

A Survey of Fuzzy Based ARM Clustering on Crime Pattern Discovery

Prof. J. Mohana Sundaram, Dr. T. Karthikeyan, R. Karthik Raj

Abstract -Fuzzy set theory is an excellent mathematical tool to handle the uncertainty. Fuzzy set theory exhibits immense potential for effective solving of the uncertainty in the problem. A systematic process of gathering, classifying, investigating, and publicizing timely, exact, and useful evidence that describes crime patterns, crime trends, and potential suspects. The randomness, non-linear nature, and seeming disorder of criminal activity often makes it difficult to employ traditional prediction tools such as geographic information systems (GIS). Fuzzy is flexible which will randomly initialized parameters, this make them particularly appropriate for criminal activity forecasting. ARM algorithm is used to save time which make execution progress faster and quickly. Cluster is grouping of a number of similar things into single. It is collection of data objects.

Keyword-- Fuzzy set theory, Geographic information system (GIS), ARM, Cluster.

1 INTRODUCTION

Crime data mining is receiving increased attention to discover underlying patterns in crime data. The need to act quickly and response to the crime activity happened and identify the interconnection between different crimes. GIS is used to identify the crimes happened in the cities which is older data mining techniques. Real-time solutions can save significant resources and push the capability of law enforcement closer to the pulse of criminal activity. Modern computing systems provide a unique opportunity to study this vast amount of data in ways that were previously not feasible. The volume of data being digitally recorded about crimes, suspicious activities, and suspect records is at an all-time high. More critical crimes should give first preferences to identified which may be data storage warehousing and privacy. Current manual inspection of crime data by analysts and only limited staff are presents to investigate the crime held, but it is insufficient since there are large number of data present so there should be a simultaneous action needed. Further, complex relationships between various crime attributes can be overlooked and it should be analysed by means of human interruption. Providing automated knowledge discovery tools becomes attractive to enhance and accelerate the efforts of local law enforcement. Similarly, certain crimes may be more probable in locations with higher populations and dense housing. Regional crime patterns can be discovered which enable law enforcement personnel and criminal investigators to address large-scale trends. Moreover, evidence can be loosely coupled while being geospatially sparse, forcing a more widespread analysis effort. Borrowing data mining techniques provide us better analyse, predict, prepare for, and respond to criminal acts and potential security risks. By which criminal victims will be reduced easily with reduced amount of time which was less than the investigation done by a human. In this paper, we study the application of fuzzy association rule mining [9] for community crime pattern discovery.

2 CLUSTER ANALYSIS

Cluster analysis, also known as classification, pattern recognition, numerical taxonomy, or morphometric, is used to

identify natural groupings of data set individuals or variables [2]. Three types of data clustering as follows:

- D-dimensional,
- Proximity matrix,
- Sorting data.

The size of the matrix is one of the major limiting factors. Pairs of items are compared for presence or absence of certain characteristics. Hierarchic methods can be further categorized as disagree with a majority. Divisive hierarchies are formed by grouping individual objects by similarity, forming subgroups. Algorithm is used to establish distances between analysed clusters and the rest of the clusters. There are three types of connection they are single linkage, complete linkage and average linkage. In single linkage, groups are merged with the nearest neighbour. Complete linkage functions the same as single linkage except that similarity between objects is calculated via the longest distance between members, resulting in compact clusters. Average linkage, in which the distance between two clusters is regarded as the average distance between all pairs of items where one member of a pair belongs to each cluster.

3 CRIME ANALYSIS

The central focus of crime analysis is the study of crime, chaos, and information related to the nature of incidents, criminals, and sufferers. Crime analysts also study other police-related operational issues, such as insufficient staffs. There are many different types of crime are in crime analysis. The three important kinds of crime analysis are socio-demographic, spatial and temporal.

First Goal: Normally after a robbery incident the police will start investigating the case based on the mode of crime. But in aspect of crime analyst this case is investigated based on the previous robbery case which is similarly happened.

Second Goal: Based on the crime identified by investigated in first goal it must be prevented by analyzing and responses of those problems.

There are many problem that cannot be solved because it is not criminal in nature. Since this is based on quality in life. Some of the examples are false burglar alarm, traffic control. The main goal of crime analysis is to evaluate the effort of police to assess the success of programs which are all implemented during the prevention of crime. Now measure the police teams are organized during the crime prevention. The main motto of crime analysis is to support the police by minimizing the crime happened.

3.1 Crime pattern

Two or more crime is discovered by police which is same then it is called crime pattern, because they meet each of the following conditions:

1. Similar type of crime; characters involved; location where it occurred.
2. Victims and Offenders have no relationship between them.
3. Crime occurred within the date range
4. Crime activity happened between the weeks or month.
5. Crime should be analysed and focused with the help of police strategies

3.2 Crime Pattern Types

There are seven common types of crime patterns. Therefore, when examining a crime pattern in which classification uncertainty occurs, the crime analyst should group the pattern based on the crimes nature and police responses to that crime occurred. The seven primary crime pattern types are: [7]

1. **Series:** Same victims are committing the some group of crime.
2. **Spree:** High frequency crimes occurred in short span of time continuously.
3. **Hot Prey:** Crime occurred by the two or more victims with similar physical properties.
4. **Hot Product:** Crime occurred by the two or more victim's unique property.
5. **Hot Spot:** Crime occurred by the two or more victim's within the locations nearer to each other.
6. **Hot Place:** Crime occurred by the two or more victim's at same location
7. **Hot Setting:** Crime occurred by the two or more victim's which basically related with the place where the crime are occurred.

4 FUZZY SET THEORY

For a discrete membership function

Here, n represents the number of element in the sample, x_i 's are the element, and $\mu(x_i)$ is its membership function.

Classical set theory supports the element in the form of binary terms that is a bivalent condition which may belongs or does not belongs to the set. Fuzzy set theory allows continuous assessment of the element in the given set which should be a real unit interval [0, 1]. Fuzzy set theory is the increment of classical set theory. Which is extended in the way how the membership element are treated. Fuzzy is used to describe the human reasoning. This reasoning is done thru the truth value which may be true or either be false. Fuzzy set allows different members for describing element then it is called membership function. Fuzzy logic is deduced for fuzzy set theory which deals with reasoning it is derived from classical predicate logic.

4.1 Fuzzy inference system

Fuzzy inference also referred to as approximate reasoning refers to computational procedures used for evaluating linguistic description. The two important inferring procedures are [1]

1. Generalized Modus Ponens (GMP)
2. Generalized Modus Tollens (GMT)

Defuzzification: The conversion of a fuzzy set into its crisp value is called Defuzzification and is the reverse process of Fuzzification.

There are many methods present in the literature some of them are.

- i. Centroid method
- ii. Centre of sums
- iii. Means of maxima

Centroid Method:

Centroid method is also known as centre of gravity. It is given by the expression

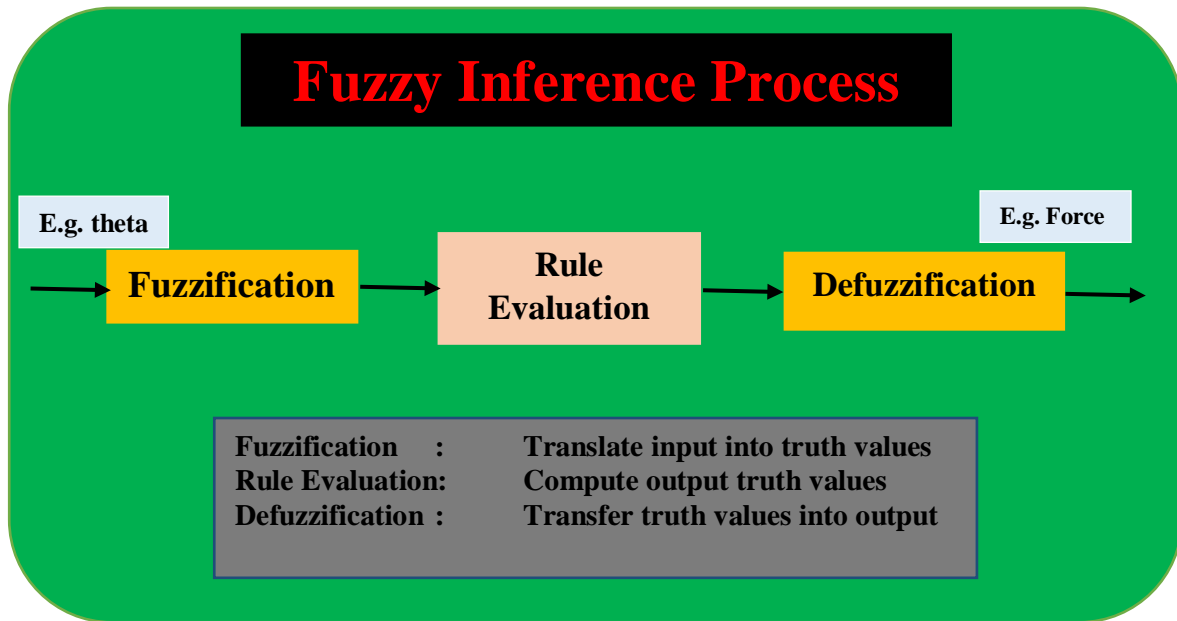
$$x^* = \frac{\int \mu(x) x dx}{\int \mu(x) dx}$$

For a continuous membership function, and

$$x^* = \frac{\sum_{i=1}^n x_i \mu(x_i)}{\sum_{i=1}^n \mu(x_i)}$$

Centre of sums (COS) method:

Centroid method has some overlapping areas so by using centre of sums the overlapping area is counted by number of times it is overlapped. The expression is given by.



The Communities and Crime Data Set, available from major crime trends - tamilnadu, was

$$x^* = \frac{\sum_{i=1}^n x_i \cdot \sum_{k=1}^n \mu_{\tilde{A}_k}(x_i)}{\sum_{i=1}^n \sum_{k=1}^n \mu_{\tilde{A}_k}(x_i)}$$

Mean of maxima (MOM) Defuzzification:

Defuzzification is used to get the output from that detect the crisp value which has the highest value of membership. The expression is given by.

$$x^* = \frac{\sum_{x_i \in M} x_i}{|M|}$$

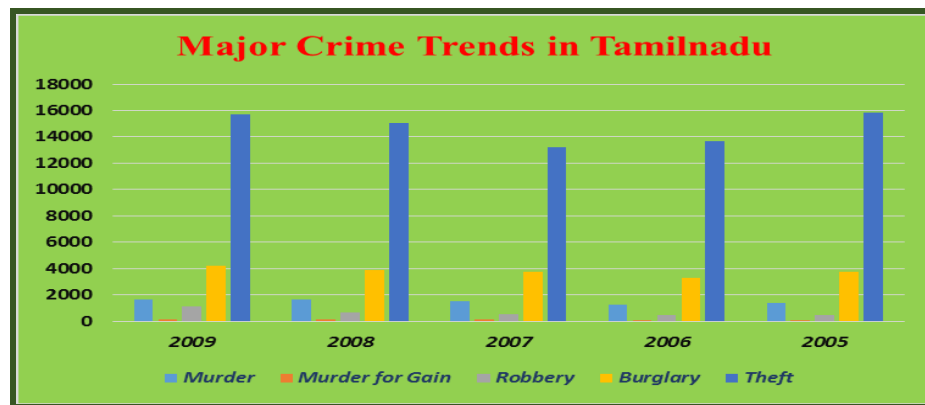
B.Data Set

Utilized for this study. This data set focuses on crime happened at tamilnadu. The data set is comprised of 103389 total instances and 5 attributes for communities the state. Some communities were omitted based on occurrence of significant missing or known incorrect crime statistics. Certain attributes contain a significant number of missing values for which the data was unavailable or not recorded for particular communities.

Attributes include information across a variety of crime-related facets, to population density and percent considered urban, to median household income. Also included are measures of crimes considered violent, which are murder, burglary, theft, murder for gain and robbery.

Major crime trends in Tamilnadu

Sl.no.	Crime head	2009	2008	2007	2006	2005
1	Murder	1644	1630	1521	1273	1365
2	Murder for Gain	123	105	102	89	74
3	Robbery	1144	662	495	450	437
4	Burglary	4221	3849	3717	3300	3738
5	Theft	15712	15019	13217	13651	15851
	TOTAL	22844	21265	19052	18763	21465
GRAND TOTAL		103389				



Average crime held in particular year

5 EXPERIMENTAL SETUP

The following mechanisms were used to select the final set of attributes: All attributes with a large number of missing values were removed. Odds ratios between each remaining attribute and Murders, Robberies, and theft were computed. This methodology yielded 5 attributes, enumerated below, that were utilized for this study. Those attributes whose value represents a percentage are indicated with (%).

1. Murder
2. Murder for gain
3. Robbery
4. Burglary
5. Theft

5.1 Data Pre-processing:

The majority of the 5 selected attributes are percentages, or are computed per 100K population. However, certain are absolute values. Our first attempt to generate rules with consequents such as Murders, Robberies, and theft gave us rules that simply pinpointed a state in India. In the remainder of this paper, when referring to the variables computed per 100k population, we will use their proper names. For example, we will refer to Murder instead of Murder per 100K Population.

5.2 Data Fuzzification:

A set of membership functions were defined for each of the 5 attributes. Membership functions were defined using subject matter expert knowledge and statistical measures such as mean and standard deviation. Three membership functions have been defined Low, Medium, and High. Four similar membership functions have been defined for

Robberies and Murder. Murders has the following membership functions defined: None, Low, Medium, and High.

6 METHODOLOGY

The developed methodology has the following primary steps:

1. Variable Fuzzification. This includes defining membership functions for each of the variables, and computing the membership values for each data item.
2. Running the Fuzzy Apriori algorithm on the data set. This includes initial pruning of the generated rules.
3. Rule post-pruning.

We have developed a similar method that we call antecedent-constraint rule pruning, in which an item constraint is used that requires the antecedents of the rules to satisfy a given constraint. This is the second technique used in our work. This technique requires prior knowledge of which items are of interest in the antecedent. In an application such as the crime domain, the user usually knows very well which attributes are of interest as antecedents or consequents. Rule post-pruning is concerned with pruning rules after they have been generated by an algorithm, such as Fuzzy Apriori. We are post-pruning rules based on 60% fuzzy confidence. We have also developed a new Relative Fuzzy Support measure. The reduction or increase of support is significant because of the square in the denominator. The Relative Fuzzy Support is well suited for applications in which the user knows the consequents of interest. This is the case in this crime application, as the user is most interested in Murders, Robberies and theft being high.

7 CONCLUSION

Fuzzy association rule mining has proven useful for this crime application, and has utility for other crime-related data sets. Results presented in this paper suggest that further analysis is required to gain a closer understanding of crime at both the community and national levels. Crime patterns were discovered which are consistent across a regions, subsets of regions, and single state. The attributes of interest were computed to measure their occurrence per 100K population, so as to remove the element of community and state size during the rule generation process. Rules discovered as part of this study therefore offer utility for use from the national level down to the state and community level. A novel relative support measure was proposed to prune the set of rules and to extract rare rules from the larger original set. The use of relative support achieves a 95.2% reduction in the final number of rules. The data set used in this study has resolution down to the community (town) level. Several attributes of the dataset has insufficient values acquiring accurate data for these attributes will enable the process to produce rules which relate directly to the police force. Utilizing data that contains precise locations of crimes or block-level demographics would help produce more meaningful rules for local law enforcement jurisdictions. These higher resolution data could then result in rules that are applicable to certain areas of a city, especially those with widespread crime of various types.

REFERENCES

- [1] "Neural Networks, Fuzzy Logic, and Genetic Algorithms Synthesis and Applications" by S. Rajasekaran, G.A. Vijayalakshmi Pai (2011), PHI Learning Private Limited.
- [2] "Crime Analysis Geographic Information Systems Services: Advanced Tools Report" by Ezra B. Zubrow, Ph.D., Philip C. Mitchell, M.S.Monika Bolino, M.A. Completed for environmental Systems Research institute (ESRI) 1999
- [3] R.J. Bayardo, R. Agrawal, and D. Gunopulos, "Constraint-Based Rule Mining in Large, Dense Databases," *Data Mining and Knowledge Discovery*, 4(2/3), pp. 217-240, 2000.
- [4] D. Brown, "The Regional Crime Analysis Program (RECAP): A Framework for Mining Data to Catch Criminals," In *Proceedings of the International Conference on Systems, Man, and Cybernetics*, pp. 2848-2853, 1998.
- [5] J. de Bruin, T. Cocx, W. Kusters, J. Laros, and J. Kok, "Data Mining Approaches to Criminal Career Analysis," In *Proceedings of the International Conference on Data Mining*, pp. 171-177, 2006, Washington, D.C., IEEE Computer Society Press.
- [6] R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rules between Sets of Items in Large Databases," In *Proceedings of the International Conference on Management of Data*, Washington, D.C., pp. 207-216, May 1993.
- [7] International Association of Crime Analysts (IACA) "Crime Pattern Definitions for Tactical Analysis", Standards, Methods, & Technology (SMT) Committee White Paper 2011
- [8] "Crime in India 2011" by National Crime Records Bureau Ministry of Home Affairs Government of India June 2012.
- [9] Dr. T. Karthikeyan, J. Mohana Sundaram, R. Karthik Raj. (2013). "A Recent Approach on Identifying the ARM Based Clustering Algorithm on Text and Image Context" *IJCS International Journal of Computer Science*, Volume 1, Issue 3, August 2013.

IJSER