Locating Optimal Coordinates of Police Deployment and Outpost in Santa Maria Bulacan Using Centroid 360: An Enhanced Centroid Initialization for K means Algorithm

Jovy Jay D. Cabrera, Ariel M. Sison, Ruji P. Medina

Abstract: The number of police in the Philippines is way below than the declared number of police per population based on the Philippines Republic Act No. 6975, Chapter 3, Section 27 which states that one (1) policeman per five hundred (500) persons. Santa Maria Bulacan only have fifty-two (52) police personnel, and one (1) police outpost which is located in the town of Poblacion which is in the far south east of the whole municipality. This produces different speed when responding to incident reports. The police force’s response time vary based from the location of the incident. This research determined the optimum location to where police will be deployed as a mobile unit to be able to respond faster to more areas than staying in the outpost. Another thing that will be determined by this research is the optimum coordinates of a perfect scenario where each town in the municipality of Santa Maria Bulacan would have their own police outpost. The coordinates will be calculated based on the populations’ location by plotting all the houses and structures located in Santa Maria Bulacan based on Google Map Images, and the optimum location will be determined in the form of the converged centroids after applying enhanced K means clustering algorithm.

Index Terms: Clustering, Enhanced K means, K means, Machine Learning.

I. INTRODUCTION

Clustering is characterized as the gathering of a specific arrangement of items dependent on their qualities, gathering the information as indicated by their likenesses. Various techniques can be utilized to segment the information and they are called clustering algorithms [16]. There are various types of clustering algorithm which can be utilized to group different kinds of information. The most famous types of clustering are the Hierarchical clustering, Partition clustering, Exclusive clustering, Overlapping clustering, Fuzzy clustering, and complete clustering [6]. K means algorithm is one of the most popular partition-based clustering algorithms. The reason behind, is the straightforwardness and flexibility of the calculation. The K means algorithm basically has two stages which incorporate initialization where given a K number of groups, a K number of random centroids would be initialized and iteratively updated where data points are assigned to a cluster and a calculation is done to update the new value of centroid. The calculation depends on the Euclidean distance of data points and the initialized centroids. Despite the fact that the calculation is exceptionally prominent, it really experiences various issues and one of them is the random initialization of centroids [14]. Initialization of centroid is one of the key factors in the successes of clustering in K means algorithm. The random initialization step works. But since it’s random, irregularity with the outcome comes hand in hand. The random initial centroid may work best, and it might likewise work. However, the arrangement of the underlying centroid might be a long way from the ideal position, and it might not work at all [3]. Clustering has many applications where intrusion detection is one of them. Intrusion detection can be done by clustering data, after which outliers can be identified as abnormal data as one of the characteristics of an intrusion. Better clustering algorithm can provide an improved detection rate of intrusion attacks [1]. Data mining algorithms such as K-means can also be used for a simple student performance monitoring as well as performance prediction [13]. These are all attainable by using the clustered data. Another application of clustering use is by converged centroid instead of using the clustered data to pinpoint something important. A clustering application of this type was done in research locating multi energy systems for the neighborhood of Geneva in which the converged centroids will serve as the position of distributed energy systems [9]. Santa Maria Bulacan is a province located thirty-two (32) kilometers north of Manila. Santa Maria has a land area of about 9,092.5 hectares [12]. The land area is subdivided into twenty-four (24) Barangays namely Barangay Bagbaguin, Balasing, Buenavista, Bulac, Camangyanan, Catmon, Cay Pombo, Caysio, Guyong, Lalakhan, Mag-asawang Sapa, Mahabang Parang, Manggahan, Parada, Poblacion, Pulong Buhangin, San Gabriel, San Jose Patag, San Vicente, Santa Clara, Santa Cruz, Silangan, Tabing Bakod and Tumana [11]. In 2015, the province had a population count of two hundred fifty six thousand four hundred fifty four (256,454) [2].

Revised Manuscript Received on July 12, 2019.
Jovy Jay D. Cabrera, Department of Information Technology at the Technological Institute of the Philippines in Quezon City, Philippines.
Ariel M. Sison, Dean, Department of Information Technology, Emilio Aguinaldo College, Metro Manila, Philippines.
Ruji P. Medina, Dean, Department of the Technological Institute of the Philippines in Quezon City, Philippines.
The province of Santa Maria Bulacan only have one police station, and it is situated at the barangay of Poblacion located far south east of the municipality [7]. Santa Maria is generally a peaceful municipality which is served by fifty-two (52) able bodied police personnel [10]. According to Philippine Republic Act No. 6975, Chapter 3, Section 27 Stipulates that there should be one (1) policeman per five hundred (500) persons [15]. This shows that the municipality is way below the standards by having a ratio of one (1) police officer to 4,931 residents which is almost ten (10) times more than the population being served than the required standard number of police as stated in the republic act. Even though the municipality of Santa Maria is generally peaceful, incidents where policemen had to respond to cases still occur. In this event, responding from the far south east of the map provides irregularity in time when responding to reports.

The main objective of this research is to locate the optimal coordinates where the police force should be deployed as their standby position in which they can be more accessible by most population. The optimal coordinates can also be used for future police station or community precinct. Theoretically, this optimal position is also balanced in terms of distance to the community. Policemen response time to different locations can be time efficient and balanced. The research will display optimal coordinates for having K number of coordinates based on 2, 8, 12, and 24. The values given represent having a new police outpost (K=2), one third of the number of barangays (K=8), half of the value of barangays (K=12), and 24 as the best case of having one coordinates for each barangays (K=24).

II. METHODOLOGY

A. Conceptual Framework

![Figure 1: IPO Diagram](image)

The concept of the study is to use the coordinates of infrastructures in Santa Maria Bulacan. The said coordinates will be clustered using the enhanced K means which use Centroid 360 method for the initialization. The converged centroids will serve as the optimal coordinates.

B. Centroid 360

The Centroid 360 method is based on the concept of separating the initial centroids to ensure separation of clusters and reduce iterations when clustering [14]. The algorithm consists of four steps to perform initialization of centroids:

Step 1. Compute for the center – the first step is to calculate for the average value of all data points to get the center all data points.

\[
\text{AreaX} = \frac{\sum (x)}{N} 
\]

Step 2. Calculate the size of circle and radius – the next step is to compute the size of the circle based on the absolute value of all data and the radius size will be half of the circle size to guarantee that the centroids position will be in the inside of the area of the data points.

\[
\text{CircleSize} = \frac{\sum (\sum 1)}{n \times 2} 
\]

\[
r = \text{CircleSize} \times .5 
\]

Step 3. Identify the angle position for the first centroid – this step involves calculation of the angle of data by using the slope formula which will identify the angle direction of the data which will also be used to the position of the first centroid.

\[
\text{Slope} = \frac{(x - \bar{x})(y - \bar{y})}{\sum(x - 3)^2} 
\]

\[
\text{Degree} = \frac{\tan^{-1}(\text{Slope}) \times 180}{\pi} 
\]

Step 4. Distribution of Centroids – This step engages in assigning the initial centroid position based on the radius and angle of the slope.

\[
\text{InitialCentroidX} = \text{radius} \times \cos\left(\frac{\text{Degree}}{180 \times \pi}\right) + \text{AveX} 
\]

\[
\text{InitialCentroidY} = \text{radius} \times \sin\left(\frac{\text{Degree}}{180 \times \pi}\right) + \text{AveY} 
\]

Step 5. Repeat step number 4 – This covers assigning the rest of initial centroids while increasing the angle value by adding the degree value for K number of centroids. The amount of angle to be added is based 360 which is the full angle of a circle divided by the number of K clusters needed.

\[
\text{Degree} = \sum_{k=1}^{n} \left(\text{Degree} + \frac{360}{K}\right) 
\]

C. Enhanced K means vs Classic K means

The enhanced K means and the classic K means algorithm both use similar steps of initialization, partitioning, and iterative number of steps to update the centroid position. The classic K means algorithm suffers from the inconsistency of the initialization method which indicates the difference [4]. In the previous research of the authors about the Centroid 360 initialization enhancement for the K means algorithm, the results show consistency, 61.21% efficiency increase with less iterations needed to reach convergence making the clustering 45.57% faster, and 14.66% more accurate [5].
The clusters are produced using the enhanced K means algorithm. The clusters are formed as the centroids reached convergence which means that it is the mean point of the cluster. This indicates that it is the best coordinate within the cluster where the centroid is near in all the data points. The converged centroid will be the optimal location for the deployment of police and the ideal location for the future police outpost or community precinct.

Figure 5: Optimal Coordinates for 8 locations (K=8)

The results shown in fig. 5 and fig. 6 shows 8 and 12 clusters. These results are produced with the idea of having multiple police stations in mind.

Figure 6: Optimal Coordinates based on half of number of barangays (K=12)

The results in fig. 5 and fig. 6 shows 8 and 12 clusters. These results are produced with the idea of having multiple police stations in mind.

Figure 7: Optimal Coordinates based on number of barangays (K=24)

The results shown in Fig. 7 shows the optimal coordinates based on the number of barangays. The output produced is the exact number of barangays but does not exactly produce a converged centroid per baranggay.
The optimal coordinates were identified; however, these coordinates are only factored based on the factor of infrastructures where people are staying. Other factors such as number of population per infrastructure were not put in mind when locating the optimal coordinates. These factors considered may not be complete, but it is enough as the optimal coordinates allow shorter distance which will result to quicker access to different datapoints where police presence might be needed.

IV. CONCLUSION

This research presents the optimal coordinates of which the police should be stationed and where to position the police outpost or community precinct in Santa Maria Bulacan. The enhanced K means clustering with Centroid 360 Initialization was used to cluster coordinates of houses and infrastructures and then use the converged centroid to identify the optimal location of police outpost. The results on using K=2 would be the optimized solution as it would be the easiest to achieve by simply adding one (1) community precinct. This means it can simply be implemented by splitting the police force or recruiting fewer police officers or simply position a police mobile as standby position. On the other hand, the ideal case would be the cluster of K=24 as it shows in the results that the coverage per station is smallest meaning shorter distance to respond faster. This is also the hardest to achieve as it would require more police officers in order to implement. This study can be further enhanced by identifying the optimal number of outposts by calculating the acceptable distance between the longest distances to response from a given outpost.

ACKNOWLEDGEMENT

The researchers would like to acknowledge the support of Technological Institute of the Philippines Graduate programs department in supporting this research.

REFERENCES