MATTER: International Journal of Science and Technology
ISSN 2454-5880

Kananke et al., 2016

Volume 2 Issue 2, pp.48-64

Year of Publication: 2016

DOI-http://dx.doi.org/10.20319/mijst.2016.22.4864

This paper can be cited as: Fabregas, A., Gerardo, B., & Tanguilig Iii, B. (2016). Modified Selection Of Initial Centroids For K- Means Algorithm. MATTER: International Journal of Science and Technology, 2(2), 48-64.

MODIFIED SELECTION OF INITIAL CENTROIDS FOR K-MEANS ALGORITHM

Aleta C. Fabregas
Graduate Programs, Technological Institute of the Philippines,Quezon City, Philippines
alet_fabregas@yahoo.com

Bobby D. Gerardo
Institute of Information and Communication Technology, West Visayas State University, Lapaz, Iloilo City, Philippines
bgerardo@wvsu.edu.ph

Bartolome T. Tanguilig III
Graduate Programs, Technological Institute of the Philippines, Quezon City, Philippines
bttanguilig_3@yahoo.com

Abstract

This study focuses on the improved initialization of initial centroids instead of random selection for the K-means algorithm. The random selection of initial seeds is a major drawback of the original K-means algorithm because it leads to less reliable result of clustering the data. The modified approach of the k-means algorithm integrates the computation of the weighted mean to improve the seeds initialization. This paper shows the comparison of K-Means and Modified K-Means algorithm, using the first simple dataset of four objects and the dataset for service vehicles. The two simple applications proved that the

Modified K- Means of selecting initial centroids is more reliable than K-Means Algorithm. Clustering is better achieved in the modified k-means algorithm,

Keywords

K-Means Algorithm, Euclidian Distance, Centroids, Clustering, Modified-K-Means Algorithm, Weighted Average Mean

---

1. Introduction

K-means algorithm is a very well-known clustering method. It involves cluster analysis, with the task of grouping a set of objects that are more similar to each other than to those in other groups. This is a prototype- based; partitional clustering technique intended to find a user-specified number of clusters (K).

The authors focused on the use of K-means algorithm for the reason that it is easy to use and implement, scalability, speed of convergence and adaptability to distribute data. This algorithm has been successfully used in various topics, including market segmentation, computer vision, geostatistics, astronomy and agriculture (K-means clustering From Wikipedia, the free encyclopedia). Since K-means algorithm is used to cluster or distribute data to similar, or meaningful groups, this concept could be easily adapted to other applications in which groupings or classification of data is highly necessary for decision making. The main objective of the "Modified K-means" is to change the methods of selecting Initial centroids. Instead of providing a random cluster point in K-mean, the modified approach is using the Weighted Mean as basis for selecting initial centroids. The purpose of this study is to eliminate the random selection of the initial centroids as one of the major limitations of the K-means algorithm because it leads to unreliable result, and to increase the accuracy of clustering. The modified version computes the highest pair and lowest pair of each point of weighted mean for each attribute prior to the selection of the initial centroids

2. Background

The K-Means algorithm is a simple repetition method to classify a given dataset into a user specified number of clusters, *k*. It is also referred to as Lloyd's algorithm, particularly in the computer science community (*k*-means clustering From Wikipedia, the free encyclopaedia). The procedure is initiated by selecting k points as the initial k cluster centroids of weighted mean. The original algorithm for selecting these initial seeds include random sampling from the dataset, setting them as the solution of clustering a small subset of the data. Then the algorithm repeats between two steps till reaching the finite limit:

- Data Assignment. Each data point is assigned to its closest centroid. This results in a separation of the data.
- Relocation of means. Each group representative is relocated to the center (mean) of all data points assigned to it.

The procedure reaches the limit when the data assignments no longer change. To quantify closeness in the assignment step, the squared Euclidean distance formula is used (Kushwah S. P. S,.Rawat K, Gupta P. 2012). This formula is used for determining the least sum of squares which is the objective function of the said algorithm.

2.1 The Advantages of and Disadvantages of K-means Algorithm

The Advantages of K-means algorithm are fast and easier to understand, relatively efficient and produces best result when data set are well separated from each other. (A Tutorial on Clustering Algorithms). And some of the disadvantages are the procedure requires apriority specification of the number of cluster centers. The use of Exclusive Assignment - If there are two highly overlapping data then k-means will not be able to resolve that there are two clusters. And the randomly selection of the cluster center cannot lead us to better result (A Tutorial on Clustering Algorithms). The authors would be proposing to improve its performance by eliminating one of the disadvantages which is the random selection of the cluster center. Partitioning result of the k-mean clustering algorithm greatly depends upon the correctness of the initial centroids, which are selected randomly. "In mathematics and physics, the centroid is the arithmetic mean ("average") position of all the points in the shape (Weighted_average (*https://*simple.m.wikipedia.org*/wiki/).*

The random selection is not based on any computation or algorithm. The way to initialize the means was not specified. The reliability of the results produced depend on the initial values for the k-means.

3. The Operational Concept Of The Original K-Means Algorithm And The Modification

This study will show the original steps of K-means algorithm and the modified version of K-means techniques. The original method follows a simple and easy way to classify a given data set (sharing common traits). The main objective is to specify *k* centers, one for each cluster or subset. The better choice is to place them as much as possible far away from each other. Next step is to take each point belonging to a given data set and associate it to the nearest center. When no point is pending, the first step is finished. At this point we need to re-calculate the *k* new centroids as barycenter of the clusters resulting from the previous step. Iteration has been generated. As a result of this loop, notice that the k centers change their location step by step until no more changes are done Finally, this algorithm aims at minimizing an objective function known as squared error function given by: (Teknomo, K. PhD, 2007).

$$J = \sum_{j-1}^{k} \sum_{i-1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \tag{1}$$

Where $\left\| x_i^{(j)} - c_j \right\|^2$ provides the distance between an entity point and the cluster's centroid.

The study will use same datasets and procedures (K-means Clustering algorithm- Data Clustering Algorithms) for the original and modified K-means with the weighted average mean.

Table 1: ***The original dataset for clustering using the original K-means algorithm***

Object	X	Y
A	1.0	1.0
B	2.0	1.0
C	4.0	3.0
D	5.0	4.0

In the original method of K-means algorithm, with the random selection of the initial centroids, we assumed that the centroid 1 is (1.0, 1.0) of the object A and the centroid 2 is (2.0, 1.0). The sample computation between the centroid 1 to each object is shown below using Euclidean distance formula.

a. $\sqrt{(1.0-1.0)^2+(1.0-1.0)^2} = 0.0$

b. $\sqrt{(2.0-1.0)^2+(1.0-1.0)^2} = 1.0$

And the sample computation for the centroid 2 is shown below.

a. $\sqrt{(1.0-2.0)^2+(1.0-1.0)^2} = 1.0$

b. $\sqrt{(2.0-2.0)^2+(1.0-1.0)^2} = 0.0$

Table 2: *The result of the K-means algorithm using the Euclidian distance based on the First Iteration*

Object	Distance to Centroid 1	Distance to Centroid 2
A	0.00	1.00
B	1.00	0.00
C	3.61	2.83
D	5.00	4.24

The result of Object Clustering based on the minimum distance is: Group 1 = Object A and Group 2 = Objects B, C and D, next the new centroid is to be computed for each group based on the new memberships. Group 1 has only one member thus the centroid remains in c1 = (1, 1). Group 2 has three members, thus the centroid is the average coordinate among the three members: c2 = (2.0+4.0+5.0)/3, (1.0+3.0+4.0)3 = (3.67, 2.67).After the second iteration, the result of clustering is: Group 1 = Object A and B, and

Group 2 = Object C and D. With new membership, another centroid cluster is computed: c1= (1.0+2.0)/2, (1.0+1.0)/2= (1.5, 1.0), c2 = (4.0+5.0)/2, (3.0+4.0)/2 = (4.5, (3.5).

Table 3. shows the result of repeating the major steps of the original method of K-means algorithm: 1. Determination of centroids, 2.Computation of Objects centroid distance using the Euclidean distance and 3. Objects clustering.

Table 3: *The result of K-means algorithm using the Euclidean Distance based on second and third iteration*

Object	2nd Iteration Distance to C1	2nd Iteration Distance to C2	3rd Iteration Distance to C1	3rd Iteration Distance to C2
A	0.00	3.14	0.50	4.30
B	1.00	2.36	0.50	3.54
C	3.61	0.47	3.20	0.71
D	5.00	1.89	4.61	0.71

The final groupings after the three (3) iterations are shown in Table 4.

Table 4: *The final clustering based on the original K-means algorithm*

Object	X	Y	Group Result
A	1.0	1.0	1
B	2.0	1.0	1
C	4.0	3.0	2
D	5.0	4.0	2

The results of the original method of K-means algorithm have three (3) iterations.

3.1 Modified K-means Algorithm

The modified algorithm does not require complicated calculation to obtain better initial centroids but relatively easy mathematical computation. Since the centroid is the average position of all the points in the dataset that indicates equal weight, weighted average is reflecting the real weight of a point from the given data set. The proposed integration of the weighted average for initial centroids shows immediately the clear separation of the clustering between points and overlapping between the groups is minimized. The weighted mean that represents corresponding weight of each point in the given attribute is used to obtain the highest and lowest pair of weighted average, these pairs will be the basis of the initial centroids. The modification is presented in figure 1 the algorithm of modified K-means.

1. To start with any initial partition that groups the data into k cluster;
- Get the highest or the perfect point for each attribute or column.
- Compute the weighted average of each point in the X and Y columns. Each point represents the two coordinate component of the object.
- The highest pair and lowest pair of weighted mean, will be the basis of initial groups of centroids. Perform the Exclusive assignment - Assign each of the remaining $N - k$ examples to the cluster with the nearest centroid. After each assignment, recalculate the new groups of centroid.
2. Take each object in sequence annot currently in the group with the

Figure 1: *The Modified K-means algorithm*

The modified K-means algorithm with improved initial centroids applying the weighted mean of the data sets reduced the iteration steps of the algorithm thus, reducing the computational complexity. Obtaining the highest pair of weighted mean and lowest pair of weighted mean from the pair of X and Y coordinate improves the K-means algorithm because it also clearly separates the clustering of the datasets with fewer iterations. With the modification, it eliminates not only the random selection of initial centroids but also the Euclidean distance computation for clustering is lessened. Thus, the stability or non-movement of objects is easier to achieve.

3.2 Mathematical definition of Weighted Mean

A weighted average is the average of values which are scaled by importance. The weighted average of values is the sum of weights times values divided by the sum of the weights (Weighted_average*https://*simple.m.wikipedia.org*/wiki/).* The following is the formula for the weighted mean

$$\text{Weighted Mean} = \left(\sum wx\right)/\left(\sum w\right) \quad (2)$$

Therefore data elements with a high weight contribute more to the weighted mean than do elements with a low weight.

The detailed application of the simple steps to find better centroids for the modified K-means algorithm is used in the same datasets applying the original K-means for comparison.

- Determine the highest point for the attribute X which is 5.0.
- And compute the weighted average by dividing each point to the highest point in the dataset.

The weighted average of each point is:

X [1] = 1.0/5.0*1.0=0.20, X [2] = 2.0/5.0* 1.0 = 0.40, X [3] = 4.0/5.0 * 1.0 = 0.80,

And X [4] = 5.0/5.0* 1.0 = 1.00

Same steps will be applied for the weighted mean of Attribute Y. The highest point is = 4.0.

The weighted average of each point for Y is:

Y[1] = 0.25, Y [2] = 0.25,Y [3] = 0.75, andY [4[=1.00.

Table 5: *Result of using computed Weighted mean from X and Y attributes*

Object	X	Y	Weighted Mean(X)	Weighted Mean (Y)
A	1.0	1.0	0.20	0.25
B	2.0	1.0	0.40	0.25
C	4.0	3.0	0.80	0.75
D	5.0	4.0	1.00	1.00
Highest point	5.0	4.0	1.00	1.00

The highest weighted mean for the two attributes is the object D (X = 1.00and Y = 1.00) and the lowest pair of weighted mean is the object A(X = 0.20 and Y =0.25) and this technique is the basis of obtaining the two (2) centroids, the highest and the lowest pair of computed weighted mean. So the initial centroids will be Centroid 1 = (1.0, 1.0) and the Centroid 2 = (5.0, 4.0). The calculation of the distance between cluster centroid to each object is traditionally achieve by using the formula of the Euclidean distance.

Iteration0. The sample computation to know the distance between the centroid 1 (1.0, 1.0) using Euclidean distance is shown below:

a. $\sqrt{(1.00 - 1.00)^2 + (1.00 - 1.00)^2} = 0.00$

b. $\sqrt{(2.00 - 1.00)^2 + (1.00 - 1.00)^2} = 1.00$

The sample computation of the Euclidean Distance of the objects for centroid 2 = (5.0, 4.0) are shown below:

a. $\sqrt{(1.00 - 5.00)^2 + (1.00 - 4.00)^2} = 5.00$

b. $\sqrt{(2.00 - 5.00)^2 + (1.00 - 4.00)^2} = 4.24$

The initial result of the first iteration using Euclidian distance is shown in Table 5.

Table 6: *Result of the first iteration using the modified k-means algorithm*

Object	Distance to C1	Distance to C2
A	0.00	5.00
B	1.00	4.24
C	3.61	1.41
D	5.00	0.00

The distance between centroid 1(1.0, 1.0) = (0.00, 1.00, 3.61, 5.00). And the distance between centroid 2 (5.0, 4.0) = (5.00, 4.24, 1.41, 0.00).Based on the computation, the Objects Clustering is determined by checking the minimum distance of the objects to the two centroids the assignment is show in Table 7.

Table 7: *Assignment based on the result of Object clustering of the Modified K-means algorithm*

Object Clustering 1	Group 1	Group 2
	A, B	C, D

In the Iteration 0, Group 1 has already 2 members, same with Group 2. In the previous computation of K-means algorithm, Group 1 has only one member and Group 2 has 3 members. Because of the additional new members for Group 1, and Group 2, the new computed centroids for Object clustering.

The new centroid 1 will be (1.0 + 2.0)/2 (1.0+1.0)2 = (1.50, 1.00).The new centroid 2 will be (4 +5)/, (3 +4)/2 = (4.50, 3.50)

a. $\sqrt{(1.00-1.50)^2+(1.00-1.00)^2} = 0.50$

b. $\sqrt{(2.00-1.50)^2+(1.00-1.00)^2} = 0.50$

The new computed distance of the objects to centroid 1 = (0.50, 0.50, 3.20, 0.50)

The sample computation of the Distance of the objects for centroid 2 = (4.5, 3.5) are shown below:

a. $\sqrt{(1.00-4.50)^2+(1.00-3.50)^2} = 4.27$

b. $\sqrt{(2.00-4.50)^2+(1.00-3.50)^2} = 2.23$

The new computed distance of the objects to centroid 2= (4.27, 2.23, 0.71, 0.31)

Table8: *Result of the Second iteration using the modified k-means algorithm*

Object	Distance to C1	Distance to C2
A	0.50	4.27
B	0.50	2.23
C	3.20	0.71
D	5.00	0.71

Comparing the grouping of last iteration of the original K-means algorithm and this iteration reveals that the objects does not move the groupings anymore. Thus, the computation of the modified k-mean clustering has reached its stability faster and no more iteration is needed. The number of iterations as compared to the previous application of the old k-means algorithm is reduced. Because of the selection of the highest pair of weighted average and lowest pair of weighted average of the X and Y coordinates, there is a significant decrease of the complexity of computing the Euclidean distance. Based on the computation of the Enhanced K-means with initial centroids obtained from the highest and lowest weighted average of the pair X and Y, the two groups are shown below in Table9.

Table 9: *Final Clustering using Modified K-means algorithm.*

Object	X	Y	Weighted Mean (X)	Weighted mean (Y)	Final Grouping
A	1	1	0.20	0.25	1
B	2	1	0.40	0.25	1
C	4	3	0.80	0.75	2
D	5	4	1.00	1.00	2
Highest point	5	4	1.00	1.00	

Using the Scatter chart for effective clustering, the final result of clustering using the modified K-means algorithm is shown below.

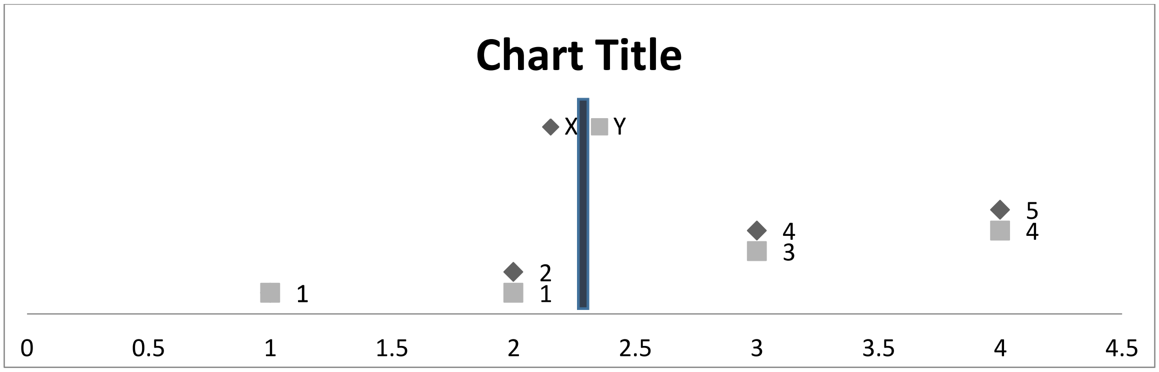

Figure 2: *The Scatter Chart showing the Clustering of the objects based on the Modified K-means algorithm*

For easy understanding, another application of the modification is the simple Fixed Asset Monitoring with Usage Factor and Age factor as the attributes of each service vehicle unit. The Service Vehicle is one of the fixed assets subject for depreciation. The Fixed Asset Monitoring uses Usage and Age factor as Depreciation factors to assess the performance of the Service Vehicle based on its remaining life and usefulness. The table below showed the application of the weighted average in determining the initial centroids with Usage factor and Age factor as the attributes of service vehicle unit. The results of computing the weighted average is also included in Table10.

Table 10: *The Table of Service Vehicles with their attributes using Modified K-means algorithm*

Fixed Asset	Age factor	Usage factor	Weighted Average of Age	WeightedAverageUsage
Service Vehicle 1	1.00	1.00	0.20	0.14
Service Vehicle 2	1.50	2.00	0.30	0.29
Service Vehicle 3	3.00	4.00	0.60	0.57
Service Vehicle 4	5.00	7.00	1.00	1.00
Service Vehicle 5	3.50	5.00	0.70	0.71
Service Vehicle	4.50	5.00	0.90	0.71

6				
Service Vehicle 7	3.50	4.40	0.70	0.63
Highest Point	5.00	7.00	1.00	1.00

The lowest weighted average pair is the Service Vehicle 1 = (1.0, 1.0) and the highest weighted average pair is the Service Vehicle 4 = (5.0, 7.0). The two pairs will be the initial centroids, c1 = (1.0.1.0) and c2 = (5.0, 7.0). Using the Euclidian Distance computation, the result is shown in the table below.

Sample Computation for c1 =(1.00, 1.00)

a. $\sqrt{(1.00-1.00)^2+(1.00-1.00)^2} = 0$

b. $\sqrt{(1.50-1.00)^2+(2.00.-1.00)^2} = 1.11$

Sample Computation for c2 = (5.0, 7.00)

a. $\sqrt{(5.00-1.00)^2+(7.00-1.00)^2} = 6.48$

b. $\sqrt{(5.00-1.50)^2+(7.00-2.00)^2} = 6.10$

Table 11: *Result of the First Iteration using Modified K-means algorithm*

Fixed Asset	Distance to C1	Distance to C2
Service Vehicle 1	0.00	6.48
Service Vehicle 2	1.11	6.10
Service Vehicle 3	3.61	3.61
Service Vehicle 4	7.21	0.00
Service Vehicle 5	4.72	2.50
Service Vehicle 6	5.31	2.06
Service Vehicle 7	4.22	3.00

The clustering based on the first iteration:Group 1 = (1.0, 1.0), (1.5, 2.0), Group 2 = (3.0, 4.0), (5.0, 7.0), (3.5, 5.0), (4.5, 5.0), and (3.5, 4.4)

Since Service Vehicle 3 = (3.0, 4.0) has an equal distance to centroid 1 = (1.0, 1.0) and centroid 2 = (5.0, 7.0), let the Service Vehicle 3 be clustered to the Second group.
The result from computing the new centroids for c1, (1.0+1.5)/2 and (1.0+2.0)/2 = (1.25, 1.50) and for c2, + (3.0 + 5.0+3.5+4.5+3.5)/5.0 and (4.0+ 7.0+5.0+5.0+4.4)/5 = (3.90, 5.08)

Sample Computation for c1 = (1.25, 1.50)

a. $\sqrt{(1.25-1.00)^2+(1.50-1.00)^2} = 0.56$

b. $\sqrt{(1.25-1.50)^2+(1.50-2.00)^2} = 0.56$

Sample Computation for c2 = (3.90, 5.08)

a. $\sqrt{(3.90-1.00)^2+(5.08-1.00)^2} = 5.01$

b. $\sqrt{(3.90-1.50)^2+(5.08-2.00)^2} = 3.90$

Table 12: *Result of the Second Iteration using Modified K-means algorithm*

Fixed Asset	Distance to C1	Distance to C2
Service Vehicle 1	0.56	5.01
Service Vehicle 2	0.56	3.90
Service Vehicle 3	3.05	1.41
Service Vehicle 4	6.66	2.21
Service Vehicle 5	4.16	0.41
Service Vehicle 6	4.78	0.61
Service Vehicle 7	3.67	0.79

Based on the result of the second iteration, the previous clustering is obtained and there is no more movement of the objects. Convergence is achieved for two (2) iterations.

Finally the groupings of the fixed asset based on the two attributes, age factor and usage factor are: Group 1 = Service Vehicle 1 =(1.0,1.0) and Service Vehicle 2 =(1.5,2.0), Group 2 are Service Vehicles 3 to Service Vehicles 7 with the following attribute values (3.0,4.0), (5.0,7.0),(3.5,5.0),(4.5,5.0) and (3.5,4.4).

Table 13: *Assignment based on the result of Object clustering of the Modified K-means algorithm*

Fixed Asset	Age factor	Usage factor	Partitioning
Service Vehicle 1	1.00	1.00	1
Service Vehicle 2	1.50	2.00	1
Service Vehicle 3	3.00	4.00	2
Service Vehicle 4	5.00	7.00	2
Service Vehicle 5	3.50	5.00	2
Service Vehicle 6	4.50	5.00	2
Service Vehicle 7	3.50	4.40	2
Highest Point	5.00	7.00	2

The partitioning of the Service Vehicles into two (2) groups based on the Age factor and Usage factor represents 1 as low performance, and 2 high performances.

Based from the two applications of the modified K-means algorithm, convergence or stability is easier to achieve, and better clustering is obtained with only limited number of computations.

The integration of the weighted mean in the modified algorithm initially reflects the real weights of the each pair of cluster, thus reducing the representation of the object in a deceiving way and reducing the Euclidian distance computation.

Representing the result of the Modified K-means algorithm with weighted average using the Scatter graph.

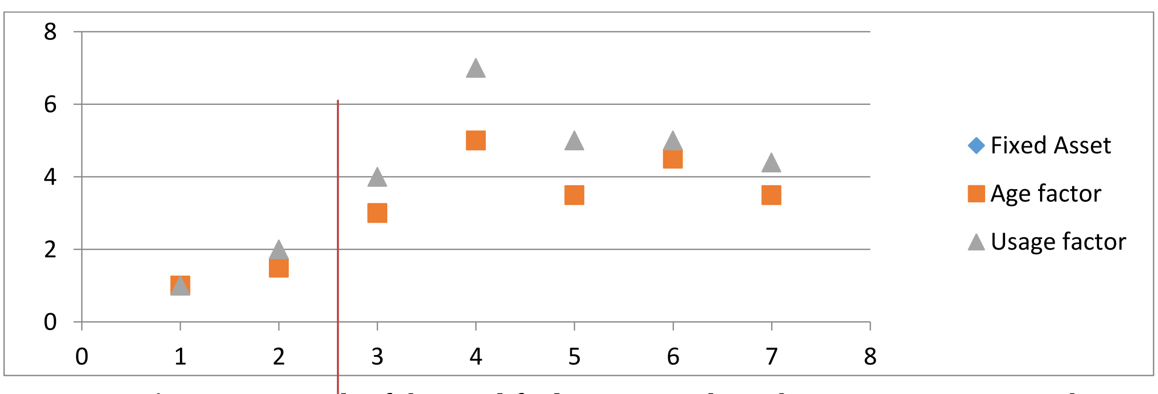

Figure 3: *Result of the Modified K-means algorithm using Scatter Graph*

4. Findings And Conclusion

The modified K-means algorithm with better initial centroids applying the weighted average mean of the data sets eliminates the random selection of the initial centroids and reduced the iteration steps of the algorithm thus, reducing the computational complexity. The separation of the clustering improves because of the selection is based on the computed highest pair and lowest pair of weighted average mean for the initial group of centroids. This is a simple mathematical method for obtaining the initial centroids. It guarantees well separated centroids. The centroids represent the "average" position of all the points of an object.The weighted average represents the real position of each point in the dataset, thus eliminating random selection that leads to unfruitful result. Since iteration is decreased, the complexity of the computation, since it uses Euclidian Distance is also decreased.

If the grouping or clustering of dataset is more than two (2), the succeeding higher average pair will be selected. This study is only limited to eliminate the random selection of initial centroids.

5. Recommendations

The integration of the weighted average mean as the basis of the selection of initial centroids eliminates the random selection. The iteration of using Euclidean distance to compute the distance between the centroids is lessened. It reduces the complexity of the computation and more effective to smaller groups of clustering.

But still, the point of each attribute in the dataset must be properly distributed or scattered to get the good result of applying the weighted average. The use of Exclusive Assignment - If there are two highly overlapping points in the dataset, then modified k-means will not be able to resolve the distinct separation orfinal groupings from the given objects. The future study must be focused on how to resolve this limitation.

6. Acknowledgements

We would like to express our deepest gratitude to all of our professors in Technological Institute of the Philippines (TIP), for their patience in teaching and monitoring us in this Dissertation paper and to our dear families who are fully supporting us from the start up to the end and of course to our God, who is the source of everything that never fail us to reach this very important endeavour. Their guidance and encouragement move our hearts to finish and complete this program and to do things right from the start to the end. To God is the glory.

References

A Tutorial on Clustering Algorithms, Retrieved from Intranet, home.deib.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html

Khedr A. E., Seddawy, A., &Idrees A... (2014) "Performance Tuning of K-Mean Clustering Algorithm a Step towards Efficient DSS" International Journal of Innovative Research in Computer Science & Technology (IJIRCST), ISSN: 2347-5552, Volume 2, Issue 6

Kushwah S. P. S,.Rawat K, Gupta P., (2012) "Analysis and Comparison of Efficient Techniques of Clustering Algorithms in Data Mining" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075,Volume-1, Issue-3

K-means Clustering algorithm- Data Clustering Algorithms (sites.google.com/site/dataclusteingalgorithms/k-means clustering-algorithmk-means)

K-means clustering From Wikipedia, the free encyclopedia en.wikipedia.org/wiki/Kmeans_clustering

Oyelade, O. J, Oladipupo, O. O, Obagbuwa, I. C, (2010) "Application of k-Means Clustering algorithm for prediction of Students' Academic Performance" International Journal of Computer Science and Information Security (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, _o. 1, 2010

Teknomo, K. PhD, Teknomo, K-Means Clustering Tutorial K-Means Clustering Tutorials. (2007)(http:\\people.revoledu.com\kardi\tutorial\kMeanWeighted_average (*https://*simple.m.wikipedia.org*/wiki/)*