Mining Regular Frequent Crime Patterns using Vertical Format

G. Vijay Kumar*, V. Indraja, K. Pratyusha and N. Sarada

Department of Electronics and Computer Engineering, School of Computing, KL University, Vaddeswaram, Guntur – 522502, Andhra Pradesh, India; gvijay_73@kluniversity.in, indraja.indu131@gmail.com, pratyushakaja@gmail.com, chinni.nallamala@gmail.com

Abstract

Background/Objectives: The goal of crime data mining is to understand various crime patterns in criminal behavior in order to predict crimes and anticipate criminal activity to avoid the crime not to happen. **Methods/Statistical Analysis:** Predicting crime is one of the global challenges facing by Law enforcement department and it requires persistent efforts in order to restrict. In this paper we are introducing a new crime pattern called regular frequent crime pattern which occurs regularly at certain time intervals using vertical data format also satisfies downward closure property. **Findings**: Crime patterns were not defined by statistics and its identification is more than just counting and summarizing crimes that are similar in characteristics and/or location on a map. Crime pattern is a group of one or more crimes reported to or discovered by the police. The approaches which are pattern based have the possibility to help the police department in discovering new type of crime patterns. **Applications/Improvement**: Our experiment results show the impact on execution time and memory. This project is also useful for police department in finding the regular-frequent crimes which are happening in today's world.

Keywords: Crime Pattern Mining, Frequent Patterns, Regular Patterns, Vertical Data Format

1. Introduction

Nowadays the most influential tool to grab patterns and relationships within in our data is Data mining[1]. The data mining technique can also be used in traditional educational system[2]. The method of data mining is also used in stock exchange in companies for retrieving large amount of data[3]. A Crime pattern is a group of two or more crimes reported to or discovered by police that are unique. Data mining can be used to model crime detection problem. A crime pattern is identified through a systematic, deductive analytical process, subsequently communicated to police agents via some form of bulletin[4]. So here in this paper we are following Regular Frequent Pattern Mining approach which is efficient in order to get the crime patterns from large data sets. Regular Frequent Pattern Mining is based on occurrence frequency and occurrence behavior of a pattern. When the occurrence frequency is greater than or equal to user given threshold value and occurrence behaviour is less than or equal to the user given regularity threshold value then that pattern is called regular frequent pattern.

In order to find such crime patterns from large data sets we are using the Regular Frequent Pattern Mining Technique. Frequent pattern mining is one of the active research areas in data mining to find interesting patterns. Frequent pattern mining basically depends on the support count i.e. number of times a pattern appears in the database. Not only occurrence frequency of a pattern but also occurrence behavior of a pattern may be treated as important criteria to measure the interestingness of the pattern in several applications online. The significance of a product may depend upon other occurrence characteristics such as regularity of the pattern. For example in a super market the user may be interested in frequently sold items which are sold at regular intervals (or) to enhance

a web site design the website administrator may be concerned in more often popular pages at regular intervals. From the above examples we observed that the occurrence behavior of a frequent pattern at regular intervals plays an important role among various applications like network monitoring, telecommunications or sensor networks.

Association Rule Mining is an important data mining model proposed by Agrawal, et al. in 1993. Apriori approaches have been employed in the past for this purpose and have been identified as having some shortcomings. Association Rule Mining is all about finding frequent patterns of item-sets. It is mainly used for market basket analysis to find how items are purchased by customers. Mining frequent patterns[5,6] was first introduced by Agrawal, et al. in 1993 need k-number of scans to generate k-item set. In[7] introduced regular frequent pattern mining in 2008 over transactional databases. Recently periodic frequent patterns[8] or frequent regular patterns[9] is playing a vital role in data mining research because of the occurrence frequency of the pattern along with occurrence behavior of a pattern i.e. regularity or interval. There is no such pattern which mines crime patterns using Regular Frequent Pattern algorithm. Therefore in this paper we are applying RFCPM approach to obtain crime patterns.

Many researchers[10–12] are investigating ways of tackling crime and improving on existing algorithms used to derive crime patterns, in order to assist public safety and security agencies in achieving their objective of deterring crime and promoting citizens safety. The most important and competitive area in data mining and knowledge discovery research is to mine various interesting patterns[13]. In the paper[14] we came to know that regularity of item set and frequency of item set both are equally important on any kind of database. A new algorithm has been proposed RFRP to find the regular frequent patterns of an item-set. In this paper[15] occurrence frequency of the pattern is measured as an important factor and is used for measuring the interestingness of the pattern in many applications. Based on min_support user given threshold value a regular frequent pattern is found. A Series Finder algorithm for detecting patterns is proposed by Tong Wang which detects patterns of crime committed by the same individual.

In our paper we compared our results with RP-tree algorithm which generates a complete set of regular patterns in a database for a user given regularity threshold value[16]. Apriori is an algorithm for association rules. To get the support count of the item sets this apriori algorithm scans the database for many times[17]. In the paper[18] we came to know how to mine frequent item sets using vertical format. Our algorithm satisfies the downward closure property. Any subset of a frequent item-set must be frequent this is called downward closure property. A survey of various frequent pattern mining algorithms has been done to find crime patterns in[19]; this paper followed three approaches with candidate generation, without candidate generation and vertical layout approach. It helps the researchers to get an idea about the frequent pattern mining algorithm in various applications.

2. Problem Description

The concepts of regular frequent pattern mining are defined in this section and the basic definitions of the problem to obtain complete set of regular frequent patterns in crime database.

Let $I = \{i_1, i_2, \ldots . i_n\}$ be a set of items. A set $X = \{i_j, \ldots . i_k\}$ subset(I), where $j \leq k$ and $j,k \in [1,n]$ is called a itemset or pattern. Crime $c = (cid, Y)$ is a couple where cid is a crime id and Y is a pattern. Let the number of items in Y be the size of t size(t). A transaction database DB over I is a set of crime transactions $C = \{c_1, \ldots . c_m\}$, $m = |DB|$ is the size of DB, i.e. total number of crime transactions in DB. If X subset (Y), which means that c contains X or X occurs in c and denoted as c_j^X, $j \in [1,m]$. Therefore, $C^X = \{c_j^X, \ldots . c_k^X\}$, $j \leq k$ and $k \in [1,m]$ is the set of all crime transactions where pattern X occurs in DB.

2.1 Definition 1 (Frequent Pattern X)

The total number crime transactions in a DB that contains pattern X is called the support of X i.e. Sup(X). Hence $Sup(X) = |C^k|$, where $|C^k|$ is the size of C^k. The pattern X is said to be frequent if its support is greater than or equal to user given minimum support threshold i.e., $Sup(X) \geq minsup(\delta)$.

2.2 Definition 2 (Regularity of Frequent Pattern X in Crime Database)

Let c_{j+1}^X and c_j^X, $j \in [1,(m-1)]$ be two successive transactions where frequent pattern X appears. The variation between these two successive transactions can be defined as a period of X, say p^X (i.e., $p^X = c_{j+1}^X - c_j^X$, $j \in [1, (m-1)]$). For ease, to calculate the period of a pattern, we consider

Table 1. An illustration of selected categorical crime attributes

Location	Incident	Victim Id	Gender	Time	Culprit Category	…….
Wood Stock	Robbery	A3	M	Night	Gang	…
Milnerton	Kidnapping	V4	F	Noon	Individual	…
Mowbray	Robbery	B1	M	Night	Gang	…
…	…	…	…	…	…	…

the first transaction in the DB as null i.e., $t_f = 0$ and the last transaction is the mth transaction i.e., $c_l = c_m$. Let C^X, P^X be the set of all periods of X i.e., $P^X = \{ p_1^X, \ldots .p_r^X\}$, where r is the total number of periods in P^X. Then the regularity of a frequent pattern X can be denoted as $Reg(X) = max(p_1^X, \ldots . p_r^X\}$. A frequent pattern X is said to be regular frequent if its regularity is less than or equal to user given minimum regularity threshold i.e., λ.

3. Methodology

3.1 Regular Frequent Pattern Mining (RFCPM) approach to Crime Pattern Mining

Consider Table 1 as a typical Crime Database (CDB) with categorical crime attributes, which comprises of incident location, suspect and victim information, day of the week, time and date, weapons used and crime scene status etc. The first step is to manage and identify the attributes of interests that are available in the crime investigation report set Table 1. Then the database is queried to extract the attributes of interest, creating the Crime the Crime Transaction Data Base (CTDB) to be mined.

To illustrate how Regular Frequent Pattern mines crime data we consider a simple example shown in Table 2. Suppose we passed a query (involving location information and corresponding crime incidents that occurred there) to obtain the Crime Transaction Data Base (CTDB) which is further encoded or transformed with Crime Transaction Identification (CTID) being unique identifier for each crime transaction and Crime Incident Information (CII). Crime Incident and crime location has been encoded with abbreviations, in order to reduce processing time.

Now we need to mine regular frequent crime patterns with regard to location from Table 2.

Firstly, we need to obtain the support and regularity count for the crime incidents in Table 2 as presented in

Table 3. The support is an indication of how frequent a crime item-set is observed. And the regularity count is an indication of how regularly a crime item-set has occurred. By following the RFCPM approach we convert our table into vertical data format and obtain the support and regularity count values.

Table 2(a). Crime Transaction Data Base (CTDB)

Crime Location	Crime Incidents
Loc1	Kidnapping, Rape
Loc2	Kidnapping, Rape, Murder, Robbery
Loc3	Murder, Robbery, Burglary, Kidnapping
Loc4	Kidnapping, Robbery, Burglary, Murder
Loc5	Rape, Murder, Burglary
Loc6	Kidnapping, Rape Murder, Robbery
Loc7	Shoplifting, Kidnapping Arson
Loc8	Rape, Robbery, Murder, Kidnapping
Loc9	Kidnapping, Rape, Robbery, Murder
Loc10	Robbery, Murder, Kidnapping, Rape

Table 2(b). Encoded CTDB

CTID	CII
CT1	kp, rp
CT2	kp. rp, md, rb
CT3	md, rb, bg, kp
CT4	kp, rb, bg, md
CT5	rp, md, bg
CT6	kp, rp, md, rb
CT7	sl, kp, ar
CT8	rp, rb, md, kp
CT9	kp, rp, rb,md
CT10	rb, md, kp, rb

3.2 RFCPM-Algorithm

Input: DB, λ, δ

Output: Complete set of regular frequent patterns.

Procedure:

Let X_i subset (I) be a k-itemset

$P^x_i = 0$ for all X_i

For each X_i

Update **Sup**

    If Sup(X_i)>=δ

    Find the period of X_i

$P^x_i = P^x_{i+1} - P^x_i$

    Reg(X_i) = max(P^x_i)

        If reg(X_i)<=λ

        X_i is a regular frequent itemset

            Else

            Delete X_i

        Else

Delete X_i

Repeat

Now we prune our table based on minimum support threshold value (δ = 5) and maximum regularity value (λ = 4). We obtain our result in Table 4. The crime items which do not satisfy the minimum support and maximum regularity values would be discarded. This process continues until a Regular Frequent is obtained.

Now associations are made between kp ,rp, md, rb as they satisfy the condition and are suitable to obtain length-2 item-sets which are shown in Table 5.

Similarly again the associations are made between the 2-item-sets which satisfy the condition and a length-3 item-set is obtained which is shown in Table 6.

Table 6 both support and regularity satisfy the user given threshold values so we need to continue this process until no candidate is generated for mining regular frequent patterns in database. And at last finally we get the regular frequent pattern which satisfies both the conditions which is shown in Table 7.

Table 3. Converting into vertical data format following RFCPM approach

Crime Items	CTID	Support	Regularity
kp	1,2,3,4,6,7,8,9,10	9	1
rp	1,2,5,6,8,9,10	7	3
md	2,3,4,5,6,8,9,10	8	2
rb	2,3,4,6,8,9,10	7	2
bg	3,4,5	3	5
sl	7	1	7
ar	7	1	7

Table 4. Crime item set table with support (δ = 5) and regularity (λ = 4) for 1-item set

Crime Items	CTID	Support	Regularity
kp	1,2,3,4,6,7,8,9,10	9	1
rp	1,2,5,6,8,9,10	7	3
md	2,3,4,5,6,8,9,10	8	2
rb	2,3,4,6,8,9,10	7	2
~~bg~~	3,4,5	~~3~~	~~5~~
~~sl~~	7	~~1~~	~~7~~
~~ar~~	7	~~1~~	~~7~~

Table 5. Crime item set table with support (δ = 5) and regularity (λ = 4) for 2-item set

Item set	CTID	Support	Regularity
kp,rp	1,2,6,8,9,10	6	4
kp,md	3,4,6,8,9,10	5	3
kprb	2,3,4,6,8,9,10	6	2
rp,md	2,5,6,8,9,10	5	3
rp,rb	2,6,8,9,10	5	4
md.rb	~~2,3,8,10~~	4	5

Table 6. Crime item set table with support (δ = 5) and regularity (λ = 4) for 3-item set

Item set	CTID	Support	Regularity
kp,rp,md	2,6,8,9,10	5	4
kp,rp,rb	2,6,8,9,10	5	4
kp,md.rb	2,3,4,6,8,9,10	7	2
rp,md,rb	2,3,6,8,9,10	7	3

Table 7. Crime item set table with support (δ = 5) and regularity (λ = 4) for 4-item set

Item set	CTID	Support	Regularity
kp,rp,md,rb	2,6,8,9,10	5	4

Therefore the item set {kp, rp, md, rb} satisfy the user given support and regularity, hence this item-set is the regular frequent crime pattern.

4. Experimental Results

Our experimentation results are performed over synthetic dataset (T1014D100K) and real datasets (Chicago crime database). We compare the results of RP-tree with our algorithm RPFM which shows that our algorithm is more efficient and fast in finding the regular frequent

crime pattern. All experiments are done in java on windows XP containing 2.7 GH with 2 GB of main memory.
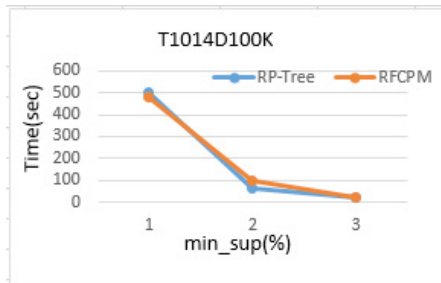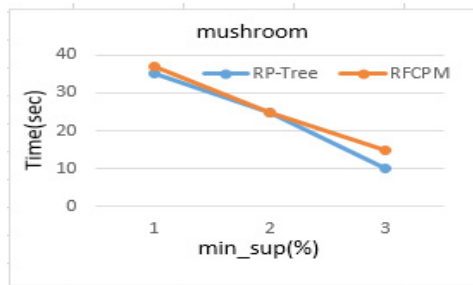

Figure 1(a). Execution time over T1014100K.

Figure 1(b). Execution time over mushroom.

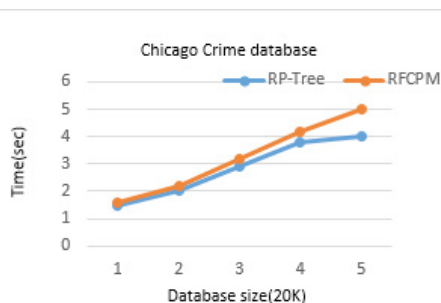**Figure 2(a).** Execution time over Chicago crime database.

Figure 2(b). Memory usage over Chicago crime database.

Firstly we compared our results with RP-tree on T1014D100K which contains 20K transactions as shown in Figure 1(a) and Figure 1(b). By observing the graphs we can say that RFCPM algorithm relatively takes equal time.

The time and memory specify the total execution time and memory required with the database size. Scalability of our algorithm is shown in Figure 2(a) and Figure 2(b). The performance of our algorithm (RFCPM) is more efficient and as we are using vertical data format this method is efficient and scalable over large databases.

5. Conclusion

In this paper we presented a new algorithm RFCPM which mines to acquire a regular frequent pattern in crime databases using vertical data format. Finding regular frequent patterns is very important as the information would be much more useful. The advantages of vertical data format that needs simple calculations like unions, intersections, subtraction etc. are used by our algorithm. This method is scalable and efficient over large databases. Our algorithm is very useful in finding regular frequent patterns not only for crime database but also in various applications.

6. References

1. Vedanayaki M. A study of data mining and social network analysis. Indian Journal of Science and Technology. 2014 Nov; 7(S7):185–7.
2. Murugananthan V, Shiva Kumar BL. An adaptive educational data mining technique for mining educational data models in e-learning systems. Indian Journal of Science and Technology. 2016 Jan; 9(3):1–5. Crossref
3. Azad N, Ranjbar V, Khani D, Moosavi ST. Information disclosure by data mining approach. Indian Journal of Science and Technology. 2012 Apr; 5(4):2593–602.
4. Bruce C, Santos RB. Crime pattern definitions for Tactical Analysis. Standards Methods and Technology (SMT) Committee; White paper-2011.
5. Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases. ACM SIGMOD International Conference on Management of Data; 1993. p. 207–16, Crossref
6. Agrawal R, Srikanth R. Fast algorithms for mining association rules. VLDB; 1994. p. 489–99.
7. Tanbeer SK, Farhan CA, Jeong BS, Lee YK. Mining regular patterns in transactional database. IEICE Transactions; 2008. p. 2568–77. Crossref
8. Tanbeer SK, Farhan CA, Jeong BS, Lee YK. Discovering periodic-frequent patterns in transactional databases. PAKDD; 2009. p. 242–53. Crossref

9. Ruggieri S. Frequent regular itemset mining. ACM KDD; 2010. Crossref

10. Chen H, Chung W, Xu J, Wang G, Qin Y, Chau M. Crime data mining: A general framework and some examples. IEEE Computer Journal. 2004; 37(4):50–6. Crossref

11. Dandu S, Deekshatulu B, Chandra P. Improved algorithm for frequent item sets mining based on apriori and fp-tree. Computer Science and Technology Software and Data Engineering Global Journal. 2013; 13(2):1–5.

12. Khan NG, Bhaga V. Effective data mining approach for crime-terror pattern detection using clustering algorithm technique. Engineering Research and Technology International Journal. 2013; 2(4):2043–8.

13. Sreedevi M, Reddy LSS. Mining regular closed patterns in transactional databases. 2013 7th International Conference on Intelligent Systems and Control (ISCO); 2013. p. 380–3,

14. Vijay Kumar G, Valli Kumari V. Parallel and distributed frequent-regular pattern mining using vertical format in large databases. IEEE Xplore, IET; 2012. P. 110–4.

15. Rashid MM, Karim MR, Jeong BS, Chai HJ. Efficient mining regularly frequent patterns in transactional databases. Springer Lecture Notes in Computer Science; 2012. p. 258–71.

16. Tanbeer SK, Chowdhury FA. RP-Tree: A tree structure to discover regular patterns in transactional database. Intelligent Data Engineering and Automated Learning-IDEAL 2008, Volume 5236 of the series Lecture notes in Computer Science; 2008. p. 193–200. Crossref

17. Ming GY, Zhi-jun W. A vertical format algorithm for mining frequent itemsets. IEEE Transactions; 2010. p. 11–3.

18. Usha D, Ramesh Kumar K. A complete survey on application of frequent pattern mining and association rule mining on crime pattern mining. International Journal of Advances in Computer Science and Technology. 2014; 3(4):264–75.

19. Sherly KK, Nedunchezhian R. A improved incremental and interactive frequent pattern mining techniques for market basket analysis and fraud detection in distributed and parallel systems. Indian Journal of Science and Technology. 2015 Aug; 8(10):1–12. Crossref