

3D PARTITION-BASED CLUSTERING FOR SUPPLY CHAIN DATA MANAGEMENT

A. Suhaibah^{a*}, U. Uznir^a, F. Anton^b, D. Mioc^b and A. A. Rahman^a

^aGeospatial Information Infrastructure (GeoI²) Research Lab.,
Faculty of Geoinformation and Real Estate,
Universiti Teknologi Malaysia,
81310 Skudai, Johor, Malaysia
norsuhaibah@gmail.com, mduznir@utm.my, alias@utm.my

^bDept. of Geodesy, National Space Institute, Technical University of Denmark,
Elektrovej 328, 2800 Kgs. Lyngby, Denmark
fa@space.dtu.dk, mioc@space.dtu.dk

KEY WORDS: Supply Chain Management, 3D Spatial Data Clustering, 3D Spatial Database, 3D GIS, Data Management, Information Retrieval

ABSTRACT: Supply Chain Management (SCM) is the management of the products and goods flow from its origin point to point of consumption. During the process of SCM, information and dataset gathered for this application is massive and complex. This is due to its several processes such as procurement, product development and commercialization, physical distribution, outsourcing and partnerships. For a practical application, SCM datasets need to be managed and maintained to serve a better service to its three main categories; distributor, customer and supplier. To manage these datasets, a structure of data constellation is used to accommodate the data into the spatial database. However, the situation in geospatial database creates few problems, for example the performance of the database deteriorate especially during the query operation. We strongly believe that a more practical hierarchical tree structure is required for efficient process of SCM. Besides that, three-dimensional approach is required for the management of SCM datasets since it involve with the multi-level location such as shop lots and residential apartments. 3D R-Tree has been increasingly used for 3D geospatial database management due to its simplicity and extendibility. However, it suffers from serious overlaps between nodes. In this paper, we proposed a partition-based clustering for the construction of a hierarchical tree structure. Several datasets are tested using the proposed method and the percentage of the overlapping nodes and volume coverage are computed and compared with the original 3D R-Tree and other practical approaches. The experiments demonstrated in this paper substantiated that the hierarchical structure of the proposed partition-based clustering is capable of preserving minimal overlap and coverage. The query performance was tested using 300,000 points of a SCM dataset and the results are presented in this paper. This paper also discusses the outlook of the structure for future reference.

1 INTRODUCTION

According to (Dinu Popa, 2014), Supply Chain Management (SCM) main functions can be categorized as: locating activities in order to get the best goods for the manufacturing capacity (inbound) and the distribution of finished goods to the final customer (outbound). By implementing an efficient SCM, the management operations can optimize or minimize the inventory expenditure, reducing the supply chain cost and improving the delivery time of products to the final consumer. As for an example, the logistic operation is required to manage the execution of transportation, warehousing, handling and delivery at the right time, in the best quality and in the most cost-effective way in order to fulfill customers' requirements.

SCM deals with geographical location, for instance, the location of raw material supplies, warehousing sites and the position of the final consumers. By knowing the location of these sites, the decision can be investigated and evaluated based on the most effective and efficient route and later it can optimize the cost of delivering the goods. One of the most common analyses used before making the decision is the shortest/optimal route analysis.

It uses spatial data in identifying the best route or path in delivering the products. In order to manage this information, Geospatial Information Systems (GISs) are used for storing and managing the spatial data and GISs are a common tool used in most of SCM spatial data management (Craighead et al., 2011; Delen et al., 2011).

In SCM, before analyzing the location of the spatial data, the data need to be categorized into several clusters. For an example, consumers with different annual income should be grouped and located based on their clusters. Later, by using GIS, SCM can identify what category of products should be available for the consumers based on their geographical location. However, a serious overlapping issue in categorizing the clusters leads to an inefficient geospatial analysis. Moreover, SCM in urban areas requires clusters to be organized in a three-dimensional (3D) way (i.e. multi-level shop lots, residential apartments) which is a constraint with the current GIS framework. Therefore, this research introduces a new approach in SCM for categorizing the clusters. The approach efficiently minimizes the overlapping among clusters and increases the accuracy of spatial object groupings in 3D.

This paper is organized as follows: problems and motivation regarding the 3D clustering in supply chain data management are discussed in the next section. In Section 3, the concept of the proposed method is explained with its implementation. Section 4 presents the analysis and results of the experiment. Finally, the conclusions are presented in Section 5.

2 RESEARCH PROBLEM AND MOTIVATION

Clusters in SCM can be seen in various viewpoints. Delivering the products to the final customers in a due date in time is the major objective. However, clusters in SCM are geographically interconnected. Each cluster includes socially homogenous neighborhood. The clustering involving in socially homogenous neighborhood ensures an optimization based on proximity and lowers the transaction costs. It builds long-term relationships among firms involved from the source through to end-users (DeWitt et al., 2006). In addition, clusters can attract suppliers and they can locate the companies next to their largest demand centers. This will create economic efficiencies for both the suppliers and the companies located in a cluster.

As discussed in the previous section, a serious overlapping between clusters can produce inaccurate result. According to the first law of geography, everything is related to everything else, but near things are more related than distant things. Thus, inaccurate result will not optimize the assessment of SCM in determining the best location for supplies and transporting their products or services. Considering transporting the products in urban areas with high rise multi-level buildings, a 3D method is prominently required in order to define the clusters before further spatial analysis can be performed. The overlapping issue can be seen as a serious problem in data retrieval and analysis. It will affect the query efficiency in terms of response time, database storage and information accuracy. Furthermore, with the immense number of urban datasets, the analysis and the efficiency of SCM data retrieval will become more complex and crucial.

Overlap usually occurs when data or information is constellated in a tree structure. In a database management system (DBMS) environment, several tree structures are used to constellate data and information such as R-Tree (Guttman, 1984), QuadTree and kd-Tree. However, these structures are still facing the issue of overlap between nodes. In spatial database overlap between nodes is the main reason for the low efficiency of query performance due to multi-path queries. Since the aim of this paper is to tackle the issue of SCM in 3D space, the overlap among node is expected to be serious compared with 2D structure. For instance, the Oracle commercial database provides a 3D R-Tree structure to deal with 3D data (Murray, 2009; Ravada et al., 2009). However, when the R-Tree is extended into 3D space, the MBV of sibling nodes tends to frequently overlap, and MBVs among nodes can even contain other MBV.

Due to critical overlap of sibling nodes and uneven size of nodes in 3D R-Trees, a research was conducted by (Zhu et al., 2007) to minimize the overlap and optimize the clustering algorithm by introducing k -means clustering algorithm to put forward an improved 3D R-Tree. From his experiment, the performance of 3D R-Trees indicates that, by using an improved algorithm the overlapping of node is minimized while balancing the volume of

parallelepipeds. However, using k -means, would not drastically minimize the overlap among nodes. This is due to the random selection of the initial seed or cluster center which will lead to an unequal cluster number of dataset. This condition will increase the risk of having serious overlap in the tree structure. Thus, an improvised method of clustering is needed to push the limits of the 3D R-Tree structure for the application of supply chain data management.

3 SUPPLY CHAIN DATA MANAGEMENT IN THREE-DIMENSIONAL (3D) SPACE

3.1. Partition-Based Clustering

Partition-based clustering is a method of clustering that requires a pre-set number of clusters from the user. The commonly used algorithm from this type of cluster is k -means algorithm. By using this algorithm, data will be partitioned into k clusters (C_1, C_2, \dots, C_k). The means for all instances in each cluster are then calculated as a cluster centres. The algorithm of k -means is described as follows.

Input: P (object set), k (number of cluster)
Output: clusters
 Step 1: Initialize k cluster centers.
 Step 2: **while** termination condition is not satisfied **do**
 Step 3: Assign objects to the nearest cluster center.
 Step 4: Update cluster centers.
 Step 5: **end while**

The algorithm starts by randomly finding initial cluster centres. For each one of the iterations, an object or instance is assigned to the nearest cluster centre based on Euclidean distance. Then, cluster centres are recalculated until the coordinates are constant.

The centre of each cluster is calculated as the mean of all objects or instances belonging to that cluster:

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} X_q \quad (1)$$

where N_k is the number of instances belonging to cluster k and μ_k is the mean of the cluster k .

According to (Selim and Ismail, 1984), a sample size of m instances or objects does affect the complexity of T iterations of the k -means algorithm. Thus, the complexity of each characterized search by N attribute is:

$$O(T \cdot k \cdot m \cdot N) \quad (2)$$

The linear complexity is one of the reasons of k -means popularity. Even if the size of instances is large, this algorithm is computationally attractive. Besides that, k -means offer simplicity, speed and adaptability to sparse data (Dhillon and Modha, 2001). k -means algorithm is sensitive to the presence of noise and

outliers (Hasan et al., 2009; Kaufman and Rousseeuw, 2008). A single outlier can increase the squared error dramatically. However, this disadvantage is exceptional in our case, since data used in this research is static and stagnant such as buildings and customer locations. The only concerning issue using this method is the initial seed of the cluster centre in *k*-means algorithm. Besides that, one of the functions in *k*-means is categorized as NP (Non-deterministic Polynomial-time) hard problem. This function is inclined to focus cluster centers at one point or area. Thus, any selection is very sensitive to the group cluster and may make a difference for the end result.

Even though the algorithm has been invented more than twenty years ago, surprisingly, there is no approach that has been suggested to resolve this issue. A few years ago, (Arthur and Vassilvitskii, 2007) proposed a variant of *k*-means that chooses cluster centers by using weight. Data points are weighted by squared distance from the closest cluster centers that were initially defined. The initial seed of cluster centers were defined and combined with the *k*-means algorithm, the resulting algorithm is known as *k*-means++. The algorithm of *k*-means++ is described as follows.

Input: P (object set), k (number of cluster)
Output: clusters

Step 1: Initialize k cluster centers.
 Step 2: Choose one center C_1
 Step 3: Repeat for $2 \leq i \leq k$
 Step 4: Choose C_i to be $x \in X$ with D^2 weighting

$$\frac{D(x)^2}{\sum_{x \in X} D(x)^2}$$

Step 5: **while** termination condition is not satisfied **do**
 Step 6: Assign objects to the nearest cluster center.
 Step 7: Update cluster centers.
 Step 8: **end while**

To compare the accuracy of random initialization of cluster centers between these two algorithms, a set of 3D points representing customer location in urban areas was tested. The same number of clusters, k and number of instances N was used for approaches. The Figure 1 presents the result from the tests. From the figure, there are four random initialization of cluster center (points in yellow). The result produced by *k*-means++ algorithm indicates that the initial seeding picked evenly spread cluster centers compared to the original *k*-means algorithm. Based on the result from this test, it is obvious that the *k*-means++ algorithm offers a better solution than the original algorithm. Besides that, this algorithm offers up to 70% faster clustering than the original algorithm especially on the larger dataset (Arthur and Vassilvitskii, 2007). As mentioned in the previous section, the immense number of urban datasets does affect the complexity of SCM analysis in terms of data size, storage and performance. Thus, this algorithm suits the application of SCM due to its simplicity, accuracy and practicality.

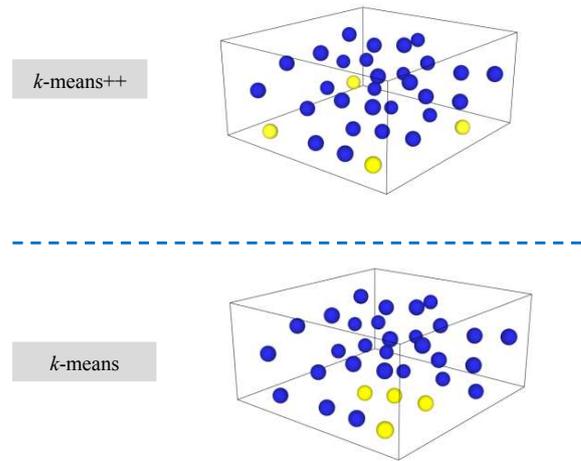


Figure 1. Comparison of random initialization cluster centers.

3.2. Hierarchical Tree of Supply Chain Management

Hierarchical trees are one of the methods to improve data analysis and management. In SCM application, tree structures could be used to improve the information retrieval and analysis such as nearest point location. To accommodate the SCM datasets into a tree structure, several categories need to be identified and attached to the tree. According to (Gunasekaran and Ngai, 2004; Lancioni et al., 2000; Power, 2005) there are several categories or parties involved during the process of SCM. Figure 2 describes an example of SCM's process from one party to another. From the figure, we could see that the products need to be considered in several processes such as logistic, transportation, marketing etc. before deliver to customers. The roles of these three parties in SCM are described as follows:



Figure 2. The process and concept of supply chain management.

Manufacturer: The transportation of goods or products starts with the manufacturer or warehousing management. Manufacturers carry the most valuable role in operation. They need to provide a perfect storage and office with the convenient facility. They also have to make sure the dispatching department to deliver the goods or products on time and loading or unloading it at the right location. All of these activities are important to the manufacturer to reduce the company's cost and expenses.

Supplier: A supplier is a middleman between the manufacturer and retailer or customer. According to (Miocevic and Crnjak-Karanovic, 2012) the key of successful SCM depends on the relationship between suppliers and customers. Thus, it is very important for the company to monitor every detail of the goods and products before they hand them over to the customer. Some people may think that it is unnecessary to have a supplier as a middleman. If the middleman is eliminated, the cost for the end consumer could be reduced. However, for the wholesale trade businesses, suppliers are needed in serving customer as a regional center of products distribution.

Shipping Services: The movement of finished goods from the manufacturer or distribution center to a customer. Shipments from one source are separated into a number of parts and distributed to many receivers.

Products: An output produced from the manufacturer. Product is the main subject in SCM. Products are manufactured using raw materials. After the items have been completed and tested, they are stored back in the warehouse prior to delivery to the supplier and customer.

Customer: In SCM application, a customer is the end consumer or user who purchases or receives product. This is the final or end of the SCM process.

In this paper, we focus on the three main parties or categories; manufacturer, suppliers and customers. Based on these groups, all of the information are gathered and linked with the other two groups; shipping services and products. Then, the information is tabulated in the database schema or view. The data model diagram for the manufacturer, supplier, shipping services, products and customer is designed and described as in Figure 3.

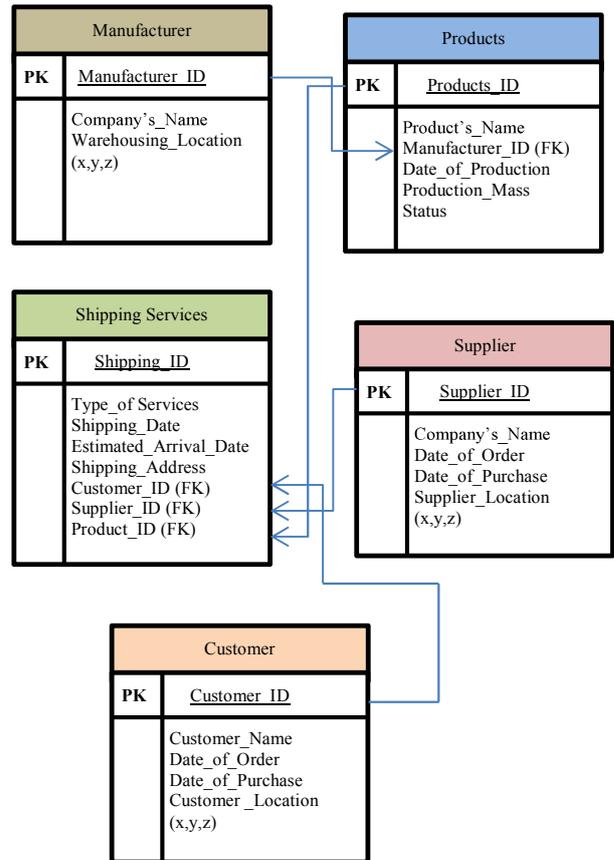


Figure 3. Data model diagram of the SCM dataset

Based on the data model diagram in Figure 3, three main groups of dataset, i.e. $C_{Manufacturer}$ (Manufacturer), $C_{Supplier}$ (Supplier) and $C_{Customer}$ (Customer) will be clustered with k number of cluster center C_i , k is defined based on the total number of instances N divided with the maximum number of entry M in page size. Information is stored by forming a bounding volume as the delimitation between clusters. For example, an instance (i.e. 3D point) with a minimum value of coordinate x_{min} , y_{min} and z_{min} and the maximum value x_{max} , y_{max} and z_{max} is used to form the volume delimitation. Each instance is identified by the cluster pointer $cluster_id$ and volume delimitation (x_{min} , y_{min} , z_{min} and x_{max} , y_{max} , z_{max}). Figure 4 presents the hierarchical tree of $C_{Manufacturer}$ (Manufacturer), $C_{Supplier}$ (Supplier) and $C_{Customer}$ (Customer).

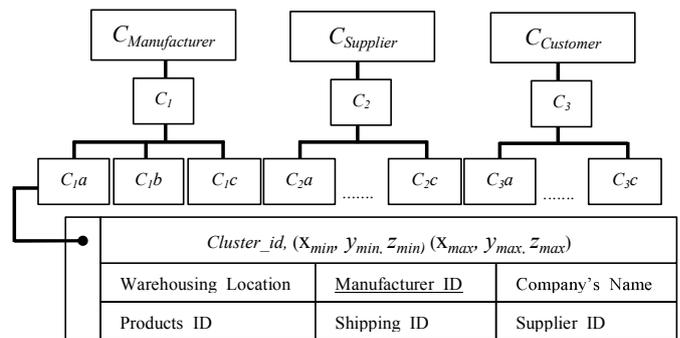


Figure 4. Hierarchical structure of SCM dataset.

The significance of nearest neighbor information for the SCM has been mentioned in several researches such as (Akhbari et al., 2014; Kiekintveld et al., 2007; Rodger, 2014). From this information, the manufacturer would capacitate the best vehicle routing for products delivery and shipping. Besides that, the manufacturer could also plan and schedule the tour procedure for loading the products from one center to another (Bellman, 1962).

Since the SCM application handles a lot of travelling, the Euclidean minimum spanning tree, which is a subgraph of the Delaunay graph is to be considered to control the physical flow of the supply chain. It is also a key point to increase service productivity and control the cost. In this paper, the nearest neighbor information is retrieved from the structure using the following algorithm.

Input: q (query point)
Output: p (nearest point)
 Step 1: find the nearest group cluster C_i
 Step 2: get the *cluster_id*
 Step 3: list all instance in the cluster $p (p_1, p_2, p_3, \dots, p_n)$
 Step 4: for each p calculate distance D from q
 Step 5: find the minimum value of D

4 EXPERIMENT AND ANALYSIS

4.1 SCM Data Retrieval

The proposed structure of hierarchical tree for SCM application is aimed to cater spatial information of SCM dataset. In this test, a set of supplier location is tested to find the nearest supplier from a customer location. In SCM application, the nearest point query is important to save cost and time. Besides that, marketing analysis could also be performed by analysing the location of customers and suppliers. Figure 5 presents a 3D dataset of customer locations and a query point q of supplier location.

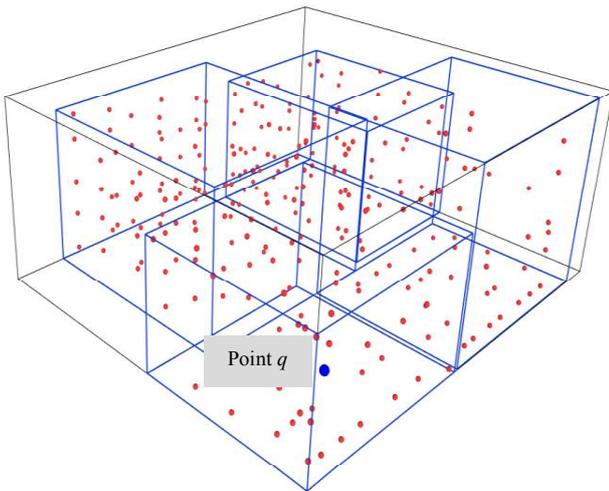


Figure 5. SCM datasets of customers and supplier q .

Based on the location of the supplier (point q), the nearest cluster of customer is identified. Then, all the candidates in the group cluster are listed. From the lists, the distance is calculated and the nearest customer is identified. The result in Figure 6 shows that customer with ID 5214 (in yellow) is the nearest point location from supplier q . The total distance between these two points is 1.2 km.

Result.

Customer ID = 5214, Distance = 1.2 km

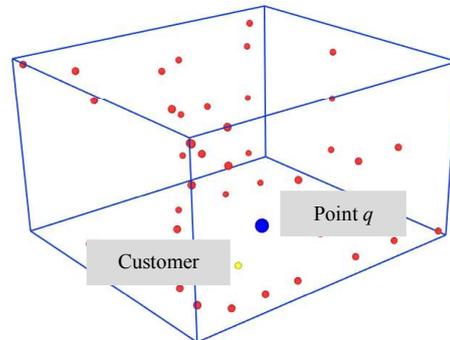


Figure 6. Nearest point from supplier q .

4.2 Overlap Percentage among Nodes and Volume Coverage Percentage

Overlap is a key factor to produce an efficient tree structure. Overlap occurs during the construction of tree structure or during the data updating operation. When objects are frequently updated (for inserted and deleted operation) the structure of the tree will be revised and the structure will tend to have serious overlap between nodes. In this test, the structure is tested by using several groups of SCM datasets; Group A = 100,000, Group B = 300,000 and Group C = 500,000. The overlap percentage is calculated and compared with the existing hierarchical data structure in spatial database 3D R-Tree and hierarchical structure based on original k -means clustering. The result of this test is demonstrated in Figure 7. From the result in Figure 7, it shows that the proposed hierarchical structure produced a lower percentage (about 20 to 30 percent reduction) than the other two approaches.

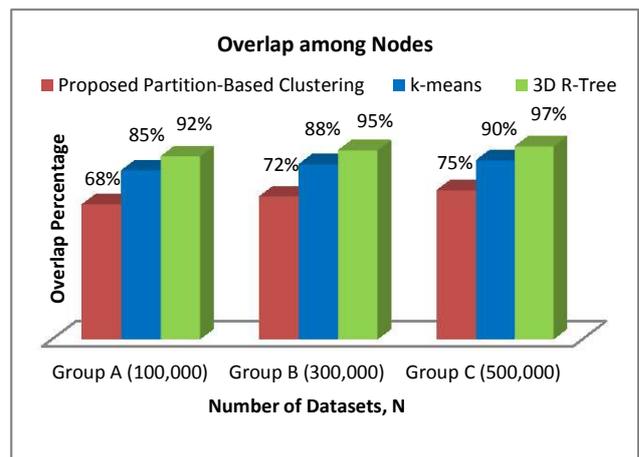


Figure 7. Overlap percentage analysis for varying number of datasets.

Another benchmark of efficient tree structure is its coverage area. In this paper, we are using the term of volume coverage since we are dealing with 3D datasets. Volume coverage is produced from the total volume of boundary delimitation. Minimum volume coverage indicates the efficiency of tree structure since the search area is small and the performance of information retrieval can be increased. In this test the coverage percentage is calculated by accumulating the total volume among the nodes (generated by clusters) and divide with the boundary delimitation of the whole dataset. A 500,000 points are used in this test to calculate the coverage area for three different methods, proposed partition-based clustering, *k*-means and 3D R-Tree. The result of this test (Figure 8) indicates that the percentage of volume coverage of the proposed structure is minimal.

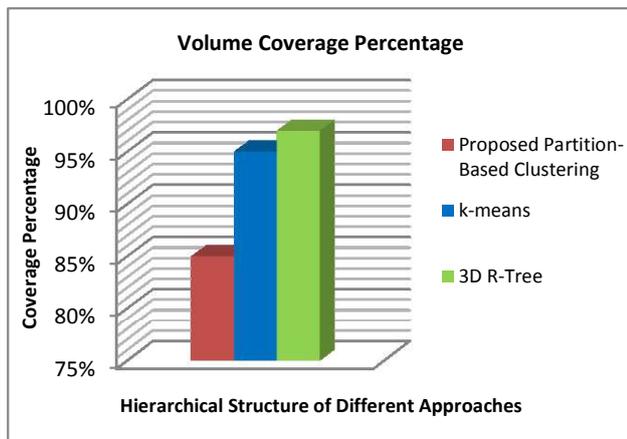


Figure 8. Coverage percentage analysis.

4.3 Effect of SCM Data Updating towards the Hierarchical Structure

The SCM dataset is also tested for the data updating processes such as insertion and deletion. To analyse the effect of these operations towards the overlap percentage, pre-existing datasets are updated with several groups of SCM datasets. The percentage of overlap based on proposed approach is compared with 3D R-Tree.

4.3.1 SCM Data Updating: Insert Operation

A group of 300,000 customer dataset is structured with the proposed hierarchical structure and stored in the database. These datasets are then updated with several groups of customer data. Each group contains different numbers of entries; Group A = 10,000 datasets, Group B = 50,000 and Group C = 100,000. Result of this test is described in Figure 9. From the result, proposed partition-based clustering offer lower overlap percentage which is 15 to 18 percent than 3D R-Tree.

4.3.2 SCM Data Updating: Delete Operation

A group of 500,000 supplier dataset is structured with the proposed hierarchical structure and stored in the database. Several groups of records from these datasets are then removed from the database. Each group contains different number of entries; Group A = 50,000 datasets, Group B = 100,000 and Group C = 150,000. Result in Figure 10 substantiated that proposed approach produce lower percentage after bulk deletion compared to existing index structure, 3D R-Tree.



Figure 9. Effect of bulk insertion towards the overlap among node.

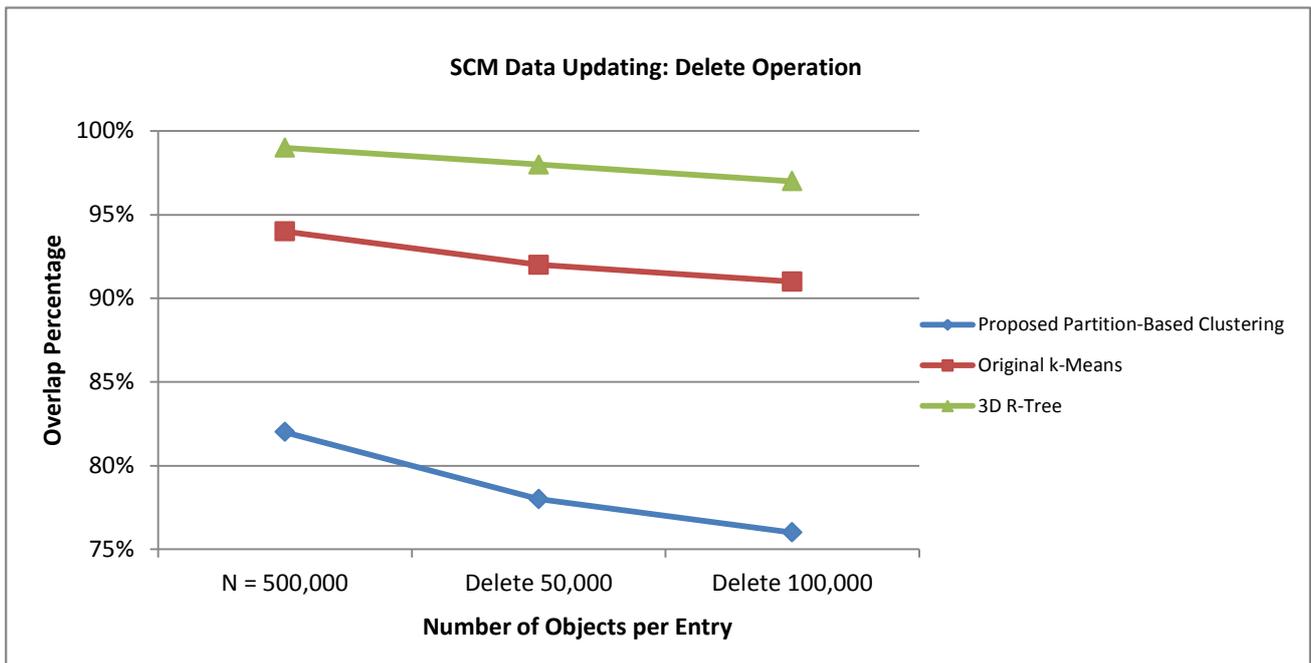


Figure 10. Effect of data deletion towards the overlap among node.

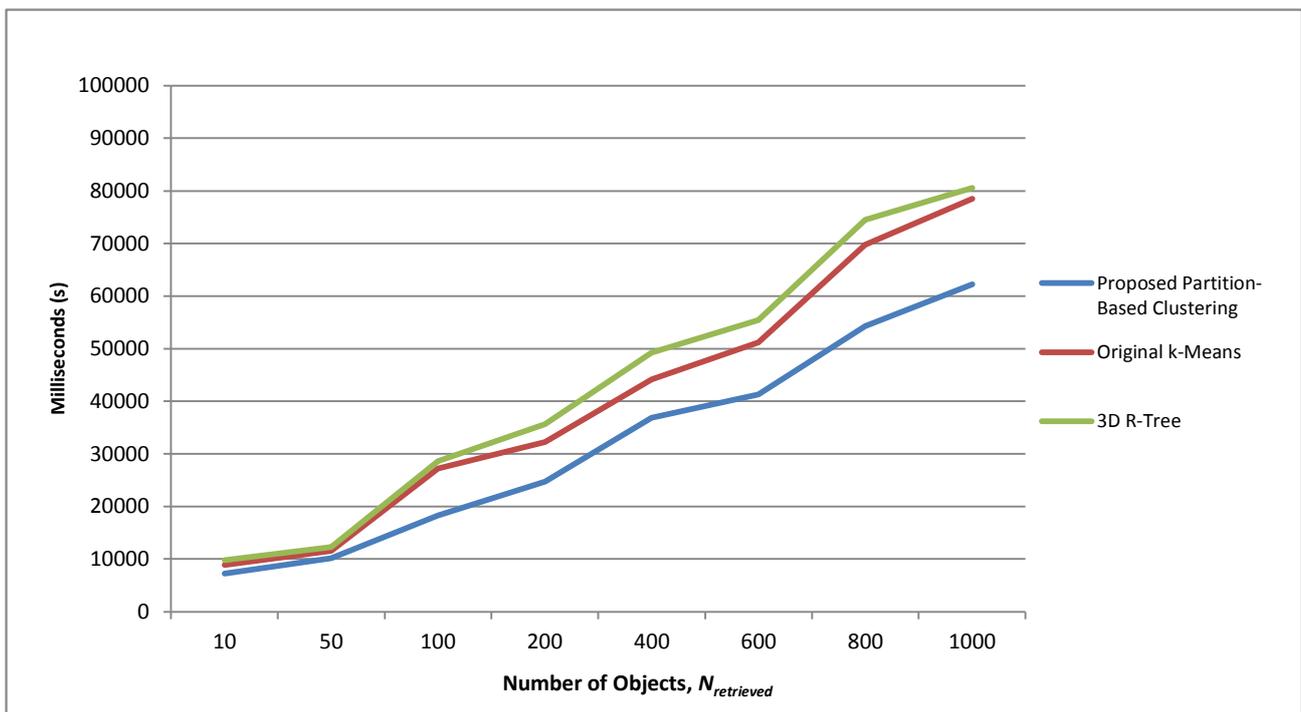


Figure 11. Time retrieval for several groups of data.

4.4 Response Time Analysis

Since indexing improves the data retrieval capabilities, search analysis is performed in order to show the improvement. In this test, the search operation is performed based on *cluster_id* and location. Data retrieval is measured in millisecond (s) and it was tested in a windows operating system with single Intel Xeon running at 2.2GHz of 4GB Random Access Memory (RAM). The graph shown in Figure 11 shows a set of 300,000 data of SCM is used to retrieve a different number of records $N_{retrieved}$. From the result, the proposed hierarchical tree structure offers a lower data retrieval time, which is 30% faster than the original 3D R-Tree and 20% faster than hierarchical structure based on original *k*-means.

5 CONCLUSIONS

This paper proposed an efficient version of partition-based clustering algorithm to constellate SCM data in geospatial databases for efficient data retrieval and analysis. The structure is constructed based on the group of clusters. To the best of our knowledge, this is the first data constellation structure based on a clustering approach for SCM case or scenario.

The findings and applications, resulting from the comprehensive tests and analyses of the proposed structure are discussed below. First, the SCM datasets were identified based on cluster id and volume delimitation. On the basis of this identification, data and information were directly retrieved from specific groups. Compared to non-constellated data, each record in the database table needed to be scanned or visited for the filtering process. In the overlap test, the capability of the proposed clustering algorithm in reducing overlap was examined. The results show that the percentage of overlap between nodes are substantially reduced compared with the existing tree structure in the database and tree structure based on original *k*-means algorithm. In the final test, the query response time of for data retrieval was performed using the structure. Based on the test, the proposed hierarchical tree structure shows an advantage in fast acquiring information retrieval for SCM dataset.

On the basis of these findings, we believe that the proposed hierarchical structure of partition-based clustering is suitable for handling SCM datasets. However, we suggested further testing to facilitate further improvements. To extend the analysis of this structure, we suggest to perform network distance analysis between manufacturer, supplier and customer. The objective is to acquire accurate distance calculation. This may help the community in the SCM process to plan each process within their time frame.

ACKNOWLEDGEMENT

This work is supported by the Ministry of Education, Malaysia under the MyPhD, MyBrain15 initiatives and UTM Research University Grant, Vote J.130000.2427.02G77.

REFERENCES

- Akhbari, M., Zare Mehrjerdi, Y., Khademi Zare, H., Makui, A., 2014. A Novel Continuous kNN Prediction Algorithm to Improve Manufacturing Policies in a VMI Supply Chain. *International Journal of Engineering, Transactions B: Applications* 27, 1681-1690.
- Arthur, D., Vassilvitskii, S., 2007. *k*-Means++: The Advantages of Careful Seeding, *Proceedings of The Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, Pp. 1027-1035.
- Bellman, R., 1962. Dynamic Programming Treatment of The Travelling Salesman Problem. *Journal of ACM* 9, 61-63.
- Craighead, C.W., Ketchen, D., Dunn, K.S., Hult, G., 2011. Addressing Common Method Variance: Guidelines for Survey Research on Information Technology, Operations, and Supply Chain Management. *Engineering Management, IEEE Transactions on* 58, 578-588.
- Delen, D., Erraguntla, M., Mayer, R., Wu, C.-N., 2011. Better Management of Blood Supply-Chain With GIS-Based Analytics. *Ann Oper Res* 185, 181-193.
- Dewitt, T., Giunipero, L.C., Melton, H.L., 2006. Clusters And Supply Chain Management: The Amish Experience. *International Journal of Physical Distribution & Logistics Management* 36, 289-308.
- Dhillon, I.S., Modha, D.S., 2001. Concept Decompositions for Large Sparse Text Data Using Clustering. *Machine Learning* 42, 143-175.
- Dinu Popa, M.D., 2014. Supply Chain Management—Key Factors. *Network Intelligence Studies*, 32-37.
- Gunasekaran, A., Ngai, E.W.T., 2004. Information Systems In Supply Chain Integration and Management. *European Journal of Operational Research* 159, 269-295.
- Guttman, A., 1984. R-Trees: A Dynamic Index Structure for Spatial Searching. *Sigmod Rec.* 14, 47-57.
- Hasan, M.A., Chaoji, V., Salem, S., Zaki, M.J., 2009. Robust Partitional Clustering By Outlier and Density Insensitive Seeding. *Pattern Recognition Letters* 30, 994-1002.
- Kaufman, L., Rousseeuw, P.J., 2008. *Divisive Analysis (Program Diana), Finding Groups In Data*. John Wiley & Sons, Inc., Pp. 253-279.
- Kiekintveld, C., Miller, J., Jordan, P.R., Wellman, M.P., 2007. Forecasting Market Prices in a Supply Chain Game, *Proceedings of The International Conference on Autonomous Agents*, Pp. 1323-1330.
- Lancioni, R.A., Smith, M.F., Oliva, T.A., 2000. The Role of The Internet in Supply Chain Management. *Industrial Marketing Management* 29, 45-56.

Miocevic, D., Crnjak-Karanovic, B., 2012. The Mediating Role of Key Supplier Relationship Management Practices on Supply Chain Orientation-The Organizational Buying Effectiveness Link. *Industrial Marketing Management* 41, 115-124.

Murray, C., 2009. Oracle Spatial Developer's Guide, 11g Release 1 (11.1), Oracle Spatial User's Guide And Reference, Pp. 1 - 870.

Power, D., 2005. Supply Chain Management Integration and Implementation: A Literature Review. *Supply Chain Management: An International Journal* 10, 252-263.

Ravada, S., Kazar, B., Kothuri, R., 2009. Query Processing In 3D Spatial Databases: Experiences With Oracle Spatial 11g, In: Lee, J., Zlatanova, S. (Eds.), *3D Geo-Information Sciences*. Springer Berlin Heidelberg, Pp. 153-173.

Rodger, J.A., 2014. A Fuzzy Nearest Neighbor Neural Network Statistical Model for Predicting Demand for Natural Gas and Energy Cost Savings in Public Buildings. *Expert System Application* 41, 1813-1829.

Selim, S.Z., Ismail, M.A., 1984. *k*-Means-Type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality. *Pattern Analysis and Machine Intelligence, IEEE Transactions On Pami-6*, 81-87.

Zhu, Q., Gong, J., Zhang, Y., 2007. An Efficient 3D R-Tree Spatial Index Method For Virtual Geographic Environments. *ISPRS Journal of Photogrammetry and Remote Sensing* 62, 217-224.