RESEARCH ARTICLE OPEN ACCESS

Predicting of Crime Detection Using K-Means Clustering Algorithm

Than Win [1], Ei Ei Phyo [2]
Department of Information Technology
Technological University (Mawlamyine)
Myanmar

ABSTRACT
Nowadays, the increasing use of computerized systems, crime data analysis can help the Law enforcement officers to speed up the process of solving crimes. Administration and maintenance in police station can be made more efficient and effective with the use of computers that are enormous help in many jobs. Crime Detection is analysed using K-means Clustering algorithm, our approach supports the polices, the detectives and the law enforcement officers to speed up the process of crimes detection and criminals identification. Crime analysis is one of the important applications of data mining. K-means Clustering algorithm is done by partitioning data into the process of identification of crime patterns. Data mining is the appropriate field to apply on high volume crime dataset and knowledge gained. This easy to implement data mining framework works with the geospatial plot of crime and helps to improve the productivity of the detectives and other law enforcement officers.
Keywords :– *k-means clustering, initial centroid, Euclidean distance, data mining.*

I. INTRODUCTION

Criminals are nuisance for the society in all corners of world for a long time now and measures are required to eradicate crimes from our world. Current policing strategies work towards detecting the criminals, basically after the crime has occurred. But, with the help of technological advancement, it can use historic crime data to recognize crime patterns and use these patterns to detect crimes beforehand [1]. Convert crime information into a data-mining problem such that it can help the detectives in solving crimes faster. These crime reports have the following kinds of information categories namely – date, type of crime, quantity, gender and location etc. The increasing use of the computerized systems to track crimes, computer data analysts have started helping the law enforcement officers and detectives to speed up the process of solving crimes [2].

The clustering algorithms in data mining [3] are equivalent to the task of identifying groups of records that are similar between themselves but different from the rest of the data. In our case some of these clusters will useful for identifying a crime type committed by one or same group of suspects. Clustering techniques converts dataset to clusters which are further examined for determining crime types. These clusters visually represent group of crimes overlaid on map of police jurisdiction. Clusters store location of crimes along with other credentials of crime like type and date. Detection measures are implemented according to crime type in crime areas.

To use clustering technique over any supervised technique such as classification, since crimes vary in nature widely and crime database often contains several unsolved crimes. Therefore, classification technique that will rely on the existing and known solved crimes, will not give good predictive quality for future crimes. The large volumes of crime datasets and the complexity of relationships between these kinds of data have made criminology an appropriate field for applying data mining techniques [4]. Criminology is an area that focuses the scientific study of crime and criminal behaviour and law enforcement and is a process that aims to identify crime characteristics. It is one of the most important fields where the application of data mining techniques [10] can produce important results. Identifying crime characteristics is the first step for developing further analysis. The knowledge gained from data mining approaches is a very useful tool which can help and support police forces.

II. AIMS AND OBJECTIVES

The main aim and objectives are:

- To produce accurate record of crimes.
- To help identification of crime patterns.
- To speed up the process of solving crime detection.
- To improve the productivity of the detectives and other law enforcement offices.
- To increase the efficiency and replace the traditional paper-based crime report with computerized system crime report.
- To reduce the time for searching one or more types of crime records and manage records efficiently and effectively.
- To reduce the crime and solve the crimes with less time.
- To formulate crime pattern detection as machine learning task and to use data mining to support police detectives in solving crimes.
- To display the result graphically.

III. BACKGROUND THEORY

A. Algorithm of K-Means

K-means clustering is one of the methods of cluster analysis. K-means clustering technique [5] is simple, and it begins with a description of the algorithm. First, choose K initial centroids, where K is a user specified parameter, namely, the number of clusters desired. Each point is assigned to the closet centroid, and each collection of points assigned to a centroid is a cluster. The centroid of each cluster is then update based on the points assigned to the cluster. Repeat the assignment and update steps until no point changes clusters, or equivalently, until the centroids remain the same.

K-means clustering is one of the methods of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. The K-means clustering algorithm is popular because it can be applied to relatively large sets of data. The user specifies the number of clusters to be found. The algorithm then separates the data into spherical clusters by finding a set of cluster centers, assigning each observation to a cluster, determining new cluster centers, and repeating this process.

Process:

- Decide on a value for *K*, the number of clusters.
- Initialize the *K* cluster centers (randomly, if necessary).
- Decide the class memberships of the *N* objects by assigning them to the nearest cluster center.
- Re-estimate the *K* cluster centers, by assuming the memberships found above are correct.
- Repeat and until none of the *N* objects changed membership in the last iteration.

K means algorithm complexity where n is instances, c is clusters, and t is iterations and relatively efficient. It often terminates at a local optimum. Its disadvantage is applicable only when mean is defined and need to specify c, the number of clusters, in advance [6].

B. The Euclidean Distance

The Euclidean distance is the ordinary straight line. It is the distance between two points in Euclidean space. With this distance, Euclidean space becomes a metric space. The Euclidean distance measure corresponds to the shortest geometric distance between two points. The Euclidean distance is often used for data points in Euclidean space. The Euclidean distance or simply distance examines the root of square differences between coordinates of a pair of objects [7].

$$\text{Distance} \quad d = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{3.1}$$

$$\text{Distance} \quad d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots}$$

IV. METHODOLOGY

A. Design of Crime Detection System

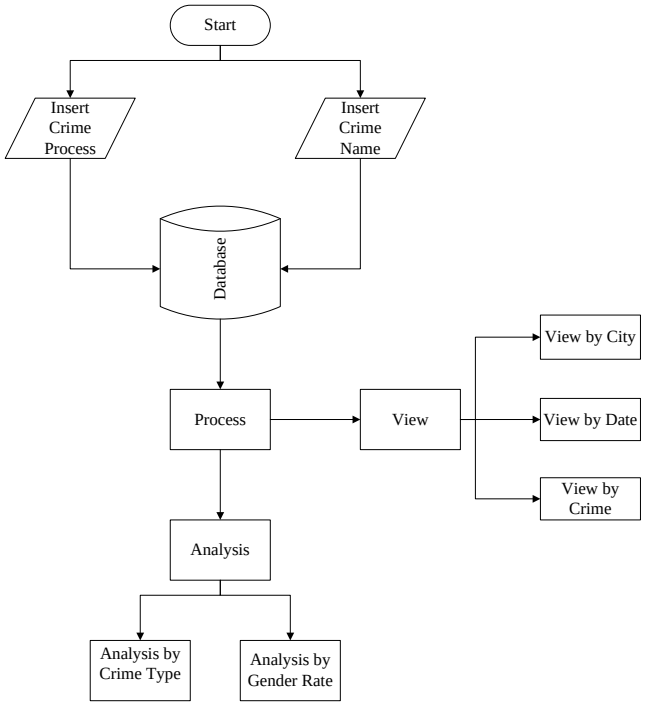

Fig. 1 Design of Crime Detection System

B. Crime Analysis

Crime analysis [8] is defined as analytical processes which provide relevant information relative to crime patterns and trend correlations to assist personal in planning the deployment of resources for the detection and suppression of criminal activities. It is important to analyse crime due to following reasons:

- Analyse crime to inform law enforcers about general and specific crime trends in timely manner.
- Analyse crime to take advantage of the plenty of information existing in justice system and public domain.

Crime rates are rapidly changing and improved analysis finds hidden patterns of crime, without any explicit prior knowledge of these patterns.

The main objectives of crime analysis include:

- Extraction of crime patterns by analysis of available crime and criminal data.
- Detection of crime based on spatial distribution of existing data and anticipation of crime rate using different data mining techniques [9].
- Crime detection.

C. Flowchart of Crime Detection

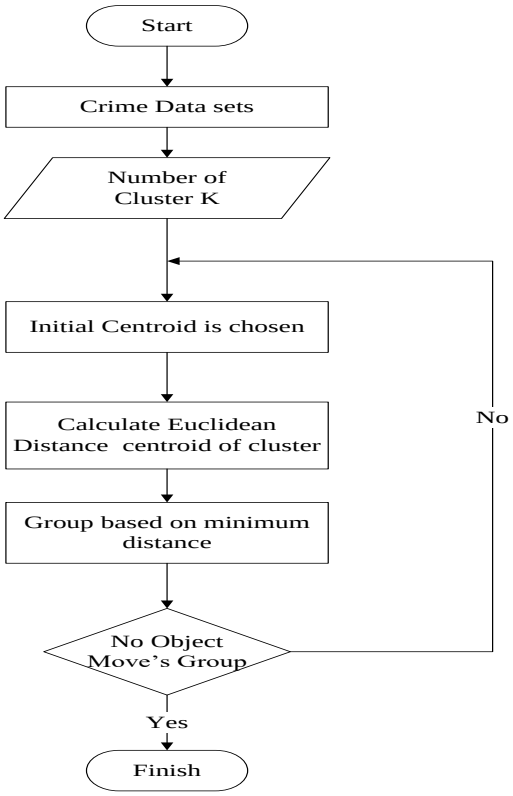

Fig. 2 Flowchart of Crime Detection

V. RESULTS OF K-MEANS CLUSTERING PROPOSED SYSTEM

Table 1. K-Means Clustering Proposed System Assigned City

City	Variable
Ye	1
Thanphyuzayat	2
Mudon	3
Chaungzon	4
Kyaimayaw	5
Mawlamyine	6
Paung	7
Thaton	8
Belin	9
Kyaitho	10

Table 2. K-Means Clustering Proposed System Assigned Crime Type

Crime Type	Variable
Theft	1
Robbery	2
Drugs	3
Vehicle	4
Rape	5
Murder	6

Table 3. K-Means Clustering Proposed System Assigned Crime Datasets

City	Crime Type	Quantity
1	1	4
2	2	5
3	3	6
4	4	5
5	5	1
6	6	2
7	3	3
8	1	3
9	4	3
10	5	3

Initially, two centroids are assigned randomly. No: of cluster k=2. The two centroids k1 and k2, so that k1 (1, 4) and k2 (2, 5). The Euclidean Distance is used to find out which centroid is close to each data point and the data points are assigned to the corresponding centroids.

Table 4. Clustering Result with First Iteration of Crime Datasets

City	Point	Dist Mean1 (1,4)	Dist Mean2(2,5)	Cluster
1	(1,4)	0	2	1
2	(2,5)	2	0	2
3	(3,6)	4	2	2
4	(4,5)	4	2	2
5	(5,1)	7	7	1
6	(6,2)	7	7	1
7	(3,3)	3	3	1
8	(1,3)	1	3	1
9	(4,3)	4	4	1

10	(5,3)	5	5	1

Cluster 1- (1, 4), (5,1), (6,2), (3,3), (1,3), (4,3), (5,3)
Cluster 2- (2, 5), (3,6), (4,5)

All this point, it assign each object to the cluster it is closer to (that is taking the minimum between the two computed distance for each object).

The two different clusters is blue and red color. The yellow color indicates the position of the respective centroids.

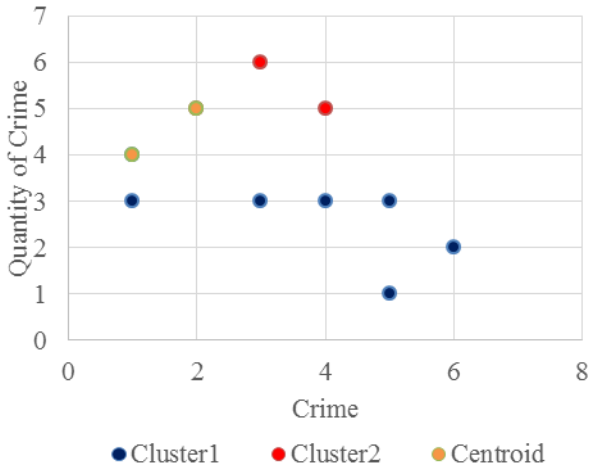

Fig. 3 Iteration 1

Let's iterate, which means to redefine the centroids by calculating the means of the members of each of the two clusters. It again recalculates the centroids.

Table 5. Clustering Result with Second Iteration of Crime Datasets

City	Point	Dist Mean 1 (3.6,2.7)	Dist Mean 2 (3,5.3)	Cluster
1	(1,4)	3.9	3.3	2
2	(2,5)	3.9	1.3	2
3	(3,6)	3.9	0.7	2
4	(4,5)	2.7	1.3	2
5	(5,1)	3.1	6.3	1
6	(6,2)	3.1	6.3	1
7	(3,3)	0.9	2.3	1
8	(1,3)	2.9	4.3	1
9	(4,3)	0.7	3.3	1
10	(5,3)	1.5	4.3	1

Cluster 1 include: (5,1), (6,2), (3,3), (1,3), (4,3), (5,3)
Cluster 2 include: (1,4), (2,5), (3,6), (4,5)

The two different clusters labelled with two different color blue and red. Cluster 1 is blue color and cluster 2 is red color. The position of the centroids are change given by yellow color are shown in Figure 4.

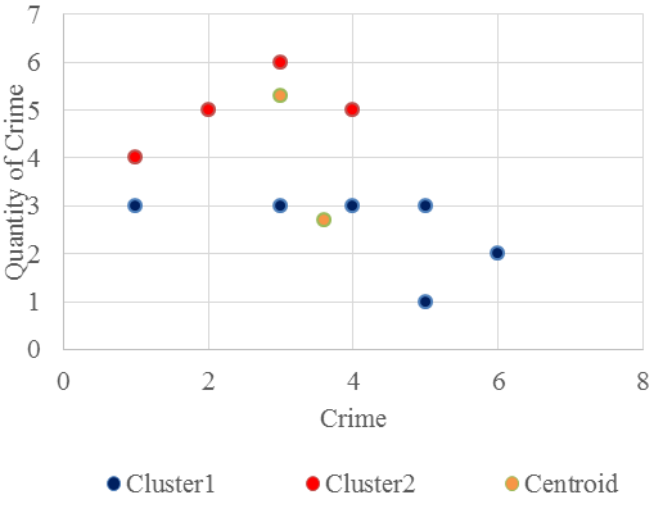

Fig. 4 Iteration 2

Let's iterate, which means to redefine the centroids by calculating the means of the members of each of the two clusters. It again recalculates the centroids. Assign each point to closet centroids. The process is iteratively repeated until the centroids become static.

Table 6. Clustering Result with Third Iteration of Crime Datasets

City	Point	Dist Mean 1 (4,2.5)	Dist Mean 2 (2.5,5)	Cluster
1	(1,4)	4.5	2.5	2
2	(2,5)	4.5	1.5	2
3	(3,6)	4.5	1.5	2
4	(4,5)	2.5	1.5	2
5	(5,1)	2.5	6.5	1
6	(6,2)	2.5	6.5	1
7	(3,3)	1.5	2.5	1
8	(1,3)	3.5	3.5	1
9	(4,3)	0.5	3.5	1
10	(5,3)	1.5	4.5	1

Cluster 1 include: (5,1), (6,2), (3,3), (1,3), (4,3), (5,3)
Cluster 2 include: (1,4), (2,5), (3,6), (4,5)

The process is iteratively repeated until our centroids become static. Algorithm has converged recalculating distances, reassigning cases until clusters results in no change. This is the final solution. The two different clusters labelled

with two different color blue and red. Cluster 1 is blue color and cluster 2 is red color. The position of the centroids change given by yellow color. The final iteration result is given in the Figure 5.

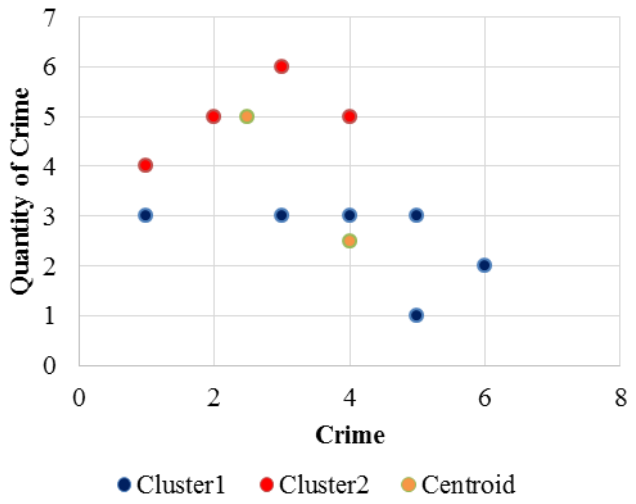

Fig. 5 Iteration 3 (Final iteration)

VI. CONCLUSIONS

Computers are important notes in business, government, military and police. This paper focuses on crime detection by implementing clustering algorithm on crime data set. The use of data mining for predicting of crime detection patterns using the K-means clustering techniques. This paper was to formulate crime detection as machine learning task and data mining to support police detectives in solving crimes. The modelling technique was able to identify the crime patterns from a large number of crimes making the job for crime detectives easier. The proposed method has promising value in the current complex crime and can be used as an effective tool by Myanmar police and enforcement of law organizations for crime detection.

REFERENCES

[1] Vineet Jain, Yogest Sharma, Augush Bhatia, Vaibhav Arora. "Crime Prediction using K-means Algorithm". Global Research and Development journal for engineering Volume2, issue5, April, 2017.

[2] Shyam Varan Nath, Oracle Corporation, Shyam. Nath, @Oracle.com "Crime Pattern Detection Using Data Mining".

[3] JERZY Ste FANOUSKI Institute of computing. Sciences Poznon University at technology, Polan "Data Mining- Clustering".

[4] K.S.Arthisree, M.E, A.Jaganraj, M.E, CSE Department. "Identify Crime Detection Using Data Mining Techniques".

[5] Zijun Zhang, "K-means Algorithm and Cluster Analysis in Data Mining."

[6] "Clustering". 15-381 Artificial Intelligence Henry Lin, Modified From Excellent slide of Eamonn Keogh, Ziv Bar-Josept, and Andrew Moore.

[7] "Cluster Analysis : Basic Concepts and Algorithms".

[8] Jyoti Agarwal, Renuka Nagpal, Rajni Sehgal, "Crime Analysis using K-means Clustering". International Journal of Computer Applications(1975-8887) volume83, No-4, December, 2013.

[9] Hsinchum chen, Wingyan Chug, Yin Qin, Michael Chau, "Crime Data Mining: An Overview and Case Studies".

[10] Ms.Aruna J.Chamatkar, Dr.PK.Butey "Important of Data Mining with different Types of data Applications and Challenging Areas".