A Constrained Clustering Algorithm for Spare Parts Segmentation

Manuel Rossetti, Ph.D., P.E., Ashish Achlerkar, M.S.I.E.
Department of Industrial Engineering
University of Arkansas
Fayetteville, AR 72701, USA

Abstract

This paper explores the application of constrained clustering algorithms to inventory grouping based problems. Our research in this area is based upon a segmentation methodology that considers operationally relevant part attributes. In this paper, we present a brief review of literature on multi-echelon inventory models, and clustering applications to inventory segmentation. Finally, we discuss our segmentation methodology from the perspective of our experiments on a single-indentured, single-echelon, multi-item inventory system. While bringing to focus many important trade-offs in the inventory segmentation process, we also recommend future work to achieve optimal part groups and extend this segmentation methodology to multi-indentured, multi-echelon inventory systems.

Keywords

multi-indentured, multi-echelon, constrained clustering, UPGMA.

1. Introduction

This paper discusses a constrained clustering methodology applied to address the segmentation of inventory into more easily controlled categories. We propose a segmentation methodology that uses standard clustering techniques to group parts in a single-echelon, single-indentured inventory system. In our current work, we have tested our segmentation methodology on a single-echelon, single-indentured inventory system so as to bring to focus many important trade-offs involved in the inventory segmentation process. Our segmentation methodology considers several functionally and operationally important part attributes for grouping parts. The major advantages associated with inventory segmentation are reduced computational time for calculating inventory policies and enhanced convenience in managing the spare parts. We have made an effort to design a means of implementing the concept of differentiated customer service based on the relative importance of the parts in the inventory system. We used the Unweighted Pair Group Method Using Arithmetic Averages (UPGMA) clustering technique to group those parts that are similar in characteristics. The similarity or dissimilarity between parts is determined using a set of part attributes that are operationally significant. The choice of operationally relevant attributes such as part essentiality and part criticality for grouping parts make our methodology significantly different and more practical than the conventional grouping techniques like ABC analysis or group technology which only consider the cost, volume, and physical attributes of parts. We have used a multi-product backorder model proposed by Hopp and Spearman [5] to calculate the inventory control policies for the parts. This model minimizes the total inventory cost while satisfying the constraints for average order frequency and total backorder level. In our research, we have built upon the prior work by Cohen and Ernst [3] in which they have proposed an inventory segmentation methodology considering operationally relevant part attributes.

In order to test our segmentation methodology we have also designed a data generation procedure that can generate datasets of different characteristics. This data generation procedure is designed to produce different attribute values for every item in the inventory system. By varying the proportion of items with a given characteristic for each attribute, we can obtain datasets of different characteristics for testing purposes. We used the single-echelon, single-indentured multi-product backorder model and the UPGMA clustering algorithm to compare the computational time, cost, fill-rate and customer wait time for individual items and the groups of items formed using our segmentation methodology. The results obtained prove the feasibility of our methodology and are discussed in the last section of this paper. In section 3 of this paper we briefly describe our segmentation methodology. In the next section, we present a brief literature review on multi-echelon inventory models and applications of clustering techniques to inventory segmentation problems.

2. Literature Review

Research in multi-echelon inventory models dates back to 1968 when Sherbrooke [8] developed a mathematical model called METRIC for deciding the control policies of multi-echelon, multi-item systems with repairable items. The METRIC model calculates inventory control policies for repairable items at both the retail and the wholesale levels while minimizing the expected backorders. There were attempts to apply more restrictive assumptions and develop more exact solutions to the multi-echelon problem that led to the development of VARIMETRIC. Some of the key assumptions that these models make are an (*s-1, S*) inventory policy, Poisson demand, uncapacitated repair, and no lateral resupply between bases. These models entail a high computational time which becomes a reckonable disadvantage when modeling complex multi-indentured, multi-echelon inventory systems. Sherbrooke [9] describes the improved approximation for the combination of multi-indenture and multi-echelon using the VARI-METRIC model.

In our research, we have used the multi-product backorder model proposed by Hopp and Spearman [5] to calculate and compare the inventory policies for items or groups of items formed after clustering on the basis of similarities. This inventory model can be described as shown in Exhibit 1:

Exhibit 1: Multi-product backorder model

Minimize *Inventory holding cost*
Subject to: *Average order frequency* $\leq F$
Total backorder level $\leq B$
Where F = Desired average order frequency and
B = Desired total backorder level

Hopp and Spearman [5] suggest an iterative procedure to solve this optimization problem. We refer the readers to page number 605 of Hopp and Spearman [5] for the details on this iterative procedure. From the iterative procedure suggested by Hopp and Spearman [5], we calculate optimal inventory policies for each individual item in the inventory system or groups of items obtained after using the UPGMA clustering algorithm. It seems very obvious that in order to use this iterative procedure for calculating inventory policies for huge inventory systems consisting of hundreds of thousands of items, much computational time will be expended. This necessitates the segmentation of these items into groups and a methodology to decide group policies for these groups. This is expected to reduce the computational time and managerial cost associated with deciding inventory control policies and also leads to managerial convenience while physically managing parts.

Next we discuss briefly the concept of clustering and the applications of clustering in the area of inventory management. Clustering involves the grouping of objects together based on similarities between the objects. The basic clustering procedure uses attributes to group objects together. The values of attributes for each object are compared with the values of attributes for every other object and a dissimilarity coefficient is computed based on the relative differences between the attribute values of each pair of objects. In order to obtain the most representative clusters from a given set of objects it is important that the within cluster dissimilarity between the objects be minimized while the between group dissimilarity between the objects be maximized. Various clustering algorithms such as the UPGMA, K-means algorithm etc; are used widely in areas like taxonomy, psychiatry, archaeology and agriculture.

Cohen and Ernst [3] worked out a methodology to group spare parts based on statistical clustering constrained by operational performance criteria. The clustering technique presented by them considers many attributes used in functional grouping going beyond the conventional cost and volume attributes used in ABC analysis. The paper uses a classical statistical grouping problem aimed towards the reduction of a sample of items to groups with the following properties:-

1. Minimum within group variance for each variable.
2. Maximum between group variance for each variable.
3. Limited number of groups.

The paper uses standard statistical techniques to obtain Trace (B) which is a between cluster sum-of-squares and cross products matrix and represents the between cluster variability (It measures deviations of each cluster's means from the sample means). Trace (T) is the total sample sum-of-squares and cross products matrix. The proportion of variance accounted for by the clusters as described by Cohen and Ernst [3] is given by,

$$R^2 = \frac{trace(B)}{trace(T)} \quad and \quad D = \frac{R^2}{1-R^2}$$

Cohen and Ernst [3] define D as the objective function to be optimized. D as defined in the paper is the degree of dissimilarity amongst groups. The statistical clustering problem as described in the paper then requires finding the set of clusters $\{S_1 \ldots\ldots\ldots S_Q\}$ to maximize D subject to a constraint on the maximum number of groups defined by the authors as $\bar{q}$. The optimization problem as described in the paper is:

$$Maximize\ G(q; S_1 \ldots\ldots. S_Q) = D\ ;$$

$$Subject\ to:\ q \le \bar{q}$$

In the next section we describe our segmentation methodology and its implementation for testing purposes.

3. Segmentation methodology and Implementation

The approach that we adopted to address the spare parts segmentation issue uses the multi-product backorder model and the UPGMA clustering technique discussed in the previous section. Our segmentation methodology consists of 11 steps which need to be applied in a sequence so as to ultimately get an optimal grouping of parts. These steps presented in Exhibit 2 become self explanatory with the information provided in the previous sections of this paper. We have not delved into the details of the UPGMA clustering method due to lack of space but the attributes that were used to group the parts using the UPGMA clustering technique have been mentioned and defined in Table 1. The multi-product backorder model was implemented in the Java programming language while the clustering procedure was implemented in the Statistical Analysis Software (SAS). The data generation procedure that was introduced in the Section 1 of this paper was used to obtain datasets of different sizes and characteristics for testing purposes. Initially, to prove and test the feasibility of our segmentation methodology, we have considered items characterizing a single-echelon, single-indentured inventory system.

Exhibit 2: Segmentation methodology

1. Select a dataset of items from a single-echelon, single-indentured inventory system.
2. Decide optimum individual control policies using the multi-product backorder model for each of these items.
3. Calculate the total holding cost, fill rate, customer wait time and execution time for the solution obtained in Step-2.
4. Identify important decision variables in the multi-product backorder model as the attributes of the items.
5. Apply the UPGMA clustering technique to cluster the items based on the attributes identified in Step-4. Find the final clustering tree by deciding on the maximum value of the Euclidean distance coefficient at which the items can merge.
6. Using the multi-product backorder model to decide optimum group control policies for each of those items which merge at the highest value of the Euclidean distance coefficient in the final clustering tree.
7. Calculate the total holding cost, fill rate, customer wait time and execution time for the solution obtained in Step-6
8. Calculate the penalty cost for grouping by taking the difference between the costs obtained in Steps 3 and 7.
9. If the penalty cost is less than the allowable value stop and go to Step-10, else reduce the maximum value of the Euclidean distance coefficient at which the items can merge by one level and go to Step-6.
10. Validate the Cluster solution.
11. Document the results.

The Euclidean distance coefficient mentioned is used as a measure of dissimilarity between items while grouping them together. The higher the value of the coefficient at which the items merge, the higher is the dissimilarity between them. Hence the items that merge at a higher level of the Euclidean distance coefficient show a loss of their identity. Hence this allowable dissimilarity between the items in the same cluster or group constrains the benefits that can be obtained via clustering items. This within-group dissimilarity is evident via an increased penalty cost for grouping. Our segmentation methodology provides a way to achieve a trade-off between these conflicting objectives. For details on the UPGMA clustering technique we refer the readers to Romesberg [6]. In our implementation of the multi-product backorder model, we have used computational methods so as to achieve computational fairness and efficiency while comparing the clustered solution with the unclustered one. Also we need a justifiable starting value of the order cost and the backorder cost while performing iterations to satisfy the average order frequency and the total backorder level constraints in the multi-product backorder model. This is necessary to ensure that the clustered and the unclustered datasets are being tested for their computational supremacy on an unbiased platform. We have also designed a procedure to obtain reasonable starting values for these costs for a given dataset. This ensures a fair comparison between the unclustered solution giving individual policies for each item with the clustered solution giving a group policy for groups of items formed using our segmentation methodology.

Once the items are grouped using the clustering algorithm, we need to have a method to decide the values of the attributes for these groups. There can be several ways of doing this. The values of each of the attributes for the groups can be the mean of the values of the corresponding attributes for each individual item in that particular group. Instead of considering the mean, we could consider the minimum, maximum, mode or median. Selecting each of these policies may have different implications and we are still researching on this aspect of segmentation. In our current tests, we have considered the mean as the group policy deciding criteria. Investigating how our grouping methodology would perform in the scenario of minimum and maximum attribute value policy and also considering the median and the mode attribute values policy is one of the prime areas we are focusing on in future research.

Table 1: List of attributes used for UPGMA clustering

No.	Attribute	Description	Type
1	Unit cost of the item	This is the cost in dollars of the item	Quantitative
2	Replenishment lead time	This is the average annual demand for the item	Quantitative
3	Mean lead-time demand	This is the average time required for the item to be delivered to the highest echelon from the manufacturing capacity	Quantitative
4	Variance of lead-time demand	This is the average time required for the item to be delivered to a lower echelon from its higher echelon	Quantitative
5	Desired Fill rate	This is the desired fill rate for a given item	Quantitative
6	Item essentiality code	This is a code which indicates the essentiality of the item for the system under consideration	Qualitative
7	Item criticality code	This is a code which indicates the criticality of the item for the operations of the system	Qualitative
8	Part commonality	This is the number of items that are children to a particular item	Quantitative
9	Part Size	This is the volume occupied by the item in cubic meters	Quantitative

In the next section, we present a summary of important results obtained by testing our segmentation methodology on a dataset of items or parts characterizing a single-echelon, single-indentured and multi-item inventory system. The results bring to light many important trade-offs involved in the inventory segmentation process. Some of the intuitive results expected are:

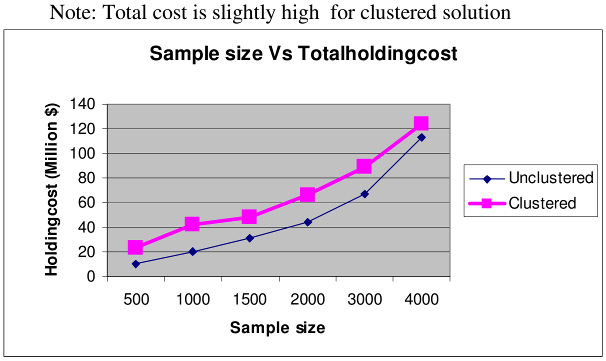
- The computational time required to calculate group policies should be much lower than that required for individual policies.
- The total cost associated with the clustered solutions should be slightly higher than of the unclustered solution.
- Certain performance metrics such as fill rate and customer wait time should show an undesirable trend for the clustered solution.

The reduction in computational time for the group policy decision is logical as a much smaller dataset has to be handled for the clustered solution. As previously mentioned, grouping items leads to a loss of identity for each individual item. Hence a slight increase in the total cost is expected when we calculate policies for groups of items instead of individual items. This difference in the cost between the clustered and the unclustered solutions is defined as the penalty cost for grouping. Incurring a penalty cost is an obvious sacrifice that one has to make so as to get computational and managerial convenience in managing parts which over a long horizon is a considerable but not an easily quantifiable benefit for huge inventory systems. The undesirable trend in the performance metrics observed after grouping items can also be attributed to the loss of identity that the items incur as a result of having group attributes which may have values either more or less than their original attribute values. The results of our experiments are presented in the next section as charts which are self-explanatory. It can be seen that the results follow the expected trends. This proves that our methodology works with a logic that is definitely feasible. The results do not indicate an optimal solution. Our ongoing research will help us obtain an ideal combination of various deciding factors to suggest a strategy for an optimal solution to the inventory segmentation problem for an inventory system of a given characteristic.

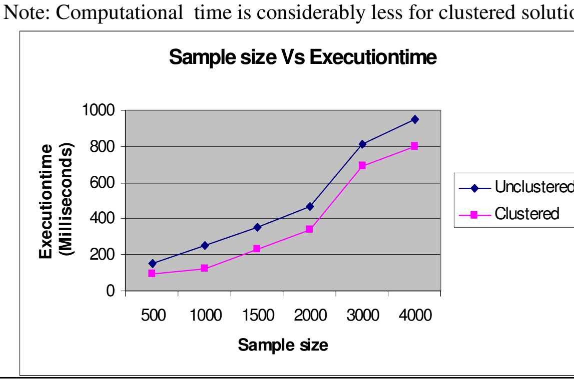
4. Experimental Results and Conclusions

The results shown in Figure 1 clearly indicate that the segmentation methodology proposed by us is feasible. The trends that were expected in the clustered and the unclustered solutions are evident in the results. The basic trade-off between cost and computational plus managerial ease is also clear. Research can be done on improving this basic segmentation methodology to obtain an effective trade-off between these conflicting objectives. Attention has to be paid in choosing the right combination of various factors involved in the process so as to give us optimal groups of parts. The benefits from the segmentation methodology are often not easily quantifiable but will lead to a great amount of operational ease and convenience. Consideration of operationally important attributes will lead to a practical and operationally feasible grouping of parts. Prospective areas of research are discussed in the next section of this paper.

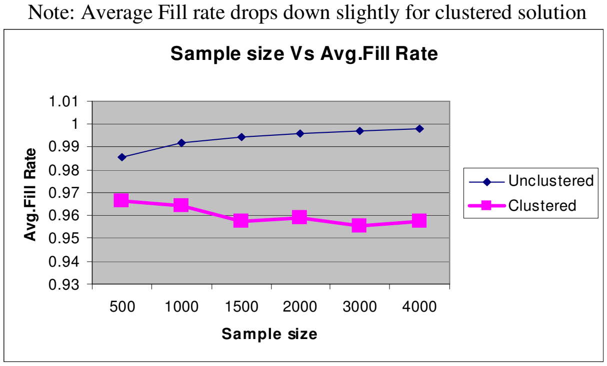
Note: Total cost is slightly high for clustered solution



Note: Computational time is considerably less for clustered solution



Note: Average Fill rate drops down slightly for clustered solution



Note: Customer Wait Time (CWT) is slightly high for clustered solution

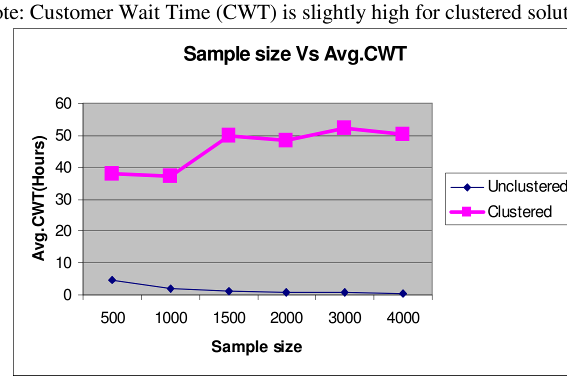


Figure 1:Experimental results

5. Future work and recommendations

Our current research suggests a segmentation methodology that is feasible but several other issues need to be addressed before we arrive at an optimal grouping solution. The type and number of attributes, dissimilarity coefficient type, number of clusters, choice of an effective clustering algorithm, an effective migration strategy to realign the groups with changing part life-cycle, an effective method to decide group policies are some of the issues which need to be addressed for obtaining an optimal solution. A right combination of these factors along with the logic of our segmentation methodology will lead to a cost effective and managerially convenient segmentation solution for a given type of inventory system. Also several additional attributes need to be considered so as to extend our segmentation methodology to the multi-indentured, multi-echelon level. We also plan to formulate the part grouping problem as an optimization problem in our future research so that the process of deciding inventory policies and deciding the groups can be simultaneously performed. Our continuing research will address these issues so as to present a segmentation strategy that promises a strategically optimal number and type of part groups for the lowest possible penalty cost.

6. Acknowledgements

This material is based upon work supported by the National Science Foundation and the US Navy Naval Supply Command. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the U.S. Navy.

References

1. Arabie, P., Hubert, L.J., and DeSoete, G. (1996), "Clustering and Classification", 1st edition, World Scientific Publishing Company Private Ltd., Singapore.
2. Cohen, M.A., Kleindorfer, P.R., Lee, H.L., and Pyke, D.F. (1992), "Multi-Item Service Constrained (s, S) Policies for Spare Parts Logistics Systems", *Naval Research Logistics* 39, 561-577.
3. Ernst, R., and Cohen, M.A. (1986), "Operations Related Groups (ORGs): A Clustering Procedure for Production / Inventory Systems", *Journal of Operations Management* 9(4), 574-598.
4. Hartigan, J.A. (1975), "Clustering Algorithms", 1st edition, John Wiley and Sons, New-York.
5. Hopp, W.J., and Spearman, M.L. (1995), "Factory Physics", 2nd edition, McGraw-Hill, New-York.
6. Romesburg, H.C. (1984), "Cluster Analysis for Researchers", 1st edition, Lifetime Learning Publications, California.
7. Samatova, N.F., Potok, T.E., and Leuze, M. (1999), "A Vector Perturbation Approach to the Generalized Aircraft Spare Parts Grouping Problem", *International Journal of Flexible Automation and Integrated Manufacturing*.
8. Sherbrooke, C.C. (1968), "METRIC: A Multi-Echelon Technique for Recoverable Item Control", *Operations Research* 16, 122-141.
9. Sherbrooke, C.C. (1986), "VARI-METRIC: Improved Approximations for Multi-Indenture, Multi-Echelon Availability Models", *Operations Research* 34 (2), 311-340.
10. Spath, H. (1980), "Cluster Analysis Algorithms for Data Reduction and Classification of Objects", Elis Horwood Publishers, England.