International Journal of Engineering & Technology, 7 (4.5) (2018) 713-715

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



An enhanced rule mining algorithm to detect suspects of crime against women in the state of Tamil Nadu

D. Usha [1] *, D. Chitradevi [2]

[1] *Associate Professor*
[2] *Assistant Professor Dr. M. G. R. Educational and Research Institute, Hindustan Institute of Technology & Science*
**Corresponding author E-mail: ushahits@gmail.com*

Abstract

Crime against women in India has become a prominent topic of argument in the recent years and the issue has been brought to the foreground for concern due to the increasing trend in crimes performed against women. It is the major challenge to the investigators to detect and prevent crimes, particularly crime against women. Most of the crimes get reported and a massive dataset is being generated every year. Analyzing the crime reports can help the law enforcement officers to take preventive measures for reducing the crime, but processing this voluminous data is backbreaking and error prone. So, the application of various data mining techniques can help in visualizing the crime trend. Crime is one of the interesting applications where data mining plays an important role in terms of prediction and analysis in the interest of society. This paper covers in detail analysis of modus operandi of committing crimes and effective use of data mining techniques and algorithms in narrow down to identify the criminals at a short span of time.

Keywords: *Crime Pattern; Data mining; Rule Mining Algorithm; Modus Operandi.*

1. Introduction

India is a vast country with diversified societies. Position of women has been of great importance since ancient times in Indian culture. Unfortunately the current scenario depicts a different story. According to National Crime Records Bureau, crime against women has significantly increased in recent years. It has become the greatest challenge to the law enforcement agencies to enforce law & order to reduce this increasing rate of the crime against women. This is where criminology comes into the picture. Criminology [5] is the scientific study of crime and criminal behavior in order to detect crime characteristics. Use of data mining techniques can produce important results from crime dataset. Crime analysis [4] is exploring, interrelating and detecting relationships between various crimes and characteristics of crimes. Police department maintains crime data at the record. This data contains a huge amount of data set with complex relationships which needs use of data mining techniques [7] in order to be transformed into useful information. The knowledge extracted from the dataset can be a great tool & support to the police department to prevent crimes. Nowadays crime against women has become the great problem of every nation around the world. Most of the countries are giving top priority to this major issue. The crime data are usually kept confidential, whereas as data will be available for access. For example, data about crimes related to juvenile cases is usually more restricted because of its sensitive nature. However, the crime data related to sex offenders is made public to warn others in the area and to protect people themselves. Thus, the analyst has to deal with both public and private issues so that analysis-modeling process does not infringe on these legal boundaries. Thus crime analysis requires data mining techniques to extract abstract data hidden from large volumes of raw data.

2. Challenges encountered

The major challenge encountered in this paper is, in case of large amount of crime information to be stored and analyzed, it is very difficult to do with the existing available techniques. The records maintained with relates to crime data are of different methods and structures, so it is difficult to store and analyze the data. Further the data available is inconsistent and are incomplete, thus making the job of formal analysis is far more difficult. The above criteria results in investigation of crime take longer duration due to complexity of issues [13]. All the above challenges moti- vated this paper to focus on providing solutions that can enhance the process of crime analysis for identifying and reducing crime against women in the state of Tamil Nadu.

3. Problem statement

In olden days women occupied a high place of respect in the society. Many books were written about the status of women and their heroic deeds from Vedic period to the modern times. But later on, due to the cause of social, political and economic changes, women lost their status. Lots and lots of evil customs and traditions stepped in, which confined women and tied them to the boundaries of the house. The statistics revealed the declining rate of sex ratio, health status, literacy rate, work participation rate and political participation among women. On the other the spread of social evils like dowry deaths, child marriage, domestic violence, rape, sexual harassment, exploitation of women workers are rampant in different parts of India. Humiliation, rape, kidnapping, molestation, dowry death, torture, wife-beating etc. have grown up over the years.

Copyright © 2018 D. Usha, D. Chitradevi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

3.1. Victims affected

Crime against women [11] has been a crucial issue in today's Indian society. Women in the state of Tamil Nadu constitute near about half of its population and most of them are crushing under the socio-cultural and religious structures. Only in words women are treated equally to men, but when seeing the number of crimes happening against women is no par with equality. Women are the victims [15] in both inside and outside the house. The old age women are targeted by the suspects when they stay alone in their house or also when they come out of the house. The women cannot travel alone and cannot walk alone in public places all the time.

4. Data collection

The enormous amount of crime data is collected at the end of year at police records [13]. This data is made available in the National Crime Bureau of Records and State Crime Record Bureau [14]. This data is in the form of number of cases recorded all over the nation throughout the year. The data is in bulk number and it is very difficult to handle. So, with the help of a pattern mining algorithm, the required data can be mined and that particular data can be given as an input to the next process.

5. Modus operandi of committing crimes

Modus Operandi [8] is a term used by law enforcement authorities to describe the particular manner or style in which a crime is committed. This term is commonly used in criminal cases. It is also referred to by its initials M.O. Modus Operandi of the suspects will always helpful to the prosecution to find the suspect. The crimes need not be identical, but the prosecution can make a strong and convincing showing of similarity between the crime charged and the other crimes. The prosecution may introduce evidence from previous or subsequent crimes to prove modus operandi only if the other crimes share peculiar and distinctive features with the crime charged [10]. The features must be unusu- al and rarely visible in other crimes, but they can be recognized by their modus operandi.

Table 5.1: Different Modus Operandi

Weapons Used	Knife	Riffle	Sharp weapons
Time of com- mitting crime	Morning	Afternoon	Night
Place of com- mitting crime	During Journey	Commercial Areas and Public places	Lonely House
Age of victims	Young	Middle	Old

6. Crime analysis

The table 6.1 shows the National Status review of our country [6]. The crime rate in Tamil Nadu in 2016 was down when compared to the corresponding figures of the previous two years, according to statistics tabled by the state government on Thursday in the assembly. The number of rape cases decreased from 471 in 2014 to 450 in 2015 and further came down to 336 in 2016. The number of dowry deaths too came down from 95 in 2014 to 65 in 2015 while only 58 such cases were reported in 2016. The crimes against woman saw a declining trend in Tamil Nadu during the last three years. In 2014, a total of 3,771 crimes was committed against women [12]. The number marginally came down to 3,578 next year and last year it came down significantly to 2,502 cases. The total number of crimes, which stood at 1,93,197 in 2014 fell to 1,87,558 in 2015 and it further fell to 1,79,893 in 2016. The crime rate also came down from 255 per 1,00,000 population to 244 to 230 respectively at a time when the population steadily went up from 7.55 crore through 7.67 crore to 7.79 crore.

Table 6.1: National Status Review

CRIME	2009	2010	2011	2012	2013
Rape	469	507	572	706	1330
Molestation	552	601	657	727	2884
Eve-teasing	238	126	165	236	793
Abduction of women	1655	1740	2085	2210	2906
Cruelty by hus- band, in-laws	1297	1410	1585	2046	2487
Dowry death	141	143	142	134	123

7. Detecting crime against women using data mining data mining techniques

Data mining techniques such as classification, clustering, outlier, association, prediction, etc. can be used for crime detection [1] and prediction. Each technique will have its own strength. Classi- fication is used to classify the crime, according to their type. Clustering is used for grouping of similar attributes. The offence committed in the same manner can be put in one group. Offences not in a similar manner are categorized as outliers. The remaining data can be associated with various combinations of attributes so that the investigating team can predict the suspects.

Data Mining algorithms

Various data mining algorithms can be used for crime detection [8] and prediction. Data mining [9] is a catchy term which can be applied to many applications. One of the well-known applica- tions is crime. Frequent pattern mining algorithm [2] can be used to detect the crimes in short span of time because of its nature. FPM algorithm identifies the frequently identified patterns in vari- ous possible combinations of attributes (clustering technique) and with the help of rule generation these of patterns are associated and finally frequently occurred patterns can be find out which helps to narrow down the searching process. Also, rather than finding the suspects from the large group, it is easy to find out in smaller groups which arrived from frequent pattern mining al- grithm..

Enhanced Rule Mining Algorithm

Association rule mining is the process of generating association rules from those large items set with the constraints of minimal confidence. The main idea behind proposing the above algorithm is to generate an association among the mined item set. The enhanced rule mining algorithm generates rules from frequent patterns which are mined from Tamil Nadu Crime dataset relating to crime against women. The various combinations of the rules which is produced by the rule mining algorithm may be efficient or inefficient. It is a waste of time to run all the rules. So, in order to validate the most efficient rule, the proposed algorithm applied the measures support, confidence measures and additionally one more measure information gain to add more values to the rule generation and validation. The frequent item set are generated and it is given as input to the Improved Rule Mining algorithm. Final-ly, it is applied to the Tamil Nadu Crime dataset relating to crime against women.

Table 7.1: Number of Rules Mined from Crime Dataset Using Cbarm and Erma

Information Gain %									
Algorithm	10	20	30	40	50	60	70	80	90
CBARM	450	420	380	365	320	270	265	250	220
ERMA	420	390	350	320	280	255	235	200	180

The Enhanced rule mining algorithm (ERMA) increases the efficiency through the process of reducing the computational time as well as cost. It can be succeeded by reducing the number of passes over the database, by adding additional constraints on the pattern. In legal applications, some rules will have less weightage and inefficient and some rules will have more weightage. Genera- tion of the entire constructed rule will lead to waste of time. So, researcher likes to validate the rule to find the most efficient rule. Association rules have two basic measures support (Supp) and confidence (Conf). Support is defined as the percentage or fraction of records that contain XUY to the total number of records in

the database. Confidence is defined as the percentage of the number of transactions that contain XUY to the total number of records that contain X. Confidence is the measure of strength of the association rule. The additional measure Information Gain is a statistical property that measures the validity of the generated rules. The Information Gain calculates lift value and with the input of the lift value it calculates information gain by taking all the possible combination of rules that is generated from the pattern mining algorithm.

The drawback of the existing rule mining algorithm is that the number of rules can be extremely large. So, it takes more time to execute the process. The proposed rule mining algorithm minimize the number of rules mined but proves the efficiency because of its limited and exactly predictable space overhead and is faster than other existing methods.

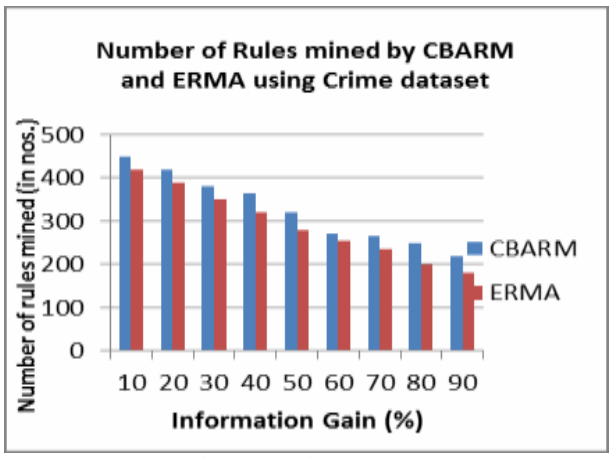


Fig. 7.1: Comparison of Number of Rules Mined by Cbarm and Erma Using Crime Dataset.

It also increases the efficiency through the process of reducing the computational time.

The above figure shows the comparison of number of rules mined by existing CBARM and the proposed ERMA with the help of crime dataset, which is collected from various sources. The ER-

Table 7.2: Sample of dataset collected

Accused Fin- ger print slip	Offence count	Worth of Val- uables (Rs)	Age of wom- en victims	Marks & Ex- hibits	Data of of- fence
FSN5 671	1	10000	40	Foot print, Finger print, tool marks	25/09/14
FSN5 671	2	25000	36	Paint marks, Cloth, Hair	10/10/15
FSN5 680	1	51000	30	Foot print, Finger print, Hair	03/01/15
FSN5 680	2	63000	39	Foot print, Finger print, Hair	12/05/16
FSN5 683	1	110000	54	Article cut marks, Shoe marks	01/09/09
FSN5 683	2	120000	60	Article cut marks, Hair	09/11/10
FSN5 885	1	30000	35	Nail clippings, Cloth, Hair	12/12/12
FSN5 885	2	35000	38	Nail clippings, Cloth, Hair	03/04/14

MA proves better than CBARM in varying measure Information Gain. When the Gain percentage is between 10 to 40, the difference is less. Whereas when it is above 40 it shows that ERMA mines a less number of rules when compared to CBARM.

The above table shows the sample dataset collected from various sources. It consist of Finger Print Number to identify the suspects, Offence count, Worth of valuables, and all other attributes in the dataset.

Sample mined rules

Some of the sample interesting rules are:

R1: instance1=2 instance3=5 instance5=4 ⇒instance7=4 [Supp=23.5%, Conf=52%, Gain=50%]

R2: nstance1=2 instance3=5⇒instance5=4 instance7=4 [Supp=35%, Conf=50%, Gain=50%]

R3:instance1=2 instance3=5 instance5=4 ⇒instance7=4 [Supp=25%, Conf=50%, Gain=50%]

R4: instance4=6 instance3=5instance5=7⇒ instance8=10 ⇒ instance 3=6 [Supp=35%, Conf=50%, Gain=50%]

R5:instance5=7 instance4=6instance3=5 instance 3=6 ⇒ instance 8=10 [Supp=25%, Conf=50%, Gain=40%]

8. Conclusion

This paper concludes that the Enhanced Rule Mining Algorithm has applied the measures support, confidence and information gain thresholds that are efficiently calculated in the dataset which collected from various sources. The Enhanced Rule mining algorithm adapted the input from pattern mining algorithm and found the results from frequent patterns rather than from bulk number of dataset. Experimental results have shown that the Enhanced rule mining algorithm can always find most of the possible association rules with satisfying minimum thresholds. It performed well in high and small density datasets. So, the suspects in crime against women cases can be identified at the short span of time and also alert can be given to the victims to avoid future crimes.

References

[1] Phua, Clifton. 2012. "Resilient identity crime detection", Knowledge and Data Engineering, IEEE Transactions, Vol.24, Issue.3, pp.533-546.

[2] O. Jamsheela; Raju G. 2015. "Frequent itemset mining algorithms: A literature survey", IEEE International Advance Computing Conference (IACC) pp. 1099 – 1104.

[3] Raymond B. Fosdick. 1916. "Modus operandi system in the detection of criminals", Journal of Criminal Law and Criminology, Vol.6, Issue.4, pp.560-570.

[4] Wortley, R., & Mazerolle, L. 2008. "Environmental criminology and crime analysis", Willam Publishing.

[5] Krishnamurthi's a handbook of criminal law", Revised by Justice G.Ramanujam (Retd.) Madras High Court, 2009.

[6] National review of status, source from: affidavit submitted by Delhi Police in Court.

[7] Revathy Krishnamurthy, J.Satheesh Kumar. 2012 Survey of data mining techniques on crime data analysis", International Journal of Data Mining Techniques and Applications, Vol.1, Issue.2, pp.117-120.

[8] Aniruddha, Lalit Dole. 2014. "A review on data mining methods for identity crime detection", International Journal of Electrical, Electronics and Computer Systems, Vol.2, Issue.1, pp. 51-55.

[9] P.Dhakshinamoorthy, T.Kalaiselvan. 2013. "Crime pattern detection using data mining" International journal of advanced research in computer science and applications, Vol.1, Issue 1. Pp.46-50.

[10] Tushar Sonawanev. 2015. "Crime pattern analysis, Visualization and prediction using Data mining", IJARIIE, Vol.1, Issue.4.

[11] Aarti Bansal, 2015. "Performance comparison of data mining techniques to analyse crime against women", International Journal of Scientific Research and Education, Vol.3, Issue.9, pp.4494-4512.

[12] Rachel Boba Santos, "Crime analysis with crime mapping", 3rdedition.

[13] Gupta. 2008. "Crime data mining for Indian Police Information System", Proceeding of the 2008 Computer Society of India.

[14] Stages of Crime, [online] Available at http://www.lawnotes.in/Criminology.

[15] Crime victims [online] available at www.countercurrents.org.