

Contents list available at JMCS

Journal of Mathematics and Computer Science

Journal Homepage: www.tjmcs.com



Offenders Clustering Using FCM & K-Means

Sara Farzai ¹, Davood Ghasemi ², Seyed Saeed Mirpour Marzuni

¹Adib High Education Institute of Mazandaran, Sari, Iran
Sara.farzai@gmail.com - (The corresponding author)

²Shomal University, Amol, Iran
d.ghasemi@shomal.ac.ir

Article history:

Received March 2015

Accepted July 2015

Available online July 2015

Abstract

One of the most applicable and successful methods to provide security in society is to use data mining techniques to recognize patterns of crimes. Data mining is a field that discovers hidden patterns of large amount of data in large data bases, and also extracts useful knowledge in every field which uses it. Clustering is a technique of data mining that divides data points into many groups so that the members of each group have the most similarity and the members from different groups have the least similarity. In this paper we cluster 100 offenders according to crime they have committed, using Fuzzy C-Means and K-Means algorithms in Matlab and Weka environments. Then we studied the intersections in efficient elements in crime occurrence in each cluster. We obtained interesting results coincided our real data. Hence we have created a pattern which is able to detect crime with considering other attributes, and reversely. It is clear that these detections can help to decrease the effects of crime. Note that Fuzzy C-Means algorithm has provided more accurate results in comparison with K-Means algorithm, because of considering fuzzy point of view and natural uncertainty in the real world.

Keywords: Crime, Offender, Data Mining, Clustering.

1. Introduction

There is no society in which crime does not occur. Since humans came together and constitute a community, crime has been consequently gathered to them and increased. Every community tries to guarantee a stable security for all the people, since security and safety is an important element to provide peace for people, and all of the scientific, economic, social and cultural developments are surely more accessible by having a desired security and safety. The main concern of legislators in every society is to provide stable definite security for all the citizens, that was never possible from first generations of human, and always a variety of crimes have been occurred in different points of the world. However it is not possible to provide perfectly guaranteed security, various strategies about reduction of crime and guilt may be studied. This goal is followed in different countries by ordaining criminal laws.

Applying Data mining techniques is of the most successful strategies in crime's patterns recognition. Data mining is the field of recognizing hidden patterns in a great amount of data which are stored in huge databases. Also it extracts "useful knowledge" in each area in which, it being used. In field of criminal law, a definition of useful knowledge would be: a knowledge that recognizes the patterns of criminal activities and aims to crime reduction and crime's consequences with on-time recognition. Data mining performs this mission well by pattern discovery and relations recognition among huge data. For example by discovery of cyber-crime's patterns, it is capable to detect similar cases.

In this paper we cluster offenders using Fuzzy C-Means¹ and K-Means algorithms, so that offenders to be divided into many classes that members of each class have the most similarity in personal characteristics, mental characteristics and the crime they have committed, and members of two different classes have the most dissimilarity. Discovery of similarity pattern that is mission of data mining and clustering, can help us to recognize new cases and to do on-time preventive reactions consequently. Note that FCM algorithm trends to fuzziness. It means instead it to say "the property (a) *surely* will result in occurrence of crime (b)" it checks *how much* the property (a) would be effective in occurrence of crime (b).

The rest of paper is arranged as follows: In section 2 we have an overview on related works. Section 3 is dedicated to required literature and a discussion on how we perform clustering. In section 4 we represent the results of our implementations and consequently we will have a comparison between results. And finally in section 5 we conclude.

2. Related works

Intra-field researches among psychology, sociology and criminology around data mining applications in criminal offense try to study about the reasons of crime occurrence and its reduction, that many of them are mentioned here.

Corapcioglu & Erdogan [1] studied about behavioral and sociological characteristics of offenders whom recommitted crime. The goal of their research was to find some special characteristics which redound crime reoccurrence. They gathered all the behavioral and sociological information of under supervision prisoners. As a main result they concluded if prisoners to be learnt and guided to control their anger, the probability of crime reoccurrence will reduce by a high ratio. Li, Kuo & Tsai [5] represented a fuzzy based decision-support model to process recognize and pattern analyze in crime occurrence. Their represented model has been implemented on Taiwan's international police's data. The results have aided the managers of the police force to achieve preventive approaches against crime and villainy. Karlis & Meligkotsidou [3] represented a model based on clustering technique for recognition and grouping of variety of crimes. Oatley & Ewart [8] worked on a project aiming police force to inspect crimes and robberies that were occurred with a high frequency. They designed a software which was working based on exact information of criminals. Moon, McCluskey & McCluskey [7] used regression to predict cyber-crimes. According to their results, increasing the time of using computer and having membership in internet networks and groups have increased cyber-crimes and have been introduced as the main cyber-crime predicting variables. Liu, Brown & Donald [6] represented a model to predict the location of crime occurrence in next week regarding to information of current week that could be effective in crime occurrence preventing. Chung et al. [2] introduced data mining as a most efficient tools in cyber-crimes and had a survey study in this area. Khan & Shaikh [4] used data mining techniques to predict and preventing social-networks-crimes in internet environment.

¹ FCM

3. Literature and method

A prerequisite to distinguish an offender is to have knowledge about the nature of crime. Because the term “offender” is assigned to whom has committed a “crime”. Several elements may boost the probability of showing criminal behaviors. Therefore firstly we should have a perfect knowledge about the meanings and concepts of crime, to analyze the criminal behaviors. Then we can study about influence of different elements on performing a crime.

Crime literally means guilt [12]. Among lots of presented definitions of crime in different societies, there is no a unified definition yet which be acceptable for everyone in all the situations. A common definition of crime is “performing any act, or not performing an act, that a punishment has been considered for it in law” [14].

3.1. Effective attributes

Since we want to cluster offenders based on their attributes and the crimes they have committed, it is required to determine the effective elements on crime occurrence, and studying relationships between attributes and the crimes. In this section we introduce some of the most important properties in people which boost probability of crime occurrence, and we will use them in next section for clustering.

However there are ambiguities in diagnose personality abnormalities, personality abnormality diagnosing is the commonest diagnosis which is often done for offenders. *Personality* is a combination of human’s body traits and his mental properties that forms a nature for a man and makes him distinguished from others. Personality is an undividable unit fact which determines how the man is adapted with society and environment [13]. Researches show that one or more personality abnormality have been diagnosed in more than 66% of offenders [11]. Some of the people when encounter with mental pressures, show criminal behaviors, such as using different narcotics and alcohol, aggression, robbery and etc., to manage their mental problems [9].

Based on researches results and according our consulting with experts in criminology area, the most effective properties in crime occurrence are as following:

1. Gender: male – female
2. Family: peaceful – no peaceful
3. Education: guidance school – high school – university
4. Mental status: normal – depressed – aggressive – wayward
5. Revenue: no revenue – low – middle – high
6. Using alcohol before committing crime: yes – no
7. Using narcotics before committing crime: no narcotic – opium – crack – crystal – grass.

In next section we will discuss about how the properties above play role in crime occurrence.

3.2. Offenders clustering

We have selected the important effective properties in performing criminal behaviors. Now we discuss about the crimes which often take place due to properties in last section. As the named properties were the commonest personal and mental elements that have been diagnosed in offenders, we have chosen

frequently committed crimes due to properties mentioned in last section. Our selected crimes are: kill – battery – robbery – abduct – rape – narcotic – alcoholism.

Our required data is obtained from questionnaires that were distributed among one hundred prisoned offenders: fifty women and fifty men. In questionnaires we considered some psychology questions to extract the offender's mental properties and related characteristics. Then we used Microsoft Excel program to save the information in .csv file format to use for clustering implementation. Figure 1 shows a part of the named file. Our statistics show that the most frequent crime was rape (34%), narcotic was stated in second place (with 19%) and abduct is stated in last place (4%).

```
Gender,Family,Education,Mental Status,Revenue,Alcoholism,Narcotism,Crime
Male,Peaceful,University,Normal,High,Yes,No Narcot.,Rape
Male,No Peaceful,H-School,Aggressive,Middle,No,Grass,Battery
Male,No Peaceful,University,Depressed,Middle,No,Crystal,Rape
Male,No Peaceful,G-School&less,wayward,Low,No,Crack,Robbery
Male,No Peaceful,G-School&less,Depressed,No Revenue,No,Opium,Robbery
Male,No Peaceful,H-School,wayward,High,No,Grass,Kill
Male,Peaceful,H-School,Normal,Low,No,No Narcot.,Entfuehrer
Male,No Peaceful,H-School,Aggressive,Middle,No,Grass,Narcotic
Male,Peaceful,University,Normal,High,Yes,No Narcot.,Alcoholic
Male,No Peaceful,University,Depressed,Middle,No,Crystal,Narcotic
Male,No Peaceful,G-School&less,wayward,Low,No,Crack,Robbery
Male,No Peaceful,G-School&less,Depressed,No Revenue,No,Opium,Robbery
Male,No Peaceful,H-School,wayward,High,No,Grass,Rape
Male,Peaceful,H-School,Aggressive,Middle,Yes,No Narcot.,Alcoholic
Male,Peaceful,H-School,Aggressive,Middle,Yes,No Narcot.,Rape
Male,Peaceful,University,Normal,High,Yes,No Narcot.,Battery
Male,Peaceful,University,Normal,High,Yes,No Narcot.,Alcoholic
Male,No Peaceful,University,Aggressive,Middle,No,No Narcot.,Entfuehrer
Male,No Peaceful,H-School,Aggressive,Middle,No,Grass,Kill
Male,No Peaceful,H-School,Aggressive,Middle,No,Grass,Narcotic
Male,No Peaceful,G-School&less,Depressed,Low,No,Crack,Robbery
Male,No Peaceful,G-School&less,Depressed,Low,No,Crack,Narcotic
Male,Peaceful,University,Normal,High,Yes,Crystal,Rape
Male,Peaceful,University,Normal,High,Yes,Crystal,Alcoholic
Male,Peaceful,University,Normal,High,Yes,Crystal,Narcotic
Male,No Peaceful,University,Depressed,Middle,No,Crack,Rape
Male,No Peaceful,G-School&less,Depressed,Middle,No,Crack,Robbery
Male,No Peaceful,H-School,wayward,Middle,Yes,Grass,Narcotic
Male,No Peaceful,H-School,wayward,Middle,Yes,Grass,Narcotic
Male,No Peaceful,H-School,wayward,Middle,Yes,Grass,Narcotic
Male,Peaceful,University,Normal,High,Yes,No Narcot.,Rape
Male,Peaceful,University,Normal,High,Yes,No Narcot.,Alcoholic
Male,Peaceful,H-School,Normal,Low,No,No Narcot.,Entfuehrer
Male,No Peaceful,G-School&less,Aggressive,Low,No,Grass,Kill
Male,No Peaceful,G-School&less,Aggressive,Low,No,Grass,Narcotic
Male,No Peaceful,G-School&less,Aggressive,Middle,Yes,Grass,Rape
Male,No Peaceful,G-School&less,Aggressive,Middle,Yes,Grass,Alcoholic
Male,No Peaceful,G-School&less,Aggressive,Middle,Yes,Grass,Narcotic
Male,Peaceful,H-School,Normal,Middle,Yes,No Narcot.,Rape
Male,Peaceful,H-School,Normal,Middle,Yes,No Narcot.,Alcoholic
Male,No Peaceful,G-School&less,Depressed,No Revenue,No,No Narcot.,Entfuehrer
Male,No Peaceful,H-School,Depressed,Low,No,No Narcot.,Kill
Male,No Peaceful,H-School,Aggressive,Low,Yes,Grass,Rape
Male,No Peaceful,University,Normal,High,Yes,No Narcot.,Rape
Male,No Peaceful,University,Normal,High,No,Grass,Rape
Male,No Peaceful,H-School,Depressed,Middle,No,Crack,Robbery
Male,Peaceful,G-School&less,Depressed,High,No,Crack,Robbery
```

Figure 1: Information of questionnaires

3.2.1. Clustering using FCM

We implement FCM clustering in Matlab, which is a numeric matrix based software. So at first we have to transform our information into a numeric form which be able to be processed in Matlab. To do so, we assign a number to each of property value, as in follows:

Gender: 0-male / 1-female

Family: 0-peaceful / 1-no peaceful

Education: 0-guidance school & less / 1-high school / 2-university

Mental status: 0-normal / 1-depressed / 2-aggressive / 3-wayward

Revenue: 0-no revenue / 1-low / 2-middle / 3-high

Using alcohol before committing crime: 0-no / 1-yes

Using narcotics before committing crime: 0-no narcotic / 1-opium / 2-crack / 3-crystal / 4-grass.

Note that here we don't use the mathematic value of these assigned numbers in calculating or computation. These numbers are just symbols which help us to recognize the values of property variables, and also Matlab can simply understand them. For example the numeric sequence as "01021031" (which should be read from left to right) means:

A man, from a no-peaceful family, educated till guidance school or less, wayward, with low revenue, didn't use alcohol, but has used grass, killed someone.

After data transforming, we study on relation between each property and crime, and then we perform FCM clustering in Matlab according to selected crimes. General FCM code in Matlab is based on equation (1):

$$[\text{center}, U, \text{obj_fcn}] = \text{fcm}(\text{data}, \text{cluster_n}) \quad (1)$$

Where data is extracted from questionnaires, cluster_n is the number of clusters, center is the cluster centers matrix, U is the fuzzy membership coefficients matrix, and obj_fcn is the value of objective function during iterations. Clustering will stop when at least one of conditions below to be satisfied:

- 1- Maximum iterations to be performed.
- 2- Objective function enhancement between two successive iterations be less than a threshold.

Figure 2 shows a sample clustering based on using alcohol. Figure shows the trend of each cluster center to crimes. As you see, offenders who have used alcohol are potentially talented for trend to narcotic.

3.2.2. Clustering using K-means

To cluster by K-means it is not needed to work on numeric data. We can use our main csv file directly in Weka for K-means clustering. According the opinion of our consultant experts, it is convenient to select 7 clusters proportionate with goal crimes. Euclidean distance is the metric of similarity here. It means the more distance implicates the less similarity and the less distance implicates the more similarity (see figure 3). Results are represented in next section.

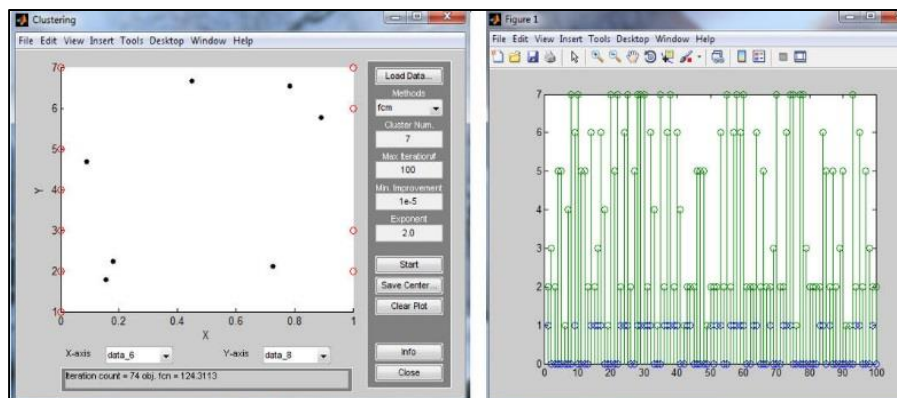


Figure 2: a sample FCM clustering based on using alcohol

Cluster centroids:								
Attribute	Full Data (100)	Cluster# 0 (27)	1 (23)	2 (12)	3 (11)	4 (2)	5 (10)	6 (15)
Gender	Male	Male	Male	Female	Male	Male	Female	Female
Family	No Peaceful	Peaceful	No Peaceful	No Peaceful	No Peaceful	Peaceful	Peaceful	No Peaceful
Education	H-School	University	H-School	University	G-School/less	H-School	H-School	G-School/less
Mental Status	Depressed	Normal	Aggressive	Depressed	Depressed	Depressed	Normal	Wayward
Revenue	Middle	High	Middle	High	Low	High	Low	No Revenue
Alcoholism	No	Yes	No	No	No	No	No	No
Narcotism	No Narcot.	No Narcot.	Grass	Grass	Crack	Crack	No Narcot.	No Narcot.
Crime	Rape	Alcoholic	Narcotic	Rape	Robbery	Robbery	Rape	Rape

Figure 3: K-means clustering

4. Results

Table 1: Results of FCM clustering

Num	properties	crimes	
1	Gender	Male	Kill – abduct – alcohol – narcotic - rape
		Female	Rape – alcohol - narcotic
2	Family	peaceful	Rape - narcotic
		No-peaceful	Rape(2) ² – narcotic – robbery - kill
3	Education	Guidance sch.& less	Rape - robbery
		High school	Rape – alcohol - narcotic
		University	Rape(2) – alcohol - narcotic
4	Mental status	Normal	Alcohol - narcotic
		Depressed	Rape – alcohol – narcotic – robbery - abduct
		Wayward	Rape –narcotic – robbery - abduct
		Aggressive	Narcotic
5	Revenue	No-revenue	Rape - robbery
		Low	Rape – robbery
		Middle	Rape(2) – narcotic - alcohol
		High	Rape(2) – narcotic - alcohol
6	Alcoholism	No	Rape(2) – robbery – abduct - kill
		Yes	Rape – narcotic - alcohol
7	narcotic	No	Rape(2) - alcohol
		Opium	Robbery - abduct
		Crack	Rape - robbery
		Grass	Rape - kill
		Crystal	alcohol

Table 1 shows the results of FCM clustering. Some of crimes are repeated in some of properties (for example rape has been appeared in both peaceful and no-peaceful families, and also in other properties values). This iteration is due to fuzziness of clustering. It means a crime has membership in different functions. Membership degree can be calculated from matrix U in equation (1). We have calculated the

² Number “2” indicates two cluster centers in a property value.

membership degree in each property by computing the distance from cluster center. For simplicity we have chosen membership degrees regardless the degrees lower than 0.5. As table 1 shows, we can fuzzily detect a particular crime occurrence regarding to properties values and mixing them. For example a woman whom lives in a no-peaceful family and uses alcohol, is generally talented for crimes below:

Gender: female	rape– narcotic- alcohol	Crime: rape(2)-narcotic- robbery- kill- alcohol
Family: no-peaceful	rape(2)- narcotic– robbery- kill	
Alcohol: yes	rape – narcotic - alcohol	

Table 2 shows the results of K-means clustering. According table 2, the following results are obtained:

- 1- The most of the crimes have occurred in no-peaceful families.
- 2- Unless in special cases, using alcohol and narcotic have reversely relation.
- 3- Most of the sexual crimes are iterated in women with low or no revenue.

Our observation from both FCM and K-means clustering results indicates existence a high similarity between our results and real world statistics. Note that FCM in comparison with K-means produces more precision results, because of considering fuzziness.

Table 2: Results of K-means clustering

num	gender	family	Edu.	Mental st.	Rev.	Narcot	Alc	crime	%
0	male	peaceful	uni	normal	high	No	Yes	Alcohol	27%
1	male	No-peac	Highsch	aggressive	mid	grass	No	Narcot.	23%
2	female	No-peac	uni	depressed	high	grass	No	Rape	12%
3	male	No-peac	Guidsch	depressed	low	Crack	No	Robbery	11%
4	male	peaceful	Highsch	depressed	high	Crack	No	Robbery	2%
5	female	peaceful	Highsch	Normal	low	No	No	Rape	10%
6	female	No-peac	Guidsch	wayward	Norv	No	no	rape	15%

5. Conclusion

In this paper we clustered offenders using FCM and K-means algorithms. Our selected goal crimes in consulting with experts of criminology are: kill– robbery– rape– abduct– battery– narcotic– alcoholism. Required data was extracted from questionnaires that we had distributed among 50 men and 50 women prisoned offenders and these offenders had committed one of the goal crimes at least. We extracted some properties such as gender, family status, revenue, mental status, education, alcoholism and narcotic related to offenders. We selected 7 clusters according expert’s opinion and proportionate the number of crimes. Interesting results were obtained from 100 under study cases as below:

- 1- Most of the crimes was seen in no-peaceful families.
- 2- Unless special cases, a reversely relation was seen between using alcohol and narcotic.
- 3- The most iteration of rape has been seen in women with low or no revenue.

- 4- Users of some particular narcotics had trend to some special crimes. For example users of opium and crack were talented for robbery, and users of grass and crystal were talented for rape and generally sexual crimes.

Both of FCM and K-means algorithms have represented results with lots of intersections, which indicates the success of clustering technique in pattern detection. Also obtained results have a great similarity with our real statistics. FCM has provided more precision results because of considering fuzziness. So use of clustering technique has enabled us to detect crime, regarding to personal and mental properties and reversely. As a future work we can use a mixture method of classification and clustering to take advantages of both methods and generally cover the low points of each method, to obtain a result near to time and cost optimality.

References

- [1] Corapcioglu A., Erdogan S. "A Cross-Sectional Study on Expression of Anger and Factors Associated With Criminal Recidivism in Prisoners with Prior Offences", *Forensic Science International*, No. 140, (2004) 167-174.2
- [2] Chung W., Chen H. Ch., Chang W., Chou SH., "Fighting Cybercrime: A Review and the Taiwan Experience", *Decision Support Systems*, No. 41 (2006) 669 - 682.
- [3] Karlis D., Meligkotsidou L., "Finite Mixtures of Multivariate Poisson Distributions with Application", *Journal of Statistical Planning and Inference*, No. 137, (2007) 1942 – 1960.
- [4] Khan J. I., Shaikh S. S., "Computing in Social Networks With Relationship Algebra", *Journal of Network and Computer Applications*, No. 31, (2008) 862–878.
- [5] Li Sh. T., Kuo SH. CH., Tsai F. CH., "An Intelligent Decision-Support Model Using FSOM and Rule Extraction for Crime Prevention", *Expert Systems with Applications*, No. 37, (2010) 7108–7119.
- [6] Liu H., Brown Donald E., "Criminal Incident Prediction Using a Point- Pattern-Based Density Model", *International Journal of Forecasting*, No. 19, (2003) 603–622.
- [7] Moon B., McCluskey J. B., McCluskey C. P., "General Theory of Crime and Computer Crime: An Empirical Test", *Journal of Criminal Justice*, No. 38, (2010) 767–772.
- [8] Oatley G. C., Ewart B. W., "Crimes Analysis Software: 'Pins in Maps'", *Clustering and Bayes Net Prediction*, *Expert Systems with Applications*, No. 25, (2003) 569–588.
- [9] Sanoroyan M., "Neural Symptoms Norm in Men Prisoners", *Notion and behavior*, No. 4 (2005).
- [10] Zadeh L. A., "Fuzzy sets", *Information and Control*, vol. 8(3): 338–353.
- [11] Dadsetan. P., "Criminal psychology", Samt publication, Tehran, (2006).
- [12] Sayyah A., "Universal Arabic-Farsi dictionary", 1st edition.
- [13] Keynia M., "Criminal psychology", 1st edition, Roshd publication, (2005).
- [14] "Iranian Human Rights" (Islamic criminal law), (2004).