We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

8

Application Similarity Coefficient Method to Cellular Manufacturing

Yong Yin

1. Introduction

Group technology (GT) is a manufacturing philosophy that 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 groups 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.
CM has been considered as an alternative to conventional batch-type manufacturing where different products are produced intermittently in small lot sizes. For batch manufacturing, the volume of any particular part may not be enough to require a dedicated production line for that part. Alternatively, the total volume for a family of similar parts may be enough to efficiently utilize a machine-cell (Miltenburg and Zhang, 1991).
It has been reported (Seifoddini, 1989a) that employing CM may help overcome major problems of batch-type manufacturing including frequent setups, excessive in-process inventories, long through-put times, complex planning and control functions, and provides the basis for implementation of manufacturing techniques such as just-in-time (JIT) and flexible manufacturing systems (FMS).

A large number of studies related to GT/CM have been performed both in academia and industry. Reisman *et al*. (1997) gave a statistical review of 235 articles dealing with GT and CM over the years 1965 through 1995. They reported that the early (1966-1975) literature dealing with GT/CM appeared predominantly in book form. The first written material on GT was Mitrofanov (1966) and the first journal paper that clearly belonged to CM appeared in 1969 (Optiz *et al*., 1969). Reisman *et al*. (1997) also reviewed and classified these 235 articles on a five-point scale, ranging from pure theory to bona fide applications.

Source: Manufacturing the Future, Concepts - Technologies - Visions , ISBN 3-86611-198-3, pp. 908, ARS/plV, Germany, July 2006, Edited by: Kordic, V.; Lazinica, A. & Merdan, M.

In addition, they analyzed seven types of research processes used by authors. There are many researchable topics related to cellular manufacturing. Wemmerlöv and Hyer (1987) presented four important decision areas for group technology adoption – applicability, justification, system design, and implementation. A list of some critical questions was given for each area.

Applicability, in a narrow sense, can be understood as feasibility (Wemmerlöv and Hyer, 1987). Shafer *et al*. (1995) developed a taxonomy to categorize manufacturing cells. They suggested three general cell types: process cells, product cells, and other types of cells. They also defined four shop layout types: product cell layouts, process cell layouts, hybrid layouts, and mixture layouts. Despite the growing attraction of cellular manufacturing, most manufacturing systems are hybrid systems (Wemmerlöv and Hyer, 1987; Shambu and Suresh, 2000). A hybrid CM system is a combination of both a functional layout and a cellular layout. Some hybrid CM systems are unavoidable, since some processes such as painting or heat treatment are frequently more efficient and economic to keep the manufacturing facilities in a functional layout.
Implementation of a CM system contains various aspects such as human, education, environment, technology, organization, management, evaluation and even culture. Unfortunately, only a few papers have been published related to these areas. Researches reported on the human aspect can be found in Fazakerley (1976), Burbidge *et al*. (1991), Beatty (1992), and Sevier (1992). Some recent studies on implementation of CM systems are Silveira (1999), and Wemmerlöv and Johnson (1997; 2000).

The problem involved in justification of cellular manufacturing systems has received a lot of attention. Much of the research was focused on the performance comparison between cellular layout and functional layout. A number of researchers support the relative performance supremacy of cellular layout over functional layout, while others doubt this supremacy. Agarwal and Sarkis (1998) gave a review and analysis of comparative performance studies on functional and CM layouts. Shambu and Suresh (2000) studied the performance of hybrid CM systems through a computer simulation investigation.

System design is the most researched area related to CM. Research topics in this area include cell formation (CF), cell layout (Kusiak and Heragu, 1987; Balakrishnan and Cheng; 1998; Liggett, 2000), production planning (Mosier and Taube, 1985a; Singh, 1996), and others (Lashkari *et al*, 2004; Solimanpur *et al*, 2004). CF is the first, most researched topic in designing a CM system. Many approaches and methods have been proposed to solve the CF problem. Among these methods, Production flow analysis (PFA) is the first one which was used by Burbidge (1971) to rearrange a machine part incidence matrix on trial and error until an acceptable solution is found. Several review papers have been published to classify and evaluate various approaches for CF, some of them will be discussed in this paper. Among various cell formation models, those based on the similarity coefficient method (SCM) are more flexible in incorporating manufacturing data into the machine-cells formation process (Seifoddini, 1989a). In this paper, an attempt has been made to develop a taxonomy for a comprehensive review of almost all similarity coefficients used for solving the cell formation problem.

Although numerous CF methods have been proposed, fewer comparative studies have been done to evaluate the robustness of various methods. Part reason is that different CF methods include different production factors, such as machine requirement, setup times, utilization, workload, setup cost, capacity, part alternative routings, and operation sequences. Selim, Askin and Vakharia (1998) emphasized the necessity to evaluate and compare different CF methods based on the applicability, availability, and practicability. Previous comparative studies include Mosier (1989), Chu and Tsai (1990), Shafer and Meredith (1990), Miltenburg and Zhang (1991), Shafer and Rogers (1993), Seifoddini and Hsu (1994), and Vakharia and Wemmerlöv (1995).

Among the above seven comparative studies, Chu and Tsai (1990) examined three array-based clustering algorithms: rank order clustering (ROC) (King, 1980), direct clustering analysis (DCA) (Chan & Milner, 1982), and bond energy analysis (BEA) (McCormick, Schweitzer & White, 1972); Shafer and Meredith (1990) investigated six cell formation procedures: ROC, DCA, cluster identification algorithm (CIA) (Kusiak & Chow, 1987), single linkage clustering (SLC), average linkage clustering (ALC), and an operation sequences based similarity coefficient (Vakharia & Wemmerlöv, 1990); Miltenburg and Zhang (1991) compared nine cell formation procedures. Some of the compared procedures are combinations of two different algorithms *A1/A2*. *A1/A2* denotes using *A1* (algorithm 1) to group machines and using *A2* (algorithm 2) to group parts. The nine procedures include: ROC, SLC/ROC, SLC/SLC, ALC/ROC, ALC/ALC, modified ROC (MODROC) (Chandrasekharan & Rajagopalan, 1986b), ideal seed non-hierarchical clustering (ISNC) (Chandrasekharan & Rajagopalan, 1986a), SLC/ISNC, and BEA.

The other four comparative studies evaluated several similarity coefficients. We will discuss them in the later section.

2. Background

This section gives a general background of machine-part CF models and detailed algorithmic procedures of the similarity coefficient methods.

2.1 Machine-part cell formation

The CF problem can be defined as: "If the number, types, and capacities of production machines, the number and types of parts to be manufactured, and the routing plans and machine standards for each part are known, which machines and their associated parts should be grouped together to form cell?" (Wei and Gaither, 1990). Numerous algorithms, heuristic or non-heuristic, have emerged to solve the cell formation problem. A number of researchers have published review studies for existing CF literature (refer to King and Nakornchai, 1982; Kumar and Vannelli, 1983; Mosier and Taube, 1985a; Wemmerlöv and Hyer, 1986; Chu and Pan, 1988; Chu, 1989; Lashkari and Gunasingh, 1990; Kamrani *et al*., 1993; Singh, 1993; Offodile *et al*., 1994; Reisman *et al*., 1997; Selim *et al*., 1998; Mansouri *et al*., 2000). Some timely reviews are summarized as follows.

Singh (1993) categorized numerous CF methods into the following sub-groups: part coding and classifications, machine-component group analysis, similarity coefficients, knowledge-based, mathematical programming, fuzzy clustering, neural networks, and heuristics.

Offodile *et al*. (1994) employed a taxonomy to review the machine-part CF models in CM. The taxonomy is based on Mehrez *et al*. (1988)'s five-level conceptual scheme for knowledge representation. Three classes of machine-part grouping techniques have been identified: visual inspection, part coding and classification, and analysis of the production flow. They used the production flow analysis segment to discuss various proposed CF models.

Reisman *et al*. (1997) gave a most comprehensive survey. A total of 235 CM papers were classified based on seven alternatives, but not mutually exclusive, strategies used in Reisman and Kirshnick (1995).

Selim *et al*. (1998) developed a mathematical formulation and a methodology-based classification to review the literature on the CF problem. The objective function of the mathematical model is to minimize the sum of costs for purchasing machines, variable cost of using machines, tooling cost, material handling cost, and amortized worker training cost per period. The model is combinatorially complex and will not be solvable for any real problem. The classification used in this paper is based on the type of general solution methodology. More than 150 works have been reviewed and listed in the reference.

2. Similarity coefficient methods (SCM)

A large number of similarity coefficients have been proposed in the literature. Some of them have been utilized in connection with CM. SCM based methods rely on similarity measures in conjunction with clustering algorithms. It usually follows a prescribed set of steps (Romesburg, 1984), the main ones being:

Step (1). Form the initial machine part incidence matrix, whose rows are ma chines and columns stand for parts. The entries in the matrix are 0s or 1s, which indicate a part need or need not a machine for a pro duction. An entry a_{ik} is defined as follows.

$$a_{ik} = \begin{cases} 1 \text{ if part } k \text{ visits machine } i, \\ 0 \text{ otherwise.} \end{cases} \quad (1)$$

where

i -- machine index (i=1,…, M)

k -- part index (k=1,…, P)

M --number of machines

P -- number of parts

Step (2). Select a similarity coefficient and compute similarity values be tween machine (part) pairs and construct a similarity matrix. An element in the matrix represents the sameness between two ma chines (parts).

Step (3). Use a clustering algorithm to process the values in the similarity matrix, which results in a diagram called a tree, or dendrogram, that shows the hierarchy of similarities among all pairs of machines (parts). Find the machines groups (part families) from the tree or dendrogram, check all predefined constraints such as the number of cells, cell size, etc.

3. Why present a taxonomy on similarity coefficients?

Before answer the question "Why present a taxonomy on similarity coefficients?", we need to answer the following question firstly "Why similarity co-

efficient methods are more flexible than other cell formation methods?".
In this section, we present past review studies on similarity coefficients, discuss their weaknesses and confirm the need of a new review study from the viewpoint of the flexibility of similarity coefficients methods.

3.1 Past review studies on similarity coefficients

Although a large number of similarity coefficients exist in the literature, very few review studies have been performed on similarity coefficients. Three review papers on similarity coefficients (Shafer and Rogers, 1993a; Sarker, 1996; Mosier *et al*., 1997) are available in the literature.
Shafer and Rogers (1993a) provided an overview of similarity and dissimilarity measures applicable to cellular manufacturing. They introduced general measures of association firstly, then similarity and distance measures for determining part families or clustering machine types are discussed. Finally, they concluded the paper with a discussion of the evolution of similarity measures applicable to cellular manufacturing.
Sarker (1996) reviewed a number of commonly used similarity and dissimilarity coefficients. In order to assess the quality of solutions to the cell formation problem, several different performance measures are enumerated, some experimental results provided by earlier researchers are used to evaluate the performance of reviewed similarity coefficients.
Mosier *et al*. (1997) presented an impressive survey of similarity coefficients in terms of structural form, and in terms of the form and levels of the information required for computation. They particularly emphasized the structural forms of various similarity coefficients and made an effort for developing a uniform notation to convert the originally published mathematical expression of reviewed similarity coefficients into a standard form.

3.2 Objective of this study

The three previous review studies provide important insights from different viewpoints. However, we still need an updated and more comprehensive review to achieve the following objectives.

- Develop an explicit taxonomy
  To the best of our knowledge, none of the previous articles has developed or employed an explicit taxonomy to categorize various similarity coefficients.

We discuss in detail the important role of taxonomy in the section 3.3.
Neither Shafer and Rogers (1993a) nor Sarker (1996) provided a taxonomic review framework. Sarker (1996) enumerated a number of commonly used similarity and dissimilarity coefficients; Shafer and Rogers (1993a) classified similarity coefficients into two groups based on measuring the resemblance between: (1) part pairs, or (2) machine pairs.

- Give a more comprehensive review
  Only a few similarity coefficients related studies have been reviewed by previous articles.
  Shafer and Rogers (1993a) summarized 20 or more similarity coefficients related researches; Most of the similarity coefficients reviewed in Sarker (1996)'s paper need prior experimental data; Mosier et al. (1997) made some efforts to abstract the intrinsic nature inherent in different similarity coefficients, Only a few similarity coefficients related studies have been cited in their paper.

Owing to the accelerated growth of the amount of research reported on similarity coefficients subsequently, and owing to the discussed objectives above, there is a need for a more comprehensive review research to categorize and summarize various similarity coefficients that have been developed in the past years.

3.3 Why similarity coefficient methods are more flexible

The cell formation problem can be extraordinarily complex, because of various different production factors, such as alternative process routings, operational sequences, production volumes, machine capacities, tooling times and others, need to be considered. Numerous cell formation approaches have been developed, these approaches can be classified into following three groups:

1. Mathematical Programming (MP) models.
2. (meta-)Heurestic Algorithms (HA).
3. Similarity Coefficient Methods (SCM).

Among these approaches, SCM is the application of cluster analysis to cell formation procedures. Since the basic idea of GT depends on the estimation of the similarities between part pairs and cluster analysis is the most basic method for estimating similarities, it is concluded that SCM based method is one of the most basic methods for solving CF problems.

Despite previous studies (Seifoddini, 1989a) indicated that SCM based approaches are more flexible in incorporating manufacturing data into the machine-cells formation process, none of the previous articles has explained the reason why SCM based methods are more flexible than other approaches such as MP and HA. We try to explain the reason as follows.

For any concrete cell formation problem, there is generally no "correct" approach. The choice of the approach is usually based on the tool availability, analytical tractability, or simply personal preference. There are, however, two effective principles that are considered reasonable and generally accepted for large and complex problems. They are as follows.

- Principle :

  Decompose the complex problem into several small conquerable problems. Solve small problems, and then reconstitute the solutions.

  All three groups of cell formation approaches (MP, HA, SCM) mentioned above can use principle◎for solving complex cell formation problems. However, the difficulty for this principle is that a systematic mean must be found for dividing one complex problem into many small conquerable problems, and then reconstituting the solutions. It is usually not easy to find such systematic means.

- Principle :

  It usually needs a complicated solution procedure to solve a complex cell formation problem. The second principle is to decompose the complicated solution procedure into several small tractable stages.

Comparing with MP, HA based methods, the SCM based method is more suitable for principle◎. We use a concrete cell formation model to explain this conclusion. Assume there is a cell formation problem that incorporates two production factors: production volume and operation time of parts.

(1). MP, HA:

By using MP, HA based methods, the general way is to construct a mathematical or non-mathematical model that takes into account production volume and operation time, and then the model is analyzed, optimal or heuristic solution procedure is developed to solve the problem. The advantage of this way is that the developed model and solution procedure are usually unique for the original problem. So, even if they are not the "best" solutions, they are usually "very good" solutions for the original problem. However, there are two disadvantages inherent in the MP, HA based methods.

• Firstly, extension of an existing model is usually a difficult work. For example, if we want to extend the above problem to incorporate other production factors such as alternative process routings and operational sequences of parts, what we need to do is to extend the old model to incorporate additional production factors or construct a new model to incorporate all required production factors: production volumes, operation times, alternative process routings and operational sequences. Without further information, we do not know which one is better, in some cases extend the old one is more efficient and economical, in other cases construct a new one is more efficient and economical. However, in most cases both extension and construction are difficult and cost works.

• Secondly, no common or standard ways exist for MP, HA to decompose a complicated solution procedure into several small tractable stages. To solve a complex problem, some researchers decompose the solution procedure into several small stages. However, the decomposition is usually based on the experience, ability and preference of the researchers. There are, however, no common or standard ways exist for decomposition.

(2). SCM:
SCM is more flexible than MP, HA based methods, because it overcomes the two mentioned disadvantages of MP, HA. We have introduced in section 2.2 that the solution procedure of SCM usually follows a prescribed set of steps:

Step 1. Get input data;
Step 2. Select a similarity coefficient;
Step 3. Select a clustering algorithm to get machine cells.

Thus, the solution procedure is composed of three steps, this overcomes the second disadvantage of MP, HA. We show how to use SCM to overcome the first disadvantage of MP, HA as follows.
An important characteristic of SCM is that the three steps are independent

with each other. That means the choice of the similarity coefficient in step2 does not influence the choice of the clustering algorithm in step3. For example, if we want to solve the production volumes and operation times considered cell formation problem mentioned before, after getting the input data; we select a similarity coefficient that incorporates production volumes and operation times of parts; finally we select a clustering algorithm (for example ALC algorithm) to get machine cells. Now we want to extend the problem to incorporate additional production factors: alternative process routings and operational sequences. We re-select a similarity coefficient that incorporates all required 4 production factors to process the input data, and since step2 is independent from step3, we can easily use the ALC algorithm selected before to get new machine cells. Thus, comparing with MP, HA based methods, SCM is very easy to extend a cell formation model.

Therefore, according above analysis, SCM based methods are more flexible than MP, HA based methods for dealing with various cell formation problems. To take full advantage of the flexibility of SCM and to facilitate the selection of similarity coefficients in step2, we need an explicit taxonomy to clarify and classify the definition and usage of various similarity coefficients. Unfortunately, none of such taxonomies has been developed in the literature, so in the next section we will develop a taxonomy to summarize various similarity coefficients.

4. A taxonomy for similarity coefficients employed in cellular manufacturing

Different similarity coefficients have been proposed by researchers in different fields. A similarity coefficient indicates the degree of similarity between object pairs. A tutorial of various similarity coefficients and related clustering algorithms are available in the literature (Anderberg, 1973; Bijnen, 1973; Sneath and Sokal, 1973; Arthanari and Dodge, 1981; Romesburg, 1984; Gordon, 1999). In order to classify similarity coefficients applied in CM, a taxonomy is developed and shown in figure 1. The objective of the taxonomy is to clarify the definition and usage of various similarity or dissimilarity coefficients in designing CM systems. The taxonomy is a 5-level framework numbered from level 0 to 4. Level 0 represents the root of the taxonomy. The detail of each level is described as follows.

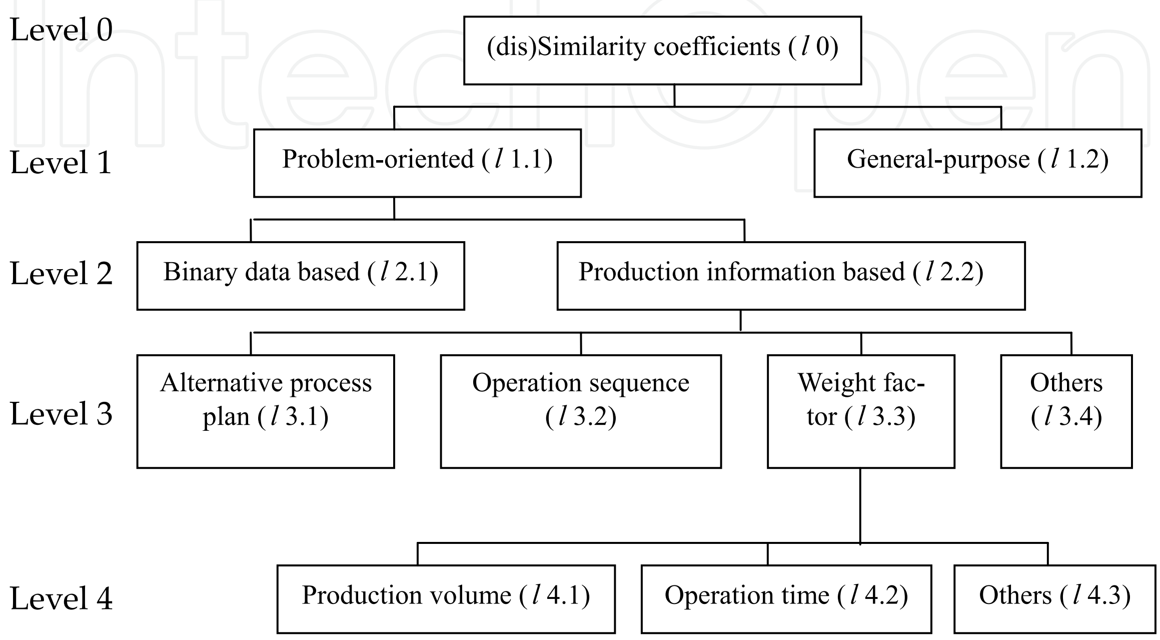

Figure 1. A taxonomy for similarity coefficients

Level 1.

l 1 categorizes existing similarity coefficients into two distinct groups: problem-oriented similarity coefficients (*l* 1.1) and general-purpose similarity coefficients (*l* 1.2). Most of the similarity coefficients introduced in the field of numerical taxonomy are classified in *l* 1.2 (general-purpose), which are widely used in a number of disciplines, such as psychology, psychiatry, biology, sociology, the medical sciences, economics, archeology and engineering. The characteristic of this type of similarity coefficients is that they always maximize similarity value when two objects are perfectly similar.

On the other hand, problem-oriented (*l* 1.1) similarity coefficients aim at evaluating the predefined specific “appropriateness” between object pairs. This type of similarity coefficient is designed specially to solve specific problems, such as CF. They usually include additional information and do not need to produce maximum similarity value even if the two objects are perfectly similar. Two less similar objects can produce a higher similarity value due to their “appropriateness” and more similar objects may produce a lower similarity value due to their “inappropriateness”.

We use three similarity coefficients to illustrate the difference between the problem-oriented and general-purpose similarity coefficients. Jaccard is the most commonly used general-purpose similarity coefficient in the literature, Jaccard similarity coefficient between machine i and machine j is defined as follows:

$$s_{ij} = \frac{a}{a+b+c}, \qquad 0 \le s_{ij} \le 1 \tag{2}$$

where

a: the number of parts visit both machines,

b: the number of parts visit machine i but not j,

c: the number of parts visit machine j but not i,

Two problem-oriented similarity coefficients, MaxSC (Shafer and Rogers, 1993b) and Commonality score (CS, Wei and Kern, 1989), are used to illustrate this comparison. MaxSC between machine i and machine j is defined as follows:

$$ms_{ij} = max[\frac{a}{a+b}, \frac{a}{a+c}], \qquad 0 \le ms_{ij} \le 1 \tag{3}$$

and CS between machine i and machine j is calculated as follows:

$$c_{ij} = \sum_{k=1}^{P} \delta(a_{ik}, a_{jk}) \tag{4}$$

Where

$$\delta(a_{ik}, a_{jk}) = \begin{cases} (P-1), & \text{if } a_{ik} = a_{jk} = 1 \\ 1, & \text{if } a_{ik} = a_{jk} = 0 \\ 0, & \text{if } a_{ik} \ne a_{jk}. \end{cases} \tag{5}$$

$$a_{ik} = \begin{cases} 1, & \text{if machine } i \text{ uses part } k, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

k: part index (k=1,… P), is the k th part in the machine-part matrix.

We use figure 2 and figure 3 to illustrate the "appropriateness" of problem-oriented similarity coefficients. Figure 2 is a machine-part incidence matrix whose rows represent machines and columns represent parts. The Jaccard coefficient s_{ij}, MaxSC coefficient ms_{ij} and commonality score c_{ij} of machine pairs in figure 2 are calculated and given in figure 3.

The characteristic of general-purpose similarity coefficients is that they always maximize similarity value when two objects are perfectly similar. Among the four machines in figure 2, we find that machine 2 is a perfect copy of machine

1, they should have the highest value of similarity. We also find that the degree of similarity between machines 3 and 4 is lower than that of machines 1 and 2. The results of Jaccard in figure 3 reflect our finds straightly. That is, $max(s_{ij})=s_{12}=1$, and $s_{12}>s_{34}$.

		Part													
		p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	p13	p14
Machine	m1	1	1	1											
	m2	1	1	1											
	m3	1	1	1	1										
	m4	1	1	1	1	1	1	1							

Figure 2. Illustrative machine-part matrix for the "appropriateness"

	Similarity values, s_{ij}, ms_{ij} and c_{ij}			
	$i=1, j=2$	$i=3, j=4$	$i=1or2, j=3$	$i=1or2, j=4$
Jaccard s_{ij}	1	4/7	3/4	3/7
MaxSC ms_{ij}	1	1	1	1
CS c_{ij}	50	59	49	46

Figure 3. Similarity values of Jaccard, MaxSC and CS of figure 2

Problem-oriented similarity coefficients are designed specially to solve CF problems. CF problems are multi-objective decision problems. We define the "appropriateness" of two objects as the degree of possibility to achieve the objectives of CF models by grouping the objects into the same cell. Two objects will obtain a higher degree of "appropriateness" if they facilitate achieving the predefined objectives, and vice versa. As a result, two less similar objects can produce a higher similarity value due to their "appropriateness" and more similar objects may produce a lower similarity value due to their "inappropriateness". Since different CF models aim at different objectives, the criteria of "appropriateness" are also varied. In short, for problem-oriented similarity coefficients, rather than evaluating the similarity between two objects, they evaluate the "appropriateness" between them.

MaxSC is a problem-oriented similarity coefficient (Shafer and Rogers, 1993b). The highest value of MaxSC is given to two machines if the machines process exactly the same set of parts or if one machine processes a subset of the parts processed by the other machine. In figure 3, all machine pairs obtain the highest MaxSC value even if not all of them are perfectly similar. Thus, in the procedure of cell formation, no difference can be identified from the four machines by MaxSC.
CS is another problem-oriented similarity coefficient (Wei and Kern, 1989). The objective of CS is to recognize not only the parts that need both machines, but also the parts on which the machines both do not process. Some characteristics of CS have been discussed by Yasuda and Yin (2001). In figure 3, the highest CS is produced between machine 3 and machine 4, even if the degree of similarity between them is lower and even if machines 1 and 2 are perfectly similar. The result $s_{34} > s_{12}$ illustrates that two less similar machines can obtain a higher similarity value due to the higher "appropriateness" between them.
Therefore, it is concluded that the definition of "appropriateness" is very important for every problem-oriented similarity coefficient, it determines the quality of CF solutions by using these similarity coefficients.

Level 2.
In figure 1, problem-oriented similarity coefficients can be further classified into binary data based (l2.1) and production information based (l2.2) similarity coefficients. Similarity coefficients in l2.1 only consider assignment information, that is, a part need or need not a machine to perform an operation. The assignment information is usually given in a machine-part incidence matrix, such as figure 2. An entry of "1" in the matrix indicates that the part needs a operation by the corresponding machine. The characteristic of l2.1 is similar to l1.2, which also uses binary input data. However, as we mentioned above, they are essentially different in the definition for assessing the similarity between object pairs.

Level 3.
In the design of CM systems, many manufacturing factors should be involved when the cells are created, e.g. machine requirement, machine setup times, utilization, workload, alternative routings, machine capacities, operation sequences, setup cost and cell layout (Wu and Salvendy, 1993). Choobineh and Nare (1999) described a sensitivity analysis for examining the impact of ignored manufacturing factors on a CMS design. Due to the complexity of CF

problems, it is impossible to take into consideration all of the real-life production factors by a single approach. A number of similarity coefficients have been developed in the literature to incorporate different production factors. In this paper, we use three most researched manufacturing factors (alternative process routing l3.1, operation sequence l3.2 and weighted factors l3.3) as the base to perform the taxonomic review study.

Level 4.
Weighted similarity coefficient is a logical extension or expansion of the binary data based similarity coefficient. Merits of the weighted factor based similarity coefficients have been reported by previous studies (Mosier and Taube, 1985b; Mosier, 1989; Seifoddini and Djassemi, 1995). This kind of similarity coefficient attempts to adjust the strength of matches or misses between object pairs to reflect the resemblance value more realistically and accurately by incorporating object attributes.
The taxonomy can be used as an aid to identify and clarify the definition of various similarity coefficients. In the next section, we will review and map similarity coefficients related researches based on this taxonomy.

5. Mapping SCM studies onto the taxonomy

In this section, we map existing similarity coefficients onto the developed taxonomy and review academic studies through 5 tables. Tables 1 and 2 are general-purpose (l1.2) similarity/dissimilarity coefficients, respectively. Table 3 gives expressions of some binary data based (l2.1) similarity coefficients, while table 4 summarizes problem-oriented (l1.1) similarity coefficients. Finally, SCM related academic researches are illustrated in table 5.
Among the similarity coefficients in table 1, eleven of them have been selected by Sarker and Islam (1999) to address the issues relating to the performance of them along with their important characteristics, appropriateness and applications to manufacturing and other related fields. They also presented numerical results to demonstrate the closeness of the eleven similarity and eight dissimilarity coefficients that is presented in table 2. Romesburg (1984) and Sarker (1996) provided detailed definitions and characteristics of these eleven similarity coefficients, namely Jaccard (Romesburg, 1984), Hamann (Holley and Guilford, 1964), Yule (Bishop *et al.*, 1975), Simple matching (Sokal and Michener, 1958), Sorenson (Romesburg, 1984), Rogers and Tanimoto (1960), Sokal and

Sneath (Romesburg, 1984), Rusell and Rao (Romesburg, 1984), Baroni-Urbani and Buser (1976), Phi (Romesburg, 1984), Ochiai (Romesburg, 1984). In addition to these eleven similarity coefficients, table 1 also introduces several other similarity coefficients, namely PSC (Waghodekar and Sahu, 1984), Dot-product, Kulczynski, Sokal and Sneath 2, Sokal and Sneath 4, Relative matching (Islam and Sarker, 2000). Relative matching coefficient is developed recently which considers a set of similarity properties such as no mismatch, minimum match, no match, complete match and maximum match. Table 2 shows eight most commonly used general-purpose (^{l}1.2) dissimilarity coefficients.

Similarity Coefficient	*Definition* S_{ij}	*Range*
1. Jaccard	$a/(a+b+c)$	0-1
2. Hamann	$[(a+d)-(b+c)]/[(a+d)+(b+c)]$	-1-1
3. Yule	$(ad-bc)/(ad+bc)$	-1-1
4. Simple matching	$(a+d)/(a+b+c+d)$	0-1
5. Sorenson	$2a/(2a+b+c)$	0-1
6. Rogers and Tanimoto	$(a+d)/[a+2(b+c)+d]$	0-1
7. Sokal and Sneath	$2(a+d)/[2(a+d)+b+c]$	0-1
8. Rusell and Rao	$a/(a+b+c+d)$	0-1
9. Baroni-Urbani and Buser	$[a+(ad)^{1/2}]/[a+b+c+(ad)^{1/2}]$	0-1
10. Phi	$(ad-bc)/[(a+b)(a+c)(b+d)(c+d)]^{1/2}$	-1-1
11. Ochiai	$a/[(a+b)(a+c)]^{1/2}$	0-1
12. PSC	$a^2/[(b+a)*(c+a)]$	0-1
13. Dot-product	$a/(b+c+2a)$	0-1
14. Kulczynski	$1/2[a/(a+b)+a/(a+c)]$	0-1
15. Sokal and Sneath 2	$a/[a+2(b+c)]$	0-1
16. Sokal and Sneath 4	$1/4[a/(a+b)+a/(a+c)+d/(b+d)+d/(c+d)]$	0-1
17. Relative matching	$[a+(ad)^{1/2}]/[a+b+c+d+(ad)^{1/2}]$	0-1

Table 1. Definitions and ranges of some selected general-purpose similarity coefficients (***l* 1.2**). a: the number of parts visit both machines; b: the number of parts visit machine i but not j; c: the number of parts visit machine j but not i; d: the number of parts visit neither machine

The dissimilarity coefficient does reverse to those similarity coefficients in table 1. In table 2, d_{ij} is the original definition of these coefficients, in order to show the comparison more explicitly, we modify these dissimilarity coefficients and use binary data to express them. The binary data based definition is represented by d_{ij}

Dissamilarity Coefficient	*Definition* d_{ij}	*Range Definition*	d'_{ij}	*Range*
1. Minkowski	$\left(\sum_{k=1}^{M}\lvert a_{ki}-a_{kj}\rvert^{r}\right)^{1/r}$	Real	$(b+c)^{1/r}$	Real
2. Euclidean	$\left(\sum_{k=1}^{M}\lvert a_{ki}-a_{kj}\rvert^{2}\right)^{1/2}$	Real	$(b+c)^{1/2}$	Real
3. Manhattan (City Block)	$\sum_{k=1}^{M}\lvert a_{ki}-a_{kj}\rvert$	Real	$b+c$	0- M
4. Average Euclidean	$\left(\sum_{k=1}^{M}\lvert a_{ki}-a_{kj}\rvert^{2}/M\right)^{1/2}$	Real	$\left(\frac{b+c}{a+b+c+d}\right)^{1/2}$	Real
5. Weighted Minkowski	$\left(\sum_{k=1}^{M}w_k\lvert a_{ki}-a_{kj}\rvert^{r}\right)^{1/r}$	Real	$[w_k(b+c)]^{1/r}$	Real
6. Bray-Curtis	$\sum_{k=1}^{M}\lvert a_{ki}-a_{kj}\rvert / \sum_{k=1}^{M}\lvert a_{ki}+a_{kj}\rvert$	0-1	$\frac{b+c}{2a+b+c}$	0-1
7. Canberra Metric	$\frac{1}{M}\sum_{k=1}^{M}\left(\frac{\lvert a_{ki}-a_{kj}\rvert}{a_{ki}+a_{kj}}\right)$	0-1	$\frac{b+c}{a+b+c+d}$	0-1
8. Hamming	$\sum_{k=1}^{M}\delta(a_{kl},a_{kj})$	0- M	$b+c$	0- M

Table 2. Definitions and ranges of some selected general-purpose dissimilarity coefficients. (l 1.2) $\delta(a_{kl},a_{kj})=\begin{cases}1, \text{if } a_{kl}\neq a_{kj};\\ 0, \text{otherwise.}\end{cases}$; r : a positive integer; d_{ij} : dissimilarity between i *and* j ; d'_{ij} : dissimilarity by using binary data; k : attribute index (k =1,…, M).

Table 3 presents some selected similarity coefficients in group l2.1. The expressions in table 3 are similar to that of table 1. However, rather than judging the similarity between two objects, problem-oriented similarity coefficients evaluate a predetermined “appropriateness” between two objects. Two objects that have the highest "appropriateness" maximize similarity value even if they are less similar than some other object pairs.

Coefficient/Resource	*Definition* S_{ij}	*Range*
1. Chandrasekharan & Rajagopalan (1986b)	$a / Min[(a+b),(a+c)]$	0-1
2. Kusiak *et al.* (1986)	a	integer
3. Kusiak (1987)	$a+d$	integer
4. Kaparthi *et al.* (1993)	$a^{'} /(a+b)^{'}$	0-1
5. MaxSC / Shafer & Rogers (1993b)	$max\,[a/(a+b), a/(a+c)]$	0-1
6. Baker & Maropoulos (1997)	$a / Max[(a+b),(a+c)]$	0-1

Table 3. Definitions and ranges of some selected problem-oriented binary data based similarity coefficients (l 2.1). $a^{'}$ is the number of matching ones between the matching exemplar and the input vector; $(a+b)^{'}$ is the number of ones in the input vector

Table 4 is a summary of problem-oriented (l 1.1) similarity coefficients developed so far for dealing with CF problems. This table is the tabulated expression of the proposed taxonomy. Previously developed similarity coefficients are mapped into the table, additional information such as solution procedures, novel characteristics are also listed in the "*Notes/KeyWords*" column.

Finally, table 5 is a brief description of the published CF studies in conjunction with similarity coefficients. Most studies listed in this table do not develop new similarity coefficients. However, all of them use similarity coefficients as a powerful tool for coping with cell formation problems under various manufacturing situations. This table also shows the broad range of applications of similarity coefficient based methods.

No	*Resource/Coefficient* — *Author(s)/(SC)*	*Year*	*Binary data based* (**l2.1**)	*Production Information* (**l2.2**) — *Alternative Proc.* (**l3.1**)	*Operation sequ.* (**l3.2**)	*Weights* (**l3.3**) — Prod. Vol.(l4.1)	*Oper. Time(l4.2)*	*Others (l4.3)*	*Others (l3.4)*	*Notes/KeyWords*
1	De Witte	1980				Y			MM	3 SC created; Graph theory
2	Waghodekar & Sahu (PSC & SCTF)	1984	Y							*l* 1.2; 2 SC created
3	Mosier & Taube	1985b				Y				2 SC created
4	Selvam & Balasubramanian	1985			Y	Y				Heuristic
5	Chandrasekharan & Rajagopalan	1986b	Y							*l* 2.1; hierarchical algorithm
6	Dutta *et al.*	1986							CS; NC	5 D developed;
7	Faber & Carter (MaxSC)	1986	Y							*l* 2.1; Graph
8	Kusiak *et al.*	1986	Y							*l* 2.1; 3 distinct integer models
9	Kusiak	1987	Y							*l* 2.1; APR by p-median
10	Seifoddini	87/88			Y	Y				
11	Steudel & Ballakur	1987					Y			Dynamic programming
12	Choobineh	1988			Y					Mathematical model
13	Gunasingh & Lashkari	1989						T		Math.; Compatibility index
14	Wei & Kern	1989	Y							*l* 2.1; Heuristic
15	Gupta & Seifoddini	1990			Y	Y	Y			Heuristic

Table 4. Summary of developed problem-oriented (dis)similarity coefficients (SC) for cell formation (*l* **1.1**)

No	*Resource/Coefficient*		*Binary data based (l2.1)*	*Production Information (l2.2)*						*Notes/KeyWords*
	Author(s)/(SC)	*Year*		*Alternative Proc. (l3.1)*	*Operation sequ. (l3.2)*	*Weights (l3.3)* Prod. Vol.(l4.1)	*Oper. Time(l4.2)*	*Others (l4.3)*	*Others (l3.4)*	
16	Tam	1990			Y					*k* Nearest Neighbour
17	Vakharia & Wemmerlöv	1987 ; 1990			Y					Heuristic
18	Offodile	1991						Y		Parts coding and classification
19	Kusiak & Cho	1992	Y							*l* 2.1; 2 SC proposed
20	Zhang & Wang	1992							Y	Combine SC with fuzziness
21	Balasubramanian & Panneerselvam	1993			Y	Y			M H C	D; covering technique
22	Ho *et al.*	1993			Y					Compliant index
23	Gupta	1993		Y	Y	Y	Y			Heuristic
24	Kaparthi *et al.*	1993	Y							*l* 2.1; Improved neural network
25	Luong	1993							C S	Heuristic
26	Ribeiro & Pradin	1993	Y							D, *l* 1.2; Knapsack
27	Seifoddini & Hsu	1994						Y		Comparative study
28	Akturk & Balkose	1996			Y					D; multi objective model
29	Ho & Moodie (POSC)	1996							F P R	Heuristic; Mathematical
30	Ho & Moodie (GOSC)	1996				Y				SC between two part groups
31	Suer & Cedeno	1996						C		
32	Viswanathan	1996	Y							*l* 2.1; modify p-median

Table 4 (continued)

No	*Resource/Coefficient*		*Binary data based* (**l2.1**)	*Production Information* (**l2.2**)						*Notes/KeyWords*
	Author(s)/(SC)	*Year*		*Alternative Proc.* (**l3.1**)	*Operation sequ.* (**l3.2**)	*Weights* (**l3.3**) *Prod. Volu.* (**l4.1**)	*Oper. Time* (**l4.2**)	*Others* (**l4.3**)	*Others* (**l3.4**)	
33	Baker & Maropoulos	1997	Y							*l* 2.1; Black box algorithm
34	Lee *et al.*	1997			Y	Y				APR by genetic algorithm
35	Won & Kim	1997		Y						Heuristic
36	Askin & Zhou	1998			Y					Shortest path
37	Nair & Narendran	1998			Y					Non-hierarchical
38	Jeon *et al.*	1998 b		Y						Mathematical
39	Kitaoka *et al.* (Double Centering)	1999	Y							*l* 2.1; quantification model
40	Nair & Narendran	1999							WL	Mathematical; Non-hierarchical
41	Nair & Narendran	1999			Y	Y			WL	Mathematical; Non-hierarchical
42	Seifoddini & Tjahjana	1999							BS	
43	Sarker & Xu	2000			Y					3 phases algorithm
44	Won	2000 a		Y						Modify p-median
45	Yasuda & Yin	2001							CS	D; Heuristic

Table 4 (continued). Summary of developed problem-oriented (dis)similarity coefficients (SC) for cell formation (*l* **1.1**)

APR: Alternative process routings; *BS*: Batch size; *C*: Cost of unit part, *CS*: cell size; *D*: dissimilarity coefficient; *FPR*: Flexible processing routing, *MHC*: Material handling cost; *MM*: Multiple machines available for a machine type, *NC*: number of cell; *SC*:

Similarity coefficient; T: Tooling requirements of parts, *WL*: Workload

Articles		***Similarity coefficients (SC) used***	***Description/Keywords***
Author(s)	*Year*		
McAuley	1972	Jaccard	First study of SC on cell formation
Carrie	1973	Jaccard	Apply SC on forming part families
Rajagopalan & Batra	1975	Jaccard	Graph theory
Waghodekar & Sahu	1984	Jaccard; PSC; SCTF	Propose MCSE method
Kusiak	1985	Minkowski (D)	p-median; heuristics
Chandrasekharan & Rajagopalan	1986a	Minkowski (D)	Non-hierarchical algorithm
Han & Ham	1986	Manhattan (D)	Classification and coding system
Seifoddini & Wolfe	1986	Jaccard	Bit-level data storage technique
Chandrasekharan & Rajagopalan	1987	Manhattan (D)	Develop ZODIAC algorithm
Marcotorchino	1987	Jaccard; Sorenson	Create a block seriation model
Seifoddini & Wolfe	1987	Jaccard	Select threshold on material handling cost
Chandrasekharan& Rajagopalan	1989	Jaccard; Simple matching; Manhattan (D)	An analysis of the properties of data sets
Mosier	1989	7 similarity coefficients	Comparative study
Seifoddini	1989a	Jaccard	SLC vs. ALC
Seifoddini	1989b	Jaccard	Improper machine assignment
Srinivasan *et al.*	1990	Kusiak (1987)	An assignment model
Askin *et al.*	1991	Jaccard	Hamiltonian path; TSP
Chow	1991	CS	Unjustified claims of LCC
Gongaware & Ham	1991	---*	Classification & coding; multi-objective model
Gupta	1991	Gupta & Seifoddini (1990)	Comparative study on chaining effect
Logendran	1991	Jaccard; Kusiak (1987)	Identification of key machine

Srinivasan & Narendran	1991	Kusiak (1987)	A nonhierarchical clustering algorithm
Wei & Kern	1991	CS	Reply to Chow (1991)
Chow & Hawaleshka	1992	CS	Define machine unit concept
Shiko	1992	Jaccard	Constrained hierarchical
Chow & Hawaleshka	1993a	CS	Define machine unit concept
Chow & Hawaleshka	1993b	CS	A knowledge-based approach
Kang & Wemmerlöv	1993	Vakharia & Wemmerlov (87,90)	Heuristic; Alternative operations of parts
Kusiak *et al.*	1993	Hamming (D)	Branch-Bound & A* approaches
Offodile	1993	Offodile (1991)	Survey of robotics & GT; robot selection model
Shafer & Rogers	1993a	Many	Review of similarity coefficients
Shafer & Rogers	1993b	16 similarity coefficients	Comparative study
Vakharia & Kaku	1993	Kulczynski	Long-term demand change
Ben-Arieh & Chang	1994	Manhattan (D)	Modify p-median algorithm
Srinivasan	1994	Manhattan (D)	Minimum spanning trees
Balakrishnan & Jog	1995	Jaccard	TSP algorithm
Cheng *et al.*	1995	Hamming (D)	Quadratic model; A* algorithm
Kulkarni & Kiang	1995	Euclidean (D)	Self-organizing neural network
Murthy & Srinivasan	1995	Manhattan (D)	Heuristic; Consider fractional cell formation
Seifoddini & Djassemi	1995	Jaccard	Merits of production volume consideration
Vakharia & Wemmerlöv	1995	8 dissimilarity coefficients	Comparative study
Wang & Roze	1995	Jaccard, Kusiak (1987), CS	An experimental study
Balakrishnan	1996	Jaccard	CRAFT
Cheng *et al.*	1996	Hamming (D)	Truncated tree search algorithm

Hwang & Ree	1996	Jaccard	Define compatibility coefficient
Lee & Garcia-Diaz	1996	Hamming (D)	Use a 3-phase network-flow approach
Leem & Chen	1996	Jaccard	Fuzzy set theory
Lin *et al.*	1996	Bray-Curtis (D)	Heuristic; workload balance within cells
Sarker	1996	Many	Review of similarity coefficient
Al-sultan & Fedjki	1997	Hamming (D)	Genetic algorithm
Askin *et al*	1997	MaxSC	Consider flexibility of routing and demand
Baker & Maropoulos	1997	Jaccard, Baker & Maropoulos (1997)	Black Box clustering algorithm
Cedeno & Suer	1997	---	Approach to “remainder clusters”
Masnata & Settineri	1997	Euclidean (D)	Fuzzy clustering theory
Mosier *et al.*	1997	Many	Review of similarity coefficients
Offodile & Grznar	1997	Offodile (1991)	Parts coding and classification analysis
Wang & Roze	1997	Jaccard, Kusiak (1987), CS	Modify p-median model
Cheng *et al.*	1998	Manhattan (D)	TSP by genetic algorithm
Jeon *et al.*	1998a	Jeon *et al.* (1998b)	p-median
Onwubolu & Mlilo	1998	Jaccard	A new algorithm (SCDM)
Srinivasan & Zimmers	1998	Manhattan (D)	Fractional cell formation problem
Wang	1998	---	A linear assignment model
Ben-Arieh & Sreenivasan	1999	Euclidean (D)	A distributed dynamic algorithm
Lozano *et al.*	1999	Jaccard	Tabu search
Sarker & Islam	1999	Many	Performance study
Baykasoglu & Gindy	2000	Jaccard	Tabu search
Chang & Lee	2000	Kusiak (1987)	Multi-solution heuristic
Josien & Liao	2000	Euclidean (D)	Fuzzy set theory
Lee-post	2000	Offodile (1991)	Use a simple genetic algorithm
Won	2000a	Won & Kim(1997)	Alternative process plan with p-median model

Won	2000b	Jaccard, Kusiak (1987)	Two-phase p-median model
Dimopoulos & Mort	2001	Jaccard	Genetic algorithm
Samatova *et al.*	2001	5 dissimilarity coefficients	Vector perturbation approach

Table 5. Literature of cell formation research in conjunction with similarity coefficients (SC). *: no specific SC mentioned

6. General discussion

We give a general discussion of production information based similarity coefficients (*l* **2.2**) and an evolutionary timeline in this section.

6.1. Production information based similarity coefficients

6.1.1 Alternative process routings

In most cell formation methods, parts are assumed to have a unique part process plan. However, it is well known that alternatives may exist in any level of a process plan. In some cases, there may be many alternative process plans for making a specific part, especially when the part is complex (Qiao *et al*. 1994). Explicit consideration of alternative process plans invoke changes in the composition of all manufacturing cells so that lower capital investment in machines, more independent manufacturing cells and higher machine utilization can be achieved (Hwang and Ree 1996).
Gupta (1993) is the first person who incorporated alternative process routings into similarity coefficient. His similarity coefficient also includes other production information such as operation sequences, production volumes and operation times. The similarity coefficient assigns pairwise similarity among machines with usage factors of all alternative process routings. The usage factors are determined by satisfying production and capacity constraints. The production volumes and operation times are assumed to be known with certainty.
An alternative process routings considered similarity coefficient was developed by Won and Kim (1997) and slightly modified by Won (2000a). In the definition of the similarity coefficient, if machine i is used by some process routing of part j, then the number of parts processed by machine i is counted as one for that part even if the remaining process routings of part j also use

machine i. The basic idea is that in the final solution only one process routing is selected for each part. p-median approach was used by Won (2000a) to associate the modified similarity coefficient.
A similarity coefficient that considers the number of alternative process routings when available during machine failure is proposed by Jeon *et al.* (1998b). The main characteristic of the proposed similarity coefficient is that it draws on the number of alternative process routings during machine failure when alternative process routings are available instead of drawing on operations, sequence, machine capabilities, production volumes, processing requirements and operational times. Based on the proposed similarity coefficient, p-median approach was used to form part families.

6.1.2 Operation sequences

The operation sequence is defined as an ordering of the machines on which the part is sequentially processed (Vakharia and Wemmerlov 1990). A lot of similarity coefficients have been developed to consider operation sequence.
Selvam and Balasubramanian (1985) are the first persons who incorporated alternative process routings into similarity coefficient. Their similarity coefficient is very simple and intuitive. The value of similarity coefficient is determined directly by the production volume of parts moves between machines.
Seifoddini (1987/1988) modified Jaccard similarity coefficient to take into account the production volume of parts moves between machine pairs. A simple heuristic algorithm was used by the author to form machine cells. Choobineh (1988) gave a similarity coefficient between parts j and k which is based on the common sequences of length 1 through L between the two parts. To select the value L, one has to balance the need to uncover the natural strength of the relationships among the parts and the computational efforts necessary to calculate the sequences of length 1 through L. In general, the higher the value of L, the more discriminating power similarity coefficient will have.Gupta and Seifoddini (1990) proposed a similarity coefficient incorporating operation sequence, production volume and operation time simultaneously. From the definition, each part that is processed by at least one machine from a pair of machines contributes towards their similarity coefficient value. A part that is processed by both machines increases the coefficient value for the two machines whereas, a part that is processed on one machine tends to reduce it. The similarity coefficient developed by Tam (1990) is based on Levenshtein's distance measure of two sentences. The distance between two sentences is defined

as the minimum number of transformations required to derive one sentence from the other. Three transformations are defined. The similarity coefficient between two operation sequences x and y, is defined as the smallest number of transformations required to derive y from x. Vakharia and Wemmerlov (1990) proposed a similarity coefficient based on operation sequences to integrate the intracell flow with the cell formation problem by using clustering methodology. The similarity coefficient measures the proportion of machine types used by two part families in the same order.

Balasubramanian and Panneerselvam (1993) developed a similarity coefficient which needs following input data: (1) operation sequences of parts; (2) additional cell arrangements; (3) production volume per day and the bulk factor; (4) guidelines for computing excess moves; (5) actual cost per move.

Ho *et al*. (1993)'s similarity coefficient calculates a compliant index firstly. The compliant index of the sequence of a part compared with a flow path is determined by the number of operations in the sequence of the part that have either "in-sequence" or "by-passing" relationship with the sequence of the flow path. There are two kinds of compliant indexes: forward compliant index and backward index. These two compliant indexes can be calculated by comparing the operation sequence of the part with the sequence of the flow path forwards and backwards.As mentioned in 6.1.1, Gupta (1993) proposed a similarity coefficient which incorporates several production factors such as operation sequences, production volumes, alternative process routings.

Akturk and Balkose (1996) revised the Levenshtein distance measure to penalize the backtracking parts neither does award the commonality. If two parts have no common operations, then a dissimilarity value is found by using the penalizing factor.Lee *et al*. (1997)'s similarity coefficient takes the direct and indirect relations between the machines into consideration. The direct relation indicates that two machines are connected directly by parts; whereas the indirect relation indicates that two machines are connected indirectly by other machines. Askin and Zhou (1998) proposed a similarity coefficient which is based on the longest common operation subsequence between part types and used to group parts into independent, flow-line families.

Nair and Narendran (1998) gave a similarity coefficient as the ratio of the sum of the moves common to a pair of machines and the sum of the total number of moves to and from the two machines. Latterly, They extended the coefficient to incorporate the production volume of each part (Nair and Narendran, 1999). Sarker and Xu (2000) developed an operation sequence-based similarity coefficient. The similarity coefficient was applied in a p-median model to group the parts to form part families with similar operation sequences.

6.1.3 Weight factors

Weighted similarity coefficient is a logical extension or expansion of the binary data based similarity coefficient. Two most researched weight factors are production volume and operation time.

De Witte (1980) is the first person who incorporated production volume into similarity coefficient. In order to analyse the relations between machine types three different similarity coefficients has be used by the author. Absolute relations, mutual interdependence relations and relative single interdependence relations between machine pairs are defined by similarity coefficients *SA*, *SM* and *SS*, respectively.

Mosier and Taube (1985b)'s similarity coefficient is a simple weighted adaptation of McAuley's Jaccard similarity coefficient with an additional term whose purpose is to trap the coefficient between -1.0 and +1.0. Production volumes of parts have been incorporated into the proposed similarity coefficient.

Ho and Moodie (1996) developed a similarity coefficient, namely group-operation similarity coefficient (*GOSC*) to measure the degree of similarity between two part groups. The calculation of *GOSC* considers the demand quantities of parts. A part with a larger amount of demand will have a heavier weight. This is reasonable since if a part comprises the majority of a part group, then it should contribute more in the characterization of the part group it belongs to.

The operation time is considered firstly by Steudel and Ballakur (1987). Their similarity coefficient is based on the Jaccard similarity coefficient and calculates the operation time by multiplying each part's operation time by the production requirements for the part over a given period of time. Operation set-up time is ignored in the calculation since set-up times can usually be reduced after the cells are implemented. Hence set-up time should not be a factor in defining the cells initially.

Other production volume / operation time considered studies include Selvam and Balasubramanian (1985), Seifoddini (1987/1988), Gupta and Seifoddini (1990), Balasubramanian and Panneerselvam (1993), Gupta (1993), Lee *et al.* (1997) and Nair and Narendran (1999). Their characteristics have been discussed in sections 6.1.1 and 6.1.2.

6.2 Historical evolution of similarity coefficients

Shafer and Rogers (1993a) delineated the evolution of similarity coefficients until early 1990s. Based on their work and table 4, we depict the historical evolution of similarity coefficients over the last three decades.

McAuley (1972) was the first person who used the Jaccard similarity coefficient to form machine cells. The first weighted factor that was considered by researchers is the production volume of parts (De Witte, 1980; Mosier and Taube, 1985b). Operation sequences, one of the most important manufacturing factors, was incorporated in 1985 (Selvam and Balasubramanian). In the late 1980s and early 1990s, other weighted manufacturing factors such as tooling requirements (Gunasingh and Lashkari, 1989) and operation times (Gupta and Seifoddini, 1990) were taken into consideration. Alternative process routings of parts is another important manufacturing factor in the design of a CF system. Although it was firstly studied by Kusiak (1987), it was not combined into the similarity coefficient definition until Gupta (1993).

Material handling cost was also considered in the early 1990s (Balasubramanian and Panneerselvam, 1993). In the middle of 1990s, flexible processing routings (Ho and Moodie, 1996) and unit cost of parts (Sure and Cedeno, 1996) were incorporated.

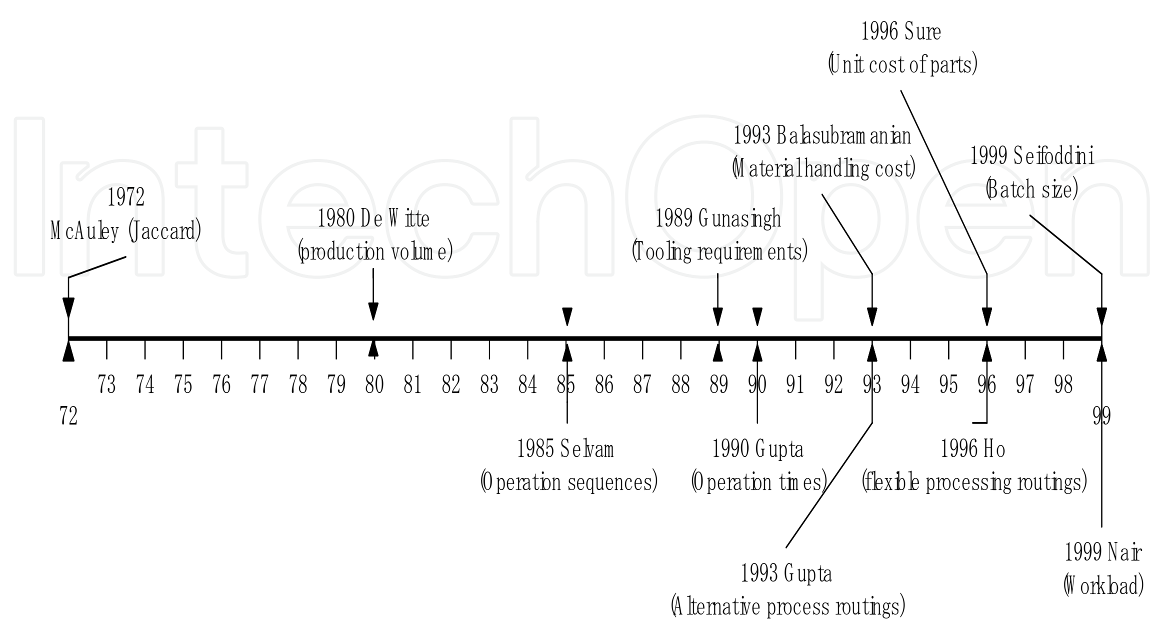

Figure 4. The similarity coefficient's evolutionary timeline

Finally, some impressive progresses that have been achieved in the late 1990s were workload (Nair and Narendran, 1999) and batch size (Seifoddini and Tjahjana, 1999) consideration in the definition of similarity coefficients.
The similarity coefficient's evolutionary timeline is given in figure 4.

7. Comparative study

7.1 The objective of the comparison

Although a large number of similarity coefficients exist in the literature, only a handful has been used for solving CF problems. Among various similarity coefficients, Jaccard similarity coefficient (Jaccard, 1908) was the most used similarity coefficient in the literature (Table 5). However, contradictory viewpoints among researchers have been found in the previous studies: some researchers advocated the dominant power of Jaccard similarity coefficient; whereas some other researchers emphasized the drawbacks of Jaccard similarity coefficient and recommended other similarity coefficients; moreover, several researchers believed that there is no difference between Jaccard and other similarity coefficients, they considered that none of the similarity coefficients seems to perform always well under various cell formation situations.
Therefore, a comparative research is crucially necessary to evaluate various similarity coefficients. Based on the comparative study, even if we cannot find a dominant similarity coefficient for all cell formation situations, at least we need to know which similarity coefficient is more efficient and more appropriate for some specific cell formation situations.
In this paper, we investigate the performance of twenty well-known similarity coefficients. A large number of numerical data sets, which are taken from the open literature or generated specifically, are tested on nine performance measures.

7.2 Previous comparative studies

Four studies that have focused on comparing various similarity coefficients and related cell formation procedures have been published in the literature.
Mosier (1989) applied a mixture model experimental approach to compare seven similarity coefficients and four clustering algorithms. Four performance measures were used to judge the goodness of solutions: simple matching

measure, generalized matching measure, product moment measure and intercellular transfer measure. As pointed out by Shafer and Rogers (1993), the limitation of this study is that three of the four performance measures are for measuring how closely the solution generated by the cell formation procedures matched the original machine-part matrix. However, the original machine-part matrix is not necessarily the best or even a good configuration. Only the last performance measure, intercellular transfer measure is for considering specific objectives associated with the CF problem.

Shafer and Rogers (1993) compared sixteen similarity coefficients and four clustering algorithms. Four performance measures were used to evaluate the solutions. Eleven small, binary machine-part group technology data sets mostly from the literature were used for the purpose of comparison. However, small and/or "well-structured" data sets may not have sufficient discriminatory power to separate "good" from "inferior" techniques. Further, results based on a small number of data sets may have little general reliability due to clustering results' strong dependency on the input data (Vakharia & Wemmerlöv, 1995; Milligan & Cooper, 1987; Anderberg, 1973).

Seifoddini and Hsu (1994) introduced a new performance measure: grouping capability index (GCI). The measure is based on exceptional elements and has been widely used in the subsequent researches. However, only three similarity coefficients have been tested in their study.

Vakharia and Wemmerlöv (1995) studied the impact of dissimilarity measures and clustering techniques on the quality of solutions in the context of cell formation. Twenty-four binary data sets were solved to evaluate eight dissimilarity measures and seven clustering algorithms. Some important insights have been provided by this study, such as data set characteristics, stopping parameters for clustering, performance measures, and the interaction between dissimilarity coefficients and clustering procedures. Unfortunately, similarity coefficients have not been discussed in this research.

8. Experimental design

8.1 Tested similarity coefficients

*T*wenty well-known similarity coefficients (Table 6) are compared in this paper. Among these similarity coefficients, several of them have never been studied by previous comparative researches.

Coefficient	**Definition** S_{ij}	**Range**
1. Jaccard	$a/(a+b+c)$	0-1
2. Hamann	$[(a+d)-(b+c)]/[(a+d)+(b+c)]$	-1-1
3. Yule	$(ad-bc)/(ad+bc)$	-1-1
4. Simple matching	$(a+d)/(a+b+c+d)$	0-1
5. Sorenson	$2a/(2a+b+c)$	0-1
6. Rogers and Tanimoto	$(a+d)/[a+2(b+c)+d]$	0-1
7. Sokal and Sneath	$2(a+d)/[2(a+d)+b+c]$	0-1
8. Rusell and Rao	$a/(a+b+c+d)$	0-1
9. Baroni-Urbani and Buser	$[a+(ad)^{1/2}]/[a+b+c+(ad)^{1/2}]$	0-1
10. Phi	$(ad-bc)/[(a+b)(a+c)(b+d)(c+d)]^{1/2}$	-1-1
11. Ochiai	$a/[(a+b)(a+c)]^{1/2}$	0-1
12. PSC	$a^2/[(b+a)*(c+a)]$	0-1
13. Dot-product	$a/(b+c+2a)$	0-1
14. Kulczynski	$1/2[a/(a+b)+a/(a+c)]$	0-1
15. Sokal and Sneath 2	$a/[a+2(b+c)]$	0-1
16. Sokal and Sneath 4	$1/4[a/(a+b)+a/(a+c)+d/(b+d)+d/(c+d)]$	0-1
17. Relative matching	$[a+(ad)^{1/2}]/[a+b+c+d+(ad)^{1/2}]$	0-1
18. Chandrasekharan & Rajagopalan (1986b)	$a/Min[(a+b),(a+c)]$	0-1
19. MaxSC	$Max[a/(a+b),a/(a+c)]$	0-1
20. Baker & Maropoulos (1997)	$a/Max[(a+b),(a+c)]$	0-1

Table 6: Definitions and ranges of selected similarity coefficients a: the number of parts visit both machines; b: the number of parts visit machine i but not j; c: the number of parts visit machine j but not i; d: the number of parts visit neither machine.

8.2 Data sets

It is desirable to judge the effectiveness of various similarity coefficients under varying data sets conditions. The tested data sets are classified into two distinct groups: selected from the literature and generated deliberately. Previous comparative studies used either of them to evaluate the performance of various similarity coefficients. Unlike those studies, this paper uses both types of the data sets to evaluate twenty similarity coefficients.

8.2.1 Data sets selected from the literature

In the previous comparative studies, Shafer and Rogers (1993), and Vakharia and Wemmerlöv (1995) took 11 and 24 binary data sets from the literature, respectively. The advantage of the data sets from the literature is that they stand for a variety of CF situations. In this paper, 70 data sets are selected from the literature. Table 7 shows the details of the 70 data sets.

8.2.2 Data sets generated deliberately

From the computational experience with a wide variety of CF data sets, one finds that it may not always be possible to obtain a good GT solution, if the original CF problem is not amenable to well-structural data set (Chandrasekharan & Rajagopalan, 1989). Hence, it is important to evaluate the quality of solutions of various structural data sets. Using data sets that are generated deliberately is a shortcut to evaluate the GT solutions obtained by various similarity coefficients. The generation process of data sets is often controlled by using experimental factors. In this paper, we use two experimental factors to generate data sets.

- Ratio of non-zero Element in Cells (REC)

Density is one of the most used experimental factors (Miltenburg & Zhang, 1991). However, in our opinion, density is an inappropriate factor for being used to control the generation process of cell formation data sets. We use following Fig.5 to illustrate this problem.
Cell formation data are usually presented in a machine-part incidence matrix such as Fig.5a. The matrix contains 0s and 1s elements that indicate the machine requirements of parts (to show the matrix clearly, 0s are usually unshown). Rows represent machines and columns represent parts.
A '1' in the i^{th} row and j^{th} column represents that the j^{th} part needs an operation on the i^{th} machine; similarly, a '0' in the i^{th} row and j^{th} column represents that the i^{th} machine is not needed to process the j^{th} part.
For Fig.5a, we assume that two machine-cells exist. The first cell is constructed by machines 2 ,4, 1 and parts 1, 3, 7, 6, 10; The second cell is constructed by machines 3, 5 and parts 2, 4, 8, 9, 5, 11. Without loss of generality, we use Fig.5b to represent Fig.5a. The two cells in Fig.5a are now shown as capital letter 'A', we call 'A' as the inside cell region. Similarly, we call 'B' as the outside cell region.

There are three densities that are called Problem Density (PD), non-zero elements Inside cells Density (ID) and non-zero elements Outside cells Density (OD). The calculations of these densities are as follows:

$$\text{PD}=\frac{\textit{total number of non-zero elements in regions } A+B}{\textit{total number of elements in regions } A+B} \tag{7}$$

$$\text{ID}=\frac{\textit{total number of non-zero elements in regions } A}{\textit{total number of elements in regions } A} \tag{8}$$

$$\text{OD}=\frac{\textit{total number of non-zero elements in regions } B}{\textit{total number of elements in regions } B} \tag{9}$$

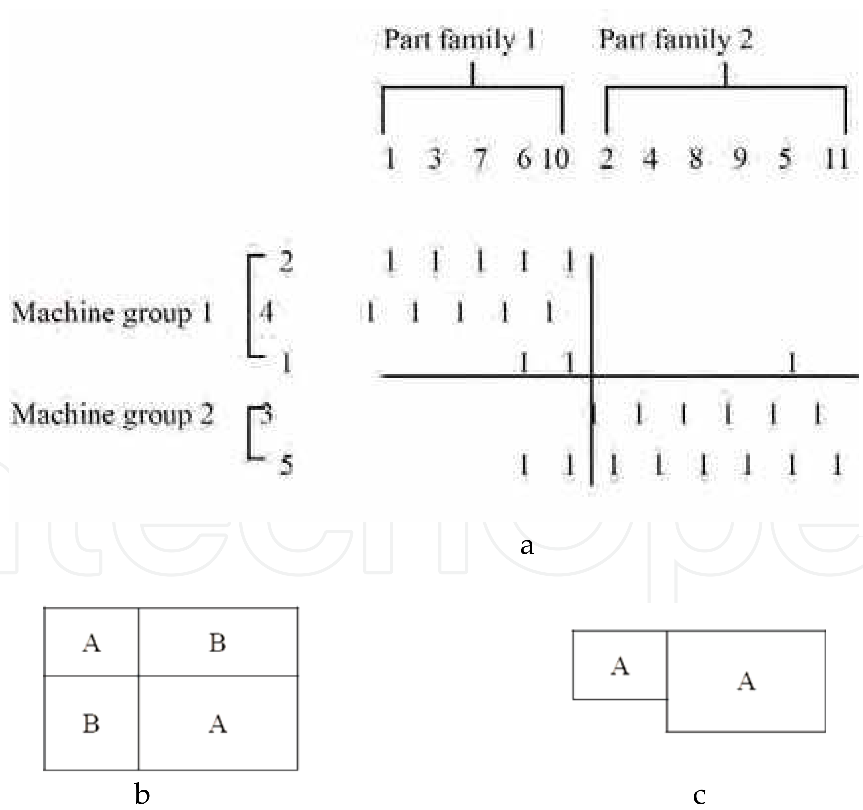

Figure 5. Illustration of three densities used by previous studies

In the design of cellular manufacturing systems, what we concerned is to find out appropriate machine-part cells – the region A. In practice, region B is only

a virtual region that does not exist in the real job shops. For example, if Fig.5a is applied to a real-life job shop, Fig.5c is a possible layout. There is no region B exists in the real-life job shop. Therefore, we conclude that region B based densities are meaningless. Since PD and OD are based on B, this drawback weakens the quality of generated data sets in the previous comparative studies.
To overcome the above shortcoming, we introduce a ratio to replace the density used by previous researchers. The ratio is called as Ratio of non-zero Element in Cells (REC) and is defined as follows:

$$REC = \frac{total\ number\ of\ non-zero\ elements}{total\ number\ of\ elements\ in\ region\ A} \qquad (10)$$

The definition is intuitive. REC can also be used to estimates the productive capacity of machines. If REC is bigger than 1, current machine capacity cannot response to the productive requirements of parts. Thus, additional machines need to be considered. Therefore, REC can be used as a sensor to assess the capacity of machines.

- Radio of Exceptions (RE)

The second experimental factor is Radio of Exceptions (RE). An exception is defined as a '1' in the region B (an operation outside the cell). We define RE as follows:

$$RE = \frac{total\ number\ of\ non-zero\ elements\ in\ region\ B}{total\ number\ of\ non-zero\ elements} \qquad (11)$$

RE is used to judge the "goodness" of machine-part cells and distinguish well-structured problems from ill-structured problems.
In this paper, 3 levels of REC, from sparse cells (0.70) to dense cells (0.90), and 8 levels of RE, from well-structured cells (0.05) to ill-structured cells (0.40), are examined. 24 (3*8) combinations exist for all levels of the two experimental factors. For each combination, five 30*60-sized (30 machines by 60 parts) problems are generated. The generation process of the five problems is similar by using the random number. Therefore, a total of 120 test problems for all 24 combines are generated, each problem is made up of 6 equally sized cells. The levels of REC and RE are shown in Table 8.

	Data set	**Size**	**NC**
1.	Singh & Rajamani	1996	4*4 2
2.	Singh & Rajamani	1996	4*5 2
3.	Singh & Rajamani	1996	5*6 2
4.	Waghodekar& Sahu	1984	5*7 2
5.	Waghodekar& Sahu	1984	5*7 2
6.	Chow & Hawaleshka	1992	5*11 2
7.	Chow & Hawaleshka	1993a	5*13 2
8.	Chow & Hawaleshka	1993b	5*13 2
9.	Seifoddini	1989b	5*18 2
10.	Seifoddini	1989b	5*18 2
11.	Singh & Rajamani	1996	6*8 2
12.	Chen *et al*.	1996	7*8 3
13.	Boctor	1991	7*11 3
14.	Islam & Sarker	2000	8*10 3
15.	Seifoddini & Wolfe	1986	8*12 3
16.	Chandrasekharan & Rajagopalan	1986a	8*20 2, 3
17.	Chandrasekharan & Rajagopalan	1986b	8*20 2,3
18.	Faber & Carter	1986	9*9 2
19.	Seifoddini & Wolfe	1986	9*12 3
20.	Chen *et al*.	1996	9*12 3
21.	Hon & Chi	1994	9*15 3
22.	Selvam & Balasubramanian	1985	10*5 2
23.	Mosier & Taube	1985a	10*10 3
24.	Seifoddini & Wolfe	1986	10*12 3
25.	McAuley	1972	12*10 3
26.	Seifoddini	1989a	11*22 3
27.	Hon & Chi	1994	11*22 3
28.	De Witte	1980	12*19 2, 3
29.	Irani & Khator	1986	14*24 4
30.	Askin & Subramanian	1987	14*24 4
31.	King	1980(machine 6, 8removed)	14*43 4, 5
32.	Chan & Milner	1982	15*10 3
33.	Faber & Carter	1986	16*16 2, 3
34.	Sofianopoulou	1997	16*30 2, 3
35.	Sofianopoulou	1997	16*30 2, 3
36.	Sofianopoulou	1997	16*30 2, 3
37.	Sofianopoulou	1997	16*30 2, 3

38.	Sofianopoulou	1997	16*30 2, 3
39.	Sofianopoulou	1997	16*30 2, 3
40.	Sofianopoulou	1997	16*30 2, 3
41.	Sofianopoulou	1997	16*30 2, 3
42.	Sofianopoulou	1997	16*30 2, 3
43.	Sofianopoulou	1997	16*30 2, 3
44.	King	1980	16*43 4, 5
45.	Boe & Cheng	1991 (mach 1 removed)	19*35 4
46.	Shafer & Rogers	1993	20*20 4
47.	Shafer & Rogers	1993	20*20 4
48.	Shafer & Rogers	1993	20*20 4
49.	Mosier & Taube	1985b	20*20 3, 4
50.	Boe & Cheng	1991	20*35 4
51.	Ng	1993	20*35 4
52.	Kumar & Kusiak	1986	23*20 2, 3
53.	McCormick *et al.* 1	972	24*16 6
54.	Carrie	1973	24*18 3
55.	Chandrasekharan & Rajagopalan	1989	24*4 7
56.	Chandrasekharan & Rajagopalan 1	989	24*40 7
57.	Chandrasekharan & Rajagopalan	1989	24*40 7
58.	Chandrasekharan & Rajagopalan	1989	24*40 7
59.	Chandrasekharan & Rajagopalan	1989	24*40 7
60.	Chandrasekharan & Rajagopalan	1989	24*40 7
61.	Chandrasekharan & Rajagopalan	1989	24*40 7
62.	McCormick *et al.*	1972	27*27 8
63.	Carrie	1973	28*46 3, 4
64.	Lee *et al.*	1997	30*40 6
65.	Kumar & Vannelli	1987	30*41 2,3,9
66.	Balasubramanian & Panneerselvam	1993	36*21 7
67.	King & Nakornchai	1982	36*90 4, 5
68.	McCormick *et al.*	1972	37*53 4,5,6
69.	Chandrasekharan & Rajagopalan	1987	40*100 10
70.	Seifoddini & Tjahjana	1999	50*22 14

Table 7. Data sets from literature

Level	1	2	3	4	5	6	7	8
REC	0.70	0.80	0.90	--	--	--	--	--
RE	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40

Table 8. Test levels of REC and RE

8.3 Clustering procedure

The most well-known clustering procedures that have been applied to cell formation are single linkage clustering (SLC) algorithm, complete linkage clustering (CLC) algorithm and average linkage clustering (ALC) algorithm. These three procedures have been investigated by lots of studies. A summary of the past comparative results is shown in Table 9.

Due to that ALC has the advantage of showing the greatest robustness regardless of similarity coefficients, in this paper, we select ALC as the clustering algorithm to evaluate the twenty similarity coefficients (Table 6).

Procedure	Advantage	Drawback
SLC	Simplicity; Minimal computational requirement; Tends to minimize the degree of adjusted machine duplication. (Vakharia & Wemmerlöv, 1995).	Largest tendency to chain; Leads to the lowest densities and the highest degree of single part cells (Seifoddini, 1989a; Gupta, 1991; Vakharia & Wemmerlöv, 1995).
CLC	Simplicity; Minimal computational requirement (does the reverse of SLC)	Performed as the worst procedure (Vakharia & Wemmerlöv, 1995; Yasuda & Yin, 2001).
ALC	Performed as the best procedure; Produces the lowest degree of chaining; Leads to the highest cell densities; Indifferent to choice of similarity coefficients; Few single part cells (Tarsuslugil & Bloor, 1979; Seifoddini, 1989a; Vakharia & Wemmerlöv, 1995; Yasuda & Yin, 2001).	Requires the highest degree of machine duplication; Requires more computation (Vakharia & Wemmerlöv, 1995).

Table 9. Comparative results of SLC, ALC and CLC

The ALC algorithm usually works as follows:

Step (1). Compute similarity coefficients for all machine pairs and store the values in a similarity matrix.

Step (2). Join the two most similar objects (two machines, a machine and a machine group or two machine groups) to form a new machine group.

Step (3). Evaluate the similarity coefficient between the new machine group and other remaining machine groups (machines) as follows:

$$S_{tv} = \frac{\sum_{i \in t} \sum_{j \in v} S_{ij}}{N_t N_v} \qquad (12)$$

where i is the machine in the machine group t; j is the machine in the machine group v. And N_t is the number of machines in group t ; N_v is the number of machines in group v .

Step (4). When all machines are grouped into a single machine group, or pre determined number of machine groups has obtained, go to step 5; otherwise, go back to step 2.

Step (5). Assign each part to the cell, in which the total number of exceptions is minimum.

8.4 Performance measures

A number of quantitative performance measures have been developed to evaluate the final cell formation solutions. Sarker and Mondal (1999) reviewed and compared various performance measures.

Nine performance measures are used in this study to judge final solutions. These measures provide different viewpoints by judging solutions from different aspects.

8.4.1 Number of exceptional elements (EE)

Exceptional elements are the 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.

8.4.2 Grouping efficiency

Grouping efficiency is one of the first measures developed by Chandrasekharan and Rajagopalan (1986a, b). Grouping efficiency is defined as a weighted average of two efficiencies η_1 and η_2:

$$\eta = w\eta_1 + (1-w)\eta_2 \tag{13}$$

where

$$\eta_1 = \frac{o-e}{o-e+v}$$

$$\eta_2 = \frac{MP-o-v}{MP-o-v+e}$$

M number of machines
P number of parts
o number of operations (1s) in the machine-part matrix $\{a_{ik}\}$
e number of exceptional elements in the solution
v number of voids in the solution

A value of 0.5 is recommended for w. η_1 is defined as the ratio of the number of 1s in the region A (Fig.5b) to the total number of elements in the region A (both 0s and 1s). Similarly, η_2 is the ratio of the number of 0s in the region B to the total number of elements in the region B (both 0s and 1s). The weighting factor allows the designer to alter the emphasis between utilization and intercell movement. The efficiency ranges from 0 to 1.
Group efficiency has been reported has a lower discriminating power (Chandrasekharan & Rajagopalan, 1987). Even an extremely bad solution with large number of exceptional elements has an efficiency value as high as 0.77.

8.4.3 Group efficacy

To overcome the problem of group efficiency, Kumar and Chandrasekharan (1990) introduced a new measure, group efficacy.

$$\tau = (1-\varphi)/(1+\phi) \tag{14}$$

where φ is the ratio of the number of exceptional elements to the total number of elements; ϕ is the ratio of the number of 0s in the region A to the total number of elements.

8.4.4 Machine utilization index (Grouping measure, GM)

Proposed by Miltenburg and Zhang (1991), which is used to measure machine utilization in a cell. The index is defined as follows:

$$\eta_g = \eta_u - \eta_m \tag{15}$$

where $\eta_u = d/(d+v)$ and $\eta_m = 1-(d/o)$. d is the number of 1s in the region A, η_u is the measure of utilization of machines in a cell and η_m is the measure of inter-cell movements of parts. η_g ranges form –1 to 1, η_u and η_m range from 0 to 1. A bigger value of machine utilization index η_g is desired.

8.4.5 Clustering measure (CM)

This measure tests how closely the 1s gather around the diagonal of the solution matrix, the definition of the measure is as follows (Singh & Rajamani, 1996).

$$\eta_c = \frac{\left\{\sum_{i=1}^{M}\sum_{k=1}^{P}\left(\sqrt{\delta_h^2(a_{ik}) + \delta_v^2(a_{ik})}\right)\right\}}{\sum_{i=1}^{M}\sum_{k=1}^{P} a_{ik}} \tag{16}$$

where $\delta_h(a_{ik})$ and $\delta_v(a_{ik})$ are horizontal and vertical distances between a non-zero entry a_{ik} and the diagonal, respectively.

$$\delta_h = i - \frac{k(M-1)}{(P-1)} - \frac{(P-M)}{(P-1)} \tag{17}$$

$$\delta_v = k - \frac{i(P-1)}{(M-1)} - \frac{(P-M)}{(M-1)} \tag{18}$$

8.4.6 Grouping index (GI)

Nair and Narendran (1996) indicated that a good performance measure should be defined with reference to the block diagonal space. And the definition should ensure equal weightage to voids (0s in the region A) and exceptional elements. They introduced a measure, incorporating the block diagonal space, weighting factor and correction factor.

$$\gamma = \frac{1 - \dfrac{qv + (1-q)(e-A)}{B}}{1 + \dfrac{qv + (1-q)(e-A)}{B}} \tag{19}$$

where B is the block diagonal space and q is a weighting factor ranges between 0 and 1. A=0 for $e \leq B$ and $A = e - B$ for $e > B$. For convenience, equation (19) could be written as follows:

$$\gamma = \frac{1-\alpha}{1+\alpha} \tag{20}$$

where

$$\alpha = \frac{qv + (1-q)(e-A)}{B} \tag{21}$$

Both α and γ range from 0 to 1.

8.4.7 Bond energy measure (BEM)

McCormick *et al*. (1972) used the BEM to convert a binary matrix into a block diagonal form. This measure is defined as follows:

$$\eta_{BE} = \frac{\sum_{i=1}^{M}\sum_{k=1}^{P-1} a_{ik} a_{i(k+1)} + \sum_{i=1}^{M-1}\sum_{k=1}^{P} a_{ik} a_{(i+1)k}}{\sum_{i=1}^{M}\sum_{k=1}^{P} a_{ik}} \tag{22}$$

Bond energy is used to measure the relative clumpiness of a clustered matrix. Therefore, the more close the 1s are, the larger the bond energy measure will be.

8.4.8 Grouping capability index (GCI)

Hsu (1990) showed that neither group efficiency nor group efficacy is consistent in predicting the performance of a cellular manufacturing system based on the structure of the corresponding machine-part matrix (Seifoddini & Djassemi, 1996). Hsu (1990) considered the GCI as follows:

$$GCI = 1 - \frac{e}{o} \tag{23}$$

Unlike group efficiency and group efficacy, GCI excludes zero entries from the calculation of grouping efficacy.

8.4.9. Alternative routeing grouping efficiency (ARG efficiency)

ARG was propose by Sarker and Li (1998). ARG evaluates the grouping effect in the presence of alternative routings of parts. The efficiency is defined as follows:

$$\eta_{ARG} = \frac{(1-\frac{e}{o'})(1-\frac{v}{z'})}{(1+\frac{e}{o'})(1+\frac{v}{z'})} = \left(\frac{o'-e}{o'+e}\right)\left(\frac{z'-v}{z'+v}\right) \quad (24)$$

where o' is the total number of 1s in the original machine-part incidence matrix with multiple process routings, z' is the total number of 0s in the original machine-part incidence matrix with multiple process routings. ARG efficiency can also be used to evaluate CF problems that have no multiple process routings of parts. The efficiency ranges from 0 to 1 and is independent of the size of the problem.

9. Comparison and results

Two key characteristics of similarity coefficients are tested in this study, discriminability and stability. In this study, we compare the similarity coefficients by using following steps.

Comparative steps

1. Computation.
   - 1.1. At first, solve each problem in the data sets by using 20 similarity coefficients; compute performance values by 9 performance measures. Thus, we obtain at least a total of δ*20*9 solutions. δ is the number of the problems (some data sets from literature are multi-problems due to the different number of cells, see the item NC of Table 7).
   - 1.2. Average performance values matrix: create a matrix whose rows are problems and columns are 9 performance measures. An element in row i and column j indicates, for problem i and performance measure j, the average performance value produced by 20 similarity coefficients.

2. Based on the results of step 1, construct two matrixes whose rows are 20 similarity coefficients and columns are 9 performance measures, an entry SM_{ij} in the matrixes indicates:
   2.1. Discriminability matrix: the number of problems to which the similarity coefficient i gives the best performance value for measure j.
   2.2. Stability matrix: the number of problems to which the similarity coefficient i gives the performance value of measure j with at least average value (better or equal than the value in the matrix of step 1.2).
3. For each performance measure, find the top 5 values in the above two matrixes. The similarity coefficients correspond to these values are considered to be the most discriminable/stable similarity coefficients for this performance measure.
4. Based on the results of step 3, for each similarity coefficient, find the number of times that it has been selected as the most discriminable/stable coefficient for the total 9 performance measures.

We use small examples here to show the comparative steps.

Step 1.1: a total of 214 problems were solved. 120 problems were deliberately generated; 94 problems were from literature, see Table 2 (some data sets were multi-problems due to the different number of cells). A to tal of 38,520 (214*20*9) performance values were gotten by using 20 similarity coefficients and 9 performance measures. For example, by using Jaccard similarity coefficient, the 9 performance values of the problem McCormick *et al*. (no.62 in Table 7) are as follows (Table 10):

	EE	*Grouping efficiency*	*Group efficacy*	*GM*	*CM*	*GI*	*BEM*	*GCI*	ARG
Jaccard	87	0.74	0.45	0.25	7.85	0.44	1.07	0.6	0.32

Table 10: The performance values of McCormick et al. by using Jaccard similarity coefficient

Step 1.2: The average performance values matrix contained 214 problems (rows) and 9 performance measures (columns). An example of row (problem McCormick *et al*.) is as follows (Table 11):

	EE	*Grouping efficiency*	*Group efficacy*	*GM*	*CM*	*GI*	*BEM*	*GCI*	**ARG**
Average values	94.7	0.77	0.45	0.28	8.06	0.4	1.06	0.57	0.31

Table 11. The average performance values of 20 similarity coefficients, for the problem McCormick *et al*

We use Jaccard similarity coefficient and the 94 problems from literature to explain following steps 2, 3, and 4.

Step 2.1 (discriminability matrix): among the 94 problems and for each performance measure, the numbers of problems to which Jaccard gave the best values are shown in Table 12. For example, the 60 in the column EE means that comparing with other 19 similarity coefficients, Jaccard produced minimum exceptional elements to 60 problems.

	EE	*Grouping efficiency*	*Group efficacy*	*GM*	*CM*	*GI*	*BEM*	*GCI*	**ARG**
Jaccard	60	51	55	62	33	65	41	60	57

Table 12. The number of problems to which Jaccard gave the best performance values

Step 2.2 (stability matrix): among the 94 problems and for each performance measure, the numbers of problems to which Jaccard gave the value with at least average value (matrix of step 1.2) are shown in Table 13. For example, the meaning of 85 in the column EE is as follows: comparing with the average exceptional elements of 94 problems in the matrix of step 1.2, the numbers of problems to which Jaccard produced a fewer exceptional elements are 85.

	EE	*Grouping efficiency*	*Group efficacy*	*GM*	*CM*	*GI*	*BEM*	*GCI*	**ARG**
Jaccard	85	85	85	89	69	91	75	88	73

Table 13. The number of problems to which Jaccard gave the best performance values

Step 3: For example, for the exceptional elements, the similarity coefficients that corresponded to the top 5 values in the discriminability matrix are Jaccard, Sorenson, Rusell and Rao, Dot-product, Sokal and Sneath 2, Relative matching, and Baker and Maropoulos. These similarity coefficients are considered as the most discriminable coefficients for the performance measure – exceptional elements. The same procedures are conducted to the other performance measures and stability matrix.

Step 4: Using the results of step 3, Jaccard have been selected 5/6 times as the most discriminable/stable similarity coefficient. That means, among 9 performance measures, Jaccard is the most discriminable/stable similarity coefficient for 5/6 performance measures. The result is shown in the column – literature of Table 14.
The results are shown in Table 14 and Figs. 6-8 (in the figures, the horizontal axes are 20 similarity coefficients and the vertical axes are 9 performance measures).

The tables and figures show the number of performance measures for which these 20 similarity coefficients have been regarded as the most discriminable/stable coefficients. The columns of the table represent different conditions of data sets. The column – literature includes all 94 problems from literature; the column – all random includes all 120 deliberately generated problems. The deliberately generated problems are further investigated by using different levels of REC and RE.

"Literature" and "All random" columns in Table 14 (also Fig.6) give the performance results of all 214 tested problems. We can find that Jaccard and Sorenson are two best coefficients. On the other hand, four similarity coefficients: Hamann, Simple matching, Rogers & Tanimoto, and Sokal & Sneath are inefficient in both discriminability and stability.

"REC" columns in table 9 (also Fig.3) show the performance results under the condition of different REC ratios. We can find that almost all similarity coefficients perform well under a high REC ratio. However, four similarity coefficients: Hamann, Simple matching, Rogers & Tanimoto, and Sokal & Sneath, again produce bad results under the low REC ratio.
"RE" columns in Table 14 (also Fig.8) give the performance results under the

condition of different RE ratios. All similarity coefficients perform best under a low RE ratio (data sets are well-structured). Only a few of similarity coefficients perform well under a high RE ratio (data sets are ill-structured), Sokal & Sneath 2 is very good for all RE ratios. Again, the four similarity coefficients: Hamann, Simple matching, Rogers & Tanimoto, and Sokal & Sneath, perform badly under high RE ratios.

No.	**Similarity Coefficient**	*Literature*		*All random*		*REC*						*RE*					
						0.7		*0.8*		*0.9*		*0.05-0.15*		*0.2-0.3*		*0.35-0.4*	
		D	*S*	*D*	*S*	*D*	*S*	*D*	*S*	*D*	*S*	*D*	*S*	*D*	*S*	*D*	*S*
1	Jaccard	5	6	6	9	8	9	8	9	9	9	9	9	9	9	8	9
2	Hamann	0	0	2	1	1	1	2	3	7	7	9	9	1	0	2	2
3	Yule	4	4	2	6	3	7	5	7	7	8	9	9	2	6	6	7
4	Simple matching	0	0	2	0	1	0	3	5	6	8	9	9	0	0	2	2
5	Sorenson	6	4	9	8	7	9	8	9	9	9	9	9	9	9	7	7
6	Rogers & Tanimoto	0	0	2	1	2	2	4	4	6	7	9	9	1	2	2	2
7	Sokal & Sneath	0	0	0	0	2	1	5	6	6	8	9	9	1	1	2	2
8	Rusell & Rao	4	4	5	3	5	5	9	8	8	6	9	9	9	8	6	6
9	Baoroni-Urban & Buser	5	6	1	3	3	7	9	7	7	8	9	9	4	7	2	6
10	Phi	5	5	6	6	9	7	8	8	7	8	9	9	9	8	7	7
11	Ochiai	1	4	8	7	9	7	8	8	9	9	9	9	9	9	7	7
12	PSC	2	2	9	8	9	9	9	8	9	9	9	9	9	9	8	9
13	Dot-product	3	5	9	8	7	9	8	9	9	9	9	9	9	9	7	7
14	Kulczynski	2	5	8	7	8	8	8	8	9	9	9	9	9	9	7	7
15	Sokal & Sneath 2	4	5	6	8	9	9	7	9	9	9	9	9	9	9	9	9
16	Sokal & Sneath 4	5	5	7	6	8	7	8	8	7	8	9	9	8	8	7	7
17	Relative matching	5	4	4	8	7	9	9	9	9	9	9	9	5	9	6	8
18	Chandraseharan & Rajagopalan	2	5	8	6	9	8	8	8	7	7	9	9	9	9	6	7
19	MaxSC	1	4	8	6	9	8	8	8	7	7	9	9	9	9	6	7
20	Baker & Maropoulos	5	3	6	9	7	9	8	9	9	9	9	9	6	9	6	8

Table 14. Comparative results under various conditions. *D:* discriminability; *S:* stability

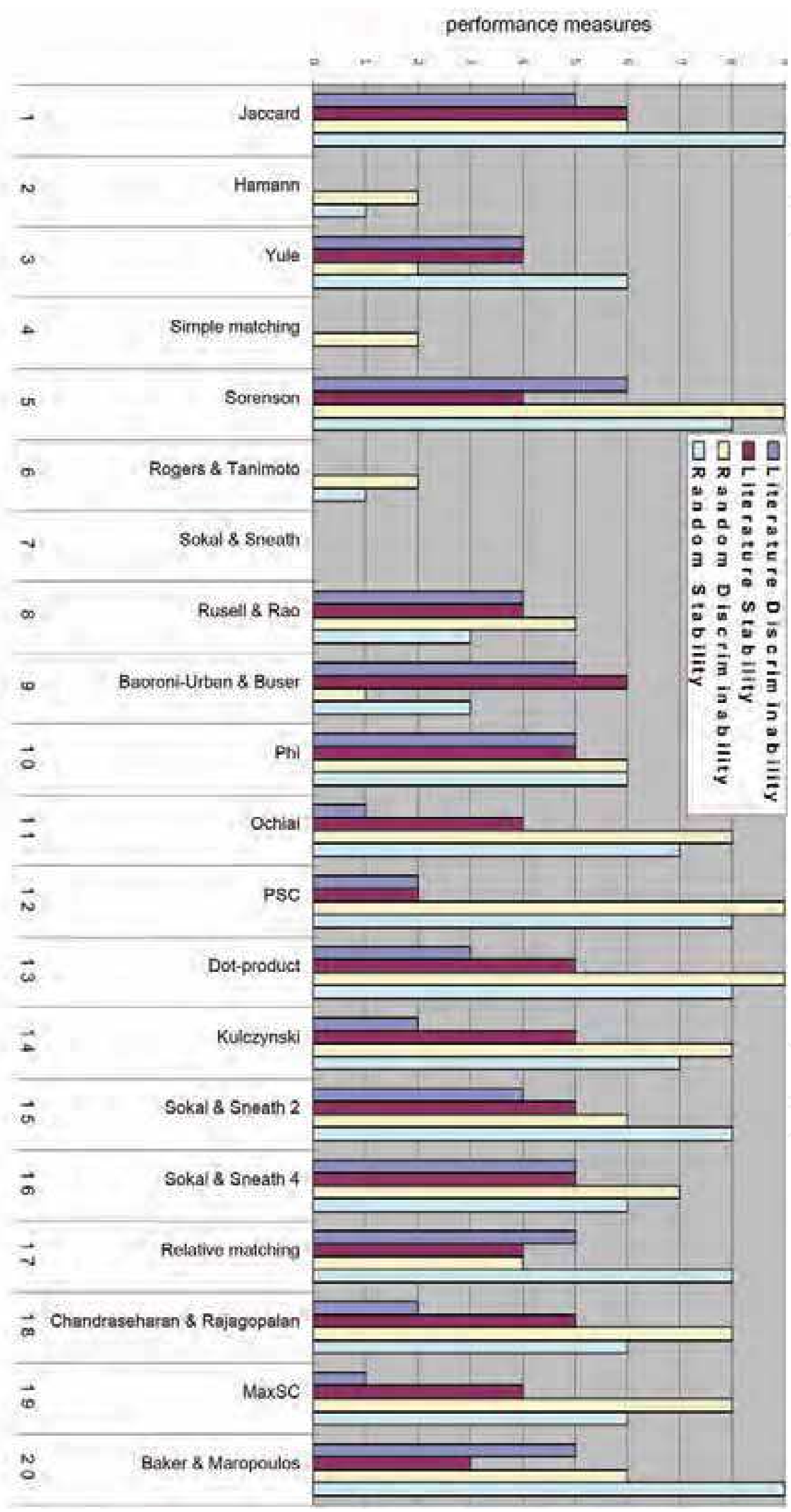

Figure 6. Performance for all tested problems

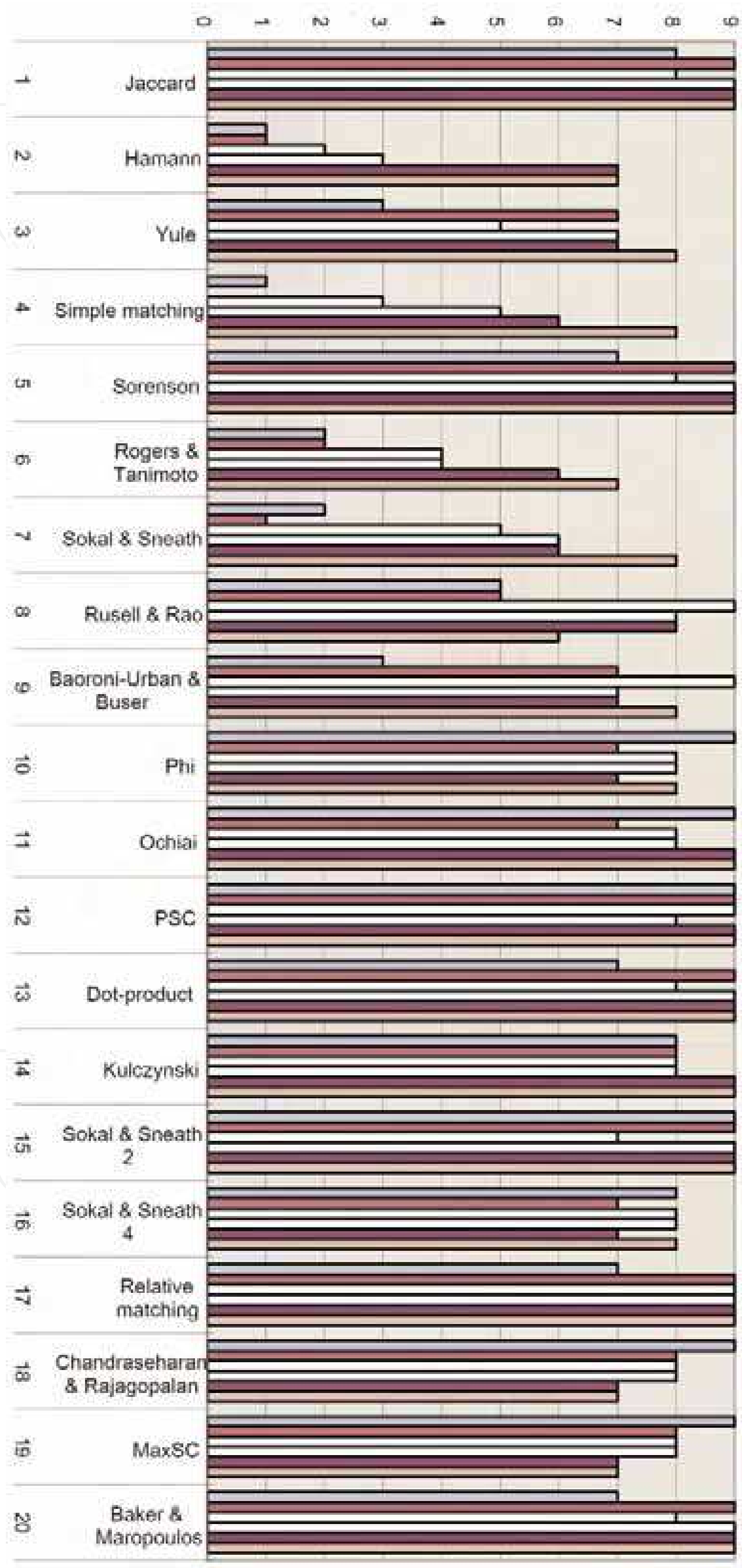

Figure 7. Performance under different REC

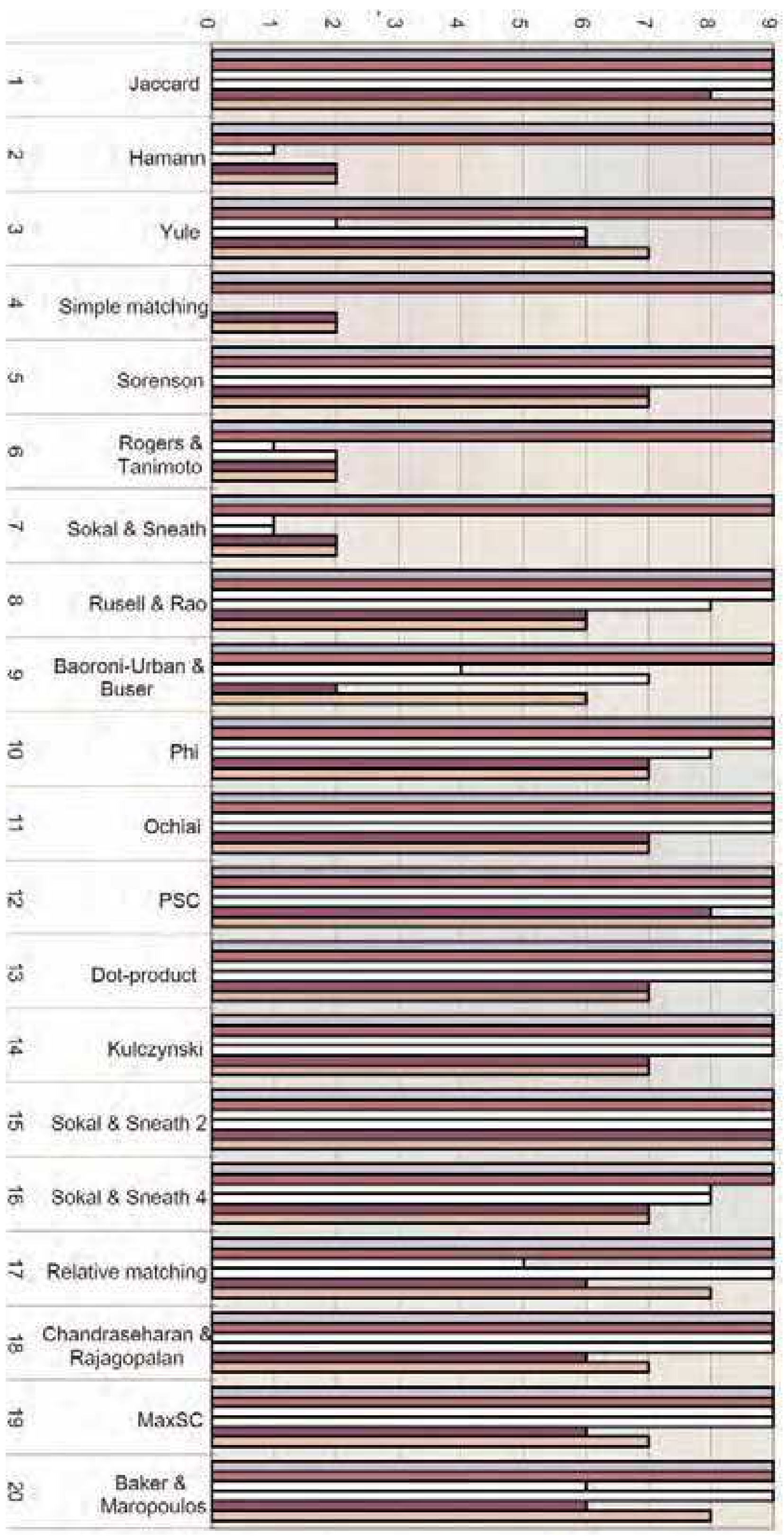

Figure 8. Performance under different RE

In summary, three similarity coefficients: Jaccard, Sorenson, and Sokal & Sneath 2 perform best among twenty tested similarity coefficients. Jaccard emerges from the twenty similarity coefficients for its stability. For all problems, from literature or deliberately generated; and for all levels of both REC and RE ratios, Jaccard similarity coefficient is constantly the most stable coefficient among all twenty similarity coefficients. Another finding in this study is four similarity coefficients: Hamann, Simple matching, Rogers & Tanimoto, and Sokal & Sneath are inefficient under all conditions. So, these similarity coefficients are not recommendable for using in cell formation applications.

9. Conclusions

In this paper various similarity coefficients to the cell formation problem were investigated and reviewed. Previous review studies were discussed and the need for this review was identified. The reason why the similarity coefficient based methods (SCM) is more flexible than other cell formation methods were explained through a simple example. We also proposed a taxonomy which is combined by two distinct dimensions. The first dimension is the general-purpose similarity coefficients and the second is the problem-oriented similarity coefficients. The difference between two dimensions is discussed through three similarity coefficients. Based on the framework of the proposed taxonomy, existing similarity (dissimilarity) coefficients developed so far were reviewed and mapped onto the taxonomy. The details of each production information based similarity coefficient were simply discussed and a evolutionary timeline was drawn based on reviewed similarity coefficients. Although a number of similarity coefficients have been proposed, very fewer comparative studies have been done to evaluate the performance of various similarity coefficients. This paper evaluated the performance of twenty well-known similarity coefficients. 94 problems from literature and 120 problems generated deliberately were solved by using the twenty similarity coefficients. To control the generation process of data sets, experimental factors have been discussed. Two experimental factors were proposed and used for generating experimental problems. Nine performance measures were used to judge the solutions of the tested problems. The numerical results showed that three similarity coefficients are more efficient and four similarity coefficients are inefficient for solving the cell formation problems. Another finding is that Jaccard similarity coefficient is the most stable similarity coefficient. For the further studies, we suggest comparative studies in consideration of some production factors, such as production volumes, operation sequences, etc. of parts.

7. References

Agarwal, A., Sarkis, J., 1998. A review and analysis of comparative performance studies on functional and cellular manufacturing layouts. Computers and Industrial Engineering 34, 77-89.

Akturk, M.S., Balkose, H.O., 1996. Part-machine grouping using a multi-objective cluster analysis. International Journal of Production Research 34, 2299-2315.

Al-Sultan, K.S., Fedjki, C.A., 1997. A genetic algorithm for the part family formation problem. Production Planning & Control 8, 788-796.

Anderberg, M.R., 1973. Cluster analysis for applications (New York: Academic Press).

Arthanari, T.S., Dodge, Y., 1981. Mathematical programming in statistics (New York: John Wiley & Sons, Inc).

Askin, R.G., Cresswell, S.H., Goldberg, J.B., Vakharia, A.J., 1991. A Hamiltonian path approach to reordering the part-machine matrix for cellular manufacturing. International Journal of Production Research 29, 1081-1100.

Askin, R.G., Selim, H.M., Vakharia, A.J., 1997. A methodology for designing flexible cellular manufacturing systems. IIE Transaction 29, 599-610.

Askin, R.G., & Subramanian, S.P., 1987. A cost-based heuristic for group technology configuration. *International Journal of Production Research, 25*(1), 101-113.

Askin, R.G., Zhou, M., 1998. Formation of independent flow-line cells based on operation requirements and machine capabilities. IIE Transactions 30, 319-329.

Baker, R.P., Maropoulos, P.G., 1997. An automatic clustering algorithm suitable for use by a computer-based tool for the design, management and continuous improvement of cellular manufacturing systems. Computers Integrated Manufacturing Systems 10, 217-230.

Balakrishnan, J., 1996. Manufacturing cell formation using similarity coefficients and pair-wise interchange: formation and comparison. Production Planning & Control 7, 11-21.

Balakrishnan, J., Cheng, C. H., 1998. Dynamic layout algorithms: a state-of-the-art survey. Omega 26, 507-521.

Balakrishnan, J., Jog, P.D., 1995. Manufacturing cell formation using similarity coefficients and a parallel genetic TSP algorithm: formulation and comparison. Mathematical and Computer Modelling 21, 61-73.

Balasubramanian, K.N., Panneerselvam, R., 1993. Covering technique-based algorithm for machine grouping to form manufacturing cells. International Journal of Production Research 31, 1479-1504.

Baroni-Urbani, C., Buser, M.W., 1976. Similarity of binary data. Systematic Zoology 25, 251-259.

Baykasoglu, A., Gindy, N.N.Z., 2000. MOCACEF 1.0: multiple objective capability based approach to form part-machine groups for cellular manufacturing applications. International Journal of Production Research 38, 1133-1161.

Beatty, C.A., 1992. Implementing advanced manufacturing technologies: rules of the road. Sloan Management Review Summer, 49-60.

Ben-Arieh, D., Chang, P.T., 1994. An extension to the p-median group technology algorithm. Computers and Operations Research 21, 119-125.

Ben-Arieh, D., Sreenivasan, R., 1999. Information analysis in a distributed dynamic group technology method. International Journal of Production Economics 60-61, 427-432.

Bijnen, E.J., 1973. Cluster analysis (The Netherlands: Tilburg University Press).

Bishop, Y.M.M., Fienberg, S.E., Holland, P.W., 1975. Discrete multivariate analysis: theory and practice (MA: MIT Press Cambridge).

Boctor, F.F., 1991. A linear formulation of the machine-part cell formation problem. *International Journal of Production Research, 29*(2), 343-356.

Boe, W.J., & Cheng, C.H., 1991. A close neighbour algorithm for designing cellular manufacturing systems. *International Journal of Production Research, 29*(10), 2097-2116.

Burbidge, J.L., 1971. Production flow analysis. Production Engineer 50, 139-152.

Burbidge, J.L., Falster, P., Rhs, J.O., 1991. Why is it difficult to sell group technology and just-in-time to industry? Production Planning & Control 2, 160-166.

Carrie, A.S., 1973. Numerical taxonomy applied to group technology and plant layout. International Journal of Production Research 11, 399-416.

Cedeno, A.A., Suer, G.A., 1997. The use of a similarity coefficient-based method to perform clustering analysis to a large set of data with dissimilar parts. Computers and Industrial Engineering 33, 225-228.

Chan, H.M., & Milner, D.A., 1982. Direct clustering algorithm for group formation in cellular manufacture. *Journal of Manufacturing Systems, 1*(1), 65-75.

Chandrasekharan, M.P., Rajagopalan, R., 1986a. An ideal seed non-hierarchical clustering algorithm for cellular manufacturing. International Journal of Production Research 24, 451-464.

Chandrasekharan, M.P., Rajagopalan, R., 1986b. MODROC: an extension of rank order clustering for group technology. International Journal of Production Research 24, 1221-1233.

Chandrasekharan, M.P., Rajagopalan, R., 1987. ZODIAC: an algorithm for concurrent formation of part families and machine cells. International Journal of Production Research 25, 451-464.

Chandrasekharan, M.P., Rajagopalan, R., 1989. GROUPABILITY: an analysis of the properties of binary data matrices for group technology. International Journal of Production Research. 27, 1035-1052.

Chang, P.T., Lee, E.S., 2000. A multisolution method for cell formation – exploring practical alternatives in group technology manufacturing. Computers and Mathematics with Applications 40, 1285-1296.

Chen, D.S., Chen H.C., & Part, J.M., 1996. An improved ART neural net for machine cell formation. *Journal of Materials Processing Technology, 61*, 1-6.

Cheng, C.H., Goh, C.H., Lee, A., 1995. A two-stage procedure for designing a group technology system. International Journal of Operations & Production Management 15, 41-50.

Cheng, C.H., Gupta, Y.P., Lee, W.H., Wong, K.F., 1998. A TSP-based heuristic for forming machine groups and part families. International Journal of Production Research 36, 1325-1337.

Cheng, C.H., Madan, M.S., Motwani, J., 1996. Designing cellular manufacturing systems by a truncated tree search. International Journal of Production Research 34, 349-361.

Choobineh, F., 1988. A framework for the design of cellular manufacturing systems. International Journal of Production Research 26, 1161-1172.

Choobineh, F., Nare, A., 1999. The impact of ignored attributes on a CMS design. International Journal of Production Research 37, 3231-3245.

Chow, W.S., 1991. Discussion: a note on a linear cell clustering algorithm. International Journal of Production Research 29, 215-216.

Chow, W.S., Hawaleshka, O., 1992. An efficient algorithm for solving the machine chaining problem in cellular manufacturing. Computers and Industrial Engineering 22, 95-100.

Chow, W.S., Hawaleshka, O., 1993a. Minimizing intercellular part movements in manufacturing cell formation. International Journal of Production Research 31, 2161-2170.

Chow, W.S., Hawaleshka, O., 1993b. A novel machine grouping and knowledge-based approach for cellular manufacturing. European Journal of Operational Research 69, 357-372.

Chu, C.H., 1989. Cluster analysis in manufacturing cellular formation. Omega 17, 289-295.

Chu, C.H., Pan, P., 1988. The use of clustering techniques in manufacturing cellular formation. Proceedings of International Industrial Engineering Conference, Orlando, Florida, pp. 495-500.

Chu, C.H., & Tsai, M., 1990. A comparison of three array-based clustering techniques for manufacturing cell formation. *International Journal of Production Research, 28*(8), 1417-1433.

De Witte, J., 1980. The use of similarity coefficients in production flow analysis. International Journal of Production Research 18, 503-514.

Dimopoulos, C., Mort, N., 2001. A hierarchical clustering methodology based on genetic programming for the solution of simple cell-formation problems. International Journal of Production Research 39, 1-19.

Dutta, S.P., Lashkari, R.S., Nadoli, G., Ravi, T., 1986. A heuristic procedure for determining manufacturing families from design-based grouping for flexible manufacturing systems. Computers and Industrial Engineering 10, 193-201.

Faber, Z., Carter, M.W., 1986. A new graph theory approach for forming machine cells in cellular production systems. In A. Kusiak (ed), Flexible Manufacturing Systems: Methods and Studies (North-Holland: Elsevier Science Publishers B.V), pp. 301-315.

Fazakerley, G.M., 1976. A research report on the human aspects of group technology and cellular manufacture. International Journal of Production Research 14, 123-134.

Gongaware, T.A., Ham, I., 1991. Cluster analysis applications for group technology manufacturing systems. Proceedings Ninth North American Manufacturing Research Conference, pp. 503-508.

Gordon, A.D., 1999. Classification, 2nd edition (US: Chapman & Hall).

Gunasingh, K.R., Lashkari, R.S., 1989. The cell formation problem in cellular manufacturing systems – a sequential modeling approach. Computers and Industrial Engineering 16, 469-476.

Gupta, T., 1991. Clustering algorithms for the design of a cellular manufacturing system – an analysis of their performance. Computers and Industrial Engineering 20, 461-468.

Gupta, T., 1993. Design of manufacturing cells for flexible environment considering alternative routeing. International Journal of Production Research 31, 1259-1273.

Gupta, T., Seifoddini, H., 1990. Production data based similarity coefficient for machine-component grouping decisions in the design of a cellular manufacturing system. International Journal of Production Research 28, 1247-1269.

Han, C., Ham, I., 1986. Multiobjective cluster analysis for part family formations. Journal of Manufacturing Systems 5, 223-230.

Ho, Y.C., Lee, C., Moodie, C.L., 1993. Two sequence-pattern, matching-based, flow analysis methods for multi-flowlines layout design. International Journal of

Production Research 31, 1557-1578.

Ho, Y.C., Moodie, C.L., 1996. Solving cell formation problems in a manufacturing environment with flexible processing and routeing capabilities. International Journal of Production Research 34, 2901-2923.

Holley, J.W., Guilford, J.P., 1964. A note on the G index of agreement. Educational and Psychological Measurement 24, 749-753.

Hon, K.K.B, & Chi, H., 1994. A new approach of group technology part families optimization. *Annals of the CIRP, 43*(1), 425-428.

Hsu, C.P., 1990. *Similarity coefficient approaches to machine-component cell formation in cellular manufacturing: a comparative study*. Ph.D. thesis. Department of Industrial and Manufacturing Engineering, University of Wisconsin-Milwaukee.

Hwang, H., Ree, P., 1996. Routes selection for the cell formation problem with alternative part process plans. Computers and Industrial Engineering 30, 423-431.

Irani, S.A.,& Khator, S.K., 1986. A microcomputer-based design of a cellular manufacturing system. In: *Proceedings of the 8th Annual Conference on Computers and Industrial Engineering*, *11*, 68-72.

Islam, K.M.S., Sarker, B.R., 2000. A similarity coefficient measure and machine-parts grouping in cellular manufacturing systems. International Journal of Production Research 38, 699-720.

Jaccard, P., 1908. Novelles recgerches sur la distribution florale. *Bull. Soc. Vaud. Sci. Nat.*, *44*, 223-270.

Jeon, G., Broering, M., Leep, H.R., Parsaei, H.R., Wong, J.P., 1998a. Part family formation based on alternative routes during machine failure. Computers and Industrial Engineering 35, 73-76.

Jeon, G., Leep, H.R., Parsaei, H.R., 1998b. A cellular manufacturing system based on new similarity coefficient which considers alternative routes during machine failure. Computers and Industrial Engineering 34, 21-36.

Josien, K., Liao, T.W., 2000. Integrated use of fuzzy c-means and fuzzy KNN for GT part family and machine cell formation. International Journal of Production Research 38, 3513-3536.

Kamrani, A.K., Parsaei, H.R., Chaudhry, M.A., 1993. A survey of design methods for manufacturing cells. Computers and Industrial Engineering 25, 487-490.

Kang, S.L., Wemmerlöv, U., 1993. A work load-oriented heuristic methodology for manufacturing cell formation allowing reallocation of operations. European Journal of Operational Research 69, 292-311.

Kaparthi, S., Suresh, N.C., Cerveny, R.P., 1993. An improved neural network leader algorithm for part-machine grouping in group technology. European Journal of Operational Research 69, 342-356.

King, J.R., 1980. Machine-component grouping in production flow analysis: an approach using a rank order clustering algorithm. *International Journal of Production Research, 18*(2), 213-232.

King, J.R., Nakornchai, V., 1982. Machine component group formation in group technology: review and extension. International Journal of Production Research 20, 117-133.

Kitaoka, M., Nakamura, R., Serizawa, S., Usuki, J., 1999. Multivariate analysis model for machine-part cell formation problem in group technology. International Journal of Production Economics 60-61, 433-438.

Kulkarni, U.R., Kiang, M. Y., 1995. Dynamic grouping of parts in flexible manufacturing systems – a self-organizing neural networks approach. European Journal of Operational Research 84, 192-212.

Kumar, C.S., Chandrasekharan, M. P., 1990. Grouping efficacy: a quantitative criterion for goodness of block diagonal forms of binary matrices in group technology. International Journal of Production Research 28, 233-243.

Kumar, K.R., Kusiak, A., & Vannelli, A., 1986. Grouping of parts and components in flexible manufacturing systems. *European Journal of Operational Research, 24*, 387-397.

Kumar, K.R., Vannelli, A., 1987. Strategic subcontracting for efficient disaggregated manufacturing. International Journal of Production Research 25, 1715-1728.

Kusiak, A., 1985. The part families problem in flexible manufacturing systems. Annals of Operations Research 3, 279-300.

Kusiak, A., 1987. The generalized group technology concept. International Journal of Production Research 25, 561-569.

Kusiak, A., Boe, W.J., Cheng, C., 1993. Designing cellular manufacturing systems: branch-and-bound and A* approaches. IIE Transactions 25, 46-56.

Kusiak, A., Cho, M., 1992. Similarity coefficient algorithms for solving the group technology problem. International Journal of Production Research 30, 2633-2646.

Kusiak, A., & Chow, W.S., 1987. Efficient solving of the group technology problem. *Journal of Manufacturing Systems, 6,* 117-124.

Kusiak, A., Heragu, S.S., 1987. The facility layout problem. European Journal of Operational Research 29, 229-251.

Kusiak, A., Vannelli, A., Kumar, K.R., 1986. Clustering analysis: models and algorithms. Control and Cybernetics 15, 139-154.

Lashkari, R.S., Boparai, R., Paulo, J., 2004. Towards an integrated model of operation allocation and material handling selection in cellular manufacturing systems. International Journal of Production Economics 87, 115-139.

Lashkari, R.S., Gunasingh, K.R., 1990. A Lagrangian relaxation approach to machine allocation in cellular manufacturing systems. Computers and Industrial Engineering 19, 442-446.

Lee, H., Garcia-Diaz, A., 1996. Network flow procedures for the analysis of cellular manufacturing systems. IIE Transactions 28, 333-345.

Lee, M.K., Luong, H.S., Abhary, K., 1997. A genetic algorithm based cell design considering alternative routing. Computers Integrated Manufacturing Systems 10, 93-107.

Leem, C.W., Chen, J. J.G., 1996. Fuzzy-set-based machine-cell formation in cellular manufacturing. Journal of Intelligent Manufacturing 7, 355-364.

Lee-post, A., 2000. Part family identification using a simple genetic algorithm. International Journal of Production Research 38, 793-810.

Liggett, R.S., 2000. Automated facilities layout: past, present and future. Automation in Construction 9, 197-215.

Lin, T.L., Dessouky, M.M., Kumar, K.R., Ng, S.M., 1996. A heuristic-based procedure for the weighted production-cell formation problem. IIE Transactions 28, 579-589.

Logendran, R., 1991. Effect of the identification of key machines in the cell formation problem of cellular manufacturing systems. Computers and Industrial Engineering 20, 439-449.

Lozano, S., Adenso-Diaz, B., Eguia, I., Onieva, L., 1999. A one-step tabu search algorithm for manufacturing cell design. Journal of the Operational Research Society 50, 509-516.

Luong, L.H.S., 1993. A cellular similarity coefficient algorithm for the design of manufacturing cells. International Journal of Production Research 31, 1757-1766.

Mansouri, S.A., Husseini S.M.M., Newman, S.T., 2000. A review of the modern approaches to multi-criteria cell design. International Journal of Production Research 38, 1201-1218.

Marcotorchino, F., 1987. Block seriation problems: a unified approach. Applied Stochastic Models and Data Analysis 3, 73-91.

Masnata, A., Settineri, L., 1997. An application of fuzzy clustering to cellular manufacturing. International Journal of Production Research 35, 1077-1094.

McAuley, J., 1972, Machine grouping for efficient production. The Production Engineer, 51, 53-57.

McCormick, W.T., Schweitzer P.J., & White, T.W., 1972. Problem decomposition and data reorganization by a clustering technique. *Operations Research, 20*(5), 993-1009.

Mehrez, A., Rabinowitz, G., Reisman, A., 1988. A conceptual scheme of knowledge systems for MS/OR. Omega 16, 421-428.

Milligan, G.W., & Cooper, S.C., 1987. Methodology review: clustering methods. *Applied Psychological Measurement, 11*(4), 329-354.

Miltenburg, J., Zhang, W., 1991. A comparative evaluation of nine well-known algorithms for solving the cell formation problem in group technology. Journal of Operations Management 10, 44-72.

Mitrofanov, S.P., 1966. Scientific principles of group technology, Part⊚ (MA, Boston: National Lending Library of Science and Technology).

Mosier, C.T., 1989. An experiment investigating the application of clustering procedures and similarity coefficients to the GT machine cell formation problem. International Journal of Production Research 27, 1811-1835.

Mosier, C.T., Taube, L., 1985a. The facets of group technology and their impacts on implementation – a state of the art survey. Omega 13, 381-391.

Mosier, C.T., Taube, L., 1985b. Weighted similarity measure heuristics for the group technology machine clustering problem. Omega 13, 577-583.

Mosier, C.T., Yelle, J., Walker, G., 1997. Survey of similarity coefficient based methods as applied to the group technology configuration problem. Omega 25, 65-79.

Murthy, Ch.V.R., Srinivasan, G., 1995. Fractional cell formation in group technology. International Journal of Production Research 33, 1323-1337.

Nair, G.J.K., Narendran, T.T., 1996. Grouping index: a new quantitative criterion for goodness of block-diagonal forms in group technology. *International Journal of Production Research, 34*(10), 2767-2782.

Nair, G.J.K., Narendran, T.T., 1998. CASE: A clustering algorithm for cell formation with sequence data. International Journal of Production Research 36, 157-179.

Nair, G.J.K., Narendran, T.T., 1999. ACCORD: A bicriterion algorithm for cell formation using ordinal and ratio-level data. International Journal of Production Research 37, 539-556.

Ng, S.M., 1993. Worst-case analysis of an algorithm for cellular manufacturing. *European Journal of Operational Research, 69*, 384-398.

Offodile, O.F., 1991. Application of similarity coefficient method to parts coding and classification analysis in group technology. Journal of Manufacturing Systems 10, 442-448.

Offodile, O.F., 1993. Machine grouping in cellular manufacturing. Omega 21, 35-52.

Offodile, O.F., Mehrez, A., Grznar, J., 1994. Cellular manufacturing: a taxonomic review framework. Journal of Manufacturing Systems 13, 196-220.

Offodile, O.F., Grznar, J., 1997. Part family formation for variety reduction in flexible manufacturing systems. International Journal of Operations & Production Management 17, 291-304.

Onwubolu, G.C., Mlilo, P.T., 1998. Manufacturing cell grouping using similarity coefficient-distance measure. Production Planning & Control 9, 489-493.

Opitz, H., Eversheim, W., Wienhal, H.P., 1969. Work-piece classification and its industrial applications. International Journal of Machine Tool Design and Research 9, 39-50.

Qiao, L.H., Yang, Z.B., Wang, H.P., 1994. A computer-aided process planning methodology. Computers in Industry 255, 83-94.

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

Reisman, A., Kirshnick, F., 1995. Research strategies used by OR/MS workers as shown by an analysis of papers in flagship journals. Operations Research 43, 731-739.

Reisman, A., Kumar, A., Motwani, J., Cheng, C.H., 1997. Cellular manufacturing: a statistical review of the literature (1965-1995). Operations Research 45, 508-520.

Ribeiro, J.F.F., Pradin, B., 1993. A methodology for cellular manufacturing design. International Journal of Production Research 31, 235-250.

Rogers, D.J., Tanimoto, T.T., 1960. A computer program for classifying plants. Science 132, 1115-1118.

Romesburg, H.C., 1984. Cluster analysis for researchers (CA: Lifetime Learning Publications (Wadsworth Inc.), Belmont).

Samatova, N.F., Potok, T.E., Leuze, M.R., 2001. Vector space model for the generalized parts grouping problem. Robotics and Computer Integrated Manufacturing 17, 73-80.

Sarker, B.R., 1996. The resemblance coefficients in group technology: a survey and comparative study of relational metrics. Computers and Industrial Engineering 30, 103-116.

Sarker, B.R., Islam, K.M.S., 1999. Relative performances of similarity and dissimilarity measures. Computers and Industrial Engineering 37, 769-807.

Sarker, B.R., Li, Z., 1998. Measuring matrix-based cell formation considering alternative routings. *Journal of the Operational Research Society, 49*(9), 953-965.

Sarker, B.R., Mondal, S., 1999. Grouping efficiency measures in cellular manufacturing: a survey and critical review. *International Journal of Production Research, 37*(2), 285-314.

Sarker, B.R., Xu, Y., 2000. Designing multi-product lines: job routing in cellular manufacturing systems. IIE Transactions 32, 219-235.

Seifoddini, H., 1987. Incorporation of the production volume in machine cells formation in group technology applications. Proceedings of the ◎th ICPR, October, pp. 2348-2356; or In A. Mital (ed), 1988, Recent Developments in Production Research (The Netherlands: Elsevier Science Publishers B.V.), pp. 562-570.

Seifoddini, H., 1989a. Single linkage versus average linkage clustering in machine cells formation applications. Computers and Industrial Engineering 16, 419-426.

Seifoddini, H., 1989b. A note on the similarity coefficient method and the problem of improper machine assignment in group technology applications. International Journal of Production Research 27, 1161-1165.

Seifoddini, H., Djassemi, M., 1995. Merits of the production volume based similarity coefficient in machine cell formation. Journal of Manufacturing Systems 14, 35-44.

Seifoddini, H., Djassemi, M., 1996. A new grouping measure for evaluation of machine-component matrices. *International Journal of Production Research, 34*(5), 1179-1193.

Seifoddini, H., Hsu, C.P., 1994. Comparative study of similarity coefficients and clustering algorithms in cellular manufacturing. Journal of Manufacturing Systems 13, 119-127.

Seifoddini, H., Tjahjana, B., 1999. Part-family formation for cellular manufacturing: a case study at Harnischfeger. International Journal of Production Research, 37 3263-3273.

Seifoddini, H., Wolfe, P.M., 1986 Application of the similarity coefficient method in group technology. IIE Transactions 18, 271-277.

Seifoddini, H., Wolfe, P.M., 1987. Selection of a threshold value based on material handling cost in machine-component grouping. IIE Transactions 19, 266-270.

Selim H.M., Askin, R.G., Vakharia, A.J., 1998. Cell formation in group technology: review, evaluation and directions for future research. Computers and Industrial Engineering 34, 3-20.

Selvam, R.P., Balasubramanian, K.N., 1985. Algorithmic grouping of operation sequences. Engineering Cost and Production Economics 9, 125-134.

Sevier, A.J., 1992. Managing employee resistance to just-in-time: creating an atmosphere that facilitates implementation. Production and Inventory Management Journal 33, 83-87.

Shafer, S.M., Meredith, J.R., 1990. A comparison of selected manufacturing cell formation techniques. *International Journal of Production Research, 28*(4), 661-673.

Shafer, S.M., Meredith, J.R., Marsh, R.F., 1995. A taxonomy for alternative equipment groupings in batch environments. Omega 23, 361-376.

Shafer, S.M., Rogers, D.F., 1993a. Similarity and distance measures for cellular manufacturing. Part◎. A survey. International Journal of Production Research 31, 1133-1142.

Shafer, S.M., Rogers, D.F., 1993b. Similarity and distance measures for cellular manufacturing. Part◎. An extension and comparison. International Journal of Production Research 31, 1315-1326.

Shambu, G., Suresh, N.C., 2000. Performance of hybrid cellular manufacturing systems: a computer simulation investigation. European Journal of Operational Research 120, 436-458.

Shiko, G., 1992. A process planning-orientated approach to part family formation problem in group technology applications. International Journal of Production Research 30, 1739-1752.

Silveira, G.D., 1999 A methodology of implementation of cellular manufacturing. International Journal of Production Research 37, 467-479.

Singh, N., 1993. Design of cellular manufacturing systems: an invited review. European Journal of Operational Research 69, 284-291.

Singh, N., 1996. Systems Approach to Computer-Integrated Design and Manufacturing (New York: Wiley).

Singh, N., Rajamani, D., 1996. *Cellular Manufacturing Systems: Design, planning and control*. London: Chapman & Hall.

Sneath, P.H.A., Sokal, R.R., 1973. Numerical taxonomy (San Francisco: W.H. Freeman).

Sofianopoulou, S., 1997. Application of simulated annealing to a linear model for the formulation of machine cells in group technology. *International Journal of Production Research, 35*(2), 501-511.

Sokal, R.R., Michener, C.D., 1958. A statistical method for evaluating systematic relationships. The University of Kansas Science Bulletin 38, 1409-1438.

Solimanpur, M., Vrat, P., Shankar, R., 2004. A heuristic to minimize makespan of cell scheduling problem. International Journal of Production Economics 88, 231-241.

Srinivasan, G., 1994. A clustering algorithm for machine cell formation in group technology using minimum spanning trees. International Journal of Production Research 32, 2149-2158.

Srinivasan, G., Narendran, T.T., Mahadevan, B., 1990. An assignment model for the part-families problem in group technology. International Journal of Production Research 28, 145-152.

Srinivasan, G., Narendran, T.T., 1991. GRAFICS – a nonhierarchical clustering algorithm for group technology. International Journal of Production Research 29, 463-478.

Srinivasan, G., Zimmers, E.W., 1998. Fractional cell formation – issues and approaches. International Journal of Industrial Engineering 5, 257-264.

Steudel, H.J., Ballakur, A., 1987. A dynamic programming based heuristic for machine grouping in manufacturing cell formation. Computers and Industrial Engineering 12, 215-222.

Suer, G.A., Cedeno, A.A., 1996. A configuration-based clustering algorithm for family formation. Computers and Industrial Engineering 31, 147-150.

Tam, K.Y., 1990. An operation sequence based similarity coefficient for part families formations. Journal of Manufacturing Systems 9, 55-68.

Tarsuslugil, M., Bloor, J., 1979. The use of similarity coefficients and cluster analysis in production flow analysis. In: *Proceedings 20th International Machine Tool Design and Research Conference, Birmingham, UK, September*, 525-532.

Vakharia, A.J., Kaku, B.K., 1993. Redesigning a cellular manufacturing system to handle long-term demand changes: a methodology and investigation. Decision Sciences 24, 909-930.

Vakharia, A.J., Wemmerlöv, U., 1987. A new design method for cellular manufacturing systems. Proceedings of the IXth ICPR, Cincinnati, Ohio, pp. 2357-2363.

Vakharia, A.J., Wemmerlöv, U., 1990. Designing a cellular manufacturing system: a materials flow approach based on operation sequences. IIE Transactions 22, 84-97.

Vakharia, A.J., Wemmerlöv, U., 1995. A comparative investigation of hierarchical clustering techniques and dissimilarity measures applied to the cell formation problem. Journal of Operations Management 13, 117-138.

Viswanathan, S., 1996. A new approach for solving the p-median problem in group technology. International Journal of Production Research 34, 2691-2700.

Waghodekar, P.H., Sahu, S., 1984. Machine-component cell formation in group technology: MACE. International Journal of Production Research 22, 937-948.

Wang, J., 1998. A linear assignment algorithm for formation of machine cells and part families in cellular manufacturing. Computers and Industrial Engineering 35, 81-84.

Wang, J., Roze, C., 1995. Formation of machine cells and part families in cellular manufacturing: an experimental study. Computers and Industrial Engineering 29, 567-571.

Wang, J., Roze, C., 1997. Formation of machine cells and part families: a modified p-median model and a comparative study. International Journal of Production Research 35, 1259-1286.

Wei, J.C., Gaither, N., 1990. A capacity constrained multiobjective cell formation method. Journal of Manufacturing Systems 9, 222-232.

Wei, J.C., Kern, G.M., 1989. Commonality analysis: a linear cell clustering algorithm for group technology. International Journal of Production Research 27, 2053-2062.

Wei, J.C., Kern, G.M., 1991. Discussion: reply to 'A note on a linear cell clustering algorithm'. International Journal of Production Research 29, 217-218.

Wemmerlöv, U., Hyer, N.L., 1986. Procedures for the part family/machine group identification problem in cellular manufacturing. Journal of Operations Management 6, 125-147.

Wemmerlöv, U., Hyer, N.L., 1987. Research issues in cellular manufacturing. International Journal of Production Research 25, 413-431.

Wemmerlöv, U., Johnson, D.J., 1997. Cellular manufacturing at 46 user plants: implementation experiences and performance improvements. International Journal of Production Research 35, 29-49.

Wemmerlöv, U., Johnson, D.J., 2000. Empirical findings on manufacturing cell design. International Journal of Production Research 38, 481-507.

Won, Y.K., 2000a. New p-median approach to cell formation with alternative process plans. International Journal of Production Research 38, 229-240.

Won, Y.K., 2000b. Two-phase approach to GT cell formation using efficient p-median formulation. International Journal of Production Research 38, 1601-1613.

Won, Y.K., Kim, S.H., 1997. Multiple criteria clustering algorithm for solving the group technology problem with multiple process routings. Computers and Industrial Engineering 32, 207-220.

Wu, N., Salvendy, G., 1993. A modified network approach for the design of cellular manufacturing systems. International Journal of Production Research 31, 1409-1421.

Yasuda, K., Yin, Y., 2001. A dissimilarity measure for solving the cell formation problem in cellular manufacturing. Computers and Industrial Engineering 39, 1-17.

Zhang, C., Wang, H.P., 1992. Concurrent formation of part families and machine cells based on the fuzzy set theory. Journal of Manufacturing Systems 11, 61-67.

Manufacturing the Future

Edited by Vedran Kordic, Aleksandar Lazinica and Munir Merdan

ISBN 3-86611-198-3

Hard cover, 908 pages

Publisher Pro Literatur Verlag, Germany / ARS, Austria

Published online 01, July, 2006

Published in print edition July, 2006

The primary goal of this book is to cover the state-of-the-art development and future directions in modern manufacturing systems. This interdisciplinary and comprehensive volume, consisting of 30 chapters, covers a survey of trends in distributed manufacturing, modern manufacturing equipment, product design process, rapid prototyping, quality assurance, from technological and organisational point of view and aspects of supply chain management.

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821