

A Complete Survey on application of Frequent Pattern Mining and Association Rule Mining on Crime Pattern Mining



D.Usha, Dr.K.Rameshkumar

Assistant Professor, Associate Professor

Hindustan University, Chennai

dusha@hindustanuniv.ac.in, krkumar@hindustanuniv.ac.in

ABSTRACT

Frequent pattern (itemset) mining plays an important role in association rule mining. The Apriori & FP-growth algorithms are the most famous algorithms which can be used for Frequent Pattern mining. This paper presents the survey of various Frequent Pattern Mining and Rule Mining algorithm which can be applied to crime pattern mining. The analysis of literature survey would give the information about what has been done previously in the same area, what is the current trend and what are the other related areas. This paper explains the concepts of Frequent Pattern Mining and three important approaches that is candidate generation approach, without candidate generation and vertical layout approach. It also explains various frequent pattern algorithms and how it can be applied to different areas particularly in crime pattern detection. This paper surely helps the researches to get clear idea about the application of frequent pattern mining algorithm in various areas.

Keywords : Frequent Pattern Mining, Apriori, FP-growth, Association Rule Mining, Crime Pattern mining.

1. INTRODUCTION

Frequent pattern mining [1] plays a major field in research since it is a part of data mining. Many research papers, articles are published in the field of Frequent Pattern Mining (FPM). This chapter details about frequent pattern mining algorithm, types and extensions of frequent pattern mining, association rule mining algorithm, rule generation, suitable measures for rule generation. This chapter describes about various existing FPM algorithms, data mining algorithm for crime pattern. By applying frequent pattern mining algorithm and suitable measures, the proposed new algorithm is applied to crime dataset in order to find out the suspects in the short span of time.

Frequent pattern mining is fundamental in data mining. The goal is to compute on huge data efficiently. Finding frequent patterns plays a fundamental role in association rule mining, classification, clustering, and other data mining tasks. Frequent pattern mining was first proposed by Agarwal et. al. [1] for market basket analysis in the form of association rule mining. Frequent Itemset mining came into existence where it is needed to discover useful patterns in customer's transaction database. A customer's transaction database is a sequence of transactions ($T=t_1...t_n$), where each transaction is an itemset ($t_i \subseteq I$). An itemset with k elements is called a k -itemset. An itemset is frequent if its support is greater than a support threshold, denoted by $\min \text{supp}$. The frequent itemset problem is to find all frequent itemset in a given transaction database. The first and most important solution for finding frequent itemsets, is the Apriori algorithm. The fundamental frequent pattern algorithms are classified into three ways as follows:

1. Candidate generation approach (E.g. Apriori algorithm)
2. Without candidate generation approach (E.g. FP-growth algorithm)
3. Vertical layout approach (E.g. Eclat algorithm)

1.1 Candidate Generation Approach

1.1.1 Apriori Algorithm

Agrawal and Srikant (1994) developed the classical Apriori algorithm. This algorithm relies on generate and test approach and an important property: the Apriori property. This property is also known as anti-monotone property, and it is a basic pillar of the Apriori algorithm. It states that all non-empty subsets of a frequent itemset must be frequent. For example, if itemset 1,2,3 is a frequent itemset, then all of its subsets 1,2,3,1-2,2-3 and 1-3 must be frequent. In the other view, if an itemset is not frequent, then none of its supersets can be frequent. As a result, the list of potential frequent itemsets eventually gets smaller as mining progresses.

1.2 Without Candidate Generation Approach

1.2.1 FP Growth [2]

In order to count the supports of all generated itemsets, FP-growth uses a combination of the vertical and horizontal database layout to store the database in main memory. Instead of storing the cover for every item the database, it stores the actual transactions from the database in a tree structure and every item has a linked list going through all transactions that contain that item. This new data structure is denoted by FP-tree (Frequent-Pattern tree). FP-growth is a key frequent itemset mining algorithm, which is based on the pattern growth paradigm. It adopts a prefix tree structure, FP-tree, to represent the database (or conditional databases).

1.3 Vertical Layout Approach

1.3.1 Eclat Algorithm

The first algorithm developed to generate all frequent itemsets in a depth-first manner is the Eclat (Equivalence CLAss Transformation) algorithm. If the database is stored in the vertical layout, the counting of support can be much easier by simply intersecting the covers of two of its subsets that together give the set itself. The Eclat algorithm essentially used this technique inside the Apriori algorithm. Always this is not possible since the total size of all covers at a certain iteration of the local set generation procedure could exceed 28 main memory limits. It is usually more efficient to first find the frequent items and frequent 2-sets separately and use the Eclat algorithm only for all larger sets.

2. ASSOCIATION RULE CONSTRUCTION

Association rules are created by analyzing data for frequent if/then patterns and using the criteria support and confidence to identify the most important relationships. Support is an indication of how frequently the items appear in the database. Confidence indicates the number of times the if/then statements have been found to be true. In data mining, association rules are useful for analyzing and predicting customer behavior. They play an important part in shopping basket data analysis, product clustering, catalog design and store layout.

Association rules are if/then statements that help uncover relationships between seemingly unrelated data in a relational database or other information repository. An example of an association rule would be "If a customer buys a dozen eggs, he is 80% likely to also purchase milk." An association rule has two parts, an antecedent (if) and a consequent (then). An antecedent is an item found in

the data. A consequent is an item that is found in combination with the antecedent.

2.1 Types of Association Rule Mining

1. Positive Association Rule
2. Negative Association Rule
3. Constraint Based Association Rule

2.2 Measures

2.2.1 Categories of Measures

1. Objective Measures
 - a. Based on probability (generality and reliability)
 - b. Based on the form of the rules
 - Peculiarity
 - Surprisingness
 - Conciseness
 - Nonredundant rules
 - Minimum description length
2. Subjective Measures
 - a. Surprisingness
 - b. Novelty
3. Semantic Measures
 - a. Utility
 - b. Actionability

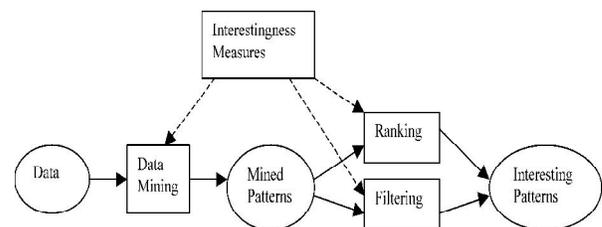


Figure. 1 Roles of interestingness measures in the data mining process

2.2.2 Role of Measures

The Figure.1 shows the roles of interestingness measures in the data mining process. The measures are used in the following three ways:

1. They are helped to classify each pattern as either interesting or uninteresting.
2. The measures are used to determine one pattern is more interesting than another.
3. Also the measure helps to rank the interesting or useful patterns.

3. APPLICATION OF FPM & ARM

Frequent Pattern Mining applications have since been applied to many different domains including market basket and risk analysis in

commercial environments, epidemiology, clinical medicine, fluid dynamics, astrophysics, crime prevention, and counter-terrorism - all areas in which the relationship between objects can provide useful knowledge.

3.1 Crime pattern functions

Crime analysis can occur at various levels, including tactical, operational, and strategic. Crime analysts study crime reports, arrests reports, and police calls for service to identify emerging patterns, series, and trends as quickly as possible. They analyze these phenomena for all relevant factors, sometimes predict or forecast future occurrences, and issue bulletins, reports, and alerts to their agencies. They then work with their police agencies to develop effective strategies and tactics to address crime and disorder. Other duties of crime analysts may include preparing statistics, data queries, or maps on demand, preparing information for community or court presentations, answering questions from the public and the press, and providing data and information support for a police department's Complete process.

Socio demographics, along with spatial and temporal information, are all aspects that crime analysts look at to understand what's going on in their jurisdiction. Crime analysis employs data mining, crime mapping, statistics, research methods, desktop publishing, charting, presentation skills, critical thinking, and a solid understanding of criminal behavior. In this sense, a crime analyst serves as a combination of an information systems specialist, a statistician, a researcher, a criminologist, a journalist, and a planner for a local police department.

3.2 Crime prevention theory

Crime Pattern Theory is a way of explaining why crimes are committed in certain areas. Crime is not random, it is either planned or opportunistic. According to the theory crime happens when the activity space of a victim or target intersects with the activity space of an offender. A person's activity space consists of locations in everyday life, for example home, work, school, shopping areas, entertainment areas etc. These personal locations are also called nodes. The course or route a person takes to and from these nodes are called personal paths. Personal paths connect with various nodes creating a perimeter. This perimeter is a person's awareness space.

Crime Pattern Theory claims that a crime involving an offender and a victim or target can only occur when the activity spaces of both cross paths. Simply put crime will occur if an area provides opportunity for crime and it exists within an offender's awareness space. Consequently an area that provides shopping, recreation and restaurants such as a shopping mall has a higher rate of crime. This is largely due to the high amount of potential victims and offenders visiting the area and the various targets in the area. It is also probable that people may fall victim of purse snatching or pick pocketing because victims typically carry cash with them. Therefore crime pattern theory provides analysts an organized way to explore patterns of behavior.

4. IN LEGAL FIELD

1. Shyam Varan Nath [3] proposed an idea to solve crime detection problems using Data mining. Crimes are a social nuisance and cost our society dearly in several ways. Author look at the use of clustering algorithm (k-means clustering) to detect the crimes patterns and speed up the process of solving crimes. This clustering technique are applied to real crime data to validate the results. Author also used semi-supervised learning technique for knowledge discovery from the crime records to increase the predictive accuracy.
2. Yagnik Ankur N., Dr .Ajay Shanker Singh [4] are presented a new method to detect outliers by discovering frequent pattern from the data set. The outliers are defined as the data transactions that contain less frequent pattern in their item sets. It can be applied to law enforcement to discover unusual patterns from multiple actions of an criminal entity, especially fraud committed in financial transactions, trading activity or insurance claims.
3. Anna L. Buczak , Christopher M. Gifford [5] are applied fuzzy association rule mining for community crime pattern discovery. Discovered rules are presented and considered at regional and national levels. To extract rare and novel rules from thousands of discovered rules, support metric was defined. It helps the law enforcement personnel to find out interesting and meaningful crime patterns rather than wasting time in uninteresting and rare (rules) patterns.

4. A.Malathi, Dr. S. Santhosh Baboo [6] are used missing values and clustering algorithm for crime data. They also concentrate on MV algorithm and Apriori algorithm with some enhancements in the process of filling missing value and to identify crime patterns. They applied the above techniques to real crime data collected from city police department. They also used semi-supervised learning technique for knowledge discovery from the crime records in order to increase the accurate prediction.
5. Aniruddha Kshirsagar, Lalit Dole [7] are used data mining methods to identify crime detection in both transaction domain and in application domain. Outlier detection is used which deviates from other observations. The authors give the summary on how to find out and overcome fraud identity by using data mining techniques which helps in both transaction and credit card application domain in financial field.
6. Manish Gupta, B. Chandra and M. P. Gupta [8] proposed a crime analysis tool for CCIS (Crime and Criminal Information System) to help the police to respond and serve the people better. This tool provides a user-friendly tool to analyze huge crime database as building up in CCIS. This tool also eliminates the rescanning of the database for every new query and well the user's needs. It provides the user an interactive and fastest way to carry out process of identification of crime hot spots and crime zones. The crime analysis tool as an adaptive query interface has been designed to make best use of the existing CCIS. Police station level analysis is also possible through the interface.
7. Revatthy Krishnamurthy, J. Satheesh Kumar [9] present a detailed study on clustering techniques and its role on crime applications. Authors present clustering one of the data mining techniques to combine data objects into groups. The data objects within the group are very similar and very dissimilar as well when compared to objects to other groups. Authors whole idea is this study will surely help the crime branch for better prediction and classification of crimes. It also helps the crime analysts and law enforcers to precede the case in the investigation and help solving unsolved crimes faster. Authors also present partition clustering algorithm one of the best methods for finding similarity measures.
8. Divya Bansal, Lekha Bhambhu [10] explain the use of association rule mining in extracting patterns that occur frequently within a dataset. Association algorithm can be applied to the conversion of quantitative data into qualitative data. Author considers two Association Rule algorithms namely Apriori algorithm and Predictive Apriori algorithm. They implement Apriori algorithm in mining association rules from dataset of crime against women collected from session court and compare the result of both the algorithms using data mining tool called WEKA.
9. Devesh Bajpai [11] made about the various contemporary and evolving techniques in data mining extensively used in criminal analysis and investigation. The data mining techniques are applied to the range of functions in a crime investigation organization which can assist in the areas of decision support, prediction, resource handling, forecasting and estimation. Author also discusses about the current techniques like Entity Extraction, Social Network Analysis, String Comparator, etc which can aid the study and analysis of crime networks and thereby extract useful information.
10. Jitendra Kumar, Sripati Mishra, Neeraj Tiwari [12] presents study aims to identify the area suffering major crimes named hotspot and the area with fewer crimes named safe zone with respect to different heads of crime against body. The data are collected by State Crime Record Bureau, Uttar Pradesh, and taken for study, using the cluster analysis.
11. John David Elijah Sandig, Ruby Mae Somoba, Ma. Beth Concepcion and Bobby D. Gerardo [13] proposed to develop a mechanism in mining an Online Graphical Information System (GIS) for crime rate and models using frequent pattern analysis. It is a web-based system that includes GIS for robbery, homicide and physical injury incidents within a (Iloilo) City. This system can help police identify where and what time crime frequently happen. It generates who are the frequent victims in the community. The main focus of the proposed system is the analysis of frequent crime patterns and its association rules and crime hot spots. Finally result shows the hot spots which displays clustered crimes in the Google Map. It uses the Google Heat Map layer in clustering plotted crimes. These hot spots are showing the

- density of crime. Moreover, the system is able to generate frequent patterns of crimes. This uses Apriori algorithm in generating the results.
12. Dr. M. VijayaKumar 1, Dr. S. Karthick 2 and N.Prakash [14] proposed GIS system to grasp spatial and temporal patterns of crime offenses. GIS makes possible to reduce the crime and extend the security of residents. It offers a vital method that is identification of crime hotspots with a high crime rate. Spatial-Temporal crime analysis plays a fundamental role in lots of security related crime forecasting applications. Authors carried out the research by efficiently connecting spatial and temporal crime factors for repeated events in particular high density crime area. The main objective of the paper is to explore a combined approach of the spatial and temporal patterns of particular crimes and forecasting the results through simulation. Input has given from Tamil Nadu cases which is gathered from Chennai City Police from January 2008 to December 2008.
 13. Malathi. A, Lt. Dr. S. Santhosh Baboo [15] concerned that national security has been increased after Mumbai Taj Hotel attack on 26.11.2008. This was happened as a criminal and terrorist activities. So, to analyze and solve such kind of issues author applied data mining in the context of law enforcement and intelligence analysis. Author used clustering/classification based model to anticipate crime trends. The paper also carry various data mining approaches and techniques which can be applied for crime pattern and various analysis of city crime such as property crime, violent crime, crime against women and other crimes. The results of this data mining could potentially be used to lessen and even prevent crime for the forth coming years.
 14. Li Ding, Dana Steil, Matthew Hudnall, Brandon Dixon, Randy Smith, David Brown, Allen Parrish [16] proposed an integrated system called Perp-Search that will gives the important description of suspects as input. To detect suspects, the system will process these inputs through four integrated components: geographic profiling, social network analysis, crime patterns, and physical matching. Essentially, geographic profiling determines “where” the suspects are, while other components determine “who” the suspects are. Finally the results can be processed using a score engine to give investigators a ranked list of individuals.
 15. Jian Pei, Jiawei Han, Hongjun Lu, Shojiro Nishio, Shiwei Tang, Dongqing Yang [17] proposed a simple and novel hyperlinked data structure, H-struct which overcomes some performance bottlenecks of the existing algorithm when mining databases with different data characteristics, such as dense vs. sparse, long vs. short, memory-based vs. disk-based, etc. The new mining algorithm, H-mine, works in a very limited and precisely predictable space overhead and runs really fast in memory-based setting. So, as a result the study shows that H-mine has high performance in various kinds of data, outperforms the previously developed algorithms in different settings, and is highly scalable in mining large databases.
 16. Charu C. Aggarwal, et.al. [18] explained the problem of frequent pattern mining with uncertain data. Particularly author studies about candidate generate and test algorithms, hyper-structure algorithms and pattern growth based algorithms. By testing the real and synthetic data sets, it shows that in the deterministic case, the FP-growth algorithm is well known to be an extremely efficient approach. However, after testing, author found that the extensions of the candidate generate-and-test as well as the hyper-structure based algorithms are much more effective.
 17. Goswami D.N., Chaturvedi Anshu., Raghuvanshi C.S. [19] proposed three different frequent pattern mining approaches that is Record filter, Intersection and Proposed algorithm which is based on classical Apriori algorithm. Author made comparative study by keeping dataset of 2000 transactions and as a result Record filter proved better than classical Apriori algorithm, Intersection approach proved better than Record filter approach and finally proposed algorithm proved that it is much better than other frequent pattern mining algorithm.
 18. M Suman , T Anuradha , K Gowtham, A Ramakrishna [20] proposed Apriori-Growth algorithm which is based on Apriori algorithm and FP-Growth algorithm. The advantage of the Apriori-Growth algorithm is that it doesn't need to generate conditional pattern bases and sub-conditional pattern tree recursively. The proposed Apriori Growth algorithm overcome the disadvantages of Apriori algorithm and efficiently mine association rules without generating candidate itemsets, and also the disadvantage of FP-Growth i.e. consumes more

memory and performs badly with long pattern data sets.

19. S.Yamuna, N.Sudha Bhuvaneshwari [21] proposed data mining techniques to analyze and predict the future crime. The prediction of future crime trends involves tracking crime rate changes from one year to the next and used data mining to project those changes into the future. The basic method involves cluster the states having the same crime trend and then using "next year" cluster information to classify records. This is combined with the state poverty data to create a classifier that will predict future crime trends. To the clustered results, a classification algorithm was applied to predict the future crime pattern. The classification was performed to find in which category a cluster would be in the next year. This allows us to build a predictive model on predicting next year's records using this year's data. The decision tree algorithm was used for this purpose. The generalized tree was used to predict the unknown crime trend for the next year. The experimental results proved that the technique used for prediction is accurate and fast.
20. A.Malathi, Dr. S. Santhosh Baboo. [22] presents the use of missing value and clustering algorithm for a data mining approach to help predict the crimes patterns and fast up the process of solving crime. Author concentrates on MV algorithm and Apriori algorithm with some enhancements to aid in the process of filling the missing value and identification of crime patterns. This technique is applied to real crime data. The proposed tool, applied to crime data, can be used as a knowledge discovery tool that can be used to review extremely large datasets and incorporate a vast array of methods for accurate handling of security issues. The development of the crime analysis tool has four steps, namely, data cleaning, clustering, classification and outlier detection. The data cleaning stage removed unwanted records and predicted missing values. The clustering technique is used to group data according to the different type of crime. From the clustered results it is easy to identify crime trend over years and can be used to design precaution methods for future. The classification of data is mainly used predict future crime trend. The last step is mainly used to identify future crimes that are emerging newly by using outlier detection on crime data. Experimental results prove that the tool is effective in terms of analysis speed, identifying common crime patterns and future prediction. Thus, it will help out Indian police and enforcement of law organizations for crime detection and prevention.
21. Malathi. A, Dr. S. Santhosh Baboo and Anbarasi. A [23] proposed data mining technique applied in the context of law enforcement and intelligence analysis helps in solving many problems. Author use clustering/classify based model to anticipate crime trends. The data mining techniques are used to analyze the city crime data from Police Department. The experimental results shows that data mining could potentially used to lessen and even prevent crime for the forth coming years.
22. Prof. Hanmant N. Renushe, Prof. Prasanna R. Rasal, Prof. Abhijit S. Desai [24] presented the importance of data mining technology to design proactive services to reduce crime incidences in the police stations jurisdiction. Crime investigation plays a significant role of police system in any country. Almost all police stations use the CIPA system to store and retrieve the crimes and criminal data and subsequent reporting. It become useful for getting the criminal information but it does not help for the purpose of designing an action to prevent the crime. It has become a major challenge for police system to detect and prevent crimes and criminals. In this paper author highlights the use of data mining techniques for effective investigation of crimes and also to prevent crimes.
23. Divya Bansal, Lekha Bhambhu [25] revealed that Apriori Algorithm is the most popular and useful algorithm of Association Rule Mining of Data Mining. The main objective of taking Apriori is to find frequent itemsets and to uncover the hidden information. Association Rule Mining technique can be applied to all field like business, medical, on line transaction, and here it is applied to legal field, particularly in detecting crime against women. In this paper association rule mining is used to extract patterns that occur frequently within a dataset and showcases the implementation of the Apriori algorithm in mining association rules from a dataset containing crimes data concerning women. Author takes the input dataset from UCI repository and other data is collected manually from the session court of

sirsa. The main motive to use UCI is to first check the proper working of dataset and then apply Apriori on real dataset against crimes on women which extracts hidden information that what age group is responsible for this and to find where the real culprit is hiding. As a comparison made between Apriori and Predictive Apriori, the result shows that Apriori is better and faster than Predictive Apriori Algorithm.

24. Robin Singh Bhadoria, Rohit Bansal, Henry Alexander [26] developed a hyper structure based pattern growth method for frequent itemset mining from uncertain data and also developed a maximal clique based candidate pruning method for uncertain data. Author implemented and analyzed the performance of the well known algorithms for frequent itemset mining for both binary and uncertain data model. The outcome shows that in case of dense binary datasets, FP-growth outperforms all other algorithms, whereas in case of sparse data hyper structure based algorithm P-Hmine outperforms other algorithms.
25. TongWang and Cynthia Rudin, DanieWagner and Rich Sevieri [27] proposed a pattern detection algorithm called Series Finder to figure out which crimes are committed by the same individual or groups. Series Finder algorithm incorporates the common characteristics of all patterns and the unique aspects of each specific pattern. To evaluate the results author compares the Series Finder algorithm with clustering and classification models and same was applied to crime analysis.
26. Clifton Phua, et. al. [28] categorise, compares and summarize various technical and review articles in automated fraud detection for the last years. Author defines the professional fraudster, formalize the main types and subtypes of known fraud, and presents the nature of data evidence collected within affected industries. This research paper presents methods and techniques for business context of mining the data to in order to achieve higher cost savings. As a survey results this paper covers more technical articles and proposed alternative data and solutions from related domains.
27. Chae Chang Lee, Ji Won Yoon [29] proposed a technique for card companies to distinguish between the rightful user and illegal users in order to minimize damage resulting from unauthorized transactions. Author presented a method for learning the individual patterns of card user's transaction amount and the region in which he uses the card, for a given period and determining whether the specified transaction is allowable in accordance with these learned user transaction patterns. Finally, classify legitimate transactions and fraudulent transactions by setting thresholds based on the learned individual patterns.
28. Dawei Wang et. al. [30] proposed a spatial data mining framework to study crime hotspots through their related variables. Author used Geospatial Discriminative Patterns (GDPatterns) to capture the significant difference between two classes (hotspots and normal areas) in a geo-spatial dataset. Utilizing GDPatterns, a novel model was developed Hotspot Optimization Tool (HOT)—to improve the identification of crime hotspots. Finally, based on a similarity measure, GDPattern clusters were grouped and visualize the distribution and characteristics of crime related variables. The results of this approach was evaluated using a real world dataset collected from a northeast city in the United States.
29. Kadhim B. Swadi Al-Janabi [31] presents a proposed framework for the crime and criminal data analysis and detection using Decision tree Algorithms for data classification and simple K Means algorithm for data clustering. The paper tends to help specialists in discovering patterns and trends, making forecasts, finding relationships and possible explanations, mapping criminal networks and identifying possible suspects. The classification technique is based mainly on grouping the crimes according to the type, location, time and other attributes and clustering is based on finding relationships between different Crime and Criminal attributes having some previously unknown common characteristics. The results of both classifications and Clustering are used for prediction of trends and behavior of the Crimes and Criminals.
30. Donald E. Brown [32] provides two technologies to turn data into information in order help the law enforcement agencies for handling large amount data and processed those data into useful information. Data fusion is the first technique which is used for organizing, combining and interpreting information from multiple sources and it also overcomes

confusion from problematic reports, noisy backgrounds etc. Data mining is used to discover patterns and relationships with large databases.

31. Yifei Xute and Donald E. Brown [33] analyzed hot spot models which shows clusters of criminal events based on past locations. Criminal incidents are treated as spatial choice processes. Spatial choice analysis can be used to discover the distribution of people's behaviors in space and time. Two adjusted spatial choice models that include models of decision making processes are presented. The comparison results show that adjusted spatial choice models provide efficient and accurate predictions of future crime patterns and can be used as the basis for a law enforcement decision support system. Author also extends spatial choice modeling to include the class of problems where the decision makers' preferences are derived indirectly through incident reports rather than directly through survey instruments.
32. P.Dhakshinamoorthy, T.Kalaiselvan [34] states that main objective of the paper is to analyze the approach between the police department computer science department. Author implements pattern detection technique to solve the crimes faster in order to help the policemen. Author quoted the Jiawei Han and Michelin Kamber words as "Need is the mother of invention". The most three important needs of the police department are 1) Finding the solution for the unsolved crimes, 2) Solve the crimes at present in faster manner, 3) Predicting the future crimes. As a conclusion, data mining is the better option to fulfill the above needs.
33. Javad Hosseinkhani *et. al.* [35] provides a review about mining useful information by means of Data mining. Author specified even though data mining can be applied to various fields, one of the crucial field is criminology where data mining utilized for identifying crime characteristics. Detecting and exploring crimes and investigating their relationship with criminals are involved in the analyzing crime process. Author reveals that criminology is a suitable field for using data mining techniques which shows the high volume and the complexity of relationships between crime datasets. The paper aims to provide useful information by means of Data Mining, so that

crime hot spots are identified and crime trends are predicted.

5. OTHER FIELDS

Network Forensic Analysis

XIUYU ZHONG [36] proposed network forensic analysis by applying Apriori algorithm. To secure the products in the network against intrusion methods, network forensic is needed. The large number of data are captured and analyzed in network forensics and after capturing and filtering network data package, the Apriori algorithm is used to mine the association rules according to the evidence relevance to build and update signature database of offense, and further it reduce the number of matching times greatly and improve the efficiency of crime detection. Simulation results show that the application of Apriori algorithm can raise the speed, exactitude and intelligence of data analysis for network forensics, the application can help to resolve the real-time, efficient and adaptable problems in network forensics.

Network Cyber Attacks

S.S.Garasia, D.P.Rana, R.G.Mehta [37] mentions about Botnet is one of the most widespread and serious threat in cyber-attacks. A botnet is a group of compromised computers which are remotely controlled by hackers to launch various network attacks, such as DDoS attack, spam, click fraud, identity theft and information phishing. Recently malicious botnets evolve into HTTP botnets out of typical IRC botnets. Data mining algorithms can be used to automate detecting characteristics from large amount of data, which the conventional heuristics and signature based methods could not apply. Here, author presents a new technique for botnet detection that makes use of Timestamp and frequent pattern set generated by the Apriori algorithm. The main advantage of the proposed technique is that prior knowledge of Botnets like Botnet signature is not required to detect the malicious botnets.

Animal behavior analysis

Susan P. Imberman, Michael E. Kress, Dan P. McCloskey [38] presents the housing environment with animals. **Author used** housing environment that has been equipped with a system of RFID sensors. RFID transponders were implanted into the study animal, the naked mole rat. The resulting data was analyzed using principal component analysis and frequent pattern mining. The results showed that these methods can identify time periods of high behavioral activity from that of low activity, along

with which groups of animals interacted with one another.

Educational Data

Dr. Vijayalakshmi M N, S. Anupama Kumar, Kavyashree BN [39] presents the application of association mining on educational data to understand the knowledge and performance of students. Author implemented Apriori algorithm on student log data to bring out the interesting rules. Those rules can be used to infer the performance of the students and to impart the quality of education in the educational institutions. The algorithm generated frequent item sets using support measure in order to understand the interest of the students in the course. Interesting rules are generated based on frequent item sets using confidence factor of the dataset. The rule helps the tutor to understand the knowledge and performance of the students in answering the questionnaire and hence understand the interest of the students in the course.

Digital Forensic

K. K. Sindhu, B. B. Meshram [40] states about forensic analysis steps in the storage media, hidden data analysis in the file system, network forensic methods and cyber crime data mining. Author proposes a new tool which is the combination of digital forensic investigation and crime data mining. The proposed system is designed for finding motive, pattern of cyber attacks and counts of attacks types happened during a period. The proposed tool enables the system administrators to minimize the system vulnerability.

Socio-Economic Impact

R. Sujatha, D. Ezhilmaran [41] states about analyzing the crime data based on various socio economic key factors using proposed model along with the data collected from various sources for crime detection and prevention. Clustering method is used to identify and grouping the crime hotspots and educate the public about the dangers of crimes and criminals. Author combines historical data with current data which gives new ideas and also helps in solving pending cases. The paper definitely helps to bring down the crime rates in the crime prone area.

Banking Sector

Dr. K. Chitra, B. Subashini [42] presents data mining techniques and its applications in banking sectors like fraud prevention and detection, customer retention, marketing and risk management. Early data analysis techniques were oriented towards extracting quantitative and statistical data. These techniques facilitate useful data interpretations for the

banking sector to avoid customer attrition. Customer retention is the most important factor to be analyzed in today's competitive business environment. Fraud is a significant problem in banking sector. Also detecting and preventing fraud is difficult. With the help of data mining algorithms, it is able to detect suspicious activity within the data in a non prescriptive way. While the system observes the user's transactions, it discovers common behavior patterns by the way of clustering. In order to discover anomalous transactions, new transactions are compared with the user's common behavior patterns. A transaction that does not correspond with one will be treated as a suspicious activity and thus precautionary steps can be taken in advance.

6. CONCLUSION

This paper presents the review on various research papers pertaining to applications of frequent patterns mining and association rule mining in the field of crime pattern detection. It gives knowledge about various frequent pattern mining algorithm and extensions of the same. It also explains about the different application areas where these frequent patterns can be used other than crime pattern. We sure that this paper will help the researches and data miners to obtain knowledge and reveals the advantages of applying frequent pattern mining algorithm along with rule mining in various fields.

REFERENCES

- [1] R. Agrawal, T. Imielinski & A. Swami Mining. 1993. "Association Rules between Sets of Items in Large Databases", Proceedings of the 1993 ACM SIGMOD Conference, pp.1-10.
- [2] Kuldeep Malik, Neeraj Raheja, Puneet Garg. 2011. "Enhanced FP-Growth Algorithm", International Journal Of Computational Engineering & Management, Vol. 12, Pp.54-56.
- [3] Shyam Varan Nath. 2006. "Crime Pattern Detection using Data Mining", Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 41-44.
- [4] Yagnik Ankur N., Dr .Ajay Shanker Singh. 2014. "Outlier Analysis Using Frequent Pattern Mining – A Review", International Journal of Computer Science and Information Technologies, Vol. 5, Issue.1, pp.47-50.
- [5] Anna L. Buczak , Christopher M. Gifford. 2010. "Fuzzy Association Rule Mining for Community

Crime Pattern Discovery”, ISI-KDD, ACM ISBN 978-1-4503-0223-4/10/07 Washington, July 25.

[6] A.Malathi, Dr. S. Santhosh Baboo. 2011. “Algorithmic Crime Prediction Model Based on the Analysis of Crime Clusters”, Global Journal of Computer Science and Technology, Vol.11, Issue.11, pp.47-51.

[7] Aniruddha Kshirsagar, 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.

[8] Manish Gupta, B. Chandra and M. P. Gupta. “ Crime Data Mining for Indian Police Information System”, PP.388-397

[9] 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.

[10] Divya Bansal, Lekha Bhambhu. 2013. “Usage of Apriori Algorithm of Data Mining as an Application to Grievous Crimes against Women”, International Journal of Computer Trends and Technology, Vol.4, Issue.9, pp.3194-3199.

[11] Devesh Bajpai. 2012. “Emerging Trends in Utilization of Data Mining in Criminal Investigation: An Overview”, Journal of Environmental Science, Computer Science and Engineering & Technology, Vol.1, Issue.2, pp.124-131.

[12] Jitendra Kumar, Sripati Mishra, Neeraj Tiwari. 2012. “ Identification of Hotspots and Safe Zones of Crime in Uttar Pradesh, India: Geo-spatial Analysis Approach”, International Journal of Remote Sensing Applications, Vol.2, Issue.1, pp.15-19.

[13] John David Elijah Sandig, Ruby Mae Somoba, Ma. Beth Concepcion and Bobby D. Gerardo. 2013. ” Mining Online GIS for Crime Rate and Models based on Frequent Pattern Analysis”, Proceedings of the World Congress on Engineering and Computer Science, Vol.2, pp.23-27.

[14] Dr. M. VijayaKumar, Dr. S. Karthick and N.Prakash. 2013. “The Day-To-Day Crime Forecasting Analysis of Using Spatial temporal Clustering Simulation”, International Journal of Scientific & Engineering Research, Vol.4, Issue.1, pp. 1-6.

[15] A.Malathi, Lt. Dr. S. Santhosh Baboo. “An Intelligent Analysis of Crime Data for Law Enforcement Using Data Mining”, International Journal of Data Engineering, Vol.1, Issue.5, pp.77-83.

[16] Li Ding, et.al. 2009. “PerpSearch: An Integrated Crime Detection System”, IEEE International Conference on Intelligence and Security Informatics, ISI 2009, Dallas, Texas, USA, June 8-11, 2009, Proceedings; 01/2009

[17] Jian Pei, Jiawei Han, Hongjun Lu, Shojiro Nishio, Shiwei Tang, Dongqing Yang. “H-Mine: Hyper-Structure Mining of Frequent Patterns in Large Databases”.

[18] Charu C. Aggarwal, et.al. 2009. “Frequent Pattern Mining with Uncertain Data”, *KDD'09*, June 28–July 1, 2009, Paris, France.

[19] Goswami D.N., Chaturvedi Anshu., Raghuvanshi C.S. 2010. “An Algorithm for Frequent Pattern Mining Based On Apriori”, International Journal on Computer Science and Engineering, Vol.02, Issue.04, pp.942-947.

[20] M Suman , T Anuradha , K Gowtham, A Ramakrishna. 2012. “ A Frequent Pattern Mining Algorithm Based On FP-Tree Structure and Apriori Algorithm”, International Journal of Engineering Research and Applications, Vol.2, Issue.1, pp.114-116.

[21] S.Yamuna, N.Sudha Bhuvanewari. 2012. “ Datamining Techniques to Analyze and Predict Crimes”, The International Journal of Engineering And Science, Vol.1, Issue.2, pp.243-247.

[22] A.Malathi, Dr. S. Santhosh Baboo. 2011. “An Enhanced Algorithm to Predict a Future Crime using Data Mining”, International Journal of Computer Applications, Vol.21, Issue No.1, pp.1-6.

[23] Malathi. A, Dr. S. Santhosh Baboo and Anbarasi. A. 2011. “An intelligent Analysis of a City Crime Data Using Data Mining”, International Conference on Information and Electronics Engineering, Singapore, Vol.6, pp.130-134.

[24] Prof. Hanmant N. Renushe, Prof. Prasanna R. Rasal, Prof. Abhijit S. Desai. 2012. “ Data Mining Practices for Effective Investigation of Crime”, International Journal of Computer Technology & Applications, Vol.3, Issue.3, pp.865-870.

- [25] Divya Bansal, Lekha Bhambhu. 2013. "Execution of APRIORI Algorithm of Data Mining Directed Towards Tumultuous Crimes Concerning Women", International Journal of Advanced Research in Computer Science and Software Engineering, Vol.3, Issue.9, pp.54-62.
- [26] Robin Singh Bhadoria, Rohit Bansal, Henry Alexander. 2011. "Analysis of Frequent Item set Mining on Variant Datasets", Internal Journal of Computer Technology Application, Vol.2, Issue.5, pp.1328-1333.
- [27] TongWang and Cynthia Rudin, DanielWagner and Rich Sevieri. 2013. "Detecting Patterns of Crime with Series Finder", Twenty-Seventh AAAI Conference on Artificial Intelligence, pp.140-142.
- [28] Clifton Phua, et. al. 2005. "A Comprehensive Survey of Data Mining-based Fraud Detection Research", pp.1-14.
- [29] Chae Chang Lee, Ji Won Yoon. 2013. "A data mining approach using transaction patterns for card fraud detection", Seoul, Republic of Korea, pp.1-12.
- [30] Dawei Wang et. al. 2013. "Understanding the spatial distribution of crime based on its related variables using geospatial discriminative patterns", Journal of Computers, Environment and Urban Systems 39, pp.93-106.
- [31] Kadhim B. Swadi 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, Issue.3, pp.8-24.
- [32] Donald E. Brown. 1998. "The Regional Crime Analysis Program (RECAP): A Framework for to Catch Criminals", pp.2848-2853.
- [33] Yifei Xute and Donald E. Brown.2003. "A Decision Model for Spatial Site Selection by Criminals: A Foundation for Law Enforcement Decision Support", IEEE Transactions On Systems, Man, And Cybernetics, Vol. 33, Issue.1, pp.78-85.
- [34] 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.
- [35] Javad Hosseinkhani et. al. 2014. "Detecting Suspicion Information on the Web Using Crime Data Mining Techniques", International Journal of Advanced Computer Science and Information Technology, Vol. 3, Issue.1, pp.32-41.
- [36] Xiuyu Zhong. 2013." The Application Of Apriori Algorithm For Network Forensics Analysis", Journal of Theoretical and Applied Information Technology, Vol.50, Issue.2, pp.430-434.
- [37] S.S.Garasia, D.P.Rana, R.G.Mehta. 2012. "HTTP Botnet Detection Using Frequent Patternset Mining", International Journal of Engineering Science & Advanced Technology Vol.2, Issue.3, pp.619-624.
- [38] Susan P. Imberman, Michael E. Kress, Dan P. McCloskey. 2012. "Using Frequent Pattern Mining to Identify Behaviors in a Naked Mole Rat Colony", Proceedings of the Twenty-Fifth International Florida Artificial Intelligence Research Society Conference, pp.394-399.
- [39] Dr.Vijayalakshmi M N, S.Anupama Kumar, Kavyashree BN. 2014. " Investigating Interesting Rules Using Association Mining for Educational Data", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3, Issue 2, pp.268-271.
- [40] K. K. Sindhu, B. B. Meshram. 2012. "Digital Forensics and Cyber Crime Datamining", Journal of Information Security, Vol.3, pp.196-201.
- [41] R.Sujatha, D.Ezhilmaran. 2013. "A Proposal for Analysis of Crime Based on Socio – Economic Impact using Data Mining Techniques", International Journal of Societal Applications of Computer Science, Vol.2, Issue.2, pp.229-231.
- [42] Dr. K. Chitra, B. Subashini. 2013. " Data Mining Techniques and its Applications in Banking Sector", International Journal of Emerging Technology and Advanced Engineering, Vol.3, Issue.8, pp.219-226.

Authors



D.Usha is currently working as Assistant Professor in Department of Information Technology, Hindustan University, Chennai. She has 6.5 years of teaching experience. She is pursuing Ph.D in the area of Frequent Pattern Mining from Hindustan University. She is a member in IEEE. Her areas of interest are Data mining, Frequent Pattern mining and Association rule mining.



Dr. K. Ramesh Kumar is currently working as Associate Professor in Department of Information Technology, Hindustan University, Chennai. He received his Ph.D degree in Computer Science and Engineering from Alagappa University, Karaikudi, India in 2011. His Ph.D involved the development of algorithms to frequent pattern and association rule mining, and constructs new dataset for AIDS/HIV infected patients' case history. He had published 32 research papers in various journals and conferences.