A comprehensive clustering algorithm for strategic analysis of supply chain networks

Mukundan Srinivasan, Young B. Moon*

Manufacturing Intelligence Laboratory, Department of Mechanical, Aerospace and Manufacturing Engineering, Syracuse University, Syracuse, NY 13244, USA

Abstract

Inventory management in supply chain networks involves keeping track of hundreds of items spread across multiple locations with complex interrelationships between them. However, it is not computationally feasible to consider each item individually during the decision making process. The use of clusters of items is preferred for the evaluation of these decisions. In addition, the use of groups of items provides management with more effective methods for characterizing and controlling system performance and results in cost savings such as group discounts. In this research, we introduce a comprehensive clustering methodology for supporting inventory management in supply chain networks. All product characteristics which have a significant impact on the performance of the supply chain are taken into account. The nodes in the network are split into subnodes prior to clustering to reduce the complexity. The average linkage clustering algorithm and the Calinski and Harabasz index are used to identify clusters of similar items. In addition, a set of heuristics is used to capture the relationships between items as specified in the bill of materials for the products. Examples are presented to demonstrate the effectiveness of the clustering methodology as well as the performance of the heuristics, by comparing the results obtained with the optimal solution. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords: Hierarchical clustering; Supply chain networks; Inventory management; Heuristics

* Corresponding author. Tel: +1-315-443-2649; fax: +1-315-443-9099.
  E-mail address: ybmoon@syr.edu (Y.B. Moon)

1. Introduction

A supply chain is a network of suppliers, manufacturing sites, assembly locations, distribution centers and customer locations through which stock keeping units (SKUs) such as components, subassemblies and finished products flow. In recent years, the advent of new markets, shortening product life cycles and stringent requirements for customer service have resulted in the globalization of companies with activities such as supply, purchasing, production and distribution being assigned to the most economical locations in the world. As a result, these companies have become more vertically integrated and their supply chains have grown into a complex network, consisting of many sites organized in multiple levels. Fig. 1 shows a supply chain network with interactions between suppliers of materials, manufacturers, distributors, transportation links and customers.

Supply chain networks are generally characterized by multiple products, several manufacturing sites and distribution warehouses, and a wide variety of equipment and operations. Production management in such networks involves keeping track of hundreds of items and encompasses a large number of decisions that affect several organizational levels. These factors, along with the uncertainty at the various levels of a supply chain such as demand, processing, supply and transportation result in high complexity in the management of material flows. A number of strategic planning and operational decisions should be made, such as (i) design and location of manufacturing sites, (ii) design of logistics systems, (iii) the appropriate inventory levels for components and finished products, (iv) the stocking policies, (v) the safety factors for each SKU, (vi) specification of customer service and performance priorities etc. However, it is not computationally or conceptually feasible to consider each SKU individually while determining the above decision variables.

From a computational point of view, the problem arises because the evaluation of alternative policies typically involves attention to complex interactions between items and the

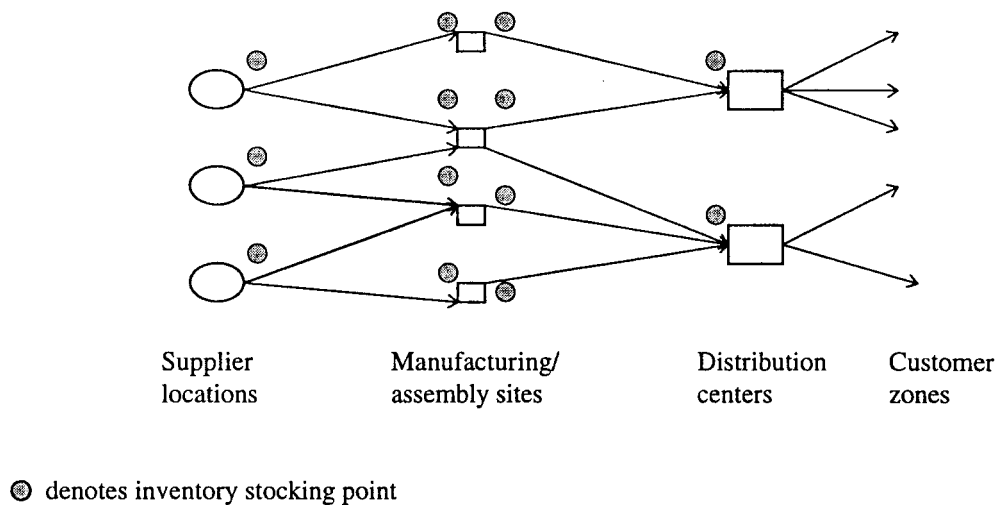| Supplier locations | Manufacturing/ assembly sites | Distribution centers | Customer zones |

◉ denotes inventory stocking point

Fig. 1. A typical supply chain network spanning multiple levels.

use of optimization and simulation. In most cases, it is not feasible to acquire all the relevant data, estimate the parameter variables and carry out the computations when the entire population of SKUs is treated on an individual basis. From a conceptual standpoint, it is difficult to monitor and evaluate system performance from a strategic perspective when the analysis is based on an individual SKU. In supply chain models involving detailed formulation with hundreds of items, managers may have difficulties in interacting with the model and comprehending the results and may get lost in the details. In addition, coordinating policies for groups of similar SKUs may result in reduced costs such as group purchase discount savings, fixed order-placing cost savings etc. The use of clusters of SKUs is suggested in this research for the evaluation of such decisions both from a computational as well as conceptual point of view, as well as for reducing costs.

The clusters of SKUs that are generated for strategic planning in a supply chain should take into consideration the following. First, all product characteristics which have a significant impact on the performance of the supply chain should be taken into account. These include the demand for the end products and the components, the manufacturing lead time, the transit time, the usage of the components in the subassemblies and assemblies to which it belongs, and the service level of the suppliers. Second, the similarity measures used to determine the clusters should be based on the application at hand. For instance, if clustering is done to determine the appropriate inventory levels for all the SKUs in a supply chain, the similarity measure should be such that items with similar inventory levels are grouped together. Finally, SKUs should be aggregated into groups such that the resulting generic policies are sufficiently close with respect to a performance criteria (e.g. total inventory cost, service performance), to policies that would have been generated if every SKU in the supply chain was treated on an individual basis.

This paper is organized as follows. Section 2 provides a survey of current literature in the application of clustering and other aggregation techniques in production and inventory management. Section 3 provides a formal problem statement while Section 4 describes a comprehensive algorithm for clustering items to support inventory management in supply chain networks. We then show the effectiveness of the methodology with an example and compare the results obtained by using clusters of items with those obtained without using clusters of items in Section 5. Finally, conclusions are presented in Section 6.

2. Literature review

Cluster analysis involves identifying groups in objects which are characterized by a set of attributes. A number of techniques are available for cluster analysis including hierarchical clustering [19,29], neural networks [21], and conceptual clustering [13,31]. These techniques have been successfully applied to a number of problems in the areas of design [30], engineering, manufacturing [18],[26] and marketing [27].

In the area of production planning and inventory management, the grouping problem has been discussed in a variety of contexts. Herron [16] focused on the use of various aggregate inventory control techniques for achieving uniform service levels. The method is based on graphical relations between item characteristics, demand predictability and cost. Mitchell [25] studied the problem of determining single-item service objectives in a multi-item inventory

system that is subject to an overall service level constraint. It was shown that sampling techniques which partition item types (e.g. by ordering them in descending order by annual dollar demand) can be used to predict stock control policies for all the items in the inventory system.

In a general context of hierarchical planning, Axsater [2] considered aggregation procedures in a K item, N machine facility in order to form K_1 ($< K$) product groups and N_1 ($< N$) machine groups to reduce the size of the planning problem. The author showed that, in general, it is not possible to find perfect aggregation and proposes an approximation scheme. The problem is solved as an optimization problem where the objective function to be minimized is the difference between the true component and capacity requirements and the aggregate component and capacity requirements.

Toczylowski and Pienkosz [32] considered the aggregation of groups of similar items for a general multi-stage lot-size scheduling problem in the presence of non zero initial stocks and variable safety stock levels. Items that have identical cost coefficients and productivity factors are grouped together into a single family. The authors present a two-stage aggregation scheme where an approximate detailed scheduling model P^* is constructed in the first stage. In the second stage, the approximate model P^* is aggregated in such a way that the optimal solution of the aggregated model can always be disaggregated into the optimal solution of the approximate detailed problem.

Ernst and Cohen [11] developed a clustering procedure for production and inventory systems and applied the technique for developing optimal inventory stocking policies in a single location distribution system. The clustering procedure adopted in their approach considers all relevant attributes such as demand, lead time, cost etc. A small sample of items is first chosen from the original list of items and an optimization clustering technique is used to generate clusters based on these items. Since the number of clusters to be generated is not known in advance, they start with the number of clusters equal to 1. Based on the clusters generated, the penalty cost for using groups is computed. The penalty cost is defined as the difference in cost of using group based inventory stocking policies versus using individual inventory stocking policies. If the penalty cost is within a prespecified threshold, clustering stops. If not, the number of clusters is incremented by 1 and the procedure is repeated all over again. The entire clustering procedure is carried out only using the small sample of items and not the original set of items. The items not belonging to the small sample are then classified into their closest clusters.

In practice, most industries use the ABC classification system [28] or one of its variants for managing inventories. The ABC classification scheme is a priority ranking system based on the annual dollar purchases of an inventoried item and classifies items into categories A, B and C. Items belonging to category A are subject to the tightest inventory control as they contribute the maximum towards the total inventory value. Items belonging to category B are subject to normal control while C items are subject to little control.

While simple to implement, the ABC method often provides unacceptable performance when evaluated with respect to cost and service measures in complex inventory systems such as those in supply chains. In addition, the ABC groups will typically exhibit poor discrimination with respect to operational attributes other than cost and sales volume.

3. Problem statement

Due to intensive competition in the present global markets, supply chain management and performance is considered an important strategic weapon to achieve and maintain competitive strength. Management of supply chains involves determining a number of strategic planning and operational decisions such as (i) the appropriate inventory levels for components and finished products, (ii) the stocking policies, (iii) the safety factors for each SKU and (iv) the specification of customer service and performance priorities. Given a supply chain network, a detailed analysis is necessary to evaluate its performance in a wide variety of scenarios. In a large enterprise with a supply chain involving a number of locations and components, analysis runs are typically executed several hundred times [1]. Due to the large number of items spread across multiple locations in a supply chain, management and analysis becomes virtually impossible both computationally and conceptually, when each SKU is considered separately. The use of groups of similar SKUs becomes imperative in this regard.

Two common assumptions among existing approaches for aggregating similar items into groups are that (i) the items are independent of each other, or (ii) the items under consideration belong to a single location. In a supply chain network, both the assumptions are invalid. The items in a supply chain network are related to one another through the bill of materials. The inventory levels and the fill rates of the items are also related by the bill of materials structure. The bill of materials for products and components may consist of multiple levels. In other words, raw components may be assembled into more complex components or subassemblies, which may be further combined to create yet more complex subassemblies and so on until the final level the bill of materials is brought together to manufacture the end products. In addition, the items are distributed across multiple locations and the same item may be manufactured in more than one location with different demands, costs, lead times and transit times.

The essence of the clustering problem is to develop a methodology for defining groups of similar SKUs which can be treated as a generic unit for strategic analysis of supply chain networks. Once the groups have been determined, the approach should provide a means for determining the individual item policies from the group values. Such a methodology should also take into account all product attributes which impact the performance of a supply chain and should incorporate constraints arising due to the bill of materials and the multiple location nature.

However, grouping similar items together is not without its disadvantages. During grouping, information regarding each SKU is lost and is replaced by aggregate information that is common to all SKUs in the group. The performance of the supply chain as measured by considering group information will always be different from that using the individual SKU information. Nevertheless, a clustering or grouping procedure is necessary in order to reduce the complexity of the supply chain and to simplify the decision making process. Moreover, for strategic planning purposes, a number of different alternatives must be examined, and faster and less expensive solutions which do not require the consideration of all factors are necessary. Thus the methodology must strike a balance between the two by providing solutions that are close to those obtained with individual SKU policies but at the same time being computationally efficient.

4. Methodology

In this section, we introduce a comprehensive clustering methodology for supporting the strategic analysis of supply chain networks where the items have attributes that are not necessarily deterministic in nature. The clustering of SKUs is carried out for one end product at a time, since the demand streams for each end product are different. The nodes in a supply chain network are split into subnodes prior to clustering and a set of heuristics are used to capture the relationships between items as specified in the bill of materials. The overall methodology is first provided in Fig. 2 and the individual steps are then described in detail.

4.1. Modification of the supply chain network

A supply chain network consists of multiple locations and items in each location form part of the bill of materials of items at other locations. The same item may be used to manufacture a number of different items at different locations with different usage ratios for each item. The usage ratio of item i with respect to item j, u_{ij}, is simply the number of units of item i required to manufacture one unit of item j. Because of the complex relationship between items, it is not possible to perform clustering while considering all the items in the supply chain at once. For instance, it is not feasible to group an item at one end of the supply chain network with

```
a)  Modify existing network such that each node contains exactly one outgoing link.

b)  While list of non-leaf nodes not empty

    For each non-leaf node
    {
        1.  Part_list = { };
        2.  Add to Part_list, all SKUs in node i, except end product, if any;
        3.  Add to Part_list, list of individual SKUs and composite SKUs from nodes that contain a link
            to node i;
        4.  Compute distance matrix of SKUs in Part_list;
        5.  Perform hierarchical clustering on SKUs in Part_list;
        6.  Determine number of clusters from the hierarchy using the Calinski and Harabasz index;
        7.  For every cluster generated
            {
                (i) Determine group average values for each cluster;
                (ii)Replace all SKUs in each cluster by composite part with group average values;
            }
        8.  Determine aggregate bill of materials structure for composite SKUs
    }
```

Fig. 2. Overall procedure for clustering SKUs in a supply chain network.

another item at the other end of the network, bypassing a number of intermediate items even if the items are similar (however unlikely it may be).

Due to the above mentioned reasons, the locations in a supply chain, which correspond to nodes in a supply chain network, are split such that each location supplies components to only one location downstream. Thus each node in the supply chain network has exactly one outgoing link.

To illustrate the concept of splitting locations, consider the example in Fig. 3. A supply chain consisting of 7 locations and 8 components is shown, along with the items supplied by or manufactured at each location. There is one finished product, 1 (indicated by the level 0 in the bill of materials in Table 1), which is assembled using 3 subassemblies 2–4 (denoted by level 1 in the bill of materials), and these subassemblies are in turn made up of other components. Note that location A supplies two components 5 and 8 to locations D and E, respectively, while location C supplies component 7 to both location D and location E. This supply chain thus consists of two locations (A and C) which supply items to more than one location. Accordingly, nodes A and C are split into subnodes as follows: Node A is split into A1 and A2 with A1 supplying item 5 to location D and A2 supplying item 8 to location E. Node C is split into C1 and C2 with C1 supplying item 7 to location D and C2 supplying item 9 to location E. The resulting network is shown in Fig. 4. A formal procedure is provided in Fig. 5.

4.2. Determination of distance matrix

The purpose of clustering is to develop a methodology for defining groups of similar SKUs which can be treated as a generic unit for evaluating the performance of supply chains. The performance of a supply chain is usually specified in terms of the overall inventory capital, the level of customer service attained, cycle time, etc. In this research, the total inventory capital held in the supply chain is used as the performance measure.

The contribution to the overall inventory capital for each SKU is a function of the average cost of the SKU $C(i)$ and the reorder point. The reorder point of each SKU is simply a function of the mean and variance of the demand over the lead time and the safety factor.
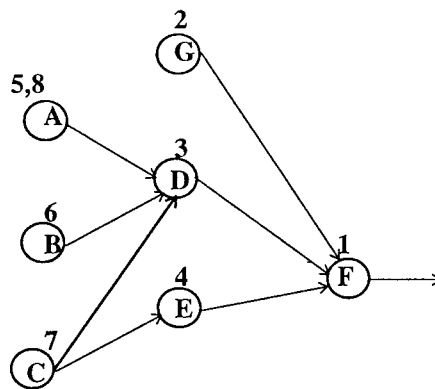

Fig. 3. Supply chain network to illustrate splitting of nodes.

Table 1
Bill of materials to illustrate splitting of locations

Level	Item			Location
0	1			F
1		2		G
1		3		D
2			5	A
2			6	B
2			7	C
1		4		E
2			7	C
2			8	A

Hence, the attributes that impact the similarity between two SKUs are (i) the usage of the SKU with respect to the finished product (which captures the demand), (ii) the manufacturing lead time of the SKU, (iii) the transit time of the SKU, and (iv) the average cost.

Suppose the lead time of SKU i is a random variable with mean $\mu_1(i)$ and variance $\sigma_1^2(i)$. Furthermore, suppose that the demand per unit time has mean $\lambda(i)$ and variance $v^2(i)$. For SKU i, it can be shown [14] that the demand during lead time has mean and variance

$$\mu_i = \lambda(i)\mu_1(i), \tag{1}$$

$$\sigma_i^2 = \mu_1(i)v^2(i) + \lambda^2(i)\sigma_1^2. \tag{2}$$

The reorder point for SKU i, R_i, can then be determined as

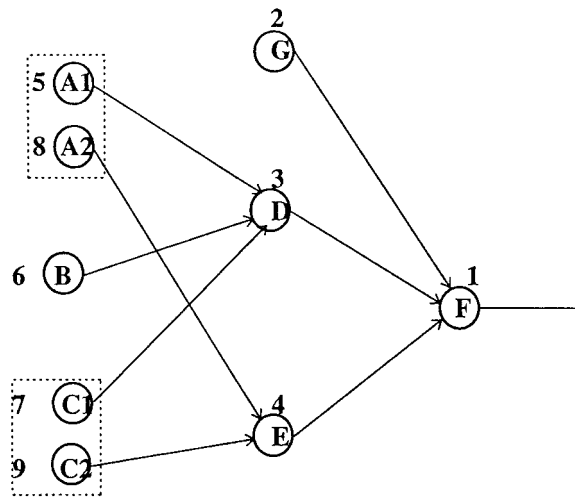Fig. 4. Modified supply chain network.

For each node i in the supply chain network

   if there exists more than one outgoing link

    {

        1. split node i into sub-nodes such that

            number of sub-nodes = number of outgoing links;

       2. assign SKUs of node i to each sub-node j such that

            all SKUs in sub-node j have the same outgoing link;

    }

Fig. 5. Procedure for modifying supply chain network prior to clustering.

$$R_i = \mu_i + s_k \sigma_i, \tag{3}$$

where s_k is the safety factor that is common for all the SKUs in a cluster.

Since the inventory capital is dependent on the cost of the SKU and the reorder point, we use the following measure $D(ij)$ to determine the distance between SKU i and j:

Given:

list of finished products and their associated bill of materials:

$L[i]$ = location of SKU i;

$\mu_{\mathrm{u}}[ij]$ = mean of number of units of SKU i required for one unit of SKU j;

$\sigma_{\mathrm{u}}^2[ij]$ = variance of number of units of SKU i required for one unit of SKU j;

$\mu_{\mathrm{ue}}[i]$ = mean of number of units of SKU i required with respect to the end product;

$\sigma_{\mathrm{ue}}^2[i]$ = variance of number of units of SKU i required with respect to the end product;

$\mu_{\mathrm{l}}[i]$ = mean of manufacturing lead time for SKU i;

$\sigma_{\mathrm{l}}^2[i]$ = variance of manufacturing lead time for SKU i;

$\mu_{\mathrm{t}}[i]$ = mean of transit time for SKU i;

$\sigma_{\mathrm{t}}^2[i]$ = variance of transit time for SKU i;

$C[i]$ = average cost of SKU i.

Define

$$d(ij) = \left| \mu_{\mathrm{ue}}[i](\mu_{\mathrm{l}}[i] + \mu_{\mathrm{t}}[i])C[i] - \mu_{\mathrm{ue}}[j](\mu_{\mathrm{l}}[j] + \mu_{\mathrm{t}}[j])C[j] \right| + C[i]\left((\mu_{\mathrm{l}}[i] + \mu_{\mathrm{t}}[i])\sigma_{\mathrm{ue}}^2[i] \right.$$

$$\left. + \mu_{\mathrm{ue}}^2(i)(\sigma_{\mathrm{l}}^2[i] + \sigma_{\mathrm{t}}^2[i])\right)^{1/2} + C[j]\left((\mu_{\mathrm{l}}[j] + \mu_{\mathrm{t}}[j])\sigma_{\mathrm{ue}}^2[j] + \mu_{\mathrm{ue}}^2(j)(\sigma_{\mathrm{l}}^2[j] + \sigma_{\mathrm{t}}^2[j])\right)^{1/2}. \tag{4}$$

Then,

$$D(ij) = \frac{d(ij)}{\max(d(ij))}, \tag{5}$$

where $\max(d(ij))$ is the distance between the most dissimilar pair of SKUs i and j.

The first term of $d(ij)$ captures the distance between the mean values of the attributes of SKUs i and j, while the second and third terms capture the distance between the variances of attributes of SKUs i and j.

4.3. Clustering of stock keeping units

The clustering step performs the important task of grouping the SKUs to reduce the overall size of the problem. Since one of the main purposes of clustering is to reduce the computational burden, the clustering methodology should be chosen so that it does not offset the improvement in speed obtained by considering each SKU individually. For this reason, we choose a hierarchical clustering technique instead of nonhierarchical clustering.

Given n SKUs, hierarchical clustering consists of a series of clusterings of the SKUs, from the initial situation when each SKU may be considered a singleton cluster to the other extreme when all SKUs belong to one cluster. Hierarchical clustering techniques may be subdivided into agglomerative methods which proceed by a series of successive fusions of the n SKUs into clusters, and divisive methods which partition the set of n SKUs successfully into finer partitions. Hierarchical classifications are usually represented by a hierarchic tree or a dendogram, and a slice through the dendogram defines a partition of the SKUs into clusters.

In addition to its computational efficiency, hierarchical clustering techniques offer a number of desirable qualities. Each hierarchical clustering run generates multiple solutions, with different number of clusters, and by using one of the many available procedures for determining the optimal level to partition the hierarchy, the number of clusters can be obtained. Milligan and Cooper [23] evaluated a number of different such procedures and showed that, among others, the Calinski and Harabasz index, the C index, and the Gamma index indicated excellent recovery of the true cluster structure at the correct hierarchy level. Klastorin [20] and Milligan [24] noted that a major disadvantage of nonhierarchical methods is that their performance in recovering known clusters is typically not as good as the hierarchical procedures unless prudent initializations are made.

We use the well known agglomerative hierarchical technique called the average linkage method. In this method, the two closest SKUs (that is, with smallest inter-SKU dissimilarity) are joined, leaving $n-1$ clusters, one of which contains two SKUs whereas the others still have only a single SKU. In succeeding steps, the two closest clusters are merged to form the hierarchy. However, this calls for a definition of the dissimilarity between clusters. In this research, the dissimilarity between two clusters is defined as the distance between the cluster means.

Once the hierarchy has been constructed, a stopping rule is necessary to determine the correct number of clusters in the data. The Calinski and Harabasz index has been chosen for this purpose because it has been shown to work fairly well in determining the optimal partition.

Let

n = number of SKUs, with index $1, \ldots, n$;

a = number of attributes representing each SKU, with index $1, \ldots, a$;

c = number of clusters for partitioning n, with index 1, ..., c;

$[X] = n \times a$ SKU attribute matrix;

$[\boldsymbol{X}] = c \times a$ matrix of cluster attribute means;

$[M] = n \times c$ membership matrix, with $m_{ik} = 1$ if SKU i belongs to cluster k;

$[D] = n \times n$ distance matrix with d_{ij} = distance between SKU i and SKU j;

$[B] = a \times a$ matrix of between cluster sum of squares and cross products;

$[W] = a \times a$ matrix of within cluster sum of squares and cross products;

where

$$[B] = ([M][\boldsymbol{X}])^{\mathrm{T}}([M][\boldsymbol{X}]),$$
$$[W] = [X]^{\mathrm{T}}[X] - [B]. \tag{6}$$

The Calinski and Harabasz index is defined as

$$\text{Calinski and Harabasz index} = \frac{\text{trace } B/c - 1}{\text{trace } W/n - c} \tag{7}$$

(the trace of a square matrix is defined as the sum of its diagonal elements).

4.4. Determination of aggregated bill of material structure

Once a set of clusters has been obtained for all SKUs in a nonleaf node and SKUs of all nodes which supply to this node, the aggregated bill of materials structure should be determined. A number of different scenarios can arise depending which SKUs are grouped together in a cluster. A set of heuristics are used to identify the appropriate bill of materials hierarchy. The remainder of this section illustrates the use of the heuristics using an example.

Consider the example presented in Table 2. A bill of materials structure consisting of 6 items is shown. There is one finished product 1, which is assembled using two items 2 and 3. Item 2 is assembled using items 4 and 5 while item 3 is made up of item 6. The location of each item is also provided in Table 2.

Table 2
Example bill of materials to illustrate heuristics for determining aggregate structure

Level	Item			Location
0	1			A
1		2		B
2			4	B
2			5	C
1		3		B
2			6	C

4.4.1. Case 1. The child SKUs of a parent are clustered together. All the other SKUs are retained as it is

This happens when SKUs 4 and 5 in the example are grouped together into one composite SKU [4/5]. The location of this SKU can either be B or C and can be chosen randomly when there is a tie. The choice of the location only serves as an identifier to the SKU and does not impact the values of the attributes such as cost, manufacturing lead time and transit time as they are computed based on the individual SKUs comprising the composite SKU. This composite SKU is assigned to the parent SKU 2. The aggregate BOM is now shown in Table 3.

4.4.2. Case 2. The parents SKUs are grouped. The child SKUs are retained as it is

This scenario can arise when SKUs 2 and 3 are grouped into one cluster [2/3] while SKUs 4, 5 and 6 are retained as it is. The assignment of child SKUs to their parents is straightforward with all the three SKUs 4, 5 and 6 being assigned to the composite SKU [2/3]. Table 4 shows the resulting aggregate bill of material.

4.4.3. Case 3. The parent SKUs are not grouped. The child SKUs are grouped into clusters

This can occur when SKUs 4–6 are clustered together to form a composite SKU [4/5/6] while their parent SKUs 2 and 3 are not grouped. In this case, we use one of two heuristics to determine the parent SKU to assign the composite SKU to.

Heuristic 1. Assign the composite SKU to the parent with higher cost in order to maintain the cost structure. For example, if SKU i is used to manufacture SKU j, this heuristic ensures that the cost of SKU i is never greater than that of SKU j. In this example, if the cost of SKU 2 is less than that of SKU 3, the composite SKU [4/5/6] is assigned to SKU 3. Table 5 shows the resulting bill of materials structure.

Heuristic 2. Assign composite SKUs to the parent that contains the most number of children from the composite SKU. In this example, SKU 2 is the parent of two SKUs that make up the composite SKU (4 and 5) while SKU 3 is the parent of one SKU. Hence the composite SKU is assigned to SKU 2. The ensuing bill of materials is shown in Table 6.

Table 3
Resulting bill of materials for case 1

Level	Item		Location
0	1		A
1		2	B
2			[4/5]	B
1		3	B
2			6	C

Table 4
Resulting bill of materials for case 2

Level	Item		Location
0	1		A
1	[2/3]		B
2		4	B
2		4	C
2		6	C

4.4.4. Case 4. The same SKU is used to manufacture two or more parent SKUs

Consider the example shown in Table 7. SKU 4 is used in both SKUs 2 and 3. If SKUs 2 and 3 are clustered together into one composite SKU [2/3], then SKU 4 is assigned to the composite SKU with an aggregated usage count. If SKUs 2 and 3 are not clustered together, SKU 4 is retained as it is.

The overall clustering methodology as outlined in Fig. 2 is carried out for each end product. The output of this procedure will be a reduced bill of materials where similar SKUs are replaced by the composite SKUs.

4.5. Determination of inventory levels

An inventory optimization system can be used to determine the optimal reorder points for the reduced BOM structure. In our research, we use a supply chain network model with base stock control developed by Ettl et al. [12]. We use this model and provide as input the reduced BOM, and the cost and lead times associated with the composite items. The resulting group safety factors s_k, which are generated for each composite SKU k in the BOM are used in conjunction with the individual SKU specific mean and standard deviation of demand over lead time μ_i and σ_i, respectively, to determine the SKU specific reorder point

$$R_{ik} = \mu_i + s_k\sigma_i. \tag{8}$$

The use of the safety factors of the composite SKU in determining the SKU specific reorder point reduces the computational burden since only one safety factor must be determined for all SKUs in a cluster.

Table 5
Resulting bill of materials for case 3, using heuristic 1

Level	Item		Location
0	1		A
1	2		B
1	3		B
2		[4/5/6]	B

Table 6
Resulting bill of materials for case 3, using heuristic 2

Level	Item		Location
0	1		A
1		2	B
2		[4/5/6]	B
1		3	B

Table 7
Example bill of materials to illustrate case 4

Level	Item			Location
0	1			A
1		2		B
2			4	C
1		3		B
2			4	C

In many real life situations, products are made of subassemblies that in turn consist of common components. In other words, the same component can be part of a number of higher level items. An obvious advantage of having common components is the cost savings due to the reduction in inventory levels resulting from the 'risk pooling' effect (which can be significant in a stochastic environment). However, when applying clustering methodology, it is possible that the same component can be a part of multiple groups depending on the bill of materials structure. This results in the component having more than one safety factor, depending on the number of groups to which it belongs.

To illustrate, consider the example shown in Table 8. Item 4 is common to both items 2 and 3. If, during clustering, item 4 is grouped with item 2 in one instance and with item 3 in another instance, it results in the aggregated BOM as shown in Table 9. This leads to two

Table 8
Example bill of materials to illustrate commonality

Level	Item			Location
0	1			A
1		2		B
2			4	C
2			5	B
1		3		D
2			4	C
2			6	E

Table 9
Aggregate bill of materials to illustrate commonality, with item 4 grouped with both items 2 and 3

Level	Item	Location
0	1	A
1	[2/4]	B
2	5	B
1	[3/4]	D
2	6	E

safety factors for item 4, one for the composite part [2/4] and the other for the composite part [3/4]. In such cases, we choose the maximum of the two safety factors along with the item specific mean and standard deviation of the demand over lead time to determine the reorder point. For instance, if the group safety factor for the composite part [2/4] is s_{k1} and that of [3/4] is s_{k2}, and the mean and standard deviation of the demand over lead time for item 4 is μ_4 and σ_4, respectively, the reorder point R_i for item 4 is

$$R_i = \mu_4 + \max(s_{k1}, \text{trace } s_{k2})\sigma_4. \tag{9}$$

5. Example

In this section we evaluate the proposed clustering methodology and demonstrate its ability to provide solutions that are close to those obtained by treating each item individually. We use the supply chain inventory optimizer developed by Ettl et al. [12] at IBM study the impact of
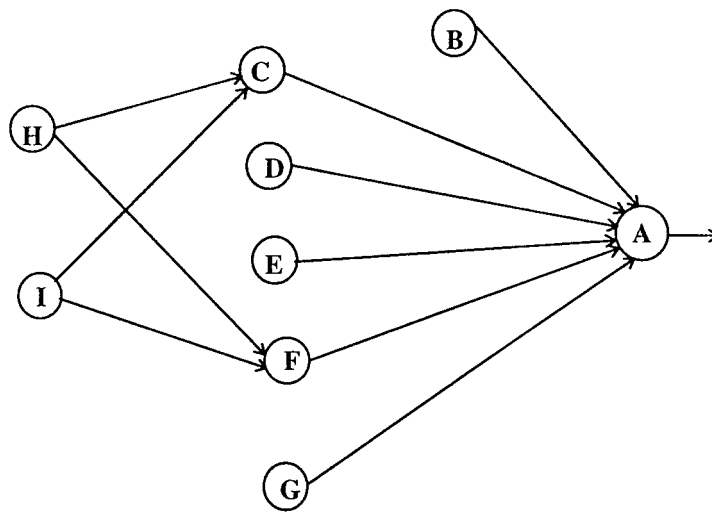

Fig. 6. Example supply chain network.

different customer service levels on the total inventory capital in the supply chain and compare the results obtained by using clusters of items versus those obtained without using clustering. We also compare the distribution of inventory across the various locations in the supply chain for both cases.

To illustrate our approach, we used a supply chain network consisting of nine locations with two finished products, both manufactured at location A as shown in Fig. 6. The first product consists of 20 components while the second one is made up of 23 components. There were 12 components that were common to both the finished products. Hence the total number of unique components was 31. The distribution characterizing the lead times and the transit times were assumed to be normally distributed for some components and constant for the others. The rationale is that the above example is derived from the computer industry and the normal distribution is most commonly used [12] to capture uncertainties in demands, lead times and transportation times.

The clustering is performed for one finished product at a time as the demand streams for the two products are different. As mentioned earlier, the Calinski and Harabasz index was used to determine the number of clusters each time the clustering methodology was used. In this example, the clustering for the two products resulted in 12 and 14 composite items.

The first experiment involved the impact of different customer service levels on the total inventory capital in the supply chain. The supply chain inventory optimizer was used to determine the overall inventory capital in the supply chain for the case when the parts were grouped into clusters and for the case when each part was treated individually. The customer service level was systematically set at 75, 80, 85, 90, 95, 98 and 99%. Fig. 7 shows a plot of the effect of customer service-level requirement on total inventory capital. The results obtained
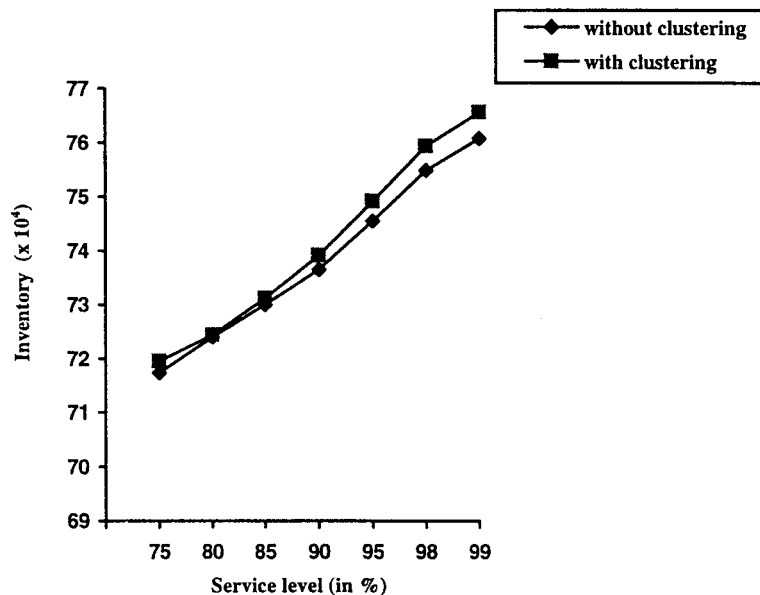

Fig. 7. Effect of customer service level on total inventory capital.

using the groups are very close to those obtained without using clustering (the optimal solution). The average deviation from the optimal solution was 3.57%. The maximum deviation was 6.38%, at a customer service-level of 99% while the minimum deviation from the optimal solution was 0.78%, achieved at a service level of 80%.

The second experiment involved an analysis of the breakdown of the inventory capital in the supply chain by location. Table 10 shows the inventory capital broken down by location, for the case when clustering was used as well as the case for which no clustering was used. The customer service level was kept constant at 95%. It can be seen that the inventory at each location for the case when clustering is used is either equal to or greater than the optimal inventory levels obtained without using clustering. This shows that the reorder points obtained by using groups of items satisfies the required customer service levels.

6. Conclusions

Supply chain management and analysis involves determining a number of decisions such as (i) the appropriate inventory levels for components and finished products, (ii) the safety factors for each item at each location, and (iii) the specification of customer service and performance priorities. In a large enterprise with a supply chain involving hundreds of items, a detailed analysis becomes virtually impossible when each item is considered individually. The use of clusters of similar items becomes necessary in this regard.

In this research, a comprehensive methodology has been presented for clustering similar items in a supply chain network. Two important aspects of supply chains viz. the multi-item multi-location nature, and the existence of complex relationships between items are taken into account during clustering. Once a collection of items have been clustered into groups, a set of heuristics are used to determine the aggregate bill of materials structure after clustering. The proposed approach is evaluated by comparing the total inventory capital for the supply chain network obtained using clustering to those obtained without clustering. Results from an

Table 10
Inventory capital by location (at 95% customer service level)

Location	Inventory capital	
	Without clustering	With clustering
A	134 511	136 373
B	7373	7409
C	47 605	48 484
D	34 150	34 150
E	78 778	78 778
F	151 473	152 328
G	9560	9560
H	289 133	289 133
I	2478	2478

example supply chain network show that the given methodology can provide solutions that are close to that of the optimal solution generated without using clustering.

Future work will focus on the impact of demand forecast accuracy as well as the impact of common components on the performance of the supply chain. One way to study the effect of forecast accuracy would be to keep the mean forecast constant and increase its variance systematically, and studying its impact on the total inventory capital. A comparison of the inventory levels for the case where there are components common across end products and for the case where there are no common components would provide a sound basis for evaluating the effect of common components.

In the example presented above, there was about a 50% reduction in the number of components (considering one end product at a time) by using clustering to form aggregate parts. For studies involving a few hundred parts or more, it would be more useful to obtain a further reduction in the number of parts. Future work would include evaluating other approaches such as the C index [17] and the Gamma statistic [3] in determining the optimal number of clusters, as well as performing sensitivity analyses on these measures.

Acknowledgements

References

[1] Arnzen BC, Brown GG, Harrison TP, Trafton LL. Global supply chain management at digital equipment cor-poration. Interfaces 1995;25:69–93.

[2] Axsater S. Aggregation of product data for hierarchical planning. Operations Research 1981;29:744–56.

[3] Baker FB, Hubert L. Measuring the power of hierarchical cluster analysis. Journal of the American Statistical Association 1975;70:31–8.

[11] Ernst R, Cohen MA. Operations related groups (ORGs): a clustering procedure for production/inventory sys-tems. Journal of Operations Management 1990;9:574–98.

[12] Ettl M, Feigin GE, Lin GY, Yao DD. A supply network model with base-stock control and service require-ments. Research report. IBM TJ Watson Research Center, USA, 1996.

[13] Fisher DH. Knowledge acquisition via incremental conceptual clustering. Machine Learning 1987;2:139–72.

[14] Hadley G, Whitin TM. Analysis of inventory systems. Englewood Cliffs, NJ: Prentice-Hall, 1963.

[16] Herron DP. A comparison of techniques for multi-item inventory analysis. The Joint ORSA/TIMS Meeting. San Francisco, CA, 1977.

[17] Hubert LJ, Baker FB. The comparison and fitting of given classification schemes. Journal of Mathematical Psychology 1977;16:233–53.

[18] Hyer NL, Wemmerlov U. Group technology in the US manufacturing industry: a survey of current practices. International Journal for Productions Research 1989;27:1287–304.

[19] Jain AK, Dubes RC. Algorithms for clustering data. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[20] Klastorin TD. The p-median problem for cluster analysis: a comparative test using the mixture model approach. Management Science 1985;31:84–95.

[21] Kohonen T. Self-organization and associative memory. New York: Springer-Verlag, 1988.

[23] Milligan GW, Cooper MC. An examination of procedures for determining the number of clusters in a data set. Psychometrika 1985;50:159–79.

[24] Milligan GW. An examination of the effect of six types of error perturbation on fifteen clustering algorithms. Psychometrika 1980;45:325–42.

[25] Mitchell JC. Multi-item (s, S) inventory systems with service objective. Report No. 26. University of North Carolina, Chapel Hill, 1983.

[26] Moon YB, Chi SC. Generalized part family formation using neural network techniques. Journal of Manufacturing Systems 1992;11:149–59.

[27] Mulhern FJ, Carpara RJ. A nearest neighbor model for forecasting market response. International Journal of Forecasting 1994;10:191–207.

[28] Peterson R, Silver EA. Decision systems for inventory management and production planning. New York: John Wiley and Sons, 1985.

[29] Romesburg CH. Cluster analysis for researchers. Belmont, CA: Lifetime Learning, 1984.

[30] Srinivasan M, Moon Y.B. A framework for a goal driven approach to group technology applications using conceptual clustering. International Journal of Production Research 1996 (to be published).

[31] Stepp III RE, Michalski RS. Conceptual clustering: inventing goal-oriented classifications of structured objects. In: Michalski RS, Carbonel JG, Mitchell TM, editors. Machine learning. Los Altos, CA: Morgan Kaufmann, 1986. p. 471–98.

[32] Toczylowski E, Pienkosz K. Restrictive aggregation of items in multi-stage production systems. Operations Research Letters 1991;10:159–63.