Applying Cluster Analysis on Medical Materials Management

Tzu Chun Tsai[1], Nai-Chun Fang[2], and Chuang-Chun Chiou[1*]
Department of Industrial Engineering and Enterprise Information, Tunghai University, Taiwan[1]
Taichung Veterans General Hospital, Taiwan[2]
Corresponding Author: cjchiou@thu.edu.tw

Received 25 February 2020; received in revised form 12 May 2020; accepted 3 June 2020

Abstract

Examining the cost structure of hospitals, the cost of materials accounts for 30-40% of the operating costs. Under the premise of maintaining a certain service level, there are still many management possibilities for the reduction of the cost of medical materials supply. The investigated data of this study were collected from a medical center in central Taiwan. The total number of medical material items is up to 7,964. In order to manage such a large number of medical materials, an effective method is needed. For this reason, the information of how to cluster and classify the medical material will help the managers to achieve a better cost control. In this study, K-Means and K-Medoid in the clustering method were used to analyze the characteristics of different items. The clustering results show that K-Medoid is better than K-Means. In the findings, it differentiates item 5571 has high volume with 7,285,200 and cluster it in an individual cluster. It is also found that medical materials in cluster 5 are high demand items, and medical materials in cluster 4 are high-priced items. Furthermore, this study recommends cluster 1, cluster 2 and cluster 3 are implemented with the same management method through the precise integration of the needs of each unit to reduce excess inventory. Cluster 4 can use the past data to accurately predict the demand for the next year to maintain the diagnosis and treatment of patients. The review time of the inventory in cluster 5 must be increased to ensure that there is no shortage. These methods will help hospital reduce related costs.

Keywords: Medical materials, clustering, K-Means, K-Medoid

1. Introduction

Examining the hospital cost structure, personnel costs are the bulk, and hospital managers are not easy to control the compression of personnel costs. In contrast, materials costs account for 30-40% of hospital operating costs. Under the premise of maintaining a certain service level, there is still a lot of room for the reduction of medical supply chain costs. Therefore, the medical system needs to re-examine the supply chain management model of medical materials to improve its business performance.

There are a variety of consumables in the hospital, including equipment, sanitary materials, cloth products, stationery, food, cleaning supplies and so on. This study focuses on medical materials, including masks, catheters, dressings, or gloves. Each medical material have its importance and indispensability for each unit or patient. If a medical material is out of stock, it will cause a loss in the work flow of a certain unit, and affect the life safety of the patient. (Moons et al., 2018) mentioned that the medical supply chain contains many links, such as the delivery of suppliers to hospitals, the allocation of different medical supplies to different units or even patients. In the above-mentioned activities related to the medical supply chain, there are different units involved. In the hospital-centered manner, there are many suppliers outside to provide medical materials and even medicines. Internally, there are different departments and wards that need these supplies. With medicines. With such a large number of participants, the entire medical supply chain must have a more precise approach to management.

In view of the importance of medical materials to hospitals and patients and the complexity of the medical supply chain, this study will use data mining methods for analysis. In the field of data mining, association rules, classification, clustering, prediction, etc. are included. Clustering can conduct a preliminary analysis of a data set, so that researchers can know the characteristics of each group for subsequent research. Zhou et al.(2016) used fuzzy clustering models to analyze household electricity consumption in Jiangsu Province, China, and to let power companies know the household electricity characteristics in different clusters to develop different operational strategies. The K-Means method was used in the study to cluster, and pointed out that the research results can enable managers to conduct similar management behaviors for the same group of items to improve efficiency (Pigman et al., 2018). The entropy weight-clustering analysis was used to analyze the maintenance consumables of CRH (Qiao & Zhang, 2019). Finally, it was pointed out that the clustering results were optimized for maintenance consumables.

This study is based on a medical center in central Taiwan, which currently has 7,964 different medical materials for use by all units. The purpose of this study is to compare the differences between different clustering. At the beginning of the study, the attributes in the data will be briefly introduced, and then the data will be cleaned up for subsequent data mining. Then, we use partitioned clustering in our data, and analyze the result of clustering. Finally, the suggestions and future issues of this research are presented.

2. Literature Review

2.1 Medical Materials Management

Moons et al. (2018) mentioned that hospitals play an extremely important role in the management of medical materials in order to meet the service level. Management includes topics such as demand forecasting, material sorting, and internal logistics. The material classification is a key role for each hospital. The reason is that the number of consumable items managed by the hospital is increasing from the size of the hospital, such as from the regional hospital to the medical center. Some more than 1,000 have risen to more than 7,000. If they are not classified at the beginning, it will cause a major difficulty in management, which in turn will lead to cost increases.

2.2 Clustering Analysis

Clustering is a method of classifying data into groups, an unsupervised learning, that is, whether the training materials are pre-defined. The main purpose is to find similar n clusters in the data, so that the member objects in the same subset have similar properties. The two common methods of clustering are hierarchical clustering and partitional clustering. The hierarchical clustering method divides and aggregates data several times, and then generates a tree structure. The common hierarchical clustering is the aggregated hierarchical clustering method, and the splitting hierarchical clustering method. The hierarchical clustering method can be started from the bottom of the tree structure, and the data or clusters are merged one by one; the split-level grouping rules are successively split from the top of the tree structure. The partitional clustering is to define the number of groups in advance, and after a series of mathematical iterations, the result of the cluster is generated. The common partitional clustering is 1.K-Means 2.K-Medoid. The advantages and disadvantages of hierarchical clustering and partitional clustering are compared by Table 1, and the researcher is made aware of which clustering method is used for the data on the opponent to obtain the best result.

Table 1: Comparison of Advantages and Disadvantages of Hierarchical Clustering and Partitioned Clustering

	Hierarchical Clustering (Aggregate Hierarchical Clustering)	Partitional Clustering(K-Means)
Advantage	1. The concept is simple, and the entire calculation process can be presented in a tree structure. 2. Only need the distance between the data points, you can construct the group structure without the actual coordinates of the data points.	1. Easy to understand 2. Automatically assigned to a cluster
Disadvantage	1. Usually only for a small amount of data, it is difficult to handle a huge amount of data	1. The number of clusters must be determined before clustering and the characteristics of each attribute must be checked 2. Too sensitive to outliers 3. Not the best solution for the overall data

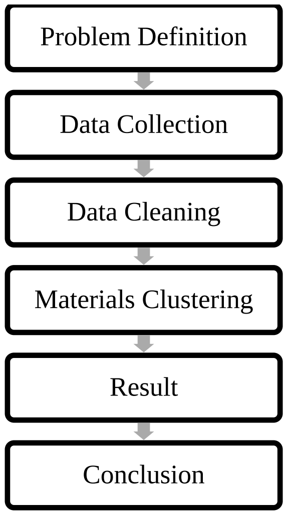

Figure 1: Research Chart

3. Methodology

3.1 Research Flowchart

Figure 1 shows the flow of this study. After first defining the problem points of the medical center, collecting relevant information and cleaning up the data to facilitate subsequent grouping, and finally analyzing the grouping results and presenting conclusions for the whole study.

3.2 Problem Definition

This study is a medical center in central Taiwan, which contains hundreds of units, each of which uses different types of medical supplies on patients, totaling 7,964. However, in order to manage such a large number of medical materials, an effective method is needed. However, the hospital has not yet implemented a special management method for these items. For this reason, it is necessary to conduct a cluster analysis of medical materials first, so that managers can know the characteristics of each cluster and propose different management. technique.

3.3 Data Collection & Cleaning

This study collected data from suppliers ordering from January to December 2017, which contains 21 attributes, such as namely application form, funding source, funding source description, application type, application reason, application reason description, order, order date, category, status, warehouse, material code, Chinese product name, English product name, brand model, order quantity, in-transit volume, accepted quantity, purchase unit price, applicant, manufacturer. Since the data attributes in the cluster analysis must be of numerical type to be executable, we will first delete the data that does not belong to this type, for example: user unit, Chinese and English product names, suppliers, etc. The most common method for medical materials classification is ABC analysis (Gupta et al., 2007). This method is based on the order quantity and order price as a basis for classification, and in accordance with the Plato 80/20 rule as a reference, where category A accounts for 20% of the number of items, 70% of the annual consumption value; category B accounts for 30% Item number, 25% of annual consumption value; Category C occupies 50% of item number, 5% of annual consumption value (Werner, 2002) But in addition to the order quantity and order price, the order cycle is also very important for the hospital, because we know the demand frequency of each medical material through the order cycle, thereby improving the order efficiency, so this study finally decided to use order quantity, order price and order cycle as the basis attributes for clustering.

3.4 Description of Clustering Method Selection

The total number of medical materials in the study population was as high as 7764. After comparing the advantages and disadvantages of the above table, we chose to use " partitional clustering" to analyze the data. The reason is that the amount of data is too large. If we use "hierarchical clustering," you may not be exhaustive in the presentation of the results. We use K-Means and K-Medoid for clustering and compare the differences between the two. Because K-Means is the most common method of clustering, but its disadvantage is that the result of cluster will be affected by outliers, plus K-Means takes the average as the center point. Conversely, the center point value of K-Medoids is the actual observation value in the dataset, so that we can understand what medical materials are represented in the center of each cluster.

3.5 K-Means＆K-Medoid

Telgarsky and Vattani (2010) pointed out that the commonly used K-Means method is proposed by Hartigan, in which the sum of intra-cluster variations is defined as the sum of squares of Euclidean distances from the center of each data point. The formula is in equation (1):

$$W(C_k) = \sum_{x_i \in C_k} (x_{i-\mu_k})^2 \quad (1)$$

where x_i represents the data point in the cluster C_k. μ_k represents the center point specified by the average of the data points in the cluster C_k. The total within-cluster variation is given by equation (2):

Total Within-Cluster Variation=

$$\sum_{i=1}^{k} W(C_k) = \sum_{i=1}^{k} \sum_{x_i \in C_k} (x_{i-\mu_k})^2 \quad (2)$$

Of course, the smaller the total variation within the group, the more similar all elements in the group.

The steps of the K-Means method are as follows:

1. First randomly select the K data points as the starting value of the K starting clusters.
2. Assign each of the remaining data to the cluster closest to the cluster center and recalculate the average of the clusters based on the data points in the cluster.
3. Calculate the distance from the data point to the center of the cluster. If the square of the total distance variation is found to decrease, it means that the cluster center has changed and the data points need to be reassigned to the new cluster.
4. Continue this process until the total distance variation no longer drops or reaches the set number of counts.

In addition, the K-Medoid is more powerful than the K-Means because when using the K-Medoid algorithm, the center point will select an observation within the group rather than the intra-group average. The K-Medoid method mentioned in (Sheng & Liu 2004) is the sum of Euclidean distance in equation (3):

$$SED = \sum_{i=1}^{n} \sum_{j=1, x_i \in C_j}^{k} d(x_i, m_j) \quad (3)$$

where m_j represents medoid of cluster C_j (j=1, 2, 3, ..., k).

The steps of the K-Medoid method are as follows:

1. Select K representative data as the starting center of the cluster.
2. Assign data to the nearest cluster based on the distance.
3. Select any non-clustered center data point to replace any cluster center, and calculate the total group and distance change S, that is, when S<0, replace the original cluster center with this data, and S>0, means that the original cluster center does not need to be replaced.
4. Repeat step three until you are sure that all data points cannot replace any of the cluster centers.

3.6 Materials Clustering

Whether using K-Means or K-Medoid for clustering, the first is to determine the K, so this study uses the Elbow Method to determine the K. It can be seen from Fig. 2 that the turning bending occurs at K=5, so that the folding line tends to be smooth, so 5 is used as our K value. The second step is to input the collated data into the R language for K-Means and K-Medoid.

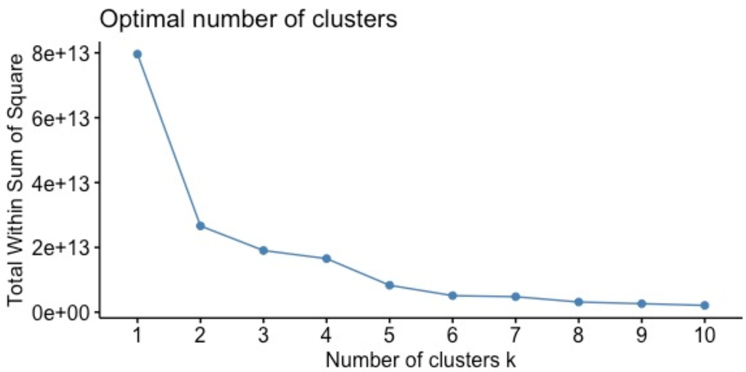

Figure 2: Elbow Method n=5

4. Result

In Figure 3, K-Means made the medical materials numbered 5571 into cluster number 4 which also had a large distance compared to other clusters; the other 4 clusters were concentrated in the upper left of Figure 3. In Figure 4, K-Medoid incorporates the medical materials numbered 5571 into the cluster number 5. Then comparing the two methods, it shows that the outliers will affect the clustering results of K-Means, so we go back to the data to observe the medical materials numbered 5571, and found that the total order quantity of the item in 2017 is 7285200. The total order quantity is quite different from other medical materials, so that K-Means will be numbered 5571.

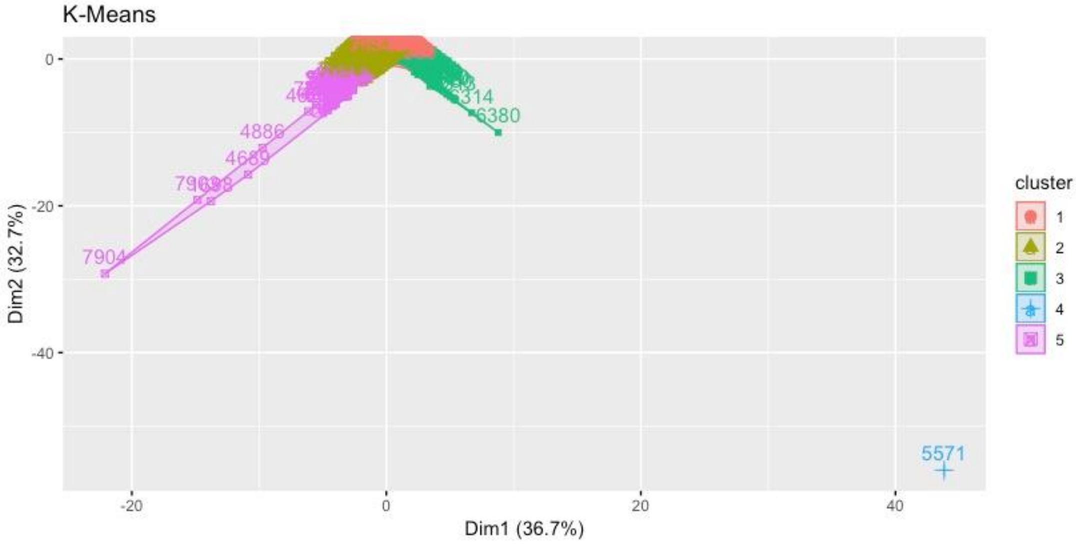

Figure 3: The Clustering Result of K-Means for Medical Materials

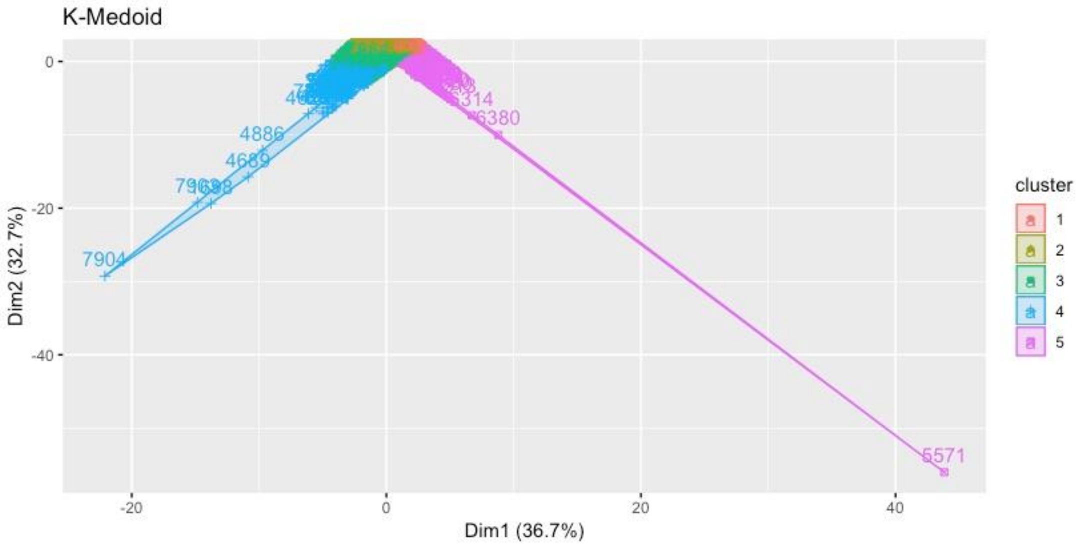

Figure 4: The Clustering Result of K-Medoid for Medical Materials

Table 2 shows the data in the K-Medoid clustering results. We can know that the medical materials in cluster 1 and cluster 2 are very similar to other clusters through the "separation" in the table. It is found that cluster 1 and cluster 2 occupy the same amount. The number of items has accounted for 88.7% of the total items.

Table 3 can be used to compare the differences of the center points of K-Medoid clusters and as a reference for implementing management methods. If we look at the number of orders, the value of cluster 5 is the largest among all groups, reaching 282400, indicating that the medical consumables in cluster 5 are high-demand items. If we look at the order price, the order price of cluster 4 is among all the clusters. The highest is 313900, which indicates that the medical materials in cluster 4 are high-priced items. If we observe the order cycle, cluster 5 is the highest in all clusters. Hospital must order it once every 8.9 days, indicating that the medical materials in clusters 5 are consumables items.

Table 2: The Analysis Outcome of K-Medoid Method

	size	av_diss	diameter	separation
Cluster 1	4900	3960.53	137980.46	50.09
Cluster 2	2167	4839.05	26451.87	50.09
Cluster 3	775	16995.93	149200.38	400.00
Cluster 4	56	140027.01	1603000.02	3800.12
Cluster 5	66	256783.11	7132200.00	15200.00

Table 3: K-Medoid Different Cluster's Center Data

	order quantity	order price	order cycle
Cluster 1	120	1250	121.67
Cluster 2	5	12960	182.5
Cluster 3	3	54550	121.67
Cluster 4	1	313900	365
Cluster 5	282400	1.03	8.9

5. Conclusion

Medical materials have their own importance and indispensable characteristics for each unit or patient. If a medical material is out of stock, it will cause a loss in the work flow of a certain unit, and affect the life safety of the patient. Therefore, the management of medical materials must be cautious and accurate. In order to allow managers of a medical center to further manage medical supplies, we are clustering through as many as 7,974 medical materials. The clustering results show that K-Medoids is better than K-Means, because K-Medoids can deal with outliers. Unlike K-Means, which will be outliers, K-Medoids is used as a clustering result. As for cluster 1, cluster 2 and cluster 3, the most obvious difference is the order quantity and order price. Medical materials for patients must be able to supply them in time when they are needed without out-of-stocks, so for the management methods of each cluster, the ordering cycle is particularly important compared to the other two attributes, so this study recommends cluster 1, cluster 2 and cluster 3 are implemented with the same management method, because their ordering period is 4 months to half a year, and the number of items accounted for reaches 98.5%, so through the precise integration of the needs of each unit to reduce excess inventory. The ordering period of cluster 4 is one year, so we can use the past data to accurately predict the demand for the next year to maintain the diagnosis and treatment of patients. The ordering period of cluster 5 is the shortest of all the clusters, and it will be ordered once every week, so the review time of the inventory must be increased to ensure that there is no shortage.

At present, this study only uses the order cycle, order quantity and order price as the attribute basis of clustering. However, the management of medical materials actually affects not only these attributes, but also the use of different units or different medical treatments will also have an impact. Obtaining more complete information to cluster the medical materials of the medical center for more appropriate results is the future issue.

Clustering can not only be applied in the medical industry, but also in manufacturing and service industries. The manager can analyze the data of the opponents and distinguish the differences of each group from the grouping results to make better decisions. But before we use clustering, we must first examine whether the problem itself can be helped by this method. If the attributes in the data cannot meet the needs of clustering, it is just a waste of time to clean the data. No matter which kind of data exploration method is used that they are not omnipotent, only by recognizing the problem itself has its value.

Acknowledgment

This research is partially supported by Ministry of Science and Technology of Taiwan, R.O.C. (Grant No. MOST 108-221-E-029-005)

References

Gupta, R., Gupta, K., Jain, B & Grag, R. (2007). ABC and VED analysis in medical stores inventory control. *Medical J Armed Forces India*,63,325-7

Moons, K., Waeyenbergh, G., & Pintelon, L. (2018). Measuring the Logistics Performance of Internal Hospital Supply Chain – A Literature Study. *Omega*, *82*, 1-13

Pigman, M., Le, H., Bhagat, U., Thompson, M., & Wang, Q. (2018). Analyzing Inventory Data Using K-Means Clustering. In *The International Conference on Data Science* (ICDATA) (pp. 117-122).

Qiao, C. Z., & Zhang, W. J. (2019, May). Research on Classification Method of CRH Maintenance Parts Based on Entropy Weight-Clustering Analysis. In *2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference* (ITAIC) (pp. 517-521).

Sheng, W., & Liu, X. (2004, June). A Hybrid Algorithm for K-medoid Clustering of Large Data Sets. In *Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No. 04TH8753)* (Vol. 1, pp.77-82). IEEE.

Telgarsky, M., & Vattani, A. (2010, March). Hartigan's Method: k-means Clustering without Voronoi. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (pp.820-827).

Werner, H. (2002). Supply Chain Management. Grundlagen, Strategien, Instrumente und Controlling.

Zhou, K., Yang, S., & Shao, Z. (2017). Household Monthly Electricity Consumption Pattern Mining: A Fuzzy Clustering-Based Model and A Case Study. *Journal of Cleaner Production*, *141*, 900-908

About Authors

Tzu Chun Tsai is a master student in the department of Industrial Engineering and Enterprise Information at Tunghai University Taichung, Taiwan.

Nai-Chun Fang is a chief logistics officer (CLO) at Taichung Veterans General Hospital, Tiachung, Taiwan. He received his Ph.D. degree from Tunghai University.

Chuang-Chun Chiou is a professor in the department of Industrial Engineering and Enterprise Information, at Tunghai University, Taichung, Taiwan. His research interests include supply chain management, healthcare logistics, lean management in service and manufacturing, and data analytics in operations management. His works have been published in European Journal of Operational Research, Annals of Operations Research, Journal of Chinese Institute of Industrial Engineering, Journal of Quality, Journal of Operations Research Society of Japan, and International journal of Production Economics、International Journal of Operational Research and many other refereed journals.