ISSN: 2321-7782 (Online)
Volume 2, Issue 5, May 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Improve Frequent Pattern mining in Spatial Database with FPAR and TFP

Krishna R. Goswami[1]
Master of Engineering Pursing in computer science department
Parul Group of Institute and Technology
Baroda – India

Ankur Shah[2]
Assistant Professor in computer science department
Parul Group of Institute and Technology
Baroda – India

Abstract: Spatial database is defined data related to space such as river, mountain, lack. Mining the spatial database is done with help of Association rule and clustering. Association rule define several techniques like apriori, FP etc. In this paper we improving the existing mining algorithm for the finding frequent pattern from spatial dataset. The proposed algorithm FPAR and TFP is improvement of the FP algorithm. In these techniques, used EVEN number to represent spatial object. Our main aim is to find frequent occurrence in minimum time.

I. INTRODUCTION

Nowadays, large quantities of data are being accumulated. The amount of data collected is said to be almost doubled in few months. Seeking knowledge from massive data is one of the most desired attributes of Data Mining Usually there is a huge gap from the stored data to the knowledge that could be construed from the data. This transition won't occur automatically, that's where Data Mining comes into picture, the database become very large, its complex to extract the meaningful information. So to overcome that data mining is used. Data mining is process of discovering useful knowledge from data [6].

This paper focus on spatial database which consist the data related to space. There are so many data which are frequently occurring in spatial data base so to find the frequent occurrence of that. We use Association rule mining technique and clustering. Here we focus on Association rule mining (ARM). The techniques of ARM are apriori, fp growth, éclat etc. From that, apriori is the basic technique but it generates more number of candidate set. In frequent pattern tree this problem is solved. So in this paper to find frequent object in space they use frequent pattern. This frequent pattern two sub types FPARM (Frequent Positive Association Rule Mining) and TFP (Transaction frequent pattern)[1][2]. In these techniques, it seen two different parameters are used, one is prime number and other Fibonacci series in previous work. I implement both number and find the different time complexity for same spatial database. I invent EVEN number which is better than these two numbers.

II. FRAMEWORK

The system work on finding frequent occurrence of space in geographical area. In this system to find frequent pattern FPAR (frequent positive Association rule) and TFP (Transaction frequent pattern) are used.

Here, we performed the test on a sample real time data base of 2000 Indian cities taken as reference to validate the framework in our study of spatial database system. It shows a sample of some spatial objects for 2000 Indian cities. The spatial objects are as Museum (A), Zoo (B), Lake (C), Monument (D), River (E), Forest (F), and Hill (G). Our main aim is to find the frequent occurrence of object. The object less than minimum supports are pruned. Analysis Processes has the following steps.

Step 1: Obtain Sample Spatial Dataset.

Step 2: Build Ordered list of objects in descending order of their frequencies.

Step3: Mapping Ordered List in form on numerical representation.

Step 4: Build a TFP Tree using numerical representation.

Step 5: Find frequent patterns and validate it against their respective support count

TID	Reference City	Positive Object
1	Bhavnagar	C,D,A,B,E,G
2	Rajkot	B,F,A,C,D
3	Junagadh	A,C,D,B,E,G
4	Vadodara	A,C,D,B,E,G
5	Ahmadabad	B,F,A,C
6	Surat	C,A,B
7	Dahod	D,B,F,E,A
8	Surendranagar	A,C,D,B,E
9	Anand	A,C,D,B,E,G
10	Jetpur	C,D,B,F,G
11	Gondal	A,C,D,B
12	Patan	E,A,D,B

Table 1: Sample Spatial Dataset

Each transaction is scanned once to find frequent object. For example, in TID (1) {C, D, A, B, E, G} is a transaction list. So the object which is less than the minimum support is pruned. Now TID (1) can be viewed as a new transaction {A, B, C, D, E} which are arranged in descending order of their frequencies. The ordered list of each transaction is mapped using even based data transformation technique as product value for each order list to reduce the size of transaction database which is shown in TABLE II.

We store the product value for each order list. Example: Suppose we take sample dense database of #1 where Bhavnagar is a reference city and its corresponding spatial objects are {A, B, C, D, E}. Using the even numbers in decreasing order for representation such as [(A:10) ,(B:8), (C:6), (D:4), (E:2)]. Therefore the ordered list can be mapped as {10, 8, 6, 4, 2} for {A, B, C, D, E}.This transaction can be mapped by product value of prime numbers i.e. (3840 = 10*8*6*4*2)

TID	Reference City	Positive Object	Even Number Representation	Product Value
1	Bhavnagar	A,B,C,D,E	10,8,6,4,2	3840
2	Rajkot	A,B,C,D,E	10,8,6,4,2	3840
3	Junagadh	A,B,C,D,E	10,8,6,4,2	3840
4	Vadodara	A,B,C,D,E	10,8,6,4,2	3840
5	Ahmadabad	A,B,C,D	10,8,6,4	1920
6	Surat	A,B,C,D	10,8,6,4	1920
7	Dahod	A,B,D,E	10,8,4,2	640
8	Surendranagar	A,B,D,E	10,8,4,2	640
9	Anand	A,C,D,E	10,6,4,2	480
10	Jetpur	A,B,C	10,8,6	480
11	Gondal	A,B,C	10,8,6	480
12	Patan	B,C,D	8,6,4	192

Table 2: Numerical Representation of Spatial dataset

A. FPAR (frequent positive association rule)

We are using numerical representation to demonstrate the object, each object is shown with individual EVEN number used to reduce the time complexity and search space.

FPAR algorithm

Analysis process has the following steps.

Step 1: Obtain Sample Spatial Dataset.

Step 2: Build Ordered list of objects in descending order of their frequencies.

Step 3:Pruned the object which is less then minimum support.

Step3: Mapping Ordered List in form on numerical representation Even.

Step 4: objects list is greater than one than increase the count and add object in path

Step 5: if step 4 not satisfy then count is equal to one and add object in path

Step 6: Find mining frequent patterns of the path which cover maximum object

TID	Object	Count	Path
1	A	11	A
1	B	10	AB
1	C	8	ABC
1	D	6	ABCD
1	E	4	ABCDE
7	ABDE	2	ABDE
9	ACDE	1	ACDE
12	BCD	1	BCD

Table 3: Mining FP from FPAR

Path	**Count**
ABCDE	**4**
ABDE	**2**

Table 4: FPAR Pattern Table

B. TFP (Transaction frequent pattern)

A TFP-Tree includes a root node and a child node that forms a sub tree as children of the root where each child stores product value of each transaction this node. While insertion of a new node takes place, the first and second transaction product values are compared and if the product value of two transactions is not divisible then it creates new descendants node. In case, it is divisible then it is inserted as a child of the existing node. If product value is equal to the current node only the count of the current node is increased by 1. In TFP, used the product value of even number to compress data.[1][2]

TFP algorithm

Analysis process has the following steps.

Step 1: Obtain Sample Spatial Dataset.

Step 2: Build Ordered list of objects in descending order of their frequencies.

Step 3:Pruned the object which is less then minimum support.

Step3: Mapping Ordered List in form on numerical representation Even.

Step 4: Build a TFP Tree using numerical representation Even

Step 5: Find frequent patterns and validate it against their respective support count.

TID	Object	Even Number	Product	Count	Path
1	A,B,C,D,E	10,8,6,4,2	3840	4	3840
5	A,B,C,D	10,8,6,4	1920	2	3840,1920
7	A,B,D,E	10,8,4,2	640	2	3840,1920,640
9	A,C,D,E	10,6,4,2	480	3	3840,480
12	B,C,D	8,6,4	192	1	3840,192

Table 5: TFP Table

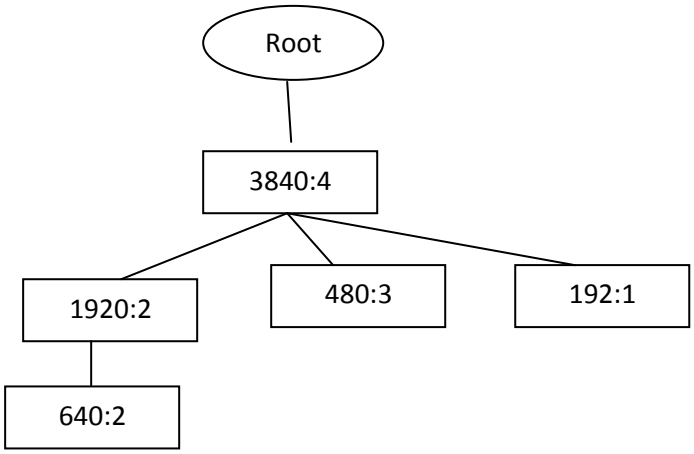

Fig 1 : TFP Tree

TID	Object	Path	Count
1	A,B,C,D,E	3840	4
2	A,B,D,E	3840,640	2
9	A,C,D,E	3840,480	3
12	B,C,D	3840,192	1

Table 6:TFP Pattern

III. Comparison of Prime number, Fibonacci Number and Even Number in FPAR based on Performance

Here I compare the frequent positive association rule mining with different parameter Prime number, Fibonacci number and Even Number. I define the number on X axis and time in seconds on Y axis.

The graph shows that the Even number require less time compare to Fibonacci series ,Prime number to find the frequent positive object in spatial database.

	Time Dependent			Data Dependent	
	Dynamic number generation	**Product of number**	**Frequent occurrence of object**	**FPAR row data**	**Mining row data**
Prime Number	25000(ms)	15000(ms)	7000(ms)	495	21
Fibonacci Number	22000(ms)	12000(ms)	5000(ms)	220	21
Even Number	15000(ms)	13000(ms)	4000(ms)	165	21

Table 7: Comparison of Prime Number, Fibonacci number and Even Number in FPAR based on performance

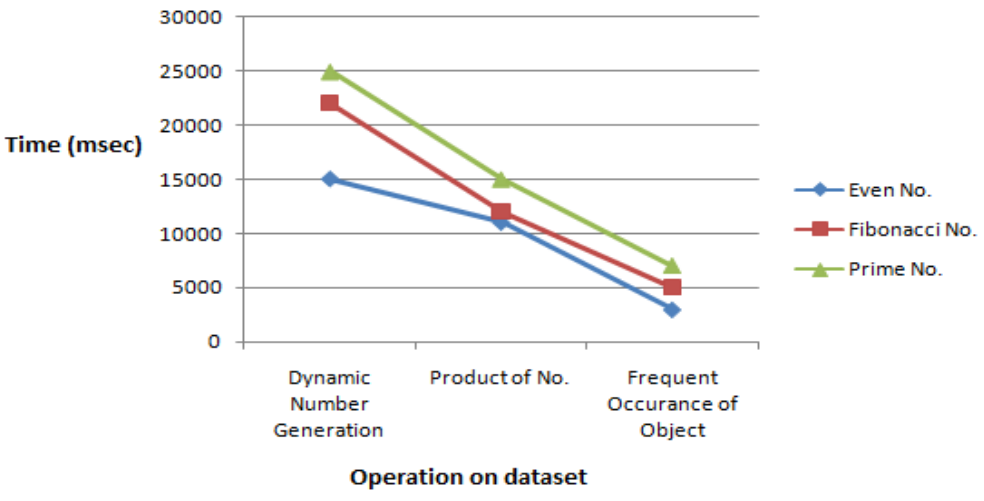

Fig 2: Time Comparison between prime, Fibonacci and even number based on performance

X axis it shows the operation on data set, Y axis it shows the time in msec

COMPARISON OF PRIME NUMBER, FIBONACCI NUMBER AND EVEN NUMBER IN FPAR BASED ON SUPPORT

	05%Support	0.3% Support	0.2% Support
Even No.	4000(ms)	6000(ms)	8000(ms)
Fibonacci No.	5000(ms)	7000(ms)	10000(ms)
Prime No.	7000(ms)	9000(ms)	11000(ms)

Table 8: Comparison of Prime Number, Fibonacci number and Even Number in FPAR with respect to their relative support

Here we defined the graphical representation of the Time performance of Even, Fibonacci and Prime number based on their support.

Support can be described by two ways absolute support and relative support. Occurrence of frequency of object is also called as absolute support. Support of X denoted by supp(X) is freq(X) / N, where N is total number of transactions received in data set is called as relative support. Here we describe base on relative support.

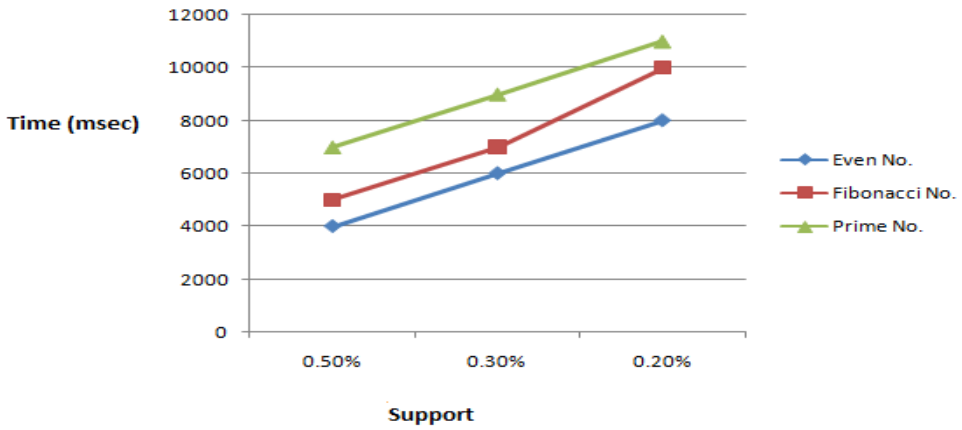

Fig 3: Time performance of Even, Fibonacci and Prime number based on their support

X axis it shows the relative support, Y axis it shows the time in msec.

IV. CONCLUSION AND FUTURE WORK

I have studied the problem of finding frequent pattern in spatial database with use of FPARM; FPARM is numerical method for finding frequent occurrence of place (like city). There are two different parameters are used for that, the Prime number and Fibonacci number. I compare both the numbers with help of FPARM and TFP Techniques , conclude from FPARM that the Fibonacci number require less time compare to Prime number. I also conclude from TFP that Fibonacci number require more number of TFP pattern then Prime number. I find the other number which is even number , even number require less time for transaction execution compare to Fibonacci num and prime number and it require less number of TFP pattern compare to Fibonacci number and prime number. Even number gives best solution for finding frequent occurrence of object in spatial dataset.

In future we can reduce the search space and time complexity of the spatial dataset with another numeric value.

References

1. Animesh Tripathy, Subhalaxmi Das, Prashanta Kumar Patra " An Intelligent Approach for Mining Frequent Patterns in Spatial Database System Using SQL "IEEE -2012
2. Kedar Nath Singh, Jitendra Agrawal, Sanjeev Sharma *Bhopal (M.P.)* Hai "New Approach to Mine Frequent Pattern in Spatial Database using TFP-Tree" International Journal of Computer Technology and Applications, IJCTA JULY-AUGUST 2011
3. Animesh Tripathy Prashanta Kumar Patra "An Intelligent Approach For Mining Frequent Spatial Objects in Geographic Information System" International Journal of Database Management Systems (IJDMS), Vol.2, No.4, November 2010

4. Donato Malerba ,Floriana Esposito, Francesca A. Lisi "A Logical Framework for Frequent Pattern Discovery in Spatial Data" Association For The Advancement of Artificial Intelligence, 201
5. Muhammad Shaheen, Muhammad Shahbaz,Aziz Guergachi" Context based positive and Negative Spatio-Temporal Association Rule Mining" Elsevier,science direct, knowledge based system -2013
6. Joyce Jackson" DATA MINING: A CONCEPTUAL OVERVIEW" Communications of the Association for Information Systems (Volume 8, 2002) 267-296
7. Shashi Shekhar, Sanjay Chawla, Andrew Fetterer, Chang-tien Lu," Spatial Databases Accomplishments and Research Needs" ieee transactions on knowledge and data engineering, vol. 11, no. 1, january/february 1999
8. Neethu C V and Subu Surendran" A New Approach for Spatial Pattern Analysis" Journal of Environmental Science, Computer Science and Engineering & TechnologySeptember 2013 – November 2013; Vol.2.No.4, 1320-1335.
9. Kevin Loney Oracle database the complete reference , Tata McGraw-Hill Edition
10. Arvind Sharma,HS Jat, RK Gupta" A Survey of Spatial Data Mining Approaches: Algorithms and Architecture" International Journal of Computer Technology and Electronics Communication ISSN 2320 – 0081
11. Donato MALERBA, Floriana ESPOSITO and Francesca A. LISI" Mining Spatial Association Rules in Census Data" Dipartimeno di Informatica, University of Bari Italy
12. Deren li, shuliang wang" concepts, principles and applications of spatial data mining and knowledge discovery" national laboratory for information engineering in surveying' mapping and remote sensing, wuhan university, wuhan 430079, china international school of software, wuhan university, wuhan 430072, china
13. Annalisa Appice, Michelangelo Ceci, Antonietta Lanza, Francesca A. Lisi and Donato Malerba" Discovery of spatial association rules in geo-referenced census data: A relational mining approach" Intelligent Data Analysis 7 (2003) IOS Press.
14. Neethu C V and Subu Surendran" A New Approach for Spatial Pattern Analysis" Department of Computer Science, SCT College of Engineering Trivandrum, Kerala, india November 2013

