International Journal of Engineering & Technology, 7 (2.7) (2018) 972-975

International Journal of Engineering & Technology

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

Research Paper

Regular frequent crime pattern mining on crime datasets

G. Vijay Kumar [1] *, M. Sreedevi [2], G. Vamsi Krishna [2], N. Sai Ram [2]

[1] Professor Department of electronics and computer science engineering, KLEF (Deemed to be university).
[2] Department of electronics and computer science engineering, KLEF (Deemed to be university).
**Corresponding author E-mail: gvijay_73@kluniversity*

Abstract

The objective of violation information mining is to comprehend different violation designs in criminal conduct in request to foresee vio-la-tions and expect criminal movement to stay away from the violation not to happen. Foreseeing violation is one of the worldwide difficulties looking by Law authorization office and it requires tireless endeavors with a specific end goal to limit. In this paper we are presenting anoth-er violation design called general incessant violation design which happens frequently at certain time interims utilizing vertical information arrange additionally fulfills descending conclusion property. Violation designs were not characterized by insights and its distinguishing proof is some-thing other than checking and abridging violations that are comparable in attributes and additionally area on a guide. Violation design is a gathering of at least one violations answered to or on the other hand found by the police.

Keywords: *Crime Pattern; Crime Dataset; Regular-Frequent Patterns; Vertical Data.*

1. Introduction

These days the most persuasive apparatus to snatch designs and connections inside in us information is Data mining [1]. The information mining system can likewise be utilized as a part of conventional instructive system [2]. The technique for information mining is additionally utilized as a part of stock trade in organizations for recovering expansive sum of data [3]. A Crime design is a gathering of at least two violations answered to or found by police that are one of a kind. Information mining can be utilized to demonstrate violation discovery issue. A violation design is recognized through an efficient, deductive diagnostic process, along these lines imparted to police operators by means of some type of bulletin [4]. So here in this paper we are following Regular Frequent Pattern Mining approach which is productive keeping in mind the end goal to get the violation designs from substantial informational collections. Customary Frequent Pattern Mining depends on event recurrence and event conduct of an example. At the point when the event recurrence is more noteworthy than or equivalent to client given edge esteem and event conduct is not exactly or equivalent to the client given normality edge esteem then that example is called regular frequent pattern. With a specific end goal to discover such violation designs from vast information sets we are utilizing the Regular Frequent Pattern Mining System. Visit design mining is one of the dynamic examine regions in information mining to discover fascinating examples. Visit design mining fundamentally relies upon the help check i.e. number of times an example shows up in the database.

Occurrence recurrence of an example as well as likewise event conduct of an example might be dealt with as vital criteria to quantify the intriguing quality of the design in a few applications on the web. The centrality of an item may rely on other event qualities for example, consistency of the example. For instance, in a market the client might be occupied with regularly sold things which are sold at standard interims (or) to upgrade a web architecture the site overseer might be concerned in more frequently prominent pages at customary interims. From the above cases we watched that the event conduct of an incessant example at standard interims assumes a critical part among different applications like organize observing, broadcast communications or sensor systems.

Affiliation Rule Mining is an imperative information mining show proposed by Agrawal, et al. in 1993. Apriori approaches have been utilized in the past for this reason and have been distinguished as having a few deficiencies. Affiliation Rule Mining is tied in with finding visit examples of thing sets. It is for the most part utilized for showcase crate investigation to discover how things are bought by clients. Mining incessant patterns [5,6] was first presented by Agrawal, et al. in 1993 need k-number of outputs to produce k-thing set. In [7] presented general continuous example mining in 2008 over value-based data-bases. . As of late occasional continuous patterns [8] or incessant standard patterns [9] is assuming a fundamental part in information mining research on the grounds that of the event recurrence of the example alongside event conduct of an ex-ample i.e. consistency or interim. There is no such example which mines violation designs utilizing Regular Frequent Pattern calculation.

Numerous researchers [10–12] are exploring methods for handling violation and enhancing existing calculations used to determine violation designs, keeping in mind the end goal to help open well-being and security organizations in accomplishing their goal of hindering violation and advancing residents wellbeing. The most imperative what's more, aggressive territory in information mining and learning disclosure look into is to mine different intriguing patterns [13]. In the paper [14] we came to realize that consistency of thing set and recurrence of thing set both are similarly imperative on any sort of database. An-other calculation has been proposed RFRP to locate the general successive examples of a thing set. In this paper [15] event recurrence of the de-sign is estimated as a critical factor and is utilized for estimating the intriguing quality of the example in numerous applications. Considering min support client given edge esteem a consistent continuous example is found. A Series Finder calculation for distinguishing designs is

Copyright © 2018 G. Vijay Kumar et. al. 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.

proposed by Tong Wang which distinguishes examples of violation conferred by the same person. In our paper we contrasted our outcomes and RP-tree calculation which produces a total arrangement of consistent examples in a database for a client given consistency limit value [16]. Apriorism is a calculation for affiliation rules. To get the help tally of the thing sets this apriori calculation filters the database for sometimes [17]. In the paper [18] we came to know how to mine incessant thing sets utilizing vertical arrangement. Our calculation fulfills the descending conclusion property. Any subset of an incessant thing set must be visit this is called descending conclusion property. A review of different successive example mining calculations has been done to discover violation designs in [19] this paper took after three methodologies with competitor age, without hopeful age and vertical design approach. It makes a difference the specialists to get a thought regarding the continuous example mining calculation in different applications.

2. Problem description

The ideas of consistent continuous example mining are characterized in this segment and the essential meanings of the issue to get finish set of normal continuous examples in violation database. Let I = {i1, i2, ...in} be an arrangement of things. A set X = {ij, .ik} subset(I), where j ≤k and j,k Є [1,n] is known as an itemset or design. Violation c = (cid, Y) is where cid is a Violation id and Y is an example. Give the quantity of things in Y a chance to be the size of t size(t). An exchange database DB over I is a set of Violation exchanges C = {cl,. . . . . cm}, m =| DB| is the size of DB, i.e. add up to number of Violation exchanges in DB. On the off chance that X subset (Y), which implies that c contains X or X hap-pens in c and meant as cj X, jϵ[1,m]. In this way, CX ={cj X, . . . ck X}, j ≤ k and kϵ[1,m] is the arrangement of all Violation exchanges where design X happens in DB. point the normality of a visit design X can be indicated as Reg(X) = max(pl X, . . . pr X}. A continuous example X is said to be standard incessant if its consistency is not exactly or equivalent to client given least consistency limit i.e., λ

3. Methodology

3.1. Regular frequent pattern mining (RFCPM) way to deal with crime pattern mining.

Think about Table 1 as a common C Data-base sets (CDB) with downright violation qualities, which involves episode area, suspect and casualty data, day of the time, week and date, weapons utilized and violation scene status and so forth. The initial step is to oversee and distinguish the qualities of premiums that are accessible in the violation examination report set Table1. At that point the database is questioned to remove the properties of enthusiasm, making the Crime the Crime Exchange Data Base (CTDB) to be mined.

To outline how Regular Frequent Pattern mines violation information we consider a straightforward illustration appeared in Table 2. Assume we passed a question (including area data furthermore, relating violation occurrences that happened there) to acquire the Crime Transaction Data Base (CTDB) which is additionally encoded or changed with Violation Transaction Identification (CTID) being extraordinary identifier for every violation exchange and Crime Incident Data (CII). Violation Incident and violation area has been encoded with shortened forms, so as to less-en preparing time.

Presently we have to mine consistent incessant violation designs with respect to area from Table 2. Right off the bat, we have to get the help and consistency include for the violation episodes Table 2 as introduced in Table 3. The help means that how visit a violation thing set is watched. Also, the consistency tally is an sign of how consistently a violation thing set has happened. By following the RFCPM approach we change over our table into vertical information arrange and get the help and normality check esteems.

Table 1: Crime Transaction Database

Crime locations1	Crime incidents1
Locat1	Kidnapp, rapee
Locat2	Kidnapp, rapee, murderr, robb
Locat3	Murderr, Robb, Burglary, Kidnapp
Locat4	Kidnapp, robb, burglary, murderr
Locat5	Rapee, murderr, burglary
Locat6	Kidnapp,rapeemurderr, robb
Locat7	Shoplifting, kidnapp
Locat8	Rapee, robb, murderr, kidnapp
Locat9	Kidnapp, rapee, robb, murderr
Locat10	Robb, murderr, kidnapp, rapee

Table 2: Encoded CT dB

CL11	CII1
Ct11	Kp1, Rp1
Ct21	Kp1, Rp1, Md1, rb1
Ct31	Md1, rb1, bg1, kp1
Ct41	Kp1, rb1, bg1, md1
Ct51	Rp1,md1,bg1
Ct61	Kp1, rp1, md1, rb1
Ct71	Sl1, kp1, ar1
Ct81	Rp1, rb1, md1, kp1
Ct91	Kp1, rp1, rb1, md1
Ct101	Rb1, md1, kp1, rb1

3.2. RFCPM-algorithm

Input: DB, λ, δ
Output: Complete set of regular frequent pat-terns.
Procedure:
Let Xi subset (I) be a k-itemset PX i=0 for all Xi For every Xi
Refresh Sup
If Sup(Xi)>=δ
Discover the time of Xi
PX i= PX i+1-PXi
Reg(Xi) = max(PXi)
if reg(Xi)<=λ
Xi is a consistent successive itemset
Else
Erase Xi
Else
Erase Xi
Rehash

Presently we prune our table in light of least help edge esteem (δ = 5) and most extreme consistency esteem (λ = 4). We acquire our outcome in Table 4. The violation things which don't fulfill the base help and most extreme consistency esteems would be disposed of. This procedure proceeds until the point that a Regular Frequent is gotten. Presently affiliations are made between kp ,rp, md, rb as they fulfill the condition and are reasonable to get length-2 thing sets which are appeared in Table 5. Correspondingly again the affiliations are made between the 2-thing sets which fulfill the condition and a length-3 thing set is acquired which is appeared in Table 6. Table 6 both help and consistency fulfill the client given limit esteems, so we have to proceed with this procedure until the point that no applicant is produced for mining normal visit designs in data-base. What's more, finally at long last we get the general regular example which fulfills both the conditions which is appeared in Table 7.

Table 3: Converting Into Vertical Data Format Following RFCPM Approach

Crime items1	CTID1	Support	Regularity1
Kp1	1,2,3,4,6,7,8,9,10	9	1
Rp1	1,2,5,6,8,9,10	7	3
Md 1	2,3,4,5,6,8,9,10	8	2
Rb1	2,3,4,6,8,9,10	7	2
Bg1	3,4,5,	3	5
Sl1	7	1	7
Ar1	7	1	7

Table 4: Crime Item Set Table with Support (Δ = 5) and Regularity (Λ = 4) for 1-Item Set

Crime items1	CTID1	Support	Regularity1
Kp1	1,2,3,4,6,7,8,9,10	9	1
Rp1	1,2,5,6,8,9,10	7	3
Md1	2,3,4,5,6,8,9,10	8	2
Rb1	2,3,4,6,8,9,10	7	2
Bg1	3,4,5	3	5
Sl1	7	1	7
Ar1	7	1	7

Table 5: Crime Item set Table with Support (Δ = 5) and Regularity (Λ = 4) for 2-Itemset

Itemset1	CTID1	Support	Regularity1
Kp1,rp1	1,2,6,8,9,10	6	4
Kp1,md1	3,4,6,8,9,10	5	3
Kp1,rb1	2,3,4,6,8,9,10	6	2
Rp1,md1	2,5,6,8,9,10	5	3
Rp1,rb1	2,6,8,9,10	5	4
Md1,rb1	2,3,8,10	4	5

Table 6: Crime Item Set Table with Support (Δ = 5) and Regularity (Λ = 4) for 3-Itemset

Itemset1	CTID1	support	Regularity1
Kp1,rp1,md	2,6,8,9,10	5	4
Kp1,rp1,rb1	2,6,8,9,10	5	4
Kp1,md1,rb1	2,3,4,6,8,9,10	7	2
Rp1,md1,rb1	2,3,6,8,9,10	7	3

Table 7: Crime Item set Table with Support (Δ = 5) and Regularity (Λ = 4) for 4-Itemset

Itemset1	CTID1	Support	Regularity1
Kp1,rp1,md1,rb1	2,6,8,9,10	5	4

Along these lines the thing set {kp1, rp1, md1, rb1} fulfill the client help and consistency, thus this thing set is the general continuous violation design.

4. Test results

Our experimentation comes about are performed over manufactured dataset (T1014D100K) and genuine datasets (Chicago violation database). We contrast the consequences of RP-tree and our calculation RPFM which demonstrates that our calculation is more proficient and quick in finding the normal frequent crime design. All analyses are done in java on windows XP containing 2.7 GH with 2 GB of primary memory.

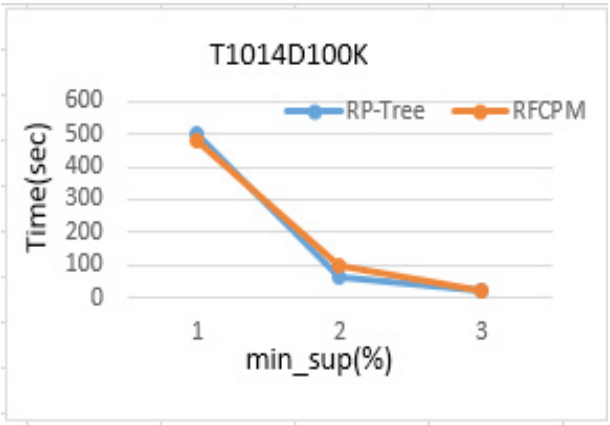


Fig. 1 :(A). Execution Time Over T1014100k.

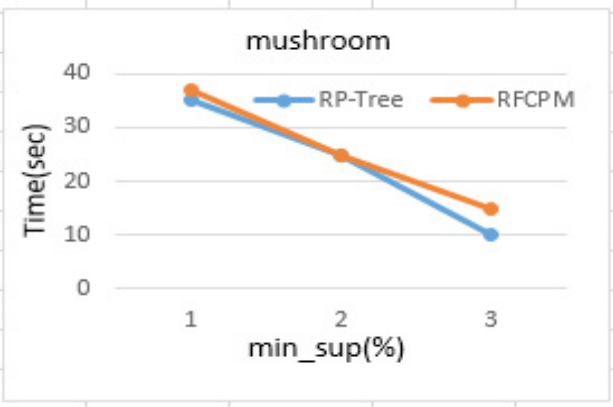


Fig. 1: (B). Execution Time Over Mushroom.

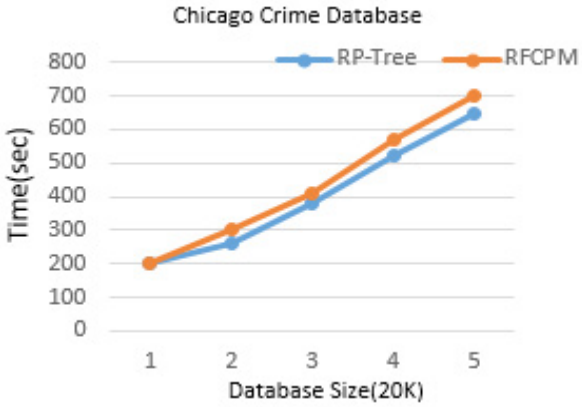


Fig. 2: (A). Execution Time Over Chicago Crime Database.

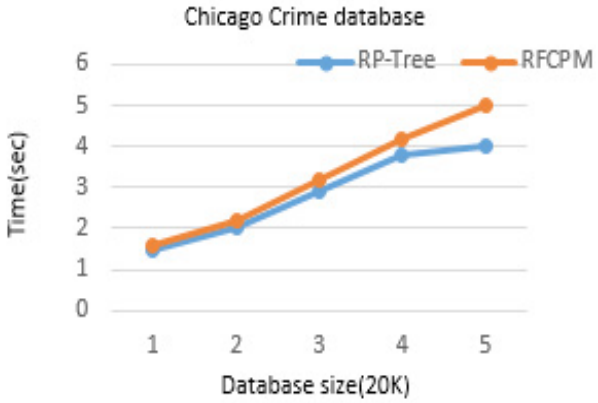


Fig. 2: (B). Memory Usage over Chicago Crime Database.

Here comparison of results is done by using rp tree on T1014D100K where it has 20k transactions in table 1(a) and table 1(b). by seeing the above graphs, we can say that the RFCM algorithm will take relatively same time

Total execution time is specified by time and memory and the memory used/required by the database. Table 2(a) and table 2(b) shows the scalability and performance of the algorithm. Our algorithm performance is more efficient, and we included the vertical data format, so it is scalable and efficient over big databases.

5. Conclusion

In this paper we introduced another calculation RFCPM which mines to get a normal successive example in violation databases utilizing vertical information arrange. Discovering normal successive designs is vital as the data would be substantially more helpful. The benefits of vertical in-formation organize that necessities straight forward computations like associations, crossing points, subtraction and so forth are utilized by our calculation. This strategy is versatile and productive over extensive databases. Our calculation is exceptionally helpful in discovering customary successive examples not just for violation data-base yet in addition in different applications

References

[1] Vedanayaki M. A study of data mining and social network analysis. Indian Journal of Science and Technology. 2014 Nov; 7(S7):185–187.

[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.

[3] Azad N, Ranjbar V, Khani D, Moosavi ST. Information disclosure by data mining ap-proach. Indian Journal of Science and Technology. 2012 Apr; 5(4):2593–2602.

[4] Bruce C, Santos RB. Crime pattern defini-tions for Tactical Analysis. Standards Meth-ods 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.207–216.

[6] Agrawal R, Srikanth R. Fast algorithms for mining association rules. VLDB; 1994. p. 489–499.

[7] G. Vijay Kumar, M. Sreedevi, NVS Pavan Kumar "Mining Regular Patterns in Transac-tional Databases using Vertical Format" In- ternational Journal of Advanced Research in Computer Science, Volume (2), Issue (5) 2011.

[8] G. Vijay Kumar, Dr. V. Valli Kumari "MaRFI: Maximal Regular Frequent Itemset Mining using a pair of Transaction-ids", International Journal of Computer Science & Engineering Technology, Volume 4 No. 7, 2013.

[9] G. Vijay Kumar, Dr. V. Valli Kumari "IncMaRFI: Mining Maximal Regular Fre-quent Itemsets in incremental databases" In-ternational Journal of Engineering Science and Technology, Volume (5) No (8) 2013.

[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):50–56.

[11] Dandu S, Deekshatulu B, Chandra P. Im-proved algorithm for frequent item sets mining based on apriori and fp-tree. Com-puter 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 Interna-tional Journal. 2013; 2(4):2043–2048.

[13] Sreedevi M, Reddy LSS. Mining regular closed patterns in transactional databases. 2013 7th International Conference on Intel-ligent Systems and Control (ISCO); 2013. p. 380–383,

[14] G. Vijay Kumar, V. Valli Kumari, Parallel and distributed frequent-regular pattern mining using vertical format in large databases. IEEE Xplore, IET; 2012. P. 110–114.

[15] Rashid MM, Karim MR, Jeong BS, Chai HJ. Efficient mining regularly frequent pat-terns in transactional databases. Springer Lecture Notes in Computer Science; 2012. p. 258–271.

[16] Rashid MM, Karim MR, Jeong BS, Chai HJ. Efficient mining regularly frequent pat-terns in transactional databases. Springer Lecture Notes in Computer Science; 2012. p. 258–271.

[17] Tanbeer SK, Chowdhury FA. RP-Tree: A tree structure to discover regular patterns in transactional database. IDEAL 2008, Volume 5236 of the series Lecture notes in Computer Science; 2008. p. 193–200.

[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):264–275.

[19] Lee JR, Kim SA, Yoo JW & Kang YK (2007), The present status of diabetes education and the role recognition as a diabetes educator of nurses in korea. Diabetes Research and Clinical Practice 77, 199–204.