Survey on Improving Frequent Pattern mining in Spatial Database

Krishna R. Goswami
Master of Engineering Pursing in computer science department
Parul Group of Institute and Technology
Baroda, India
Krishnagoswami1751@gmail.com

Asst. Prof. Ankur Shah
Assistant Professor in computer science department
Parul Group of Institute and Technology
Baroda, India
Ankur1158@gmail.com

Abstract: spatial database system as a database system that offers spatial data types in its data model and query language and supports spatial data types in its implementation, providing at least spatial indexing and spatial join methods. Spatial database systems offer the underlying database technology for geographic information systems and other applications. Here we focus on frequent positive association rule mining and frequent negative association rule mining for finding frequent and not frequent data with less time consuming.

I. INTRODUCTION

In today's world , the collection of data is larger and discover useful information and knowledge that has not seen earlier. The data should be in suitable manner and easy to access. for that we are using data mining. Data mining is the analysis of large dataset from the database ,data warehouse or other large repository, the application of random data to find relationship and summarize the data in under table and useful manner[4]. Association rule mining is one of the best data mining algorithm for searching interesting relationships among items in a large data set or database.

Spatial databases are used in a real-world applications, such as land surveying, urban planning, and environmental assessments, as well as geospatial Web services. Here we survey the research paper related to find frequently occurrence of geographic place from city with help of association rule[5].in simple way. A spatial database is a database that is used to store and query data and represents objects defined in a geometric space. that also representing simple geometric objects such as points, lines and polygons. Some spatial databases handle more complex structures such as 3D objects, topological coverage.

Here we are applying association rule mining on spatial data base in base paper for finding frequently occurrence of geographic area in city. For that we are taking Positive association rule mining and Negative association rule mining

II. LITERATURE SURVEY

According to more recent survey,it is accepted the infrequent rule is also important as the frequent rule mining because it used to show the infrequent item which not cover in frequent and these infrequent rules are often interesting. But Association among 0infrequent items and negative associations have been relatively ignored. The reason is association mining algorithm mainly due to the problem of large search space and the explosion of total number of association rules reported due to noise in data .

There are some attempts to find infrequent association that a generalize association using correlation. Correlation is measured by chi-square. association rule discovery is to find rules with strong associations between items from the training data. It focuses on detecting relationships between the items Mining

A database of transactions is the training data from which rules are generated. A rule is of the form $x \Rightarrow y$ where A is known as the antecedent and B is the consequent of the rule. Both A and B are geographic area from the database of transactions.

The most common form of association rules is implication rule which is in the form of $X \Rightarrow Y$, $X (\cap) Y = \Phi$. The support of the rule X Y is equal to the percentage of transactions in D containing $X \Rightarrow Y$. The confidence of the rule X and Y is equal to the percentage of transactions in D containing X also containing Y. Confidence is defined as below

$$\mathrm{conf}(X \Rightarrow Y) = \mathrm{supp}(X \cup Y)/\mathrm{supp}(X).$$

Here the PARM and NARM is used for finding frequent and in frquent item respectivly from spatial database.

i. Survey paper techniques

There are some techniques to find frequent pattern from spatial database as below

A. Transaction Frequent Pattern(TFP):

TFP used for finding frequent pattern from sparse and dense database. It is numerical method for finding frequent pattern

B. Closed Frequent Pattern(CFP):

A pattern α is a closed frequent pattern in a data set D if α is frequent in D and there exists no proper super-pattern β such that β has the same support as α in D.[8]

C. Maximal Frequent Pattern(MFP):

A pattern α is a maximal frequent pattern(or max-pattern) in set D if α is frequent, and there exists no super-pattern β such that $\alpha \subset \beta$ and β is frequent in D.[8]

D. Spatial Pattern Discovery Algorithm(SPADA):

Spatial Pattern Discovery Algorithm being proposed for frequent pattern discovery in spatial data implements the aforementioned level wise method

E. Share Ascending Order Frequent (SH-AOF):

This technique is used for efficiently mine the information share between the spatial pattern. It used to count the number of frequency of object

F. Frequent Positive Association Rue Mining (FPARM):

FPARM is used to identify the frequently patterns which are more close with each other.

G. Frequent Negative Association Rue Mining (FNARM):

FNARM is used to identify not frequently occurrence of object with each other.

ii. Survey paper

Paper 1: An Intelligent Approach for Mining Frequent Patterns in Spatial Database System Using SQL

Summary:

In this Paper, they represent the association rule mining for finding Frequent Positive and Negative object. For that they define Transaction Frequent Pattern Algorithm which is numerical based. Here prime number is used to represent the object. On the base of prime number, TFP(Transaction Frequent Pattern) is developed. This algorithm require minimum parameter and reduce the search space.

Research Problem:

TFP is not time efficient**.** It works for only the one occurrence of object. Suppose the object occur multiple time then there is no solution. It is not useful to find number of frequency of object.

Paper 2: A New Approach to Mine Frequent Pattern in Spatial Database using TFP-Tree

Summary :

Here, In spatial database for finding frequent pattern divide data base into two part dense and sparse database and find frequent order list for that and assign Fibonacci series to each object. On the base of Fibonacci series value generate TFP(Transaction Frequent Pattern) for numerical representation and develop TFP- tree for spatial database then find dense and sparse pattern in TFP tree with help of array .Now according to TFP pattern they consider CFP and MFP. CFP (Closed Frequent Pattern) and is used to find dense database where most of the transaction is similar. While MFP(Maximal frequent Pattern) is used to find sparse database where the length of the transaction is different from each other.

Research problem:

This paper is work for only the one occurrence of object. Suppose the object occur multiple time then there is no solution.

Paper 3 : An Intelligent Approach For Mining Frequent Spatial Objects in Geographic Information System

summary :

In this paper , they explain two method FPARM (Frequent Pattern Association Rule Mining)and SH-AOF(Share Ascending Order Frequent). FPRN is used for finding frequent pattern from the spatial database by suffix and prefix. To find the total occurrence of object in database they use share Measure concept. The combination of two technique used to reduce the search space and give more information of co related spatial objects.

Research problem:

The combination of FPARM and SH-AOF is reduce search space. SH-AOF is better than SH-FSM(Share-counted FSM) but it is complex .

Paper 4: A Logical Framework for Frequent Pattern Discovery in Spatial Data

Summary:

Here with help of ARM , They can extract the frequent pattern from spatial database with SPDA(Spatial Pattern Discovery Algorithm) which is This is efficient and scalable and Robustness technique .The task of mining spatial association rules relies on a more complex data preprocessing which is error-prone.

Research problem:

The automated extraction of symbolic descriptions from victories map

Paper 5: Context based positive and Negative Spatio-Temporal Association Rule Mining

Summary:

In this paper, Association rule mining algorithm is given with the domain of hydrocarbon prospective. The extraction of association rules on the basis of context is novel approach and has its own impediments .The comparison is done on spatial temporal data for finding positive and negative dataset

III. Methodology

In this section we propose a framework to mine frequent patterns of spatial objects. These spatial objects situated close to each other for a given sample space of geographic area.

The methodology can be described as a sequence of processes. It will extracts the spatial objects and its frequency of occurrences from the Map Database and builds a Sample Spatial Objects Datasets. The next process is to first we represent the frequent order list in form of numerical representation where each object in the transaction is represented Then the next step constructs a TFP(transacted frequent pattern) tree using the numerical representation of each transaction dataset. The final step finds frequent spatial patterns with their respective support count by intersecting its numerical ordered list.

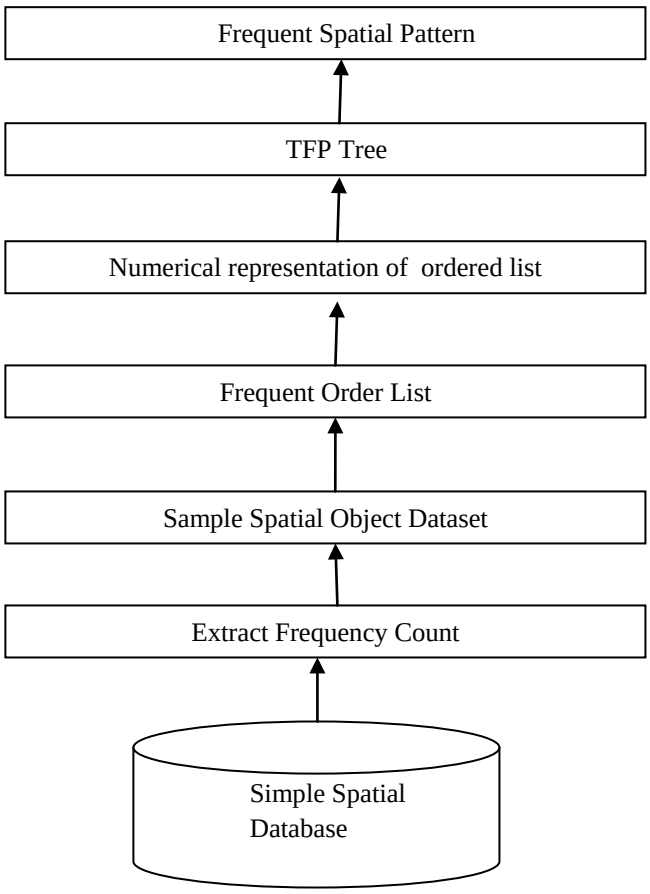

Fig: spatial framework[2]

Algorithm

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.

It shows the basic work flow for spatial database. Now define the special frequent positive association rule (FPAR) for frequently occurrence of data

Algorithm (FPAR)[1]

Construct the table FPAR

Input: Transaction table ORDER-LIST.

Output: FPAR table

Procedure:

for objects with the identical tid in table ORDER-LIST

insert into the table FPARcount=0;

path=null;

for each transaction[i]

if (list[i]>1)

object=list;

insert object into table FPAR

count++;

path+= object;

else

for each transaction

if FPAR has not an object=list[i]

insert list[i] into table FPAR

set count=1;

update table FPAR

path+= object;

Algorithm(TFP)[1]

Input: Transaction table RESULT-PN.

Output: TFP Table

Procedure:

for objects with the identical tid, object, prime, product in

table RESULT-PN

insert into the table TFP

count=1;

path=null;

for *product* in each transaction

if (*product[i+1])MOD(product[i+1])=0*

insert *product* into table TFP

count++;

path+=*product*;

else

insert *product* into table TFP

count++;

path+= *product*;

For each transaction the frequent object pattern of current transaction is inserted into the tree. Then for successive transaction if any object is not found as a node in the tree, it creates a new object node and assigns one as the frequency. Otherwise the frequency of child node adds one.If product value is equal to the current node only the count of the current node is increased by 1 otherwise create a new descendant. Similarly for every successive insertion of a new node its product value is examined against existing nodes product values. The TFP is better than the FPAR its require less space for mining spatial dataset and less time consuming [1][2]

IV. CONCLUSION

Here we go through for finding frequent pattern with help of PARM for frequent data with minimum time requirement. the future work is we should apply this algorithm for infrequent geographic area from spatial data base with help of NARM. TFP is numerical method for finding frequent. It reduce the search space and it perform operation with minimum parameter input.In future we can improve the TFP which is time efficient and work for the object which present more than one time in same transaction in other word count the frequency of object. The second is to compare the assign value of object like prime number or Fibonacci value and will find any other value which is better than these numerical value.

REFERENCE

[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] Feng Yucai, "Association Rules Incremental Updating Algorithm", Journal of Software, Sept., 1998.

[11] A. Pietracaprina and D. Zandolin.2003. "Mining frequent item sets using Patricia tries", Proceeding of IEEE FIMI (2003)

[12] Dong, X., Niu, Z., Shi, X., Zhang, X., Zhu, D.2007."Mining both Positive and Negative Association Rules from Frequent and Infrequent Item sets". ADMA 2007, LNAI 4632, Springer-Verlag Berlin Heidelberg, (2007), pp.122–133

[13] M M. Gan, M.-Y. Zhang and S.-W. Wang. 2005. "One Extended Form for Negative Association Rules and the Corresponding Mining Algorithm" Proceedings of the 4th International Conference on Machine Learning and Cybernetics, Vol. 3, (2005), pp.1716-1721.

[14] Borgelt. 2005."An Implementation of the FP-growth Algorithm," Proceedings of the 1st international workshop on open source data mining, (2005), pp.

[15] Agarwal, P., & Skupin, A. (2008). Self-organising maps: Applications in geographic information science. Chichester: Wiley. Agrawal, R., Imielinski, T., Swami, A. (1993). Mining association rules between setsof items in large databases. In ACM SIGMOD international conference on management of data (pp. 207–216).