International Journal of Computational Intelligence Research
ISSN 0973-1873 Volume 13, Number 6 (2017), pp. 1405-1417

http://www.ripublication.com

Frequent Pattern Model for Crime Recognition

Shivangee Agrawal
Department of CSE and IT
Madhav Institute of Technology and Science Gwalior, India.

Vikas Sejwar
Department of CSE and IT
Madhav Institute of Technology and Science Gwalior, India

Abstract

Crimes are a social pain and cost our society with much suffering in several ways. Data mining techniques can be used for detecting patterns from which the occurrence of crime can be analysis. Pattern based approaches can help crime experts for discovering new patterns according to criminal activity. For discovering pattern we use frequent pattern mining approach, FP- Growth model from which the pattern of crime occurrences can be detected. Also presents the execution of algorithm step wise with their result. We also show that the performance of FP-Growth algorithm can become better by applying optimization like particle swarm optimization.

Keywords: Crime mining; Frequent pattern; FP-Growth; particle swarm optimization (PSO)

I. INTRODUCTION

Historically solving crimes has been the sanction of law enforcement specialists and the criminal justice. With the rapid increase of computerized systems for tracking crime, the law enforcement officers and detectives have been helped by computer data analysts to heighten up the process of detecting and solving crimes. Here we are

trying to combine computer science and criminal justice for developing a data mining paradigm that can help crack crimes faster.[1]

Crime occurs in many forms which are categorized by police in two categories as major crime or volume crime. From many crimes most of crimes are of the type of famous and well known crimes such as armed robbery, murder and non-date rape which can be one-offs and serial. Serial crimes can easily be linked together because the similarities between description of offenders or in terms of their strategy and also because they occurrence has low volume in comparison with one-offs crime. There is a team of detectives allocated to carry out the investigation of major crimes. In distinction measure crimes such as burglary and shoplifting are far more common. They are frequently serial in environment as offenders go on to consign many such crimes. Crimes like domestic burglary offences committed by different persons are very much similar and it is rare to have a picture of the offenders (Adderley and Musgrove, 2001) [2].

Table 1 shows the crime's classification (Chen et al., 2003)[3].

Table I. Crime types at different levels. Source: (Chen et al., 2003)[3]

Crime Type	**Description**
Traffic disobedience	Driving under the effect of alcohol, killing/personal injury/property harm, traffic problem.
Sex crime.	Sexual assaults.
Fraud.	Forgery and counterfeiting, frauds, embezzlement, identity deception.
Arson.	Arson on buildings.
Gang/drug offences.	Narcotic drug offences (sales or possession).
Violent crime.	Criminal homicide, armed robbery, aggravated assault, other assaults.
Cyber crime	Internet frauds, illegal trading, network intrusion/hacking, virus spreading, hate crimes, cyber piracy, cyber pornography, cyber-terrorism, theft of confidential information.

II. CHALLENGES AND APPROACH OF CRIME DETECTION

The most barely credible challenge for the police department is investigating crime with the current technologies. Large amount of information and massive volume of

records are handled by Police organization. They need highly developed technologies to handle crime. There is a need of an ideal crime recognition system to identify the pattern of crime quickly for detection of crime pattern rapidly and action. Discovery and exploitation of knowledge can be helped by Data mining technique. Data preprocessing is the most important step in data mining processes. The main focus lies on developing a crime recognition system that can help the police in

1. Performing the crime analysis and detection of crime patterns.
2. Prevention and reduction of crime by formulating strategies with the help of information provided.
3. For reducing the occurrences of alike incidence, identification and analyzation of common crime patterns can be done[4].

III. RELATED WORK

S. Sathyadevan (2014) et al. tested the accuracy of classification and prediction by applying on different test sets. Classifications are done based on the Bayes theorem and trained numerous new articles and build a model by using apriori algorithm [5].

A. Kondaveeti (2011) have generated spatial data from GIS data and apply data mining techniques for classifying the interventions based on geographic data after preprocessing of spatial data. Supervised spatial association rules were also discovered [6].

L. Cunhua(2010) et al. done text processing by applying event ontology. They compare their performance by using two methods namely event ontology-SVM based and SVM based. Also study the various method of Web crime mining [7].

Isuru Jayaweera (2015) et al. presents a web based intelligent crime analysis system for analyzing large amount of data and performs techniques like hotspot detection, visualization of crime patterns and comparison of crime. They also use graphical user interface for representing graphs and diagrams for displaying the results which makes the task of crime analyzing easy [8].

Mugdha Sharma [2014] et al. for detecting suspicious criminal activity e-mails proposed an Enhanced decision Tree Algorithm. For generating better and faster Decision Tree, they applied enhanced feature selection method and attribute importance factor to improved ID3 Algorithm. They are trying to classify emails in various criminal activities [9].

Table II. Summary of researches

Performed researches	**Techniques**	**Tasks**	**Research Gaps**	**Research challenges**
[5]	Apriori algorithm	Analization of crime pattern	False detection	Better precise detection
[6]	Association rule and spatial techniques	Analyzing geographic data and patterns of crime	No concerned performance	Improve performance
[7]	SVM algorithm and Event ontology	Analyzing the crime patterns	No flexibility of visualization and crime model	Improve visualization and model the crime future attacks
[8]	Data base handler, crawler, entity extractor,doc ument classifier and graphical user interface	Extraction of element data and analyzing the frequency and crime	No creation of crime model and no crime prediction	Model the crime future attacks and improved performance
[9]	An improved ID3 algorithm	Classifying emails with respect to criminal activities	No collection of data,crime prediction	Model the crime future attacks and collect data

IV. FP-GROWTH

Frequent pattern mining plays an important role in mining association rules, generating sequential pattern, and many other important task of data mining. Mainly it is the process of extracting data from huge database. The support based framework is the most commonly used framework for mining process. Frequent pattern mining has many algorithm for generating frequent patterns (occurs frequently in the transaction of database), out of which FP-Growth is one of them [10]. FP-Growth generates frequent patterns without generating candidate sets, which overcomes the drawback of apriori algorithm [11]. It is a top down approach and uses tree data structure called FP-Tree [12]. The generation of frequent pattern is performed in two pass which is explained by the following example.

Consider a crime transaction database (CTDB)

Table III. Crime transaction database (CTDB)

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

In table III Crime transaction database (CTDB) is given and its encoded form is given in table IV. We created the encoded form just for simplicity.

Table IV. Encoded CTDB

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

In the first pass, the CTDB is scanned and support count for each item set is calculated.

Table V. Support count for crime item-sets

Crime items	Support
kp	8
rp	7
sl	1
rb	5
bg	3
ar	1
md	6

Now, all the crime item-sets, say C, whose support value is smaller than the minimum support threshold value, say λ=2 are taken as infrequent and discarded from the table V and _shown in table VI. This is called pruning phase.

Table VI. Pruning for each crime item-sets

Crime Items	Support
kp	8
rp	7
md	6
rb	5
bg	3

Now, sort all the frequent crime items according to their support count in descending order shown in table VII.

Table VII. CTDB' after pruning

CTID	CII
CT1	{kp, rp}
CT2	{rp, md, rb}
CT3	{kp, md, rb, bg}
CT4	{kp, rb, bg}
CT5	{rp, md, bg}
CT6	{kp, rp, md, rb}
CT7	{kp}
CT8	{kp, rp, md}
CT9	{kp, rp, rb}
CT10	{kp, rp, md}

In the second pass, the FP-Tree is constructed from the CTDB'

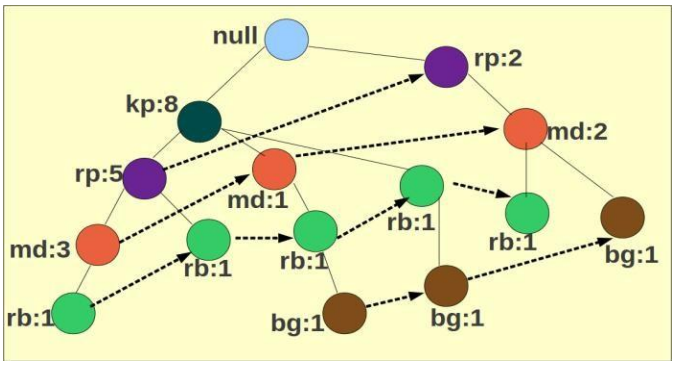
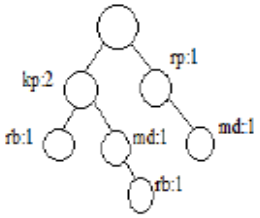
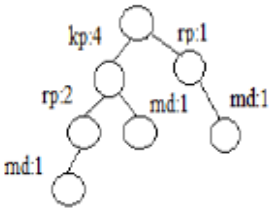
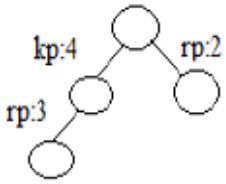
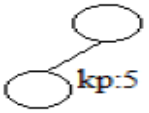


Fig.1. FP-tree after reading CT10

For construction of FP-Tree from table VII, first create a root node named null and read first transaction CT1 {kp, rp}, make the two nodes kp and rp starting from the root node and set the counts of kp and rp to 1. Read CT2 {rp, md, rb} and make the another path from the root node. It is to be noted that CT1 and CT2 have disjoints path even they have a common node named rp. The process is repeated for every CTj, j belongs to {1, 2, 3…, 10} for generating the FP-Tree. Nodes sharing the same item sets are linked with pointers as shown in figure 1.

Table 8. step by step execution

Itemset	Conditional pattern base	Frequent pattern tree	Frequent pattern
bg	{kp, rb:1} {kp, md, rb:1} {rp, md:1}		{bg}, {rb, bg}, {kp, rb, bg}, {md, bg}, {kp, bg}
rb	{kp, rp, md: 1} {kp, rp: 1} {kp: 1} {kp, md: 1} {rp, md: 1}		{rb}, {md, rb}, {rp, md, rb}, {kp, md, rb}, {rp, rb}, {kp, rp, rb}, {kp, rb}
md	{kp, rp: 3} {kp: 1} {rp: 2}		{md}, {kp, md}, {rp, md}, {kp, rp, md}
rp	{kp: 5} {rp: 2}		{rp}, {kp, rp}
kp			{kp}

V. RESULT WITH SIMULATION

This study's experiment was conducted in the environment of Microsoft Windows 7 using 1.60GHz hard disk and 512MB RAM. The algorithm was coded in MATLAB.

Figure 2 shows the FP-tree constructed from CTDB database.

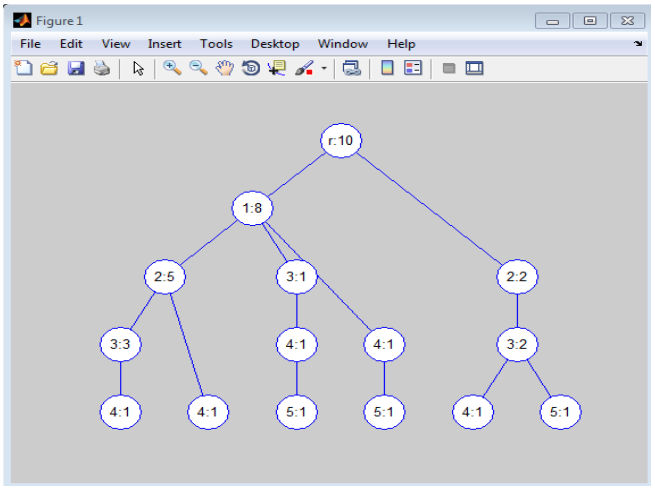


Figure 2

After FP-tree construction, conditional pattern base are generated of each item and then frequent pattern tree for each item is constructed which are shown below.

Figure 3 shows frequent pattern tree of item set bg:burglary

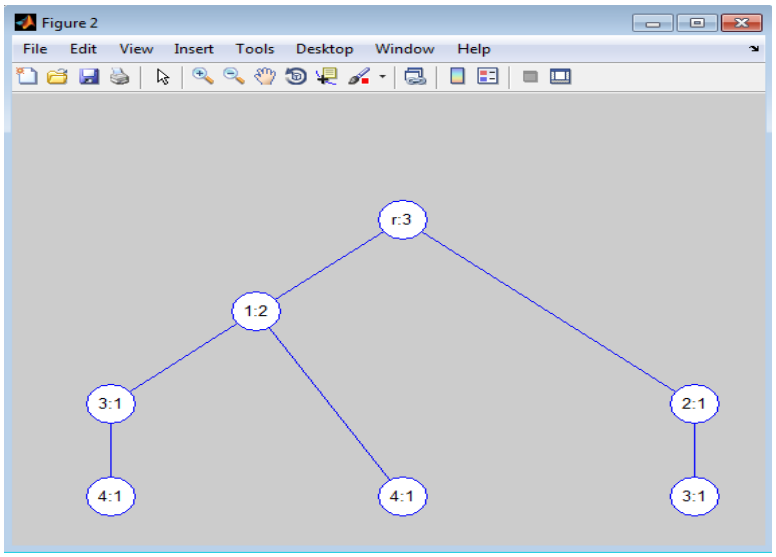


Figure 3

Frequent pattern regarding burglary are {{bg}, {rb, bg}, {kp, rb, bg}, {md, bg}, {kp, bg}}

Figure 4 shows frequent pattern tree of item set rb: robbery

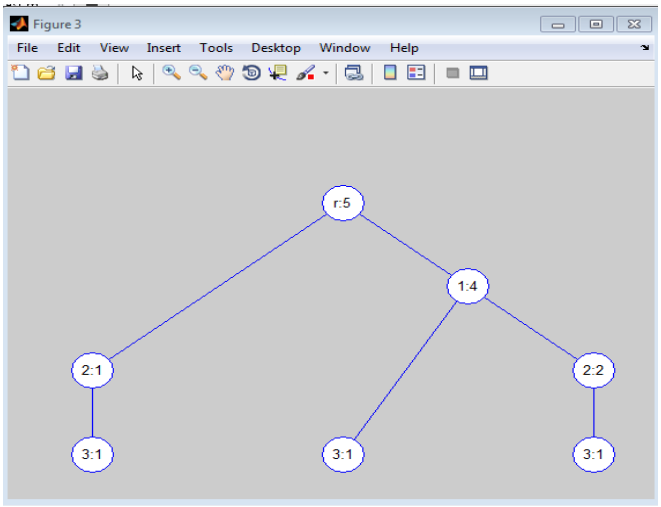


Figure 4

Frequent pattern regarding robbery are {rb}, {md, rb}, {rp, md, rb}, {kp, md, rb}, {rp, rb}, {kp, rp, rb}, {kp, rb} Figure 5 shows frequent pattern tree of item set md: murder

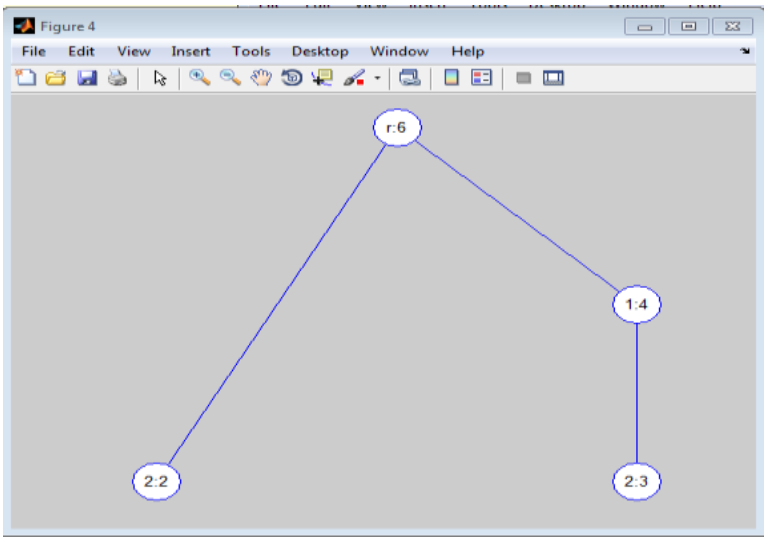


Figure 5

Frequent pattern regarding murder are {md}, {kp, md}, {rp, md}, {kp, rp, md}.

Figure 6 shows frequent pattern tree of item set rp: rape

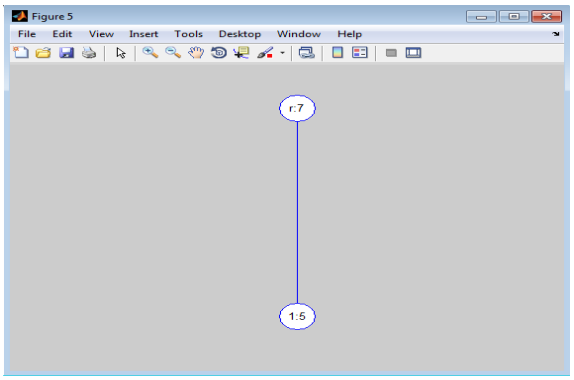


Figure 6

Frequent pattern regarding rapping are {rp}, {kp, rp}

Figure 7 shows frequent pattern tree of item set kp: kidnapping

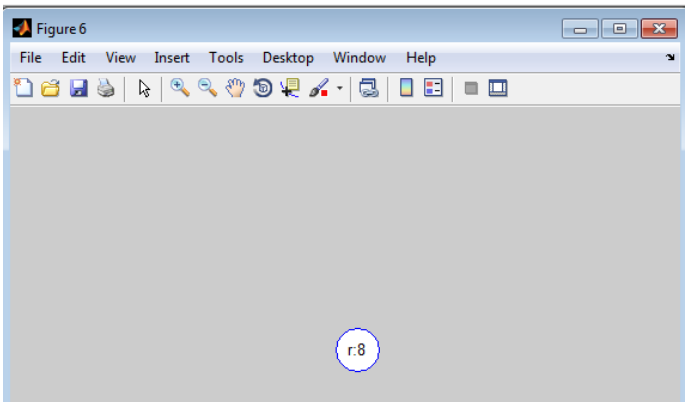


Figure 7

Frequent pattern regarding kidnapping are {kp}

Figure 8 shows the pattern generated by FP-growth algorithm

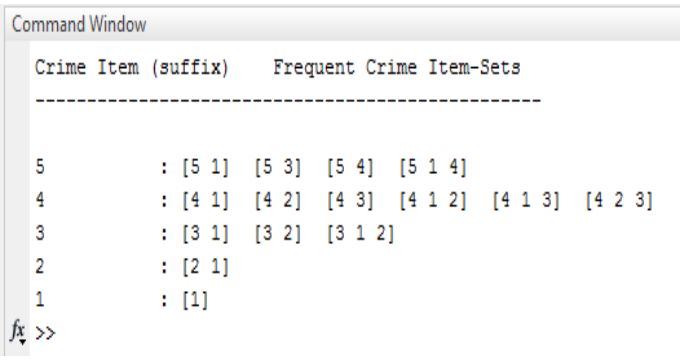


Figure 8

For simplicity we have given numeric value to itemset in code as kp:1, rp:2, md:3, rb:4, bg:5.

Since we have seen that results using FP growth algorithm requires lot of time for execution, and requires more memory for storage of data. Also some patterns are redundant and algorithm becomes complex with increasing items in data set.

For overcoming these drawbacks we can use optimization with FP-Growth algorithm. Results are not shown here.

Particle Swarm Optimization (PSO)[13]:

In 1995 Eberhart and Kennedy introduced particle swarm optimization. In particle swarm optimization particle refers to the each individual of the population. When the initialization of particle is done, each particle updates its position and velocity according to their local best position (p_{best}) and global best position (g_{best}) of all particles.

Particle swarm optimization algorithm steps:

1. Initialization of position and velocity of particles with randomly chosen value
2. Find fitness value of each particle according to fitness function.
3. If fitness value of particle i is better than the p_{best} then update p_{best} = fitness value
4. If p_{best} is updated and it is better than current g_{best} then update g_{best} = p_{best}
5. Update velocity and position of particle according to equation (a) and (b)
6. If the best fitness value or stopping criteria is reached then stop the process, otherwise repeat the process from step 2.

   For updating particle's velocity

$$V_i[t+1] = w.V_i[t] + c1.rand1\ (p^{i,best}[t] - p^{i,current}[t]) + c2.rand2\ (p^{g,best}[t] - p^{i,current}[t]) \quad \ldots.(a)$$

   For updating particle's position

$$p^i[t+1] = p^i[t] + V_i[t+1] \quad \ldots.(b)$$

VI. CONCLUSION

In this paper we studied frequent pattern mining and one of its approach, FP-Growth model and apply FP-Growth algorithm on crime transaction database. In this paper stepwise execution of FP-Growth algorithm is shown with their results. Particle swarm optimization can also be used with FP-Growth algorithm by which the drawbacks of frequent pattern growth algorithm can be removed.

REFERENCES

[1] Hsinchun Chen, Wingyan Chung, Yi Qin, Michael Chau, Jennifer Jie Xu, Gang Wang, Rong Zheng, Homa Atabakhsh, ―Crime Data Mining: An Overview and Case Studies‖, AI Lab, University of Arizona, proceedings National Conference on Digital Government Research, 2003, available at: http://ai.bpa.arizona.edu/

[2] Adderley R.William and Musgrove Peter, (2001), ―Police crime recording and investigation systems: A user's view‖, An International Journal of Police Strategies and Management, Vol. 24 No. 1, pp. 100-114.

[3] Reza Fadaei-Tehrani, Thomas M. Green, (2002) ―Crime and society‖ International Journal of Social Economics Volume 29 Number 10 pp. 781-795.

[4] M. Ramzan Begam, Dr. P.Sengottuvelan and T. Ramani ―Survey: Tools and Techniques implemented in Crime Data Sets‖ IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 6, June 2015.

[5] S. Sathyadevan, M. Devan, and S. Surya Gangadharan, ―Crime analysis and prediction using data mining,‖ in Networks Soft Computing (ICNSC), 2014 First International Conference on, Aug 2014, pp. 406–412.

[6] A. Kondaveeti, G. Runger, H. Liu, and J. Rowe, ―Extracting geographic knowledge from sensor intervention data using spatial association rules,‖ in Spatial Data Mining and Geographical Knowledge Services (ICSDM), 2011 IEEE International Conference on, June 2011, pp. 127–130.

[7]. L. Cunhua, H. Yun, and Z. Zhaoman, ―An event ontology construction approach to web crime mining,‖in Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on, vol. 5, Aug 2010, pp. 2441–2445.

[8]. Isuru Jayaweera, Chamath Sajeewa, Sampath , Adeesha Wijayasiri―Crime Analytics: Analysis of Crimes Through Newspaper Articles‖ 2015 IEEE.

[9]. Mugdha Sharma ―Z - CRIME: A Data Mining Tool for the Detection of Suspicious Criminal Activities Based on Decision Tree‖ 2014 IEEE.

[10]. J. Han, H. Pei, and Y. Yin. Mining Frequent Patterns without Candidate Generation. In: Proc. Conf. on the Management of Data (SIGMOD'00, Dallas, TX). ACM Press, New York, NY, USA 2000.

[11]. Agrawal, R. and Srikant, R. 1994. Fast algorithms for mining association rules. In Proc. 1994 Int. Conf. Very Large Data Bases (VLDB'94), Santiago, Chile, pp. 487–499.

[12]. Jiawei Han and Micheline Kamber, Data Mining: Concepts and Techniques. 2nd edition, Morgan Kaufmann, 2006.

[13]. Eberhart RC, Shi Y (2001) Particle swarm optimization: developments, applications and resources, in Proceedings of the IEEE Congress on Evolutionary Computation (CEC), Seoul, Korea.