,$P_8$,$P_9$,$P_{10}$,$P_{11}$ | 09 |
| 0.3494 | 85.6 | $P_{2,6}$,$P_1$,$P_{3,11}$, $P_{4,5}$,$P_7$,$P_8$,$P_9$,$P_{10}$ | 08 |
| 0.4762 | 80.4 | $P_{2,6,9}$,$P_1$,$P_{3,11}$, $P_{4,5}$,$P_7$,$P_8$, $P_{10}$ | 07 |
| 0.6187 | 74.5 | $P_{2,6,9}$,$P_1$,$P_{3,11,7}$, $P_{4,5}$, $P_8$, $P_{10}$ | 06 |
| 0.6463 | 73.4 | $P_{2,6,9}$,$P_1$,$P_{3,11,7}$, $P_{4,5}$, $P_{8,10}$ | 05 |
| 1.121 | 53.9 | $P_{2,6,9}$,$P_{1,3,11,7}$, $P_{4,5}$, $P_{8,10}$ | 04 |
| 1.180 | 51.4 | $P_{2,6,9}$,$P_{1,3,11,7}$, $P_{4,5,8,10}$ | 03 |
| 1.957 | 19.5 | $P_{2,6,9}$,$P_{1,3,11,7,4,5,8,10}$ | 02 |
| 2.430 | 2.8 | $P_{2,6,9,1,3,11,7,4,5,8,10}$ | 01 |

Tab.7. Percentage of similarity and Formed part family

On the basis of the formed part families the system planner first configures the manufacturing system to produce the first part family. Once it is finished, the system is reconfigured to produce the second part family and so forth. Each system reconfiguration adds cost to the production of the parts. Hence there is a need to arrive at a suitable number of part families by selecting an appropriate value of percentage of similarity as a cut off to achieve minimum cost solution of the problem.

## 5. CONCLUSION

This work has presented a novel methodology for grouping parts into families on the basis of operation sequence similarity which is a central issue in the design of reconfigurable manufacturing systems. A correlation analysis model is formulated to group the parts and operations sequence and Correlation matrix is used as the similarity coefficient matrix. Finally, Principle Component Analysis and Agglomerative Hierarchical K-means algorithm is applied to find the level of similarity in parts. The obtained part families are based on compactness of the family formation on the basis of operational sequence similarity. Furthermore it uses PCA, which is available in many commercial software packages. Although, PCA is used to reduce the dimension of data but sometimes high dimensional data may have many redundant or irrelevant features. These redundant features are of no help for clustering and may create nosie. For this, a noise reduction technique can be applied to remove the noise. However the proposed work can be further extended to accommodate other factors such as, production volume, alternative operation sequences and alternative routings.

## 6. REFERENCES

[1] Koren,Y.,Jovane,F.,Heisel,U.,Moriwaki,T.,Pritschow,G.,Ulsoy,A. G.&VanBrussel,H.(1999). Reconfigurable Manufacturing Systems. *CIRP Annals*, Vol. 48, No. 2, pp.527-540

[2] Xiaobo, Z., Jiancai, W., & Zhenbi, L.(2000). A stochastic model of a reconfigurable manufacturing system, Part 1: a framework. *International Journal of Production Research*, Vol.38, No.10, 2273–2285

[3] Lokesh,.K. & Jain, P.K.(2011). A model and optimization approach for reconfigurable manufacturing system configuration design. *International Journal of Production Research*, Vol.50, No.12, ISSN 0020-7543

[4] Abdi, M.R. & Labib, A.W.(2004). Grouping and selecting products: the design key of Reconfigurable Manufacturing Systems (RMSs). *International Journal of Production Research*, Vol.42, No.3, 521–546

[5] Galan,R., Racero, J., Eguia,I. & Garcia, J.M. (2007). A systematic approach for product families formation in Reconfigurable Manufacturing Systems. *Robotics and Computer-Integrated Manufacturing*, Vol. 23, 489–502

[6] Jolliffe, I.T.(1986). *Principal component analysis*. Springer,ISBN 0-387-95442-2, New York

[7] Pearson, K.(1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, Vol. 2, 559-572

[8] Tuncer,Y., Tanik, M.M. & Alison, D.B.(2008). An overview of statistical decomposition techniques applied to complex systems. *Computational Statistics & Data Analysis*, Vol.52, No. 5, 2292-2310

[9] Horenko, I., Dittmer, E. & Schütte, C.(2006). Reduced Stochastic Models for Complex Molecular Systems. *Computing and Visualization in Science*, Vol.9, No. 2, 89-102

[10] Lokesh, K. & Jain, P.K.(2010). Concurrent part machine group formation with important production data. *International Journal of simulation modeling*, Vol.9, No.1, 5-16

[11] Shehroz, S.K. & Ahmad,A. (2004).Cluster center initialization algorithm for K-means clustering. *Pattern RecognitionLetter*, Vol.25, 1293-1302

[12] Chaea, S.S. & Wardeb, W.D. (2005). Effect of using principal coordinates and principal components on retrieval of clusters. *Computational Statistics & Data Analysis*, Vol.50, No.6,1407-1417

[13] Gnanadesikan, R. (1997). *Methods for statistical data analysis of multivariate observations*. Wiley-Interscience, ISBN 0-471-16119-5, New York

[14] Rummel, R.J. (1988). *Applied factor analysis*. Northwestern University Press, ISBN 0-8101-0824-0, Evanston, USA