# An Improved Fuzzy C-Means Clustering Algorithm Framework for Profiling Criminal

Adeyiga J.A & Bello, A.

Department of Computer Science Bells University of Technology Ota, Ogun State, Nigeria. jadeyiga@yahoo.com, aishabello2050@gmail.com

> Adeyanju I.A Department of Computer Engineering Federal University Oye Ekiti, Ekiti State, Nigeria <ibrahim.adeyanju@fuoye.edu.ng>

Olabiyisi S.O Omidiora E.O Department of Computer Science and Engineering Ladoke Akintola University of Technology Ogbomosho, Nigeria soolabiyisi@lautech.edu.ng, eoomidiora@lautech.edu.ng

## ABSTRACT

## Abstract

Detecting criminals and solving crime is not an easy task at all and have been the prerogative of the law enforcement agencies, however with the increasing sophistication in technology, computer systems are now being used in tracking criminals and their activities and computer data analysts have started helping the law enforcement officers and detectives to speed up the process of solving crime with the use of a profiling system. Therefore, this research work will examine how real life data from the law enforcement agencies in Nigeria can be integrated and analyzed to produce "profiles" of activity and behavior of criminals. The aim is to provide investigators with rich sources of intelligence information which can be used to predict and prevent criminal activity. The proposed Improved Fuzzy Clustering algorithm for criminal profiling is an improvement over Fuzzy C-means expectation maximization (FCMEM) that uses fuzzy partition, but the proposed algorithm uses possibilistic partition which will eliminate the drawbacks of detecting outliers and the possibility of assigning them to an extra clusters. And this will be applied to the data gathered from the Nigeria Police department Elevele Ibadan and the Nigeria police force headquarters Abuja, using MATLAB software. The Improved Fuzzy C-Means clustering (Fuzzy Clustering) for the profiling system is going to be an improvement over other Clustering algorithms (Hard Clustering) based on the fact that soft clustering gives rooms for overlapping depending on the degree of membership or probability.

Keywords: Fuzzy C-Means Clustering, Algorithm, Framework, Security, Profiling and Criminal

Aims Research Journal Reference Format:

Adeyiga, J.A. Bello, A., Adeyanju, I.A., Olabiyisi, S.O. & Omidiora, E.O. (2016): An Improved Fuzzy C-Means Clustering Algorithm Framework for Profiling Criminal. Advances in Multidisciplinary Research Journal. Vol 2, No.2 Pp 123-134

## **1. INTRODUCTION**

Crime has become very sophisticated in Nigeria and fighting it with traditional means seemed not to be yielding desired results. Hardly a day goes without criminal activity being reported in our daily newspapers. Many people have had bad experiences at some point with these men of the underworld. How long shall we continue to live in fear of our lives being cut short by these criminals? No place seems safe anymore in Nigeria; not even our own little prison yards with high fences surrounding our homes and offices. However, crime is defined as "an act or the commission of an act that is forbidden or the omission of a duty that is commanded by a public law and that makes the offender liable to punishment by that law.

Security of citizens is the major concern of the police. Rather than focusing on enforcement, police can deter crime through the knowledge benefits they derive from information and its associated technologies [4]. As the rate of crime increases each year there is need to develop a computerized crime analysis tool that would assist the Nigerian Police Force and other law Enforcement Agencies to better combat crime. However, challenges to law Enforcement and Intelligence Agencies is the difficulty in analyzing large volume of data involved in criminal activities, as the crime rate is increasing [6].

Therefore, detecting patterns in crimes is the next step in predicting and subsequently responding to crime. As such, it is very important to attempt to detect patterns in crime [2]. In detecting patterns in crime, it is very important to gather data from different data sources, store and maintain the data, generate information and generate knowledge. Due to the vast use of computers and electronic devices and tremendous growth in computing power and storage capacity, there is an explosive growth in data collection. The storing of the data in data warehouse enables entire enterprise to access a reliable current database.

To analyse this vast amount of data and drawing fruitful conclusions and inferences, it needs special tools called data mining tools. According to [26], about 10% of the criminals commit about 50% of the crimes. Therefore, it is imperative to profile criminals in other to detect patterns from criminal records so as to discover clues that will help law Enforcement Agencies in the investigation of crime, because an ideal crime analysis tool should be able to identify crime patterns quickly and in an efficient manner for future crime pattern detections and action. Since Nigeria is still a developing nation little effort has been made in digitizing data and implementation of any crime analysis tool and as such most of the crime analysis is still being done manually, but due to lack of digitized data, data is still raw and as such contains many unwanted facts in which data cleaning will be performed. Hence, this research work focuses on developing a Criminal Profiling system using an improved fuzzy c-means algorithm.

## 2. FOCUS OF THE WORK/ MOTIVATION

A major challenge facing all law-enforcement and intelligence-gathering organizations is accurately and efficiently analyzing the growing volumes of crime data [6]. Another major challenge faced by the Nigerian law enforcement agencies is the lack of a central repository where all gatherings concerning crimes and criminals are stored which possess a bigger problem as there are many cases of data repetition and as such harder for law enforcers to see patterns in crimes during analysis. [14] Said, detecting crime from data analysis can be difficult because the daily activities of criminal generate large amounts of data and stem from various formats, and also the quality of data analysis depends greatly on the background knowledge of the analyst. However, intelligence and law enforcement agencies often are faced with the dilemma of having too much data, which in effect makes too little value [11], apart from which there is a lot of inconsistent, incomplete and incorrect data. In this research work we try to create a balance in looking for the best approach between the quality of data and quality of the knowledge gained after Mining. Current manual inspection of crime data by analysts is limited, primarily due to the amount of data that can be processed concurrently and in a reasonable time frame.

Also, several publications have studied the use of different clustering techniques in criminal profiling. For instance the work of [12];[23];[16] and [3]. However [12] and [23] addresses the issue of criminal career analysis by predicting unfolding careers of new criminals using hard clustering techniques. While [16] also addressed the issue of potentially identifying a criminal based on the witness/clue at the crime spot using hard clustering techniques. The techniques used lacked essential components which would have made the quality of the output more in tune with reality. Because the algorithms used were hard clustering, Fuzzy clustering is therefore proposed to address the issue of a criminal belonging to more than one cluster at a time.

The proposed fuzzy clustering algorithm to be adopted in this work is an improvement over the algorithm that was proposed by [8], Expectation Maximization Fuzzy C-Means clustering (EMFCM) which is a hybridization of FCM and Expectation Maximization algorithm with the use of fuzzy partition. It was expected that the algorithm will improve over Fuzzy C-Means clustering by avoiding the looping problems. But the performance of the algorithm has not been ascertained. But the proposed algorithm in this work uses possibilitics partition which will eliminate the drawbacks in fuzzy partition by detecting outliers and the possibility of assigning them to extra clusters.

Based on the issues identified above, this research work will examine how real life data from the law Enforcement Agencies can be integrated and analyzed to produce "profiles" of activity and behavior of criminals. The aim is develop a Criminal Profiling system using an improved Fuzzy C-Means clustering algorithm so as to provide investigators with rich sources of intelligent information which can be used to predict and prevent criminal activity.

#### 2.1 Crime Analysis, Profiling and Criminality

**Crime analysis**, a part of criminology, is a task that includes exploring and detecting crimes and their relationships with criminals [17]. Crime analysis is defined as a set of systematic, analytical processes directed at providing timely and pertinent information relative to crime patterns and trend correlations to assist operational and administrative personnel in planning the deployment of resources for the prevention and suppression of criminal activities [25]. Studying crime data includes all the intricate details of a particular crime. More specifically, crime analysis is the breaking up of acts committed in violation of laws into their parts to find out their nature and reporting statements of these findings [25]. But due to increased crime rate which leads to an increase in both crime and criminal datasets there is need to find a better and more efficient way to analyze this growing crime volumes.

Profiling is used to link a specific person or persons to a specific crime that have already occurred. Criminal profiling is the process of investigating and examining the criminal behavior in order to identify the type of person responsible [24]. A profile based on characteristics patterns or factors of uniqueness that distinguishes certain individuals from the general population. [21] explained that profiling criminals is important because not only does it help narrow down the pool of suspects in an investigation, it also helps to link related crimes, identify the potential for escalation of nuisance criminal behavior to more serious crimes, focusing the investigation, etc.

Identification of patterns may help in profiling. For example, an attacker may use a specific mode of operation, and target specific victims. Profiling can help identify the victim types, and assist in the identification and installation of sufficient deterrents. In order to develop useful profiles of different offender categories, a large amount of data is required and in order to improve the reporting of crime, there are needs to increase the trust between the public and private sectors, which will result in reporting of crimes when they occur. This will allow researchers to more precisely identify whether or not any unique patterns and characteristics actually exist [2].

#### 2.2 The Clustering Methods

Clustering is the technique that is used to group objects (crimes and criminals) without having predefined specifications for their attributes [25]. It ensures that the data objects are similar to one another within the same cluster and dissimilar to other objects in a different cluster. Clustering groups objects (crime and criminals) based on the information found in the data describing the objects or their relationships. The goal of clustering is that the objects in a group will be similar to one another and different from the objects in other groups. Clustering partitions a data set into several groups such that the similarity within a group is larger than that among different groups, the idea of clustering is similar to the way humans think. But we usually tend to have high volumes of data which is too cumbersome for the human mind to summarize into smaller groups. There is no golden standard or benchmark for the type of clustering technique to use, it depends on the goal of the researcher [14].

#### 2.2.1 Hard Clustering

These methods of clustering are based on the classical set theory, and require that an object either does or does not belong to a cluster [8]. It means partitioning the data into specified number of mutually exclusive subsets. That is, each document belongs to exactly one cluster, it only assign a value of 1 and 0. Such as, partitioning algorithm, density-based and hierarchical algorithm.



#### 2.2.2 Fuzzy Clustering

Fuzzy clustering also called soft clustering. Allow the objects to belong to several clusters simultaneously, with different degrees of membership. It is more natural than hard clustering because objects on the boundaries between several classes are not forced to fully belong to one of the classes, but rather are assigned membership degrees between 0 and 1 indicating their partial membership [8]. For example, the clusters or groups that are identified will be overlapping, meaning that one instance may fall into several clusters.

In fuzzy clustering we make a fuzzy partition of the data set. But to accommodate the introduction of fuzzy partitioning, the membership matrix (U) is randomly initialized according to equation 2.1. Each point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely to just one cluster. Fuzzy clustering uses membership function in partition data set.

$$\sum_{i=1}^{r} U_{ij} = 1, \forall j = 1....n$$
 2.1

This function is called membership function and its value is between 0 and 1. [8].

This can be applied to crime analysis because police officers commonly receive descriptions of suspects that are fuzzy in nature; the concepts of fuzzy variables will be introduced into criminal analysis to describe a fuzzy logicbased mathematical procedure that is capable of handling such variables. Fuzzy logic stipulates that an element can be a member of a given set in an uncertain manner. Unlike classical mathematical set theory in which an element can only be in two situations, a member or not a member of some set, fuzzy logic generalizes the possibilities and introduces the concept of shades of membership patterns. The concept of fuzzy logic was first put forth by [27]. Its incorporates the condition that an element is either a member or not a member of some set but also extends the condition by introducing the possibility of membership falling in mixed modes [9]. Such as Fuzzy C-Means and Expectation Algorithm.

#### 2.3 The Existing Improved Fuzzy C-means Algorithm using fuzzy partition

Step1.In these step we find the membership matrix (U) and initialise randomly in equation 2.2

$$\sum_{i=1}^{5} U_{ij} = 1 \quad \forall \quad j = 1, \dots, n$$
 2.2

This equation represent the membership matrix, it has taken the value equal to 1.

Step2. Calculate the centroids (ci) in equation 2.3

$$c_i = \frac{\sum_{j=1}^{j} u_{ij} v_{ij}}{\sum_{j=1}^{n} u_{ij}} \qquad 2.3$$

Centroid is main point of the cluster analysis system, in clustering this value of ci is depends on the member matrix function and related parameter of xi.

$$J(U, c_1, c_2, \dots, c_c) = \sum_{i=1}^{c} j_i = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m d_{ij}^2 = 2.4$$

Steps3. Using dissimilarity function to calculate the dissimilarities between centroid and data points in equation2.4 and check threshold value in equation2.5 then stop if we find the correct threshold value. In this steps check the threshold value using membership matrix and Euclidian distance between ith centroid (ci) and jth data point. If

$$|| U (k+1) - U (k) || < \epsilon$$
 2.5

In this equation we have to check the value of present classes and the next classes of the membership function. Check the value of membership matrix in next membership matrix in correct threshold value, which have been done by dissimilarities between centroid and data points. If values are satisfied, then we forwarded the next steps.

Advance	s In Multidis	sciplinary
ALIVIS Research Journal	& Scientific	Research

Step4. If threshold value is not correct we find the new mean (m i) using EM algorithm that has constraints in equation (2.6) using this equation we can find new mean, which is provided the correct threshold value for the dissimilarity function. So we can say these algorithms is deemed to have converged when the assignment no longer change. And it gave the best performance of initial means. EM algorithms commonly used initialization methods as random partition.

2.6

$$M_i^{(l+1)} = \frac{1}{|s_i^{(l)}|} \sum_{X_{j \in \mathcal{A}^{(l)}}} X_j$$

Step5. In these steps assign each observation to the cluster with the closest means in equation (2.7)  $S_i^{(l)} = \{Xp: || Xp - m_i^{(l)} || \le || Xp - m_j^{(l)} || \forall 1 \le j \le k$  2.7

In equation (2.6) we find new mean for new cluster. So we can say these algorithms is deemed to have converged when the assignment no longer change.

Source: [8]

## 3. RELATED WORK

There have been many research works in data mining especially in the area of security that has to do with criminal data analysis and profiling. For example, in order to address data mining in the area of profiling [1] proposed a method to employ computer log files as history data to search some relationships by using the frequency occurrence of incidents. Then, they analyzed the result to produce profiles, which can be used to perceive the behavior of criminal. Because criminal profiling helps in identifying crime characteristics, which is the first step in developing crime analysis. [12] Also introduced a framework for crime trends using a new distance measure for comparing all individuals based on their profiles and then clustering them accordingly.

While [12]; [23] both used data mining approaches in criminal career analysis. Four important factors play a role in the analysis of criminal career (crime nature, frequency, duration and severity). The tool that was developed in both works extracts these important characteristics of a criminal from the database and creates a digital profile for all offenders, compared all these individuals by a new distance measure and clusters them accordingly. But in the work of [23] there was significant improvement over the work of [12] by making four major key enhancements mainly to improve the semantics and the efficiency, thereby improving the existing methods of automated criminal career analysis. A new distance measure was introduced that more closely resembles the reality of policing. Instead of the previous, more rigid, comparison of career changes over time in the work of [12] they proposed the collection of crimes in a single year as a multiset, which then describes severity, nature and frequency inherently. A new distance measure called jaccard distance was employed for calculating the difference between two crime-multisets. Instead of the former method of a strict number wise comparison between years (comparing the first year of criminal A with the first year of criminal B, the second year of A with the second year of B, etc.), with the possibility of stretching or shrinking careers, they proposed a novel alignment of the mentioned multisets.

In conclusion, the enormous cloud of one-time offenders gave an unclear sketch of their distance space and the runtime of the chosen approach was not optimal yet as the clustering method (push and pull clustering) used in the former one was too intensive in a computational way causing performance delays, as this is highly inefficient in the real world where the volumes of data increase daily. This was addressed with the used of torus clustering. This was able to solve the problem of time complexity and some major problem encountered in the former method. Finally, it was suggested by [23] to equip the tool with a sub-cluster detection algorithm to provide better insights into the comparability of criminal careers. And also to set fuzzy borders between the different years' crimes within months ending or beginning such a time unit can be (partly) assigned to the next or previous year respectively as well, thus eliminating the problems arising with strict coherence to the change of calendar year. , they also suggested that other clustering techniques be adopted, if they adhere to the demand of incremental addition of single items.

According to [19], solving crimes is complex tasks that require human intelligence and experience and data mining is a technique that can assist them with crime detection problems. While [7] was able to do a survey that categorized, compared, and summarised from almost all published technical and review articles in automated fraud detection within the past 10 years. It defined the professional fraudster, formalised the main types and subtypes of known



fraud, and presents the nature of data evidence collected within affected industries. It was affirmed from their work that this field can benefit from other related fields. Specifically, unsupervised approaches from counterterrorism work, actual monitoring systems and text mining from law enforcement, and semi-supervised and game-theoretic approaches from intrusion and spam detection communities can contribute to future fraud detection research. Based on these, [3] was able to profile and analyse a criminal with the presence of evidence from the crime scene as oppose to the works of [12]; [23] who did their own analysis without evidence from the crime scene by constructing a Bayesian networks for criminal profiling from limited data with the presence of evidence from the crime scene. The method adopted in their work was that they developed a Bayesian network (BN) model of offender behaviour from a database of cleared homicides.

The BN was able to infer the characteristics of an unknown offender from the crime scene evidence, and help narrow the list of suspects in an unsolved homicide. Their research shows that 80% of offender characteristics are predicted correctly on average in new single-victim homicides, and when confidence levels are taken into account this accuracy increases to 95.6%. Their model is shown in Figure 2.1.

The knowledge that is gained from data mining approaches is a very useful tool which can help and support police forces [13]. These made [5] studied the application of fuzzy association rule mining for community crime pattern discovery. They were able to develop an approach that relieves the need of law enforcement personnel to go through uninteresting, obvious rules in order to find interesting and meaningful crime patterns of importance to their community. Rules discovered in their study offers utility for use from the national level down to the state and community level.



Figure 2.1 A BN model as decision-support tool for police investigation. Source [3]

Also, [25] was able to used decision tree and simple K-means algorithm to cluster crimes according to their attributes using an online police data in Iraq. The model of their work is shown in Figure 2.2. While [18] used a clustering/classify based model to anticipate crime trends which could be used to lessen and even prevent crime for the coming years. The first task they did was the prediction of the size of the population of a city and the basic approach used to do this was to cluster population sizes, create classes from the clusters, and then classify records with unknown population sizes, the reason why they used clustering to create classes was because classes from clusters are more likely to represent the actual population size of the cities.



WEKA mining software package was used and adopted for the clustering. The next task was the prediction of future crime trends by tracking crime rate changes from one year to the next and using data mining to project those changes into the future. The basic method they used was to cluster the cities having the same crime trend, and then using "next year" cluster information to classify records. They concluded that from the encouraging results, it is believed that crime data mining has a promising future for increasing the effectiveness and efficiency of criminal and intelligence analysis, and that there was still lots of options to explore. Visual and intuitive criminal and intelligence investigation techniques can be developed for crime pattern. While [22] used various data mining techniques to combat crime and terrorism in Nigeria, they analyzed how data mining techniques can be adopted by law enforcement agencies in tracking the activities of terrorist and their criminal activities, and also examined the limitation of data mining in fighting crime in Nigeria.



Figure 2.2: Framework Architecture for Crime Analysis Source: [25].

[18] Looked at use of missing value and clustering algorithm for crime data using data mining. They also looked at MV algorithm and Apriori algorithm with some enhancements to aid in the process of filling the missing value and identification of crime patterns. They also used semi-supervised learning technique for knowledge discovery from the crime records and to help increase the predictive accuracy. As always data pre-processing was discussed stating the various methods of treating missing values, because it is of grave importance to any data mining technique implementation. Two clustering techniques, K-means and DBScan (Density-Based Spatial Clustering Application with Noise) were used to test the data. Thus the two clustering techniques were analyzed in their efficiency in forming accurate clusters, speed of creating clusters, efficiency in identifying crime trend, identifying crime zones, crime density of a state and efficiency of a state in controlling crime rate. Experimental results showed that HYB algorithm show improved results when compared with k-means algorithm and therefore was used in further investigations.

In conclusion the research focused on developing a crime analysis tool for Indian scenario using different data mining techniques that can help law enforcement department to efficiently handle crime investigation. The tool was proven to be effective in terms of analysis speed. [16] Did an intelligent analysis of crime data using data mining methods (clustering and classification) and auto correlation models which aims to identify a criminal based on the witness/clue at the crime spot. The clustering of data was based on criminal/ crime and thereby minimizing the search space, based on the clusters the classification algorithm is then applied to classify the criminal then the auto correlation model authenticates the criminal.

[20] Proposed the hybridization of two algorithms k-means and Expectation maximization (EM) in order to improve on the drawback of k-means that does not always guarantee convergence. But with EM, it searches for a local maxima and converges very well. They therefore proposed a fuzzy K-means Expectation Maximization Algorithm (FKEM), the approach used to obtain initial cluster in their algorithm was by computing the weighted fuzzy averages instead of using simple mean which was also similar to the approached by [15]. The algorithm was tested using IRIS data set and the results compared with fuzzy K-means.



4.1

It was concluded that Fuzzy K-means was not able to cluster data correctly when the initial value of K was small. Hence, if good seed centers are not chosen, Fuzzy K-means will not perform very well. If the correct random seeds are not chosen, EM can still perform to make the K-means work correctly, since EM will iterate to find best centers for the given data. It was suggested that Fuzzy K-means along with EM gives a better clustering, than Fuzzy K-means. While [8] proposed a fuzzy clustering algorithm, Expectation maximization fuzzy c-means clustering (EMFCM) which is a hybridization of FCM and Expectation maximization algorithm. It was expected that their proposed algorithm will improve over fuzzy c-means clustering by avoiding the looping problems. But the performance of the algorithm was not ascertained and it was therefore proposed as an extension, so that the algorithm can be implemented and its performance evaluated using the mouse data set in MATLAB tool. Their algorithm uses fuzzy partition which as the limitation of detecting outliers and assigning them to extra clusters.

## 4. RESEARCH METHODOLOGY

In this research work, an improved fuzzy clustering algorithm will be developed for profiling criminals. This will be achieved by following the steps in the block diagram shown in Figure 4.1. It shows the major components of the proposed improved fuzzy clustering based framework for profiling criminals.

#### 4.1 Data Acquisition

The data acquisition involves gathering of relevant data from the law enforcement agencies in Nigeria, the data to be used will be from the Nigeria Police Department, Eleyele, Ibadan. And some pre-processing of already existing data would enable us profile criminals using Fuzzy clustering algorithm.

#### 4.2 Mathematical formation of the criminal profiles per offender

From the data set gathered, the crimes of an individual are splitted up into different types; all these crimes have got an intrinsic level of seriousness attached to them. To determine the criminal profile differences, these is done by distinguishing three types of crime seriousness; Simple offence, Misdemeanor and Felony. Each crime type must definitely fall within one of such categories with an attached weight value. And also, the crime with and without use of weapon could also be differentiated with value attached to them and then the number of crimes committed. The mathematical equation to compute the criminal profile of a criminal is shown in equation 3.1.

$$\mathsf{CP} = (\mathsf{Cv}^* \mathsf{ i}) (\mathsf{w}) + \mathsf{CFv}$$

Where,

CP is the criminal profile per offender i is the weight of the offence committed. Cv is the criminal value with an attached value of one w is the weight of the weapon used. CFv is the crime frequency value. A value of 1 is assigned to an offender without the use of a weapon, why 2 is assigned to an offender with a weapon. The CFv is calculated and gotten from Table 4.1.

## 4.3 The Improved Fuzzy C-Means clustering algorithm

The Proposed Algorithm

The proposed algorithm is an improvement on the Expectation maximization fuzzy c-means clustering algorithm that was proposed by [8]. The major improvement is on the drawback of fuzzy partitions, which makes detection of outliers difficult and also in assigning them to extra clusters. The use of possibilistic partition will be introduced in this new algorithm to overcome these drawbacks that require that the sum of memberships of each point equal to one.

## Table 4.1: Criminal Frequency Value table

Category	No of Crimes(#)	Assigned value(CFv)
0	0	0
1	1	1
2	2-4	2+(#-2)/3
3	4-7	3+(#-5)/5
4	>7	4

Source: [12]

## 4.4 The proposed improved Fuzzy C-Means algorithm using possibilistic partition.

Step1.Compute the membership matrix (U) using possibilistic partition

$$\sum_{i=1}^{c} U_{ij} > 0 \quad \forall \quad j = 1, ...., n$$
4.2

Step2. Compute the centroids which are the main point of the cluster analysis system.

$$c_{l} = \frac{\sum_{j=1}^{n} u_{lj}^{m} x_{lj}}{\sum_{j=1}^{n} u_{lj}^{m}}$$
4.3

Steps3. Dissimilarity function is used to calculate the differences between centroid and data points and the threshold value is check in equation 3.5 and terminate if we find the correct threshold value else it goes to step 4. It measures how well the clustering fits the data.

$$J(U, c_1, c_2, \dots, c_c) = \sum_{i=1}^{c} j_i = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m d_{ij}^2$$

$$4.4$$

lf

|| U (k+1) - U (k) ||< ∈

In this equation we have to check the value of present classes and the next classes of the membership function. Check the value of membership matrix in next membership matrix in correct threshold value, which have been done by dissimilarities between centroid and data points. If values are satisfied, then we forwarded the next steps. Step4. If threshold value is not correct we find the new mean (m i) using EM algorithm that has constraints in equation (4.6) using this equation we can find new mean, which is provided the correct threshold value for the dissimilarity function, So we can say these algorithms is deemed to have converged when the assignment no longer change. And it gave the best performance of initial means. EM algorithms commonly used initialization methods as random partition.

4.5

$$M_i^{(l+1)} = \frac{1}{|s_i^{(l)}|} \sum_{X_{j \in S}^{(l)}} X_j$$
4.6

Step5. In these steps assign each observation to the cluster with the closest means in equation (4.7)

$$S_i^{(l)} = \{Xp \colon \parallel Xp - m_i^{(l)} \parallel \le \parallel Xp - m_j^{(l)} \parallel \forall 1 \le j \le k$$

$$4.7$$

In equation (4.6) we find new mean for new cluster. So we can say these algorithms is deemed to have converged when the assignment no longer change.



## 4.5 Description of the proposed model for criminal profiling

Data will be gathered from different sources, such as the crime data, criminal data and historical data. And the gathered data will be transformed to a crimes/criminal data warehouse. From the crime/criminal data warehouse, attributes such as the crime nature, severity of the crime, mode of operation and the weapon used will be extracted out. Based on the attributes, a profile will be created for individual criminal. And the proposed improved Fuzzy C-Means will be applied on the profile created in order to generate different clusters and knowledge will be generated from the clusters formed which will be useful to the law enforcement agencies. The proposed model is shown in Figure 4.2.



Figure 4.2: The proposed model for criminal profiling

## 5. DIRECTION FOR FUTURE RESEARCH

The Improved Fuzzy C-Means clustering Algorithm framework should be implemented and the result evaluated with the existing Fuzzy C-Means Clustering Algorithm so as to bench mark their performance in order to justify if the proposed Algorithm is better than the existing.

#### 5.1 Conclusion

Crimes have depressed trade and weaken investors' confidence in our economy, and to that extent it is a clear danger to our national security and the prosperity of our citizens. An improved application prototype would be developed that would help the law Enforcement Agencies in the following ways:

- i. Potentially improving decision making in the various law enforcing agencies.
- ii. Potentially reducing the difficulties in managing large volume of data involved in criminal activities.
- iii. Identifying a criminal in the absence of witnesses or any clues by forensics expert
- iv. Decrease process time of crime analysis to enable quick completion of a crime investigation.

This study will also assist in making proper recommendation to the appropriate quarters, knowing that crime has a devastating effect on the nation's economy. With the use of an Improved Fuzzy C-Means clustering for the profiling system and applying it to crime data so as to establish that Fuzzy Clustering algorithm is going to be an improvement over other Clustering algorithms based on the fact that soft clustering gives rooms for overlapping depending on the degree of membership or probability.

#### REFERENCES

- [1] Abraham, T. and de Vel, O. (2006) Investigative profiling with computer forensic log dataand association rules, in Proceedings of the IEEE International Conference on Data Mining (ICDM' 02), Pp. 11 18.
- [2] Askeriya I, H. Jahankhani, S. Wee Lee & Ameer Al-Nemrat, (2010): "Education, Trainingand Awareness (ETA) Four Dimensional Cybercrime prevention model"
- [3] Baumgartner, K, Ferrari S and Palermo G (2008): "Constructing Bayesian Networks for Criminal profiling from limited data" Elsevier 21 March 2008.
- [4] Brown Mary Maureen and Brudney Jeffery. L (2003), "Learning Organization in the PublicSector, A study of Police Agencies employing Information and Technology to Advance Knowledge" Public Administration Review Vol 63, Issue 1, pp. 30- 43
- [5] Buczak, L. Anna, and Christopher M, Gifford (2010)"Fuzzy Association Rule Mining for Community Crime Pattern Discovery" Published in Proceeding ISI-KDD' 10 ACM SIGKDD Workshop on Intelligence and Security Informatics, Article No, 2.
- [6] [Chen, H. Atabakhsh. H. Petersen, T. Schroeder J. Buetow, T. Chaboya, L. & Huang, Z(2003, may).COPLINK: Visualization for Crime Analysis In proceedings of the 2003 annual national conference on Digital government research (pp, 1-6). Digital Government Society of North America
- [7] Clifton Phua, Vincent lee. Kate smith & ross Gayler (2007) "A Comprehensive survey ofData Mining based Fraud Detection Resarch"
- [8] Esh Narayan, Yogesh Birla and Gaurav Tax(2012) "Enhancement of fuzzy C-means clustering using Expectation Maximization algorithm". International journal of Computer Applications Vol 43-no 13 April 2012.
- [9] Gomez, J, and Dasgupta, D. (2002). "Evolving Fuzzy Classifiers for Intrusion Detection". InProceedings of 2002 IEEE workshop on Information Assurance, 68 – 75 West Point, NY: IEEE.
- [10] Hoppner F, Klawonn F Fruse R and Runkler(1999),"fuzzy Cluster Analysis"J. Wiley andSons, Chichester, England 1999.
- [11] Jennifer Xu, Hsinchun Chen (2005) "Criminal Network Analysis and Visualization: A Data Mining Perspective" communications of the ACM, Vol. 48 No. 6, pp, 100-107
- [12] Jeroen S, De Bruin, Tim K. Cocx, Walter A, Kosters, Jeroen F, J. Laros and Joost N, Kook (2006) "Data Mining Approaches to Criminal Career Analysis," in Proceedings of the Sith International Conference on Data Mining (ICDM' 06), pp, 171 – 177
- [13] Keyvanpour, M.R, Javideh, M. and Ebrahimi, M.R. (2010) Detecting and investigatingcrime by means of data mining : a general crime matching framework, Procedia Computer Science, World Conference on Information Technology, Elsvier B,V. Vol. 3, pp. 872 - 830
- [14] Krishnamurthy Revating (2012) "Survey of Data Mining Techniques on Crime Data", International Journal of Data Mining Techniques and Applications vol. 01, pp. 48 55
- [15] [15] Looney C. G (2002) "Interactive clustering and merging with a new fuzzy expected values . Pattern Recognition Lett. Vol 35, 2002 187-197.
- [16] [Mande Uttam, Y. Srinivas, J.V.R.Murthy (2012) An Intelligent analysis of Crime Datausing data mining and auto correlation models. "International Journal of Engineering Research and Applications" (IJERA) ISSN: 2248-9622 <u>www.ijera.com</u> Vol. 2, Issue4,July – August 2012. Pp. 149 – 153
- [17] Malathi and Santhosh, (2011) "an enhanced algorithm to predict a future crime using dat mining" International journal of computer applications (0975 - 8887) volume 21 – no. 1, May 2011.
- [18] Malathi, Dr. S. Santhosh Baboo (2011) "Algorithm crime prediction Model based on theanalysis of crime clusters" Global journal of computer science and technology volume 11 Issue 11 version 1.0 july 2011 Type:pp 46-51
- [19] Nath,S.(2007) crime data mining,advances and innovations in systems, K. Elleithy (ed.), computing sciences and software Engineering,Pp.405-409.

- [20] Naser Sara, Rawn Alkhaldi and Gregory Vert (2005) "A modified fuzzy k-means clustering using Expectation maximization algorithm.
- [21] Nykodym, N, Taylor, R, Vilela, j,(2005): "Criminal profiling and insider cybercrime" computer law and security report 21(5),pp.408-414.
- [22] Raphael O. O & Francis O.E (2011): "Combating Crime And Terrorism Using Data Mining Techniques", 10<sup>th</sup> International Conference On Information Technology for People Centred Development Nigeria Computer Science Conference Proceedings, Vol 22, pp80 – 89 2011.
- [23] Tim k. Cocx, Jeroen F. J Laros, Walter A. Kosters (2008) "Enhancing the Automated analysis of criminal careers".
- [24] Turvey, B. (2002): "Criminal Profiling, An Introduction To Behavioural Evidence". UK, Elsevier.
- [25] Swadi B, Kadhim. Al- Janabi (2011) "A Proposed Framework for Analyzing Crime Data Set Using Decision Tree and Simple K – Means Mining Algorithms" Journal of kufa for Mathematics and Computer Vol. 1. No 3, May 2011, pp 8 – 24.
- [26] [26] Shyam Varan Nath, (2006)"Crime pattern detection using Data minig" Proceeding of the International Conference on web intelligence and intelligence agent. Pg 41-44
- [27] [27] Zadeh L.A,(1975)"Fuzzy logic and approximate reasoning, Synthese , Vol, 30 pg 407-428