LOGISTICS OPTIMIZATION IN SUPPLY CHAIN MANAGEMENT USING CLUSTERING ALGORITHMS

R. MAHESH PRABHU,* M.S. HEMA,† SRILATHA CHEPURE,‡ AND NAGESWARA GUPTHA M.§

Abstract. Today's business environment, survival and making profit in market are the prime requirement for any enterprise due to competitive environment. Innovation and staying updated are commonly identified two key parameters for achieving success and profit in business. Considerably supply chain management is also accountable for profit. As a measure to maximize the profit, supply chain process is to be streamlined and optimized. Appropriate grouping of various suppliers for the benefit of shipment cost reduction is proposed. Data relating to appropriate attributes of supplier logistics are collected. A methodology is proposed to optimize the supplier logistics using clustering algorithm. In the proposed methodology data preprocessing, clustering and validation process have been carried out. The Z-score normalization is used to normalize the data, which converts the data to uniform scales for improving the clustering performance. By employing Hierarchical and K-means clustering algorithms the supplier logistics are grouped and performance of each method is evaluated and presented. The supplier logistics data from different country is experimented. Outcome of this work can help the buyers to select the cost effective supplier for their business requirements.

Key words: Supply Chain Management (SCM), Supplier logistics, Optimization, Hierarchical clustering, K-means clustering

AMS subject classifications. 90B06, 62H30

1. Introduction. Researchers have focused on most of the processes of manufacturing supply chains. More attention has been shown on both optimization and effectiveness improvement of supply chain. Gradual increase in cost for raw material, manufacturing, logistics, distribution and inventory is the main reason to focus attention on supply chain. Apart from the cost involved, time is an important criterion challenging the effectiveness of a supply chain. Two basic process of supply chain are inventory control and production planning process and logistics and distribution process. These two processes holds the responsibility of the manufacturing process design and management. First process includes scheduling and acquisition, design and scheduling of manufacturing process and design and control of material handling. Second process completely governs logistics and transportation of the products, which counts the quantity of product to be shipped from warehouse to retailers. Inventory management, shipment of products and delivering final product are major role of this process. Designing and analysis of supply chain are categorized into four namely 1. Analytical model (deterministic) which will be using specified and known variables. 2. Analytical model (Stochastic), with at least one unknown variable which is assumed to follow any one probability distribution. 3. Economic models 4. Simulation models

The main objective of supply chain management (SCM) is to maximize the total profit of the manufacturing facilities. Associated constraints are resource and production constraints and logical consistency constraints. Logical consistency constraints include demand limits, availability, variable non-negativity and feasibility. As a result, it is suggested to improve the effectiveness of SCM of a business. Instead of concentrating the entire supply chain activities it is suggested to choose a particular part and optimize it. As a continuation concentrate on all individual areas of supply chain and make the entire system as optimized one. In this work an algorithm based on clustering approach is developed to group the suppliers involved in the business. Grouped suppliers

---
*Department of Mechanical Engineering, Aurora's Scientific, Technological and Research Academy Hyderabad, India (`dr.rmaheshprabhu@gmail.com`).

†Department of Computer Science and Engineering, Aurora's Scientific, Technological and Research Academy Hyderabad, India

‡Department of Electronics and Communication Engineering, Aurora's Scientific, Technological and Research Academy Hyderabad, India

§Department of Computer Science and Engineering, Sri Venkateshwara College of Engineering, Bengaluru, India

1. Raw data collection of supplier
2. Normalize the supplier data using Z-score normalization
3. Cluster supplier data using hierarchical clustering and K-means clustering algorithm
4. Error calculation for intra and inter clusters and reassignment of elements among the clusters
5. Predict the accuracy of the modified K-means algorithm

FIG. 3.1. *Clustering Process in SCM*

are ranked in clusters. Based on the first cluster elements it is decided to be the best suppliers to be preferred first in placing the order and also the transportation route decision. The Z-score normalization is used to normalize the raw data. Hierarchical and K-means algorithms are used to cluster the supplier.

Introduction is followed by literature review, proposed methodology, results and discussion and the conclusion. Future scope for research are also pointed out.

2. Literature Survey. A detailed study on performance metrics of SCM was made [1]. Literature review on modeling a supply chain was made along with a new research agenda in SSM [2]. Quality measurement in SCM was carried out to ensure the right quantity of right product in right time [3]. An optimization based on network was proposed [5]. By using a special structure optimal inventory position was found in a reverse logistic system with a special structure [6]. Using agent oriented software architecture, the agent oriented SCM was built [7]. For the purpose of selecting and benchmarking potential supplier's artificial neural network and hybrid care based reasoning was used [8]. Inventory model was proposed with multiple supply chains options. It ensures their concrete contribution to SCM [9]. SC modeling using simulation was explained. It dealt the requirements for simulation modeling and issues of SCM. A complete casual forecasting systems using GA in SCM was illustrated. Forecasting in SCM is carried out by using a computerized system [10]. The author proposed metaheuristics based approach for SCM [11]. The author presented quantitative model using clustering for SCM [12]. Various types existing typologies in the supply chain risk management since 2000 used for analysis. Idea was proposed to from new typology [13]. Triangulrisation Clustering Algorithm and petri net were used to check vulnerabilities in supply chain networks and assessed disruption mitigation strategies in SCM [14]. In order to model, monitor and manage SCM a unified framework was proposed [15]. By conducting a questionnaire survey a framework was proposed to improve the supply chain sustainability in Indian e-waste management system [16]. Based on the existing supplier strength reconfiguration of supply base of buying firms was done [17]. Linkage between SCM and porters cluster theory was demonstrated [18]. In a different way clusters were inserted into global value chains in order to disable and enable the upgrading efforts in local level [19]. A frame work using dynamic nonlinear multi attribute decision model was developed [20]. In order to perform dynamic analyses of SCM data management a system based on information technology was developed [21]. Cost involved in total SCM was reduced by the authors with a new approach [22]. Suppliers were clustered using EM algorithms and DEA calculated the internal operation efficiency and AHP was applied to assess the importance of the external functions in a SCM [23]. A set of manufacturing firms were clustered using various clustering methods like fuzzy c means, Gaussian mixture model, self-organized map, k means and hierarchical clustering [24]. By using branch and bound algorithm route was found. Customers were clustered using random clustering, K-means and K-medoids algorithm [25].

3. Proposed Methodology. The main objective of proposed algorithms is to cluster the supplier and to reduce the overall all expenses of the SCM. The optimization will improve the performance and profit of the SCM. The proposed methodology is shown in Fig 3.1.

3.1. Normalization of data. In supply chain management, all the supplier data are in different scale. If the clustering is performed with raw data,then the accuracy of the data may be reduced. The solution to improve the accuracy of the data is to convert all into uniform scale. The Z score normalization is used to convert all data into uniform scale. The Z-score normalization formula is given in Equation 3.1:

$$Z = X - \mu/\sigma \qquad (3.1)$$

where Z is normalized value of corresponding data value. X is supplier logistic values of corresponding feature. μ is the mean of the corresponding features in the supplier logistics. In Z score nomalization, the mean is zero and standard deviation is one. σ is standard deviation of the supplier logistics feature.The standard deviation is calculated using Equation 3.2:

$$\sigma = \sqrt{(\Sigma(X - \overline{X})/n)} \tag{3.2}$$

where X is Supplier logistics values of corresponding feature. $(\overline{X})$ is the mean of the feature. n is total number of supplier logistics data sample The Z score is calculated using value of the supplier logistics data sample, standard mean and standard deviation.

3.2. Clustering using K-means algorithms. K-means clustering is unsupervised algorithm. The data are grouped based on feature similarity. The input of the algorithm is data set and number clusters. The normalized data is given as input. K-means algorithm has following three steps which are Initial centroid selection, distance calculation and assignment of elements in the cluster and calculation of cluster center and assignment of elements.

3.2.1. Initial centroid selection. The initial centroid of the cluster is calculated using formula shown in Equation 3.3:

$$D = (Kmeans[sum(i) - min])/((max - min)) + 1 \tag{3.3}$$

where D is the centroid of the cluster. $sum(i)$ is the sum of row elements for the given problem. max is the maximum values and min is the minimum values of $sum(i)$ respectively.

3.2.2. Distance calculation and assignment of elements in the cluster. The distance is calculated between data point and centroid of each cluster using formula shown in Equation 3.4:

$$D(i,l) = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{k}(L_k - Y_i)^2} \tag{3.4}$$

where $D(i,l)$ is distance between cluster(L) and i^{th} data point of the cluster. L_k is the centroid of the kth cluster. Y_i is the jth data point of the cluster. After calculation of distance, data point is allocated to one cluster which has the minimum distance from that data point.

3.2.3. Calculation cluster center and assignment of elements. After allocation of each data point to the cluster, the cluster center is calculated using formula shown in Equation 3.5.:

$$C_i = \sqrt{\sum_{i=1}^{n}x_i/n} \tag{3.5}$$

where C_i is the center of the cluster. x_i is the value of the features in the cluster. n is the total elements in the cluster. After that, the distance is calculated between remaining data points and centroid of the cluster. The assignment of data points to cluster is done. The center calculation and assignment procedure is repeated until all the elements are assigned to clusters.

3.2.4. Distance calculation and assignment of elements in the cluster. The sum Squared Error(SSE) between each observation and its group's center is calculated. It is used as a measure of variation within a cluster. If SSE is equal to 0 then all the cases within the cluster are identical. The SSE is calculated using formula shown Equation 3.6:

$$SSE = \sum_{i=1}^{p}\sum_{q\epsilon O_i}(q - C_i)^2 \tag{3.6}$$

where p is the total number of cluster. O is the number of objects in a cluster. C is the center point of a cluster. The SSE is calculated for each cluster. Minimum SSE will give more accuracy. To minimize the SSE is shown Equation 3.7, the center point is calculated using formula shown in Equation 3.8:

$$SSE = min \sum_{i=1}^{k} \sum_{q \epsilon O_i} (q - C_i)^2 \tag{3.7}$$

$$C_i = \sum_{X \epsilon C_i} (X)/|C_i| \tag{3.8}$$

where C_i is the center point of the cluster. X is the element of the cluster. $|C_i|$ is the number center points of the cluster. The center point is calculated using data points. Which data point gives minimum SSE with center point then that data point is moved that cluster. It minimizes the error within the cluster then it will reduce the SSE.

3.3. Clustering using Hierarchical clustering algorithm. In hierarchical clustering is used group the data in hierarchical manner. There are two types of hierarchical clustering namely agglomerative and divisive. In proposed, the complete link and average link agglomerative clustering is used to cluster the supplier data.

Agglomerative hierarchical clustering

1) Start with the independent cluster with level 0 and sequence is zero
2) If the clustering method is single link the the minimum distance between elements of each cluster, say pair $(r), (s)$ is calculated using Equation 3.9:

$$SLD = min\{distance(r, s) : r \epsilon C1, s \epsilon C2\} \tag{3.9}$$

where SLD is single link minimum distance between cluster points. r and s are data points. $C1$ and $C2$ are clusters. If the clustering method is complete then the maximum distance between elements of each cluster, say pair $(r), (s)$ is calculated using Equation 3.10.

$$CLD = max\{distance(r, s) : r \epsilon C1, s \epsilon C2\} \tag{3.10}$$

where CLD is complete link maximum distance between cluster points. r and s are data points. $C1$ and $C2$ are clusters.

3) Join cluster (r) and (s) combined to form new cluster in the next level clustering. The sequence number is incremented by one.
4) Remove the old distance matrix value of r and s and update new distance matrix. The new cluster distance is calculated using Equation 3.11:

$$distance(k, (r, s)) = min\{distance[(k), (r)], distance[(k), (s)]\} \tag{3.11}$$

where k is old cluster points. r and s are new cluster data points.
5) All points are merged into one cluster then stop the iteration, otherwise repeat step 2 to step5.

4. Results and discussion. The general goods of supplier logistic data are taken from the internet for experimentation. The goods are imported from foreign countries. The parameters consider for clustering are distance in KM, transportation cost and time to taken to transport. The total observation in the data set is 240. The suppliers are grouped based on the distance, time and cost using hierarchical clustering and K-means clustering. In hierarchical clustering with complete and average link is used to cluster the supplier data. The buyer can select the optimal supplier to import the goods from foreign countries based on the clustering results.

Various methods such as silhouette method (Fig. 4.1), gap statistics (Fig.4.2) and sum square (Fig.4.3) methods are used to find the number ideal number of cluster for hierarchical and K-means clustering. The analyses of methods concluded that the ideal number of cluster is 3. The range of silhouette chart is shown in Fig.4.4. It shows that reasonable structure has been formed.
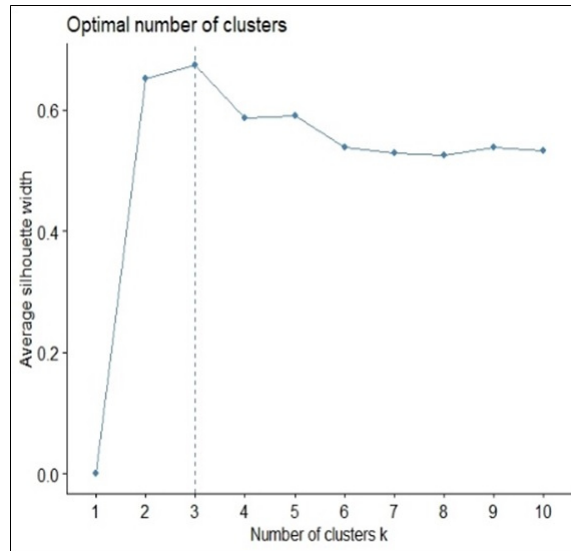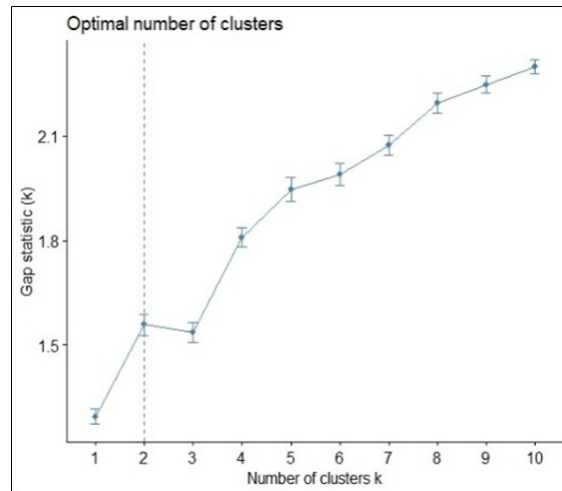
FIG. 4.1. *Silhoutte method*

FIG. 4.2. *Gap Statistic chart*

4.1. Hierarchical Clustering. In Hierarchical clustering, complete link and average link is used for clustering. The aggregation is presented in Fig. 4.5. In complete link clustering with normalized value, the aggregate means shows that the distance is positive value in cluster 1 and 2. So, it contributed more for clustering. But in cluster three cost contributed more than distance and time. Similar results for original data and average link also. Both complete linkage and average linkage are given the same result but complete linkage improves 20% distance than the average linkage. The performances show that K-means and hierarchical clustering algorithm improves the distance, cost and time 20%, 12% and 10% respectively.

4.2. K-means clustering. The performance of K-means algorithm with different K is shown Fig. 4.6.

5. Conclusion. The clustering algorithm is proposed and implemented to optimize the shipping cost from the supplier. The supplier logistic data is taken for clustering. The Z-score normalization has been used for normalizing data. It converted the data into uniform scale. The K-means and Hierarchical clustering algorithm with complete link and average link for clustering is implemented and evaluated. Outcome of this work can
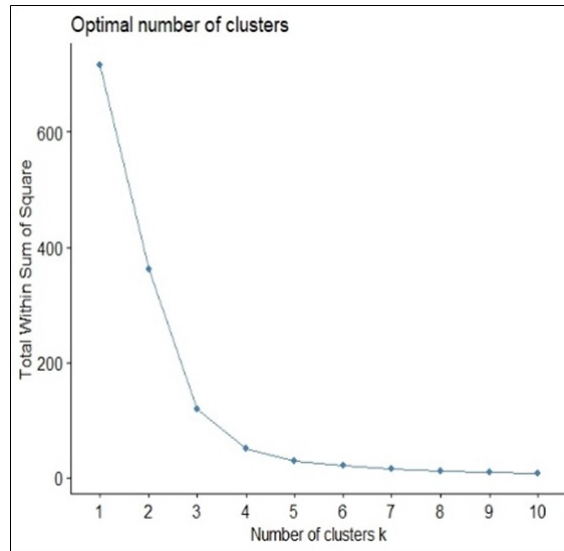
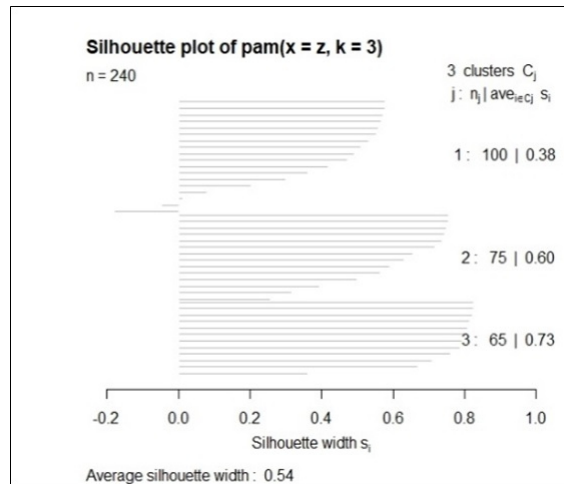Fig. 4.3. *Total width Sum of Square*
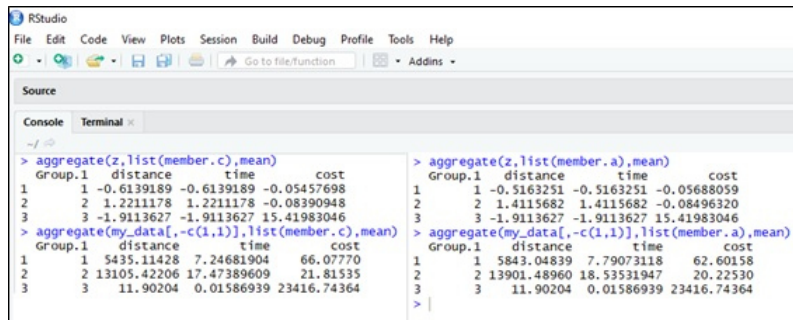

Fig. 4.4. *Range of Silhoutte Chart*

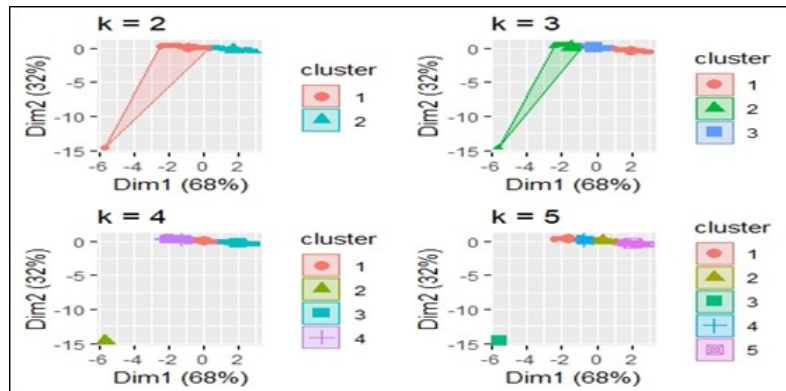Fig. 4.5. *Performance aggregation of Hierarchical clustering*

Fig. 4.6. *Performance aggregation of Hierarchical clustering*

help the buyers to select the cost effective supplier for their business requirements. In future, more parameters taken for clustering to optimize the supply chain management. The clustering algorithm can be used for all divisions of supply chain management. In future mean-shift clustering algorithm and expectation-maximization algorithm will be used for clustering.

REFERENCES

[1] Shepherd, C., Günter, H., *Measuring supply chain performance: current research and future directions. In: Behavioral Operations in Planning and Scheduling*, Springer, Berlin, Heidelberg (2010), pp. 105-121, 2010.
[2] S Beamon, B. M., *Supply chain design and analysis: Models and methods*, International journal of production economics, vol. 55, no. 3, pp. 281-294, 1998.
[3] Beamon, B. M., Ware, T. M., *A process quality model for the analysis, improvement and control of supply chain systems*, Logistics Information Management , vol. 11, no. 2, pp. 105-113, 1998.
[4] Luo, Y., Wirojanagud, P., Caudill, R. J., *Network-based optimization and simulation of sustainable e-supply chain management*,In Proceedings of the 2001 IEEE International Symposium on Electronics and the Environment. 2001 IEEE ISEE, pp. 185-190, 2001.
[5] Ganeshan, R., Jack, E., Magazine, M. J., and Stephens, P, *A taxonomic review of supply chain management research. In Quantitative models for supply chain management*, Springer, Boston, pp. 839-879, 1999.
[6] Spekman, R. E., Kamauff Jr, J. W., Myhr, N., *An empirical investigation into supply chain management: a perspective on partnerships*, . Supply Chain Management: An International Journal, vol. 3, no. 2, pp. 53-67, 1998.
[7] Barbuceanu, M., Teigen, R., Fox, M. S., *Agent based design and simulation of supply chain systems. In Enabling Technologies: Infrastructure for Collaborative Enterprises*, Proceedings., Sixth IEEE Workshops, pp. 36-41, 1997.
[8] Anderson, D. L. Ed., *Achieving Supply Chain Excellence Through Technology*, Thought Leadership Project from Montgomery Research. Montgomery Research, 1997.
[9] Minner, S., *Multiple-supplier inventory models in supply chain management*, A review. International Journal of Production Economics, vol. 81, pp. 265-279 , 2003.
[10] Jeong, B., Jung, H. S., and Park, N. K., *A computerized causal forecasting system using genetic algorithms in supply chain management*, Journal of Systems and Software, vol. 60, no. 3, pp. 223-237, 2002.
[11] Ramalhinho Dias Lourenço, H., *Supply chain management: An opportunity for metaheuristics*, 2001.
[12] Brandenburg, M., and Rebs, T, *Sustainable supply chain management: A modeling perspective*, Annals of Operations Research, vol. 229, no. 1, pp. 213-252, 2015.
[13] Ho, W., Zheng, T., Yildiz, H., and Talluri, S, *Supply chain risk management: a literature review. *, International Journal of Production Research, vol. 53, no. 16, pp. 5031-5069, 2015.
[14] Gang Wang, Angappa Gunasekaran, Eric W.T. Ngai, Thanos Papadopoulos, *Big data analytics in logistics and supply chain management: Certain investigations for research and applications*, International Journal of Production Economics, vol. 176, no. 2016, pp. 98-110, 2015.
[15] Nirupam Julka 1, Rajagopalan Srinivasan, I. Karimi, *Agent-based supply chain management framework*, Computers and Chemical Engineering, vol. 26, pp. 1755 -1769, 2002.
[16] Baidya, R., Debnath, B., Ghosh, S. K., *Analysis of E-Waste Supply Chain Framework in India Using the Analytic Hierarchy Process, In Waste Management and Resource Efficiency*, Springer, Singapore, vol. 55, no. 3, pp. 281-294, 2019.
[17] Talluri, S., Narasimhan, R., *A note on A methodology for supply base optimization*, IEEE Transactions on Engineering Management, vol. 52, no. 1, pp. 130-139, 2005.
[18] DeWitt, T., Giunipero, L. C., Melton, H. L., *Clusters and supply chain management: The Amish experience*, International

Journal of Physical Distribution and Logistics Management, vol. 36, no. 4, pp. 289-308, 2006.

[19] HUMPHREY, J., SCHMITZ, H., *How does insertion in global value chains affect upgrading in industrial clusters?*, Regional studies, vol. 36, no. 9, pp. 1017-1027, 2002.

[20] SARKIS, J., *A strategic decision framework for green supply chain management*, Journal of cleaner production, vol. 11, no. 4, pp. 397-409, 2003.

[21] CHAE, B., OLSON, D. L., *Business analytics for supply chain: A dynamic-capabilities framework*, International Journal of Information Technology and Decision Making, vol. 12, no. 1, pp. 9-26, 2013.

[22] RADHAKRISHNAN, P., PRASAD, V. M., GOPALAN, M. R., *Inventory optimization in supply chain management using genetic algorithm*, International Journal of Computer Science and Network Security, vol. 9, no. 1, pp. 33-40, 2009.

[23] CHUL PARK, S., LEE, J. H., *Supplier selection and stepwise benchmarking: a new hybrid model using DEA and AHP based on cluster analysis*, Journal of the Operational Research Society, vol. 69, no. 3, pp. 449-466, 2018.

[24] BHATNAGAR, V., MAJHI, R., JENA, P. R., *Comparative performance evaluation of clustering algorithms for grouping manufacturing firms*, Arabian Journal for Science and Engineering, vol. 43, no. 8, pp. 4071-4083, 2018.

[25] COMERT, S. E., YAZGAN, H. R., KIR, S., YENER, F., *A cluster first-route second approach for a capacitated vehicle routing problem: a case study*, International Journal of Procurement Management, vol. 11, no. 4, pp. 399-419, 2018.