International Journal of Learning and Teaching Vol. 3, No. 4, December 2017

Application of Improved Initialization of K-means Algorithm for Monitoring of Fixed Asset

Aleta C. Fabregas and Ruji P. Medina
Technological Institute of the Philippines, Graduate Programs, Quezon City, Philippines
Email: {alet_fabregas, ruji_p_medina}@yahoo.com

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

***Abstract*—A new method of initializing centroids for K-means algorithm is developed in fixed asset monitoring application. The new method utilizing improved utilization of K-means using weighted average was found to be more efficient than the standard K-means owing to its simplicity and having no need for complex computation. The Depreciation was calculated with the use of the new method through the implementation of Age and Usage factors as the major attributes. The clustering output was able to determine the level of maintenance. This paper shows that the developed algorithm is highly applicable to accounting application.**

***Index Terms*—initialization of centroids, fixed asset, depreciation, level of maintenance**

I. Introduction

K-means algorithm is famous, common clustering technique. Its mathematical computation is fast and memory efficient. This algorithm is grouping a set of datasets in such a way that items in the same cluster are more identical to each other than to those in other clusters. K-Means is a Partitioning algorithm which takes as input a positive integer number of groupings K and a data set to divide into K non-empty, non-overlapping and non-subordinated clusters [1].

This study used the improved initialization of centroids for K-means algorithm that introduced the use of easy mathematical computation which is the weighted average for the initialization of centroids. The improvements of initializing centroids will be applied to a simple Fixed Asset Monitoring Application. The fixed assets are the long-term tangible piece of property that a firm owns and uses in the generation of profit [2]. The study focuses on the depreciation of the fixed asset with the Usage factor and Age factor to assess the overall performance of the fixed assets and determine the level of maintenance to be applied [3].

The application of the improved initialization of centroids to the monitoring of fixed asset is simulated by Java program using the Eclipse platform. The study shows the effectiveness of the improved initialization of centroids for K-means as compared to the traditional K-means procedure. The purpose is to make a clear comparison of the standard and the improved initialization of the k-means algorithm in terms of processing time in reaching the convergence point of clustering, simplicity of mathematical computation and the number of iterations.

II. Background

Clustering applications most specially K-means algorithm are popularly used in various fields and this paper will focus on the accounting application of Depreciation analysis for Fixed Asset Monitoring and shows the fruitful result of this clustering algorithm.

The K-means algorithm produces the best result when dataset are well clustered from each other [3]. In this algorithm, the value of K is specified, the number of clusters, prior to initialization of centroids and partitioning. Since the initial seeds of the traditional K-means algorithm are chosen randomly that leading to more looping and computational time, the authors proposed "Enhanced Initial Centroids for K-means Algorithm" to improve the initialization of seeds. The improved k-means eliminates the unspecified selection of the initial cluster, because the separation into different group's result, lies greatly upon the reliability of the initial seeds which are selected randomly [3].

The study shows the implementation of the improvement of the initial centroids and the traditional k-means in the Fixed Asset monitoring application. Fixed assets reduced their value as they age because they are subject to be depreciated. Depreciation is a procedure of spreading the cost of a tangible asset over its useful life. The different business firms and organizations depreciate fixed assets for both tax and accounting purposes [4]. The study focuses on the two (2) depreciating factors called Age and Usage factor. The final result will be used by accounting people to apply what level of maintenance be applied in the fixed asset.

Manuscript received September 2, 2016; revised May 3, 2017.

doi: 10.18178/ijlt.3.4.336-341

III. The Standard K-means and Improved Initialization of Centroids in the Fixed Asset Monitoring Application and Simulation Results

In the standard K-means algorithm, the centroid is the mean location of all the points in the dataset that indicates equal weight while the weighted average reflects the real weight of a point from the given dataset. The weighted average is integrated into the improved initialization of centroids for K-means algorithm. The following is the formula for the weighted mean

$$\text{Weighted Mean} = \sum wx \;/\; \sum w \tag{1}$$

The weighted average of values is the sum of weights times values divided by the sum of the weights [5].

The integration of the weighted average does not change the whole process of the original K-means algorithm. It is still composed of two (2) major steps and the first step is enhanced, which is the initialization stage. With the given number K of clusters, the improved initialization of centroids is to obtain the weighted average of the pair of attributes of the object based on the highest point or perfect score of each attribute. The computed weighted average will be the initial centroid to create an initial partition. The initial partition divides the objects into K clusters with the computed initial centroids. Then the Second step is Partitioning with two-step looping steps for each object. The Assignment step, where the distances of the object from the centroids of each of K clusters are computed using the Euclidean distance. If the object is not currently in the cluster with the closest prototype, then it is reassigned to its nearest cluster [3].

The formula of the Euclidean distance is:

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

where $\left\|x_i^{(j)}-c_j\right\|^2$ a chosen distance between a data point $x_i^{(j)}$ and the cluster cj, is an indicator of the distance of the n data points from their respective cluster centers.

And the Update step: if reassignment occurs, both the clusters are updated and their seeds are recalculated using the current clustering. The Assignment and Update refinement steps are iterated until the stopping point are achieved.

The Enhanced 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 and lowest pair of the weighted mean from the pair of X and Y coordinate improves the K-means algorithm because it clearly separates the clustering of the datasets with fewer iterations [3].

The improvement is applied to a simple fixed asset monitoring. The two (2) elements of depreciation are focused on the useful life or lifespan of the fixed asset which is subject to be declined because of the systematic recording of the depreciation called the Age factor. And the physical condition caused by the depreciation is called the Usage factor. The sample of the fixed assets is buildings, furniture, leasehold improvements, and office equipment. This paper is only focused on the Service Vehicle. The Depreciation factors of Service Vehicle include two factors. First, is the Usage factor (UF) which is defined as the physical condition of the fixed asset.

Second, is the Age factor (AF) which is the remaining economic value of the property relative to its estimated service life of the equipment. It will be determined by computing first the depreciation and reckon the corresponding value of Age factor. The value of Age factor is dependent on the Depreciation [6].

Depreciation = Estimated Service Life – Actual Service/ Estimated Service life (3)

The Depreciation concept applied in the study used the principle of the Manual on Disposal of Government Property in the Philippines. If the fixed asset is not used, the Usage factor (UF) is 0.90, otherwise, it is equal to the value assigned to the condition factor [6]. As the rate of depreciation decreases, age factor (AF) decreases. If the value of AF is decreasing, the value of the UF or the physical condition will normally reduce. But, fixed assets are periodically maintained. And this is the reason why some fixed assets are still in good condition even beyond their life. The two factors are used to assess the performance of the fixed assets and determine the level of maintenance to be applied after the clustering. The purpose of maintenance is to keep equipment and systems running efficiently even beyond its lifespan [7]. The application produces the clustering of the service vehicle in two groups for the application of the maintenance type. This study is simplified to show the application of the improved initialization of K-means algorithm. The application used Java programming language and for easy implementation, the Eclipse an integrated development environment (IDE) is the based platform.

A. Fixed Asset Monitoring with the Traditional K-means Algorithm

1. Input values for usage factor and age factor.
2. Choose one pair of centroid 1 at random from among the data points of usage factor and age factor. Choose another pair of centroid 2 at random from among the data points of usage factor and age factor.
3. Since the first pair of centers are chosen, the next step used the traditional method of the algorithm
    a. Partitioning: Assignment step, where the distances of the object from the centroids of each of K clusters are computed using the Euclidean distance.
    b. Update step: if reassignment occurs, both the clusters (gaining the new object and losing the object) are updated and their seeds are recalculated using the current clustering.
    c. Convergence step: Steps a and b must be repeated until no point changes its cluster assignment or until the centroids no longer move.

Figure 1. The K-means algorithm with age factor and usage factor.

For simplicity, the clustering results used the maintenance type levels 1(heavy maintenance for the poor performance) and 2 (normal maintenance for the efficient performance), to be applied in the service vehicles to keep them running efficiently. The whole numbers are used as input values. In Fig. 1, the standard k-means algorithm applied in Java program used to simulate the input values of the Usage factor (UF) and Age factor (AF) for each service vehicle unit.

Input	Output
Age factor: [1.0, 1.5, 3.0, 5.0, 3.5, 4.5, 3.5] Usage factor: [1.0, 2.0, 4.0, 7.0, 5.0, 5.0, 4.4]	Seed1: [3.0, 4.0]vehicle: 3 Seed2: [4.5, 5.0]vehicle: 6 Distance to C1: [3.6055512, 2.5, 0.0, 3.6055512, 1.118034, 1.8027756, 0.6403125] Distance to C2: [5.315073, 4.2426405, 1.8027756, 2.0615528, 1.0, 0.0, 1.1661904] Clustering 1: [1, 1, 1, 2, 2, 2, 1] Average of cluster 1: [2.25, 2.85] Average of cluster 2: [4.3333335, 5.6666665] Distance to C1: [2.2327113, 1.1335783, 1.372953, 4.9784536, 2.4869661, 3.1120734, 1.991231] Distance to C2: [5.7348833, 4.633813, 2.1343746, 1.490712, 1.0671874, 0.6871841, 1.5162086] Average of cluster 1: [2.25, 2.85] Average of cluster 2: [4.3333335, 5.6666665] Cluster of iteration 2: [1, 1, 1, 2, 2, 2, 2] Another Iteration Distance to C1: [1.5723301, 0.4714045, 2.034426, 5.639642, 3.1446605, 3.7712362, 2.6549745] Distance to C2: [5.3561296, 4.2559514, 1.7573061, 1.8676523, 0.71632737, 0.51295704, 1.1371564] Average of cluster 1: [1.8333334, 2.3333333] Average of cluster 2: [4.125, 5.35] cluster of iteration 3: [1, 1, 2, 2, 2, 2, 2] Another Iteration Distance to C1: [0.559017, 0.559017, 3.0516388, 6.6567636, 4.160829, 4.776243, 3.6704905] Distance to C2: [5.0056367, 3.9046638, 1.4058449, 2.2127812, 0.40792164, 0.6053097, 0.7889232] Average of cluster 1: [1.25, 1.5] Average of cluster 2: [3.9, 5.08] Done **Cluster of iteration 4: [1, 1, 2, 2, 2, 2, 2]** **Iteration Done: 4**

Figure 2. The first output of the traditional algorithm.

The first sample output of the standard K-means applied in java program is shown in Fig. 2. At the left side of the Fig. 2, there are seven (7) service vehicles with the two attributes; the Age factor and the Usage factor entered in the program. After executing the java program using the standard K-means algorithm, the output at the right side is shown. Using the principle of the standard K-means, initial centroids are randomly selected by the user. In the first output, the two seeds chosen by the user are the age factor and usage factor of vehicles 3 [3.0, 4.0] and 6 [4.5, 5.0] respectively. The Distance of each object from the initial seed is computed using the formula of Euclidean Distance for the purpose of clustering. This step is called Partitioning. Then the Update step is shown in the figure by obtaining the new set of average as the new seed. Partitioning and Update steps are repeatedly done until the stopping point or no movements in the clustering is achieved.

Based on the input, the age factor (AF) and usage factor (UF) of the seven service vehicles contain values in the same behavior. The pair of values from the first to the fourth set is in increasing order and the pair from the fifth to seventh set is in decreasing order. From the concept of depreciation, when the pair of values is low, the service vehicle is old and the physical condition is low. After reaching the limit, based on the output, the result of the clustering of the service vehicles using the traditional algorithm of K-means with the two (2) attributes is: The first two Service vehicles are old and their physical conditions are not good, then it needs the type level (1-heavy) of maintenance to keep it running efficiently. And the next five (5) Service vehicles needs type level (2-normal) of maintenance. The number of iterations is four (4) before reaching the convergence point.

Input	**Output**
age factor: [1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0] usage factor: [7.0, 6.0, 5.0, 4.0, 3.0, 2.0, 1.0]	Seed1: [1.0, 7.0]vehicle: 1 Seed2: [2.0, 6.0]vehicle: 2 Distance to C1: [0.0, 1.4142135, 2.828427, 4.2426405, 5.656854, 7.071068, 8.485281] Distance to C2: [1.4142135, 0.0, 1.4142135, 2.828427, 4.2426405, 5.656854, 7.071068] Clustering 1: [1, 2, 2, 2, 2, 2, 2] Average of cluster 1: [1.0, 7.0] Average of cluster 2: [4.5, 3.5] Distance to C1: [0.0, 1.4142135, 2.828427, 4.2426405, 5.656854, 7.071068, 8.485281] Distance to C2: [4.9497476, 3.535534, 2.1213202, 0.70710677, 0.70710677, 2.1213202, 3.535534] Average of cluster 1: [1.0, 7.0] Average of cluster 2: [4.5, 3.5] Another Iteration Distance to C1: [0.70710677, 0.70710677, 2.1213202, 3.535534, 4.9497476, 6.363961, 7.7781744] Distance to C2: [5.656854, 4.2426405, 2.828427, 1.4142135, 0.0, 1.4142135, 2.828427] average of cluster 1: [1.5, 6.5] average of cluster 2: [5.0, 3.0] Another Iteration Distance to C1: [1.4142135, 0.0, 1.4142135, 2.828427, 4.2426405, 5.656854, 7.071068] Distance to C2: [6.363961, 4.9497476, 3.535534, 2.1213202, 0.70710677, 0.70710677, 2.1213202] Average of cluster 1: [2.0, 6.0] Average of cluster 2: [5.5, 2.5] Done **Cluster of iteration 4: [1, 1, 1, 2, 2, 2, 2]** **Iteration Done: 4**

Figure 3. The second output of the K-means algorithm.

The second set of data is shown in Fig. 3. The input values of age factor is increasing order and usage factor is in decreasing order. Using the same procedure, the output is shown in the figure.

The output shows that the first group of three (3) service vehicles needs type level 1(heavy) of maintenance and the next four (4) vehicles needs another type level 2(normal) of maintenance. The number of iterations is 4 for this dataset. The convergence step or the non-movement of the centroids [8] takes long processing time to achieve.

B. Fixed Asset Monitoring with the Improved Initialization of Centroids for K-means Algorithm

There is enhanced algorithm that makes k-means more efficient by removing the first limitation i.e. it limits the number of computations to some extent. The idea makes k-means more efficient, especially for dataset containing large number of clusters [9]. K-mean clustering algorithm performance depends on an initial centroids that why the algorithm doesn't have guarantee for optimal solution [10].

1	INPUT ageFactor, usageFactor;
2	highestAgeFactor← ageFactor[0] ; highestUsageFactor ← usageFactor[0]; cluster ← [0]; meanOfAge ← 0; meanOfUsage ← 0; flag ← 0;
3	FOR i ← 1; IF age Factor[i] > age Factor[0] THEN highestAgeFactor ← age Factor[i] ; INCREMENT i by 1; REPEAT UNTIL i < size of ageFactor ;
4	FOR i ← 1; IF usageFactor[i] > usageFactor[0] THEN highestUsageFactor ← usageFactor[i] ; INCREMENT i by 1; REPEAT UNTIL i < size of usageFactor ;
5	sumX ← 0; FOR i ← 1; sumX = sumX + ageFactor[i]; INCREMENT i by 1; REPEAT UNTIL i < size of ageFactor ;
6	sumY ← 0; FOR i ← 1; sumY = sumY + usageFactor[i]; INCREMENT i by 1; REPEAT UNTIL i < size of usageFactor ;
7	FOR i ←0; meanOfAge[i] ← ageFactor[i] * highestAgeFactor / sumX; meanOfUsage[i] ← usageFactor[i] * highestUsageFactor / sumY; INCREMENT i by 1; REPEAT UNTIL i < size of ageFactor ;
8	lowestMeanAge ← meanOfAge[0]; lowestMeanUsage ← meanOfUsage[0]; centroid1 ← [ageFactor[0], usageFactor[0]] ; …. FOR i ← 1; IF lowestMeanAge[i] > meanOfAge[i] AND lowestMeanUsage > meanOfUsage[i] THEN lowestMeanAge ← meanOfAge[i]; lowestMeanUsage ← meanOfUsage[i]; indexOfHighest ← I; lowestMeanUsage ← meanOfUsage[i]; centroid1←[ageFactor[indexOfLowest], usageFactor[indexOfLowest]]; highestMeanAge ← meanOfAge[0]; highestMeanUsage ← meanOfUsage[0]; centroid2 ← [ageFactor[0], usageFactor[0]]; FOR i ← 1; IF highestMeanAge[i] < meanOfAge[i] AND highestMeanUsage < meanOfUsage[i] THEN highestMeanAge ← meanOfAge[i]; highestMeanUsage ← meanOfUsage[i]; indexOfHighest ← I; centroid2←[ageFactor[indexOfHighest], usageFactor[indexOfHighest]];

Figure 4. The algorithm showing the integration of the weighted mean in K-means.

The Improved Initialization of K-means algorithm limits the number of computations to some extent and the initial centroids is obtained using the simple weighted average. The improvement is applied in fixed asset monitoring using Java program. In the improvement, given the number K of clusters, initialization means is to obtain the weighted average of the pair of attributes of the object based on the highest point or perfect score of each attribute. The computed weighted average will be the initial centroids or seeds to create an initial partition. The initial partition divides the objects into K clusters with the computed initial centroids. It is demonstrated by an algorithm shown in Fig. 4. The detailed part of the enhancement with the weighted average is included in Fig. 4 with the age factor and usage factor. The output of this algorithm is also applied in the two datasets used in the standard K-means algorithm.

The algorithm contains six (8) steps to show the improved initialization of centroids. The first two steps input the Age factor and Usage factor and sets the highest age factor and highest Usage factor to obtain the weighted average of the pair of attributes of the object based on the highest point or perfect score of each attribute. The service vehicles are acquired at the same date so the age factor rate applied as the weight is 5.00, the highest point and for the usage factor is 7.00 as the weight, since this value is the highest point.

The algorithm to compute the weighted average is shown from step 3 to step 7. The 8th step of the algorithm obtains the highest weighted pair of the Age factor and Usage factors the initial centroids to create an initial partition. The rest of the algorithm applied in java program is similar to the traditional K-means algorithm.

Input	Output
Age factor: [1.0, 1.5, 3.0, 5.0, 3.5, 4.5, 3.5] Usage factor: [1.0, 2.0, 4.0, 7.0, 5.0, 5.0, 4.4]	**Weighted mean of age factor**: [0.22727273, 0.3409091, 0.6818182, 1.1363636, 0.79545456, 1.0227273, 0.79545456] **Weighted mean of usage factor:** [0.24647887, 0.49295774, 0.9859155, 1.7253522, 1.2323943, 1.2323943, 1.0845071] [1.079306, 0.8315604] [1.079306, 0.3184934] [1.079306, 0.7380588] [1.079306, 0.30958998] [1.079306, 0.23350301] [1.079306, 0.2213149] 4: [0.7380588, 0.30958998] 5: [0.30958998, 0.23350301] 6: [0.23350301, 0.2213149] seed1: [1.0, 1.0]vehicle: 1 seed2: [3.5, 4.4]vehicle: 7 distance to C1: [0.0, 1.118034, 3.6055512, 7.2111025, 4.7169905, 5.315073, 4.2201896] distance to C2: [4.2201896, 3.1241, 0.6403125, 3.001666, 0.5999999, 1.1661904, 0.0] clustering 1: [1, 1, 2, 2, 2, 2, 2] average of cluster 1: [1.25, 1.5] average of cluster 2: [3.9, 5.08] Done cluster of iteration 2: [1, 1, 2, 2, 2, 2, 2] Iteration Done: 2

Figure 5. The first output of the improved initialization of centroids of the K-means algorithm.

The result of the improved K-means algorithm with the same datasets of the traditional K-means is shown in Fig. 5.

The initial centroids are obtained by applying first the computation of the weighted mean. Based on the first output, the weighted mean of the age factor and usage factor of each service vehicle are computed and compared to get the lowest weighted pair and the highest weighted pair. The result of the computation and comparison of the weighted mean is that; the seed 1 is the vehicle 1 with [1.0, 1.0] and seed 2 is the vehicle 4 with [5.0, 7.0]. Applying the algorithm of the improved initialization of centroids resulted to the application of level 1 of maintenance in the first two (2) Service Vehicles and the level 2 in the next five (5) vehicles. The number of iterations for the improved K-means algorithm is reduced to two as compared with the standard K-means with four iterations. The convergence step is easier to achieve in the improved algorithm.

Input	Output
Age factor: [1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0] Usage factor: [7.0, 6.0, 5.0, 4.0, 3.0, 2.0, 1.0]	**Weighted mean of age factor**: [0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75] **Weighted mean of usage factor**: [1.75, 1.5, 1.25, 1.0, 0.75, 0.5, 0.25] [1.0606601, 0.70710677] [1.0606601, 0.35355338] [1.0606601, 0.0] [1.0606601, 0.35355338] [1.0606601, 0.70710677] [1.0606601, 1.0606601] 1: [1.0606601, 0.70710677] 2: [0.70710677, 0.35355338] 3: [0.35355338, 0.0] 4: [0.0, 0.35355338] 5: [0.0, 0.70710677] 6: [0.0, 1.0606601] seed1: [1.0, 7.0]vehicle: 1 seed2: [4.0, 4.0]vehicle: 4 distance to C1: [0.0, 1.4142135, 2.828427, 4.2426405, 5.656854, 7.071068, 8.485281] distance to C2: [4.2426405, 2.828427, 1.4142135, 0.0, 1.4142135, 2.828427, 4.2426405] clustering 1: [1, 1, 2, 2, 2, 2, 2] average of cluster 1: [1.5, 6.5] average of cluster 2: [5.0, 3.0] cluster of iteration 2: [1, 1, 1, 2, 2, 2, 2] Another Iteration Done cluster of iteration 3: [1, 1, 1, 2, 2, 2, 2] Iteration Done: 3

Figure 6. The Second output enhanced initial centroids of the K-means algorithm.

The second dataset of the standard K-means is also applied in the improved initialization of the algorithm. Again the input values of the age factor is in increasing order and the usage factor is in decreasing order. The result of the second dataset is shown in Fig. 6.

The clustering result of the second dataset is similar to the previous result of standard K-means algorithm. The first three (2) Service Vehicles needs the level 1 maintenance and the next five (5) vehicles needs the level 2. The number of iterations is reduced into 3 times as compared with the previous result of 4 times.

The processing time of the improved initialization of centroids with the weighted average using the two different datasets is faster as compared with the standard k-means, and the number of iterations reduced because the weighted average showed the real weight of the point. The clustering of the dataset is well separated and convergence is achieved faster.

The Table I shows the comparison of K-means and the improved initialization of K-means in terms of the number of repetitions and time in milliseconds with the two datasets. The two datasets have different behaviors of values for age and usage factors. But the experimental results prove that the improved initialization of K-means with the weighted average is faster and more efficient compared to the standard k-means.

The first pair of seeds for the standard k-means randomly selected the 3rd service vehicle and the 6th service vehicle as initial centroids for the first dataset and the 1st and the 2nd service vehicles are the second initial centroids for the second dataset. With the improved k-means, based on the lowest and highest weighted average, the 1stservice vehicle and the 4th service vehicle are the initial pairs for the first and second datasets. The table compares the result of using the standard K-means and the improved K-means. The application proves that the program with the improved algorithm executes faster in terms of processing time in milliseconds and the number of iterations is reduced in determining the final groupings of the service vehicles.

TABLE I. COMPARISON OF K-MEANS AND ENHANCED INITIAL CENTROIDS

	K Means			Enhanced		
Trials	Process time (milliseconds)	No. of iterations	Random choice	Process time (milliseconds)	No of iterations	Based on Weighted average
Dataset 1	3	4	3,6	2	2	1, 7
Dataset 2	3	3	1,2	2	3	1, 4
Ave	4.5			2.0		

C. Simulation Results

The same datasets used in the traditional and the improved initialization of centroids for k-means resulted to an output of similar groupings but the processing time of the program decreases because the convergence step is achieved faster. The complex computational time of using the Euclidean distance is lessened in the improved initial centroids of K-means. In the standard k-means, the result of the processing using java program has four (4) iterations for the first dataset and three (4) iterations for the second dataset. The improved initial centroids using the weighted average resulted to two (2) iterations for the first dataset and three (3) iterations for the second dataset. Also, the study proves that the improved K-means algorithm is faster and efficient in determining the level of maintenance to be applied for Fixed Asset Monitoring, thus the algorithm could be also applied in accounting application.

IV. Findings and Conclusion

The improved initial centroids for K-means algorithm applying the weighted average mean for the datasets reduced the iteration steps of using the Euclidean distance algorithm thus, reducing the computational complexity and processing time of the algorithm. The separation of the clustering improves because of the initial centroids which are obtained based on the computed highest pair and the lowest pair of weighted average mean. The K-Means partitioning based clustering algorithm required to define the number of final cluster (k) beforehand. In this study, the classification is limited only to two, level 1 (heavy) and level 2 (normal) type of maintenance. The application of K-means in the Fixed Asset Monitoring, using Depreciation factors which are the age and usage proves, that the K-means algorithm is also effective in accounting procedure.

V. Recommendations

The fixed asset monitoring with the improved initial centroids of K-means algorithm is a good benchmark to apply in determining the performance level of the fixed asset and the type of maintenance to be used in the clustering of assets. This study is limited only to the two depreciation factors, the age factor and usage factor and the future study must also include the other factors of monitoring the fixed asset to make the application in accounting more effective. The point of each attribute for the object in the dataset must be properly distributed to get the good result of applying the weighted average. Highly overlapping points of the attributes for the object in the dataset is a problem in K-means, because the final clustering or convergence step will not be solved even by the improved initial centroids. The distinct separation or final clustering of the given objects is always the goal of K-means algorithm. The future study must be focused on how to resolve this limitation.

References

[1] K-means Clustering RTMath, FinMath User Guide. [Online]. Available: https://rtmath.net/help/html/dbe6855e-acc5-4421-bc80-79c0a71e97cd.htm

[2] FixedAsset: Definition-Investopedia. [Online]. Available: http://www.investopedia.com/terms

[3] A. Fabregas, B. Gerardo, and B. Tanguilig III, "Enhanced initial centroids for K-means algorithm," Doctor of Information Technology Dissertation, Technological Institute of the Philippines, Quezon City, Philippines, 2016, unpublished.

[4] Depreciation: Definition-Investopedia. [Online]. Available: http://www.investopedia.com

[5] Weighted average. [Online]. Available: https://www.mathsisfun.com

[6] Manual on Disposal of Government Property- National Budget Circular no.425 January 28, 2002 Retrieved: February 15, 2016.

[7] Types of Maintenance Chapter 5O&M Best Practices Guide, Release 3.0. [Online]. Available: http:www1.eere.energy.gov

[8] S. Ghosh and S. K. Dubey, "Comparative analysis if K-means and fuzzy-C mean algorithm," *International Journal of Advanced Computer Science and Applications*, vol. 4, no. 4, pp. 35-39, 2013.

[9] N. Aggarwal, K. Aggarwal, and K. Gupta, "Comparative analysis of K-means and enhanced K-means clustering algorithm for data mining," *International Journal of Scientific & Engineering Research*, vol. 3, no. 3, August 2012.

[10] B. Bangoria, N. Mankad, and V. Pambhar, "A survey on efficient enhanced K-means clustering algorithm," *International Journal for Scientific Research & Development*, vol. 1, no. 9, 2013.

Aleta C. Fabregas finished her Bachelor of Computer Data Processing Management (BCDPM) with honors (cum laude) in Polytechnic University of the Philippines (PUP), Manila in 1986. She obtained Master in Information Technology (MIT) from PUP Manila in 2006. She is currently working at PUP, with the rank of Associate Professor in the College of Computer and Information Sciences (CCIS), Associate Lecturer in Master of Science in Information Technology (MSIT), Graduate School and Course Facilitator at the post-baccalaureate program, master program and Expanded Tertiary Education Equivalency and Accreditation Program (ETEEAP) at the PUP Open University System. She is a module writer in Java and other programming languages.

Prof. Fabregas is a senior Member of Royal Institute of Information Technology, Singapore, member of the Philippine Schools, Universities and Colleges Computer Education and Systems Society, Inc. (PSUCCESS), and Philippine Computer Society (PCS).

Bobby D. Gerardo is currently the Vice President for Administration and Finance of West Visayas State University, Iloilo City, Philippines. He holds an academic rank of Professor VI, and he is a known innovator in the field of ICT and catalyst to prospective scientists in the country. He has published more than 70 research papers in national and international journals and conferences and received numerous international award for his research publications. He is a referee to international conferences and journal publications such as in IEEE Transactions on Pattern Analysis and Machine Intelligence and IEEE Transactions on Knowledge and Data Engineering. His research fields: / are in distributed systems, telematics systems, data mining, ubiquitous computing and mobile communications.

Ruji P. Medina is Dean of the Graduate Programs and concurrent Chair of the Environmental and Sanitary Engineering Program of the Technological Institute of the Philippines in Quezon City. He holds a Ph.D. in Environmental Engineering from the University of the Philippines with sandwich program at the University of Houston, Texas where he worked on the synthesis of nanocomposite materials. He counts among his expertise environmental modeling and mathematical modeling using multivariate analysis.