ISSN : 2231-0711

Dipali A. Sananse et al | IJCSET(www.ijcset.net) | May 2015 | Vol 5, Issue 5,131-136

Association Rules Mining Technique Based on Spatial Data Classification

Dipali A. Sananse, Prof.Ms. R. R. Tuteja

Dept:Computer Science & Engg.College: Prof. Ram Meghe Institute Of Training & Research,Badnera.India

Abstract— **Data mining is the process of extracting hidden patterns from large amounts of data and is increasingly important tool for transforming data into information. It is mostly used in a wide range of profiling practices, such as marketing, surveillance, and other practices. Association Rules mining is an important technique of data mining. Association rule mining, was proposed for market basket data, and it has many potential applications areas. The most promising application area in the spatial data is remote sensed imagery (RSI) data that can extract interesting patterns and rules from spatial data sets and combination of other data such as ground and weather data, and images. It also includes precision agriculture, resource discovery and other application areas. In this paper we discover the P-trees algorithm, based on P-trees algorithm association rule mining algorithm PARM evaluates calculation and significant pruning techniques to modify the efficiency of the rule mining process. With our algorithms, association rules mining technique is totally based on RSI spatial data.**

Keywords— **Association rule mining, data mining, remote sensed imagery (RSI), spatial data, PARM algorithm.**

I. Introduction

The task of association rule mining is to find certain association relationships among a set of data items in a database. The association relationships are described in association rules. The task of discovering association rules was first introduced in 1993.Association rule mining, is totally determined for market basket data, has potential applications in many areas. Data mining is the process of extracting hidden patterns from large amounts of data and is increasingly important tool for transforming data into information. Spatial data, such as remote sensed imagery (RSI) data, are one of the most promising areas for association rule mining. With the quantities of RSI data being collected every day from satellites, aerial sensors, telescopes, and other sensor platforms are so huge that much of this data is archived before its value can be found. RSI data are collected in different ways and are organized in different formats. BSQ, BIL, and BIP are three main formats. Association rule mining technique is totally depends on RSI spatial data. An RSI image can be viewed as a 2-D array of pixels. Associated with each pixel are various descriptive attributes, called "bands" in remote-sensing literature. Association rules from spatial data using Peano Count Tree (P-tree) structure. P-tree structure provides a lossless and compressed representation of spatial data. Based on P-trees, an efficient association rule mining algorithm PARM with fast support calculation and this techniques is introduced to improve the efficiency of the rule mining process. There were we used three basic P-tree operations such as complement, AND and OR. Spatial association rule is a rule indicating certain association relationship among a set of spatial and possibly some non-spatial predicates. Spatial data mining is the process of discovering interesting and previously unknown, but potentially useful patterns from large spatial datasets. Extracting interesting and useful patterns from spatial datasets is more difficult than extracting the corresponding patterns from traditional numeric and categorical data due to the complexity of spatial data types, spatial relationships, and spatial autocorrelation. Spatial classification methods extend the general-purpose classification methods to consider not only attributes of the object to be classified but also the attributes of neighboring objects and their spatial relations. A visual approach for spatial classification was introduced in, where the decision tree derived with the traditional algorithm is combined with map visualization to reveal spatial patterns of the classification rules. Remote sensing is one of the major areas that commonly use classification methods to classify image pixels into labeled categories. Here is the formal definition of association rules: Let I = {i1, i2, …, in} be a set of literals, called items. Let D be a set of transactions, where each transaction T is a set of items such that T subset of I. Associated with each transaction is a unique identifier, called its TID. We say that a transaction T contains X, a set of some items in I, if X subset of □T. An association rule is a relationship of the form X => Y, where X and Y are sets of items. X is called the antecedent and Y the consequence. An example of the rule can be, "customers who purchase an item X are very likely to purchase another item Y at the same time." There are two primary quality measures for each rule, support and confidence. The rule X => Y has support s% in the transaction set D if s% of transactions in D contain both X and Y. The rule has confidence c% if c% of transactions in D that contain X also contain Y. The main moto of association rule mining is to find all the rules with support and confidence exceeding user specified thresholds, i.e., minimum support and minimum confidence threshold Data mining referred a knowledge discovery in databases (KDD), that means a process of nontrivial extraction of implicit, previous unknown, and potentially useful information from data in databases. Various data mining techniques have been proposed, including association rule mining, classification, clustering, sequential pattern mining, time-series analysis, outlier detection, text mining, and web mining. Data mining techniques have also been applied to many areas, for example, market basket data, web data, DNA data, text data, and spatial data. Association rule mining is one of the important advances in the area of data mining.

II. REMOTE SENSED IMAGERY

Remote sensing can be defined as the acquisition and recording of information about an object without being in direct contact with that object. Spatial data, such as remote sensed imagery (RSI) data, are one of the most promising areas for association rule mining. With the quantities of RSI data being collected every day from satellites, aerial sensors, telescopes, and other sensor platforms are so huge that much of this data is archived before its value can be found. Application areas include in RSI are precision agriculture, resource discovery and management, and natural disaster prediction, detection, and mitigation.

Example: In precision agriculture, association rules can be mined from RSI data to identify crop yield potential, insect and weed infestations, nutrient requirements, flooding damage, and other phenomena. We use an example of the derivation of association rules from RSI data to identify high and low agricultural crop yield potential because RSI is totally depends on association rule mining technique. These high and low agricultural crop yield potential is called as precision agriculture. RSI data are used in mid-growing season to determine additional inputs such as fertilizers, herbicides, etc.

A. RSI Image View

The concept of RSI covers a broad range of methods to include satellites, aerial photography, and ground sensors. A remote sensed image typically contains several bands or columns of reflectance intensities. RSI image can be viewed as a 2-D array of pixels. Associated with each pixel are various descriptive attributes, called "bands" in remote-sensing imagery. Examples of bands include visible reflectance bands such as blue, green, and red reflectance, infrared reflectance bands includes NIR, MIR1, MIR2, and TIR, and possibly some bands of data gathered from ground sensors e.g., yield quantity, yield quality, and soil attributes such as moisture and nitrate levels, etc. The pixel coordinates in raster order constitute the key attribute. One can view such data as a relational table where each pixel is a tuple and each band is an attribute. We proposed the association rule and in this we used the some special type of rule, such as NIR[192, 255]$\wedge$Red[0, 63] =>Yield[128, 255], which is Near Infrared reflectance at least 192 and Red reflectance at most 63 implies Yield will be at least 128 and examples are bushel or acre or some normalized yield measurement. This type of rule expected in association rule. Such rules are useful to both producers and agribusiness communities for yield estimation because the RSI images are typically obtained during the growing season.

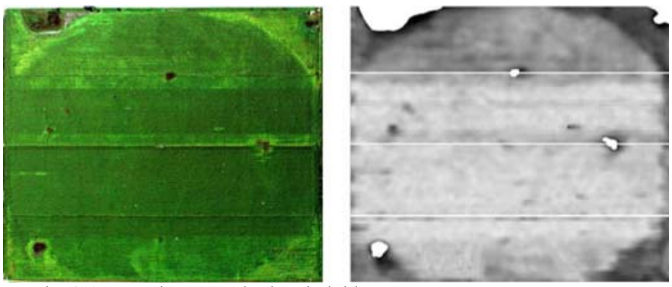

Fig. 2.1: TIFF image and related yield map.

Fig. shows a TIFF image and related yield map. A TIFF image contains three bands such as Red, Green, and Blue. Each band contains a relative reflectance intensity value in the range 0–255 for each pixel. Ground data are collected at the surface of the Earth and can be organized into images. For example, yield data can be organized into a yield map. Spatial data can be represented in two ways: raster and vector. In the raster image form, an image has pixels associated with the attribute values. In the vector representation, a spatial object is represented by its geometry, most commonly being the boundary representation along with the attributes. In this dissertation, we will use the raster image representation. Different kinds of images may have different resolutions; for example, a pixel in a TM scene represents an area of 28.5X28.5 m2; a pixel in a SPOT image represents an area of 20X20 m2; and the resolution in an AVHRR image is much higher.

B. RSI Data Formats

RSI data are collected in different ways and are organized in different formats. BSQ, BIL, and BIP are three typical formats. The Band Sequential (BSQ) format is similar to the relational format. In BSQ format, each band is stored as a separate file and each individual band uses the same raster order. TM scenes are in BSQ format. The Band Interleaved by Line (BIL) format stores the data in line-major order, i.e., the first row of all bands, followed by the second row of all bands, and so on. SPOT data, which comes from French satellite sensors, are in BIL format that shows in fig(a). Band Interleaved by Pixel (BIP) is a pixel-major format. Standard TIFF images are in BIP format. Spatial data was organized a format, called BSQ. A reflectance value in a band is a number in the range 0–255 and is represented as a byte. We split each band into eight separate files, one for each bit position. Each individual bit file is a bSQ file. bSQ files are related to the "bit planes" in image processing. In Fig(b), shows a simple example with only two data bands in a scene having only four pixels like two rows and two columns. Both decimal and binary reflectance values are given.

BAND-1	
254 (1111 1110)	127 (0111 1111)
14 (0000 1110)	193 (1100 0001)

BAND-2	
37 (0010 0101)	240 (1111 0000)
200 (1100 1000)	19 (0001 0011)

BSQ format (2 files)	BIL format (1 file)	BIP format (1 file)
Band 1: 254 127 14 193	254 127 37 240	254 37 127 240
Band 2: 37 240 200 19	14 193 200 19	14 200 193 19

bSQ format (16 files, in columns)

B11	B12	B13	B14	B15	B16	B17	B18	B21	B22	B23	B24	B25	B26	B27	B28
1	1	1	1	1	1	1	0	0	0	1	0	0	1	0	1
0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
0	0	0	0	1	1	1	0	1	1	0	0	1	0	0	0
1	1	0	0	0	0	0	1	0	0	0	1	0	0	1	1

Fig. 2.2: BSQ, BIP, BIL, and BSQ formats for a two-band2 × 2 image.

C. Advantages of BSQ Formats

There are several advantages of using BSQ formats instead of other formats: 1) Different bits make different contributions to the value. In some applications, the high-order bits alone provide the necessary information. 2)The

BSQ format facilitates the representation of a precision hierarchy from 1-bit precision up to 8-bit precision . 3) It also facilitates better compression. 4) In image data, close pixels may have similar properties. By using BSQ format, close pixels may share the same bit values in high-order bits. 5) The BSQ format facilitates high compression for high-order bit files and brings us the idea of creating P-trees.

III. P-TREES

Using Peano Count Tree (P-tree) structure we proposed an efficient approach to derive association rules from spatial data. P-tree structure provides a lossless and compressed representation of spatial data. Based on P-trees, an efficient association rule mining algorithm PARM with fast support calculation and significant pruning techniques is introduced to improve the efficiency of the rule mining process. The P-tree based Association Rule Mining (PARM) algorithm. In this strategies we discovered P-tree structure and P-tree operations.

A. P-Tree Structure

We reorganize each bit file of the bSQ format into a tree structure, called a Peano Count Tree (P-tree). A P-tree is a quadrant wise , Peano-order-run-length compressed, representation of each BSQ file. In this we just divide the entire image into quadrants and record the count of 1 bits for each quadrant, thus forming a quadrant count tree. The P-trees are based on Peano ordering. Peano ordering was selected for several reasons. Compared to raster ordering, Peano ordering has better spatial clustering properties. Peano ordering facilitates compression of the count tree structures due to the probability that neighboring pixels will be similar.

Definition**:** A basic P-tree *Pi, j* is a P-tree for the *j*th bit of the *i*th band. The complement of basic P-tree, Pi, j is denoted as Pi, j'. P-trees have the many features:P-trees contain the count of 1's for every quadrant of every dimension. The P-tree for any sub quadrant at any level is simply the sub-tree rooted at that sub-quadrant. A P-tree leaf sequence (depth-first) is a partial run-length compressed version of the original bit-band. P-trees can be partially combined to produce upper and lower bounds on all quadrant counts. Basic P-trees can be combined to reproduce the original data. P-trees can be used to smooth data by bottom-up quadrant purification such as bottom-up replacement of mixed counts with their closest pure counts.

Example**:** Fig shows an 8 × 8 BSQ file P-tree. In this example, 39 is the number of 1s in the entire image called root count. The root level is labeled level 0. The numbers 16, 8, 15, and 0 at the next level (level 1) are the 1-bit counts for the four major quadrants in raster order such as upper left, upper right, lower left, lower right. Since the first and last level-1 quadrants are composed entirely of 1 bits called pure-1 quadrant and 0 bits call pure-0 quadrant respectively, subtrees are not needed and these branches terminate. This pattern is continued recursively using the Peano also known as *Z*-ordering of the four sub-quadrants at each new level and every branch terminates. Since 8-bit data values for each band, there are eight P-trees and one for each bit position. Expanding all sub-trees, including with these for pure quadrants, the leaf sequence always in the Peano ordering of the image. The P-trees are always based on Peano ordering. Peano ordering was selected for several reasons and compared to raster ordering. Peano ordering was propose a better spatial clustering properties.

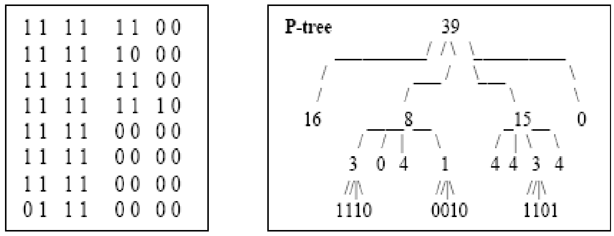

Fig. 3.1: 8 × 8 BSQ file of P-tree.

B. P-tree Variations

A variation of the P-tree data structure, the Peano Mask Tree (PM-tree), is a similar structure in which masks rather than counts are used. In a PM-tree, use a three-value logic to represent pure-1, pure-0, and mixed or called non-pure quadrants. Here 1 denotes pure-1, 0 denotes pure-0, and m denotes mixed. Thus a PM-tree is just an alternative implementation for a P-tree, simplicitly defined also we will use the same term "P-tree" for PM-tree. There are some other variations, called pure-1-tree (P1-tree) and pure-0-tree (P0-tree). In P1-tree, we use 1 to indicate the pure 1 quadrant while we use 0 to indicate others. Similarly, in P0-tree, we use 1 to indicate the pure 0 quadrant while we use 0 to indicate others. Both P1-tree and P0-tree are lossless representations of the original data.

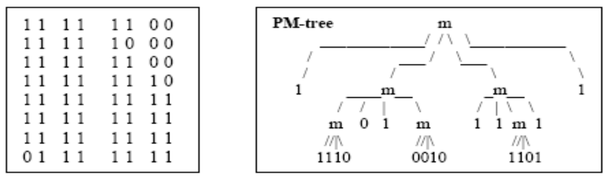

Fig. 3.2: PM-tree of an 8x8 bSQ file.

C. P-Tree Operations

There are three basic P-tree operations: complement, AND and OR. Each basic P-tree has a natural complement. The complement of a basic P-tree can be directly constructed from the P-tree by simply complementing the counts at each level shown in the fig3.3. The complement of basic P-tree *Pi ,j* is denoted as *P' i, j* . Thus the complement of a P-tree provides the 0-bit counts for each quadrant. Fig3.3 also shows AND/OR operations. AND is a very important and frequently used operation for P-trees. The AND operation is just the pixelwise AND of bits from BSQ files or their complement files. For example, a pure-1 P-tree with any P-tree X will result in X to AND, a pure-0 P-tree with any P-tree will result in a pure-0 P-tree to AND, two non-pure P-trees will result in a non-pure P-tree unless all of the four subquants result in pure-0 quadrants. OR operation can be performed as similar as in AND operation. Among the three operations, AND is the most important. Besides basic operations, a P-tree can have other operations, such as XOR. XOR is an exclusive OR operation which gives the difference of two P-trees.

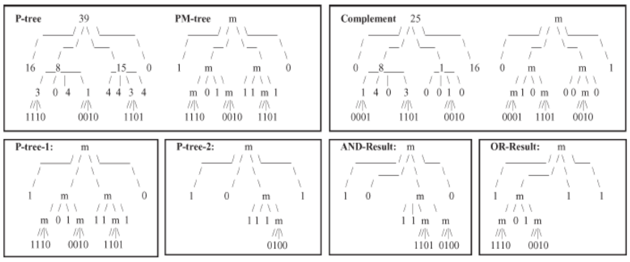

Fig. 3.3: P-tree operations (Complement, AND, and OR).

D. Definations of Value, Tuple, Interval, Cube, and Predicate P-Trees

By performing the AND operation on the appropriate subset of the basic P-trees and their complements, we can construct P-trees for values with more than one bit. These P-trees are called value P-trees.

Defination 2*:* A value P-tree *Pi*(*v*), is the P-tree of value *v* at band *i*. Value *v* can be expressed in 1- up to 8-bit precision. Value P-trees can be constructed by ANDing basic P-trees or their complements. For example, value P-tree *Pi*(110) gives the count of pixels with band-*i* bit 1 equal to 1, bit 2 equal to 1 and bit 3 equal to 0, i.e., with band-*i* value in the range of (192, 224). It can be constructed from the basic P-trees as:

Pi(110) = *Pi,*1 AND *Pi,*2 AND *P_ i,*3.

P-trees can also represent data for any value combination from any band, even including the entire tuple. In this same way, we can construct tuple P-trees.

Definition 3*:* A tuple P-tree *P*(*v*1*, v*2*, . . . , vn*), is the P- tree of value *vi* at band *i*, for all *i* from 1 to *n*.

P(*v*1*, v*2*, . . . , vn*)= *P*1(*v*1) **AND** *P*2(*v*2) **AND***, . .,***AND** *Pn*(*vn*)*.*

If value *v*j is not given, it means it could be any value in Band j.For example, P(110, ,101,001, , , ,) stands for a tuple P-tree of value 110 in band 1, 101 in band 3 and 001 in band 4 and any value in any other band.

Definition 4*:* A interval P-tree *Pi*(*v*1*, v*2), is the P-tree for value in the interval of [*v*1*, v*2] of band *i*.

Pi(*v*1*, v*2) = **OR** *Pi*(*v*)*,* for all *v* in [*v*1*, v*2]*.*

Definition 5*:* A cube P-tree *P*([*v*11*, v*12]*,* [*v*21*, v*22]*, . . . ,* [*vN*1*, vN*2]), is the P-tree for value in the interval of [*vi*1*, vi*2] of band *i*, for all *i* from 1 to *N*. Similar to a tuple P-tree, if the interval is a full range, i.e. from 0 to 255, this interval can be omitted. Any value P-tree and tuple P-tree can be constructed by performing ANDing on basic P-trees and their complements. Interval and cube P-trees can be constructed by combining AND and OR operations of basic P-trees shows in Fig(f). All the P-tree operations, including basic operations AND, OR, Complement, and other operations such as XOR, can be performed on any kinds of P-trees defined above. In general, we can define a P-tree for any condition specified by a predicate.

Definition 6*:* A predicate P-tree is the P-tree for any condition specified by a predicate *p*.

Basic, value, tuple, interval, and cube P-trees are special formats of predicate P-trees. A predicate P-tree can be constructed by performing AND, OR, and Complement operations on basic P-trees since basic P-trees have all the information in the original data that shows in fig(f). A predicate P-tree can be viewed as the result of query *p* over the original data. The query result gives the hierarchy distribution of the selected data satisfying predicate (query) *p*.

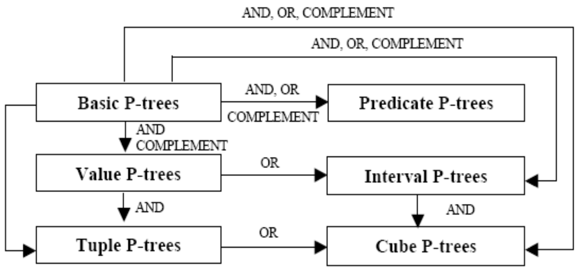

Fig. 3.4: Basic, value, tuple, interval, cube, and predicate P-trees.

IV. Deriving Association Rules Using P-Trees

For RSI data, we can formulate the association rule mining model. Let *I* be the set of all items and *T* be the set of all transactions. *I = {*(*b, v*)*|b* = band*, v* =value(1-bit*,* 2-bit*, . . . ,*or 8-bit)*}*, *T = {*pixels*}*. Admissible Itemsets also known as Asets are itemsets of the form, Int1 × Int2 ×· · ·×Int*n* = Pi*i*=1...n Int*i*, where Int*i* is an interval of values in Band*i.* A k-band Aset also read as k-Aset is an Aset in which k of the Int*i* intervals are restricted i.e. in k of the bands the intervals are not all of [0, 255].We also use the notation [a*, b*]*i* for the interval [a,b] in band *i*. For example, [00*,* 10]2 indicates the interval [00, 10] which is [0, 191] in decimal in band 2.The root count of an Aset is the root count of its P-tree. The users may be interested in some specific kinds of rules. For an agricultural producer we used precision techniques, there is simple interest in rules of the type, Red > 48 => Green < 134. A physicist might be interested in such color relationships but a producer is interested in rules with color antecedents for e.g., yield consequents i.e. observed color combinations that predict high yield or foretell low yield. Therefore, for precision agriculture applications, it makes sense to restrict our search to those rules that have a consequent in the yield band. We will refered for such rules to rules of interest to be distinct from interesting rules. Of-interest rules can be interesting or not interesting, depending on such measures as support. The candidate k-Asets are those of which (k-1)-Aset subsets are frequent. Next is a pruning technique based on the precision hierarchy. Once we find all the 1-bit frequent k-Asets, we can use the fact that a 2-bit k-Aset cannot be frequent if its enclosing 1-bit k-Aset is infrequent. A 1-bit Aset encloses a 2-bit Aset if, when the endpoints of the 2-bit Aset are shifted right 1-bit position, the 2-bit Aset is a subset of the 1-bit Aset. We proposed an algorithm called PARM algorithm, to mining association rules on RSI data using P-trees shown in fig(g). PARM algorithm was based on the classic Apriori algorithm. The Apriori algorithm uses a levelwise approach to generate all the frequent itemsets, starting with frequent 1-itemsets.The fact is if an itemset is frequent, all its subset must also be frequent, the Apriori algorithm generates candidate (*k* + 1)-itemsets from frequent k-itemsets and then calculates the support for each candidate (*k* + 1)-itemset to form frequent (*k* + 1)-itemsets. Therefore, in PARM algorithm we try to find all Asets that are frequent and of-interest. We are partitioning the data into intervals performed by the "Discretization" function. Then, we find all frequent 1-Asets by checking the root count of the corresponding P-trees. The candidate k-Asets are those whose $(k-1)$-Aset subsets are frequent. The essential difference between the PARM algorithm and the Apriori algorithm is how the candidate Asets are counted. In PARM, Asets are counted by performing AND operations on corresponding basic P-trees, while in Apriori, it is done by scanning the entire data.

```
Procedure PARM
{
    Discretization;
    F1 = { c∈1-Asets | rootcount(c) >= minsup};
    For (k=2; Fk-1 ≠∅; k++) do begin
        Ck = p-gen(Fk-1);
        Fk = {c∈Ck | rootcount(c) >= minsup}
    end
    Answer = ∪k Fk
}
```

Fig. 4(a): PARM algorithm.

The PARM algorithm assumes a fixed precision in all bands. In the Apriori algorithm, there is a function called "apriori-gen" to generate candidate k-itemsets from frequent $(k-1)$ itemsets. The p-gen function in the PARM algorithm differs from the *apriori-gen* function in the way pruning is done. We use band-based pruning in the *p-gen* function. Since no value can be in multiple intervals in the same way, joining among intervals from the same band can be avoided. For example, even if [00, 01]1 and [11, 11]1 are frequent, there is no need to join them to form a candidate Aset ([00, 00]1 × [11, 11]1). *P-gen* only joins items from different bands. Two frequent $(k-1)$-Asets will be joined into a candidate k-Aset only if the first $(k-1)$ items of both Asets are the same. The order of the last item is compared to avoid the generation of the duplicate candidate Aset. The join step in the *p-gen* function is shown in Fig(h). The rootcount function is directly used to calculate Aset counts by ANDing the appropriate basic P-trees instead of scanning the transaction databases. For example, in the Asets, *{B*1[0, 64)*,B*2[64, 127)*}*, denoted as [00, 00]1 × [01, 01]2, the count is the root count of *P*1(00) AND *P*2(01). This provides fast support calculation and is particularly useful for large data sets. It eventually improves the entire mining performance simply.

```
insert into Ck
    select p.item1, p.item2, …, p.itemk-1,
        q.itemk-1
    from Fk-1 p, Fk-1 q
    where p.item1 = q.item1,
        …,
        p.itemk-2 = q.itemk-2,
        p.itemk-1 < q.itemk-1,
        p.itemk-1.band != q.itemk-1.band
```

Fig. 4(b): Join step in *p-gen* function.

There were interest in multi-level rules, which means that the different itemsets in the rule have different precision.

V. TASKS OF SPATIAL DATA

Spatial data sharing and mapping, high-resolution remote sensing, and location-based services, more and more research domains have created or gained access to high-quality geographic data to incorporate spatial information and analysis in various studies. In this we proposed association rule mining technique for spatial data and also forms the spatial data classification. Spatial classification methods extend the general-purpose classification methods to consider not only attributes of the object to be classified but also the attributes of neighboring objects and their spatial relations.

A. Spatial Association Rule Mining

Association rule mining was originally intended to discover regularities between items in large transaction databases. Spatial association rule is a rule indicating certain association relationship among a set of spatial and possibly some non-spatial predicates. A strong rule indicates that the patterns in the rule have relatively frequent occurrences in the database and strong implication relationships. Raditional data organization and retrieval tools can only handle the storage and retrieval of explicitly stored data. The extraction and comprehension of the knowledge implied by the huge amount of spatial data, though highly desirable, pose great challenges to currently available spatial database technologies. A spatial characteristic rule is a general description of a set of spatial-related data. For example, the description of the general weather patterns in a set of geographic regions is a spatial characteristic rule. A spatial discriminant rule is the general description of the contrasting or discriminating features of a class of spatial-related data from other classes. Spatial data mining is performed with the perspective of spatial locality, that is mined patterns consider objects being close in space. Various kinds of spatial predicates can be involved in spatial association rules. They may represent topological relationships, spatial orientation/ordering, or distance information. Some studies have been done on image-content based association rule mining. In this rule, the problem is to find association rules about the size, color, texture, and shape of the images and to identify similar object in different images. Spatial co-location pattern mining is spiritually similar to, but technically very different from, association rule mining. The mining of association rules in transactional or relational databases, spatial association rules can be mined in spatial databases by considering spatial properties and predicates.

B. Spatial Data classification

Classification is about grouping data items into classes. Classification is also called supervised classification, as opposed to the unsupervised classification (clustering). ''Supervised" classification needs a training dataset to train (or configure) the classification model, a validation dataset to validate (or optimize) the configuration, and a test dataset to evaluate the performance of the trained model. In the spatial data classification, classification has the many methods such as decision trees, artificial neural networks (ANN), maximum likelihood estimation (MLE), linear discriminant function (LDF), support vector machines (SVM). The task of classification is to assign an object to a class from a given set of classes based on the attribute values of this object. The relevant attributes are extracted by comparing the attribute values of the target objects with the attribute values of their nearest neighbors. The determination of relevant attributes is based on the concepts of the nearest hit and the nearest miss. In spatial classification the attribute values of neighbouring objects may also be relevant for the membership of objects.

VI. CONCLUSION

In this seminar, presented the concept of association rule mining from spatial data such as remote sensed imagery (RSI) data ,it is one of the most promising area in the association rules mining. The task of association rule mining is to find certain association relationships among a set of data items in a database. It can also describes the bSQ data formats and their advantages and P-trees structure are used for representing remote sensed imagery (RSI) data. We reorganize each bit file of the bSQ format into a tree structure, called a Peano Count Tree (P-tree). Using Peano Count Tree (P-tree) structure we proposed an efficient approach to derive association rules from spatial data and it forms an efficient association rule mining algorithm i.e. PARM algorithm. P-tree structure provides a lossless and compressed representation of spatial data. Based on P-trees, PARM algorithm facilitates fast support calculation and significant pruning techniques is introduced to improve the efficiency of the rule mining process. P-tree can be constructed by performing AND, OR, and Complement operations on basic P-trees

In this seminar, we also proposes the spatial data classification and and their spatial association rule mining. Spatial association rule is a rule indicating certain association relationship among a set of spatial and possibly some non-spatial predicates. The P-tree can also defines the Value, Tuple , Interval, Cube, and Predicate.

REFERENCES

[1] Diansheng Guo a,1, Jeremy Mennis b,"Spatial data mining and geographic knowledge discovery—An introduction". a Department of Geography, University of South Carolina, 709 Bull Street, Room 127, Columbia, SC 29208, United States b Department of Geography and Urban Studies, Temple University, 1115 W. Berks Street, 309 Gladfelter Hall, Philadelphia, PA 19122, United States.

[2] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, Classificationand Regression Trees. Belmont, CA: Wadsworth, 1984.

[3] Qin Ding," ASSOCIATION RULE MINING ON REMOTELY SENSED IMAGERY USING P-TREES",A Dissertation Submitted to the Graduate Faculty of the North Dakota State University of Agriculture and Applied Science.

[4] Qin Ding, Qiang Ding, and William Perrizo," PARM—An Efficient Algorithm to Mine Association Rules From Spatial Data", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 38, NO. 6, DECEMBER 2008.

[5] Q. Ding, Q. Ding, and W. Perrizo, "Decision tree classification of spatial data streams using Peano Count Trees," in *Proc. ACM Symp .* Appl. Comput., 2002, pp. 413–417.

[6] Q. Ding, Q. Ding, and W. Perrizo, "Association rule mining on remotely sensed images using P-trees," in *Proc. Pacific-Asia Conf. Knowl. Discovery DataMining*. Berlin, Germany: Springer-Verlag, May 2002, vol. 2336, pp. 66–79.

[7] R. Miller and Y. Yang, "Association rule mining over interval data," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 1997, pp. 452–461.

[8] Subhasmita Mahalik 113050073 CSE," SPATIAL DATA MINING TECHNIQUES", Department of Computer Science and Engineering Indian Institute of Technology, Bombay,Mumbai.