

Frequent Pattern Analysis in Crime Detection

Sunil Kumar Sahu

Assistant Professor, Department of CSE
PCEM Bhilai CG

Umesh Kumar

BE Student, Department of CSE
PCEM Bhilai CG
India

Shrishti Sao

BE, Computer Science & Engineering
Chhattisgarh, India

Abstract—In frequent pattern mining, there were various algorithms. Objective is to find frequent itemsets and association between different item sets Apriority is the classical algorithm and has been play vital role in association rule mining. Main idea behind of this algorithm is to find useful patterns between different set of data. It is a very simple algorithm and many researchers have been done for the improvement of this algorithm. The Apriori & FP-growth algorithms are the most famous algorithms for which it can be used for Frequent Pattern mining. This survey paper presents various Frequent Pattern Mining and Rule Mining algorithm that can be applied in crime pattern detection. The analysis of survey would give the brief information about what has been done previously in this area, what is the current technology and other related areas. It also explains various frequent pattern algorithms and how it can be applied to the different areas in crime pattern detection. This survey paper surely helps the researches to get brief idea about the advantages frequent pattern algorithm in various areas.

Keywords- *Frequent Pattern Mining, Apriori, FPgrowth, Association Rule Mining, Crime Pattern mining.*

I. INTRODUCTION

Frequent itemsets play an important role in data mining and that and it has to find an interesting patterns from databases, such as association rules, correlations, sequences, , classifiers, clusters and many more of which the mining of association rules. Frequent pattern mining major field in research and development since it is a part of data mining. Many research papers, articles are published in the field of Frequent Pattern Mining (FPM). This paper exploring the details about frequent pattern mining algorithm and its types and extensions of frequent pattern mining, association rule mining algorithm, rule generation. This paper describes about various existing FPM algorithms, data mining algorithm for crime pattern detection. By applying frequent pattern mining algorithm and suitable measures to proposed a new algorithm is applied to crime dataset in order to find out the suspects in the short time. These rules can be used in many fields, such as customer shopping analysis, additional sales, goods shelves design, storage planning and classifying the users according to the buying patterns, etc. The techniques and discovering association rules from the dataset have traditionally focusing on identifying the relationships between items and telling some suspect of Human behavior, gradually buying behavior for determining items that customers buy together. All association rules of this type describe a particular local pattern. These group of association rules can be easily determined for which it is interpreted and communicated.

II. LITERATURE REVIEW

Survey analysis on Crime and Forensic department

A. Crime pattern functions

Crime pattern analysis can be occur at various levels, including strategic, tactical and operational. Crime analysts studying in crime reports, arrests reports, and police calls for service to identify emerging patterns matching, series, and trends as quickly as possible. 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.

B. 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.

C. In legal field

Shyam Varan Nath 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.

D. Network Forensic Analysis

XIUYU ZHONG 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.

E. Network Cyber Attacks

S.S.Garasia, D.P.Rana, R.G.Mehta 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.

F. Research Directions proposal

We described research under the banner of frequent pattern mining have given a solution of the most known problems related to frequent pattern mining, and the provided solutions are very good for most of the data mining tasks. But, it is required to solve several critical research problems before frequent pattern mining can become a central approach in data mining applications.

III. METHODOLOGY

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. FPGrowth algorithm)
3. Vertical layout approach (E.g. Eclat algorithm)

A. Candidate generation approach (E.g. Apriori algorithm)

Apriori algorithm

Apriori algorithm is the most basic and classical algorithm for mining frequent itemsets and it was proposed by R.Agrawal and R.Srikant in 1994. Apriori algorithm is used to find the all frequent itemsets in a given database Databases. The key idea behind the Apriori algorithm is to make multiple passes over the database. It employs an iterative approach known as a breadth-first search and also known as level-wise search through the search space, where k-itemsets are used to explore (k+1)-itemsets. This property is used to get prune the infrequent candidate of that elements. In the beginning, the set of frequent 1-itemsets is found. The basic steps to mine the frequent elements are as follows:

- **Generate and test:** In this first step finding the 1-itemset in frequent itemset elements L by scanning the database and remove all those element from candidate C which cannot be satisfy the minimum support of that criteria.
- **Join step:** To get the next level of elements Ck join the previous frequent elements by self join i.e. Lk-1* Lk-1 known as Cartesian product of Lk-1 . i.e. in his step generating a new candidate of k-itemsets based on joining Lk-1 with itself which found in the previous iteration. Let Ck candidate k-itemset and Lk be the frequent k-itemset.
- **Prune step:** Ck is the superset of Lk so the members of Ck may or may not be frequent but all K ' 1 frequent itemsets are included in Ck thus prunes the Ck to find K frequent itemsets with the help of Apriori property. It is used to get eliminate the unfrequent items in the dataset.

But as the dimensionality of the database increase with the number of items then:

- More space is required and I/O cost will increase.
- Number of scan in databases so increased the generation of candidate will increase results in increase in computational cost.
- Reduce the number of passes of transaction database scans.
- Shrink number of candidates.
- Facilitate support counting of candidates.

B. Without Candidate Generation Approach

FP Growth

In order to count the supports of all generated itemsets, FP-growth algorithm uses a combination of the vertical and horizontal databases layout to storing the database in main memory. Instead of storing the cover for every item in 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. In 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.

C. Vertical Layout Approach

Eclat Algorithm

The first algorithm developed to generate all frequent itemsets in a depth-first manner 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.

Association rules

In Association Rule mining find rules that will predict the occurrence of an item based on the occurrence of the other items in the transaction. Table shows Market-Basket.

Example of Association Rules-

TABLE I. TRANSACTION AND ITEMSETS.

Transactions	TID Items
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

{Diaper} @ {Beer},
 {Bread, Milk} @ {Egg, Coke},
 {Bread, Beer} @ {Milk},

Association rule is an implication expression of the form X @ Y, where X and Y are itemsets. Example: {Milk, Diaper} @ {Beer}

Rule Evaluation:

- Support (S): Fraction of transactions that contain both X and Y.

- Confidence (C): Measures how often items in Y appear in transactions that contain X. Example: {Milk, Diaper} @ {Beer}

$$S = s(\{Milk, Diaper, Beer\}) / [T]$$

$$S = 2/5 \quad S = 0.4$$

$$C = s(\{Milk, Diaper, Beer\}) / s(\{Milk, Diaper\})$$

$$S = 2/3 \quad S = 0.67$$

- Itemset: A collection of one or more items. Example {Milk, Diaper, Beer}. K-itemset that contains k-items.

- Frequent Itemset: An itemset whose support is greater than or equal to a min_sup threshold. In association rule

mining task from a set of transactions T, the goal of association rule mining is to find all rules having Support >= min_sup threshold and Confidence >= min_conf threshold.

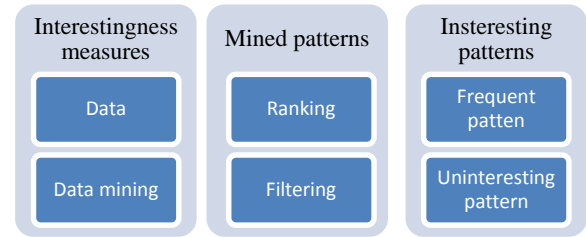


Figure 1 Roles of interestingness measures in the data mining process.

IV. CONCLUSION AND FUTURE SCOPE

This survey paper review on various fields of research pertaining to applications of frequent patterns mining and association rule mining in the field of crime pattern detection. It takes knowledge about various frequent pattern mining algorithm. It also explains about the different application areas where these frequent patterns can be used other than crime pattern. We sure that this survey paper will help to the researchers and data miners to obtain knowledge and reveals advantages on applying frequent pattern mining algorithm along with rule mining in various fields of this technology. Crime must be solved by this technology and many cases would be pended in files in the government offices to be reduced. Therefore these algorithms can be applying in other domains to bring out interesting pattern among the data present in the repository.

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. "Oulier Analysis Using Frequent Pattern Mining – A Review", International Journal of Computer Science and Information Technologies, Vol. 5, Issue.1, pp.47-
- [5] Anna L. Buczak , Christopher M. Gifford. 2010. " Fuzzy Association Rule Mining for Community.
- [6] Rakesh Agrawal, T. Imieliński, A. Swami, "Mining association rules between sets of items in large databases". In: Proceedings of the 1993 ACM SIGMOD international conference on Management of data
- [7] Sanjeev Rao, Prianka Gupta, "Implimenting Improved Algorithm Over APRIORI Data Mining Association Rule Algorithm", In: proceeding of IJCST, ISSN 0876-8491, VOL.3, Issue 1, Jan-March 2012.
- [8] Rayner Alferd , "Knowledge Discovery: Inhacing Data Mining and Dcision Support Integration", The University of York, Artificial Intelligence Group Department of Cpmputer Science, The University of York, United Kingdom.
- [9] Mamta Dhanda, Sonali Guglani, "Mining Efficient Association rules Through Apriori Algorithm Using Attributes", In: Proceeding of IJCST, ISSN 0876-8491, Vol. 2, Issue 3, September.
- [10] Jaishree Singh, Hari Ram, Dr. J.S. Sodhi, "Improving Efficiency of Apriori Algorithm Using Transaction Reducti.