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.

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

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.

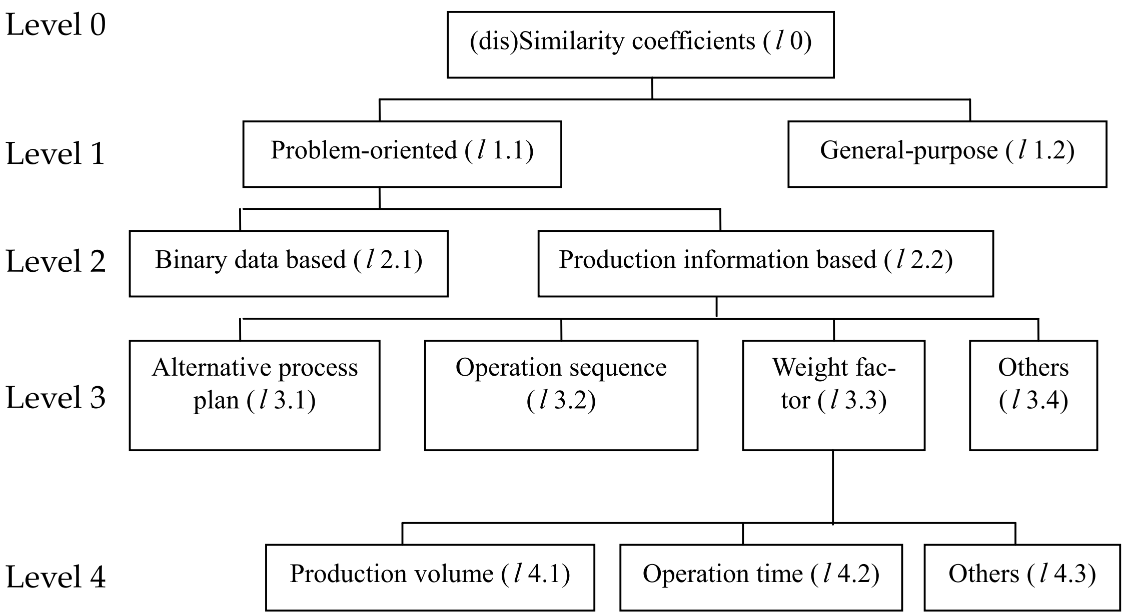

Figure 1. A taxonomy for similarity coefficients

Level 1.

l1 categorizes existing similarity coefficients into two distinct groups: problem-oriented similarity coefficients (l1.1) and general-purpose similarity coefficients (l1.2). Most of the similarity coefficients introduced in the field of numerical taxonomy are classified in l1.2 (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 (l1.1) 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 kth 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 l 2.1. 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