SAS: Implementation of scaled association rules on spatial multidimensional quantitative dataset

M. N. Doja
Professor
Faculty of Engg & Technology
Jamia Millia Islamia
New Delhi -110025,India

Sapna Jain
PhD fellow
Department of Computer Science
Jamia Hamdard
New Delhi-110062,India

M Afshar Alam
Professor
Department of Computer Science
Jamia Hamdard
New Delhi-110062,India

Abstract— **Mining spatial association rules is one of the most important branches in the field of Spatial Data Mining (SDM). Because of the complexity of spatial data, a traditional method in extracting spatial association rules is to transform spatial database into general transaction database. The Apriori algorithm is one of the most commonly used methods in mining association rules at present. But a shortcoming of the algorithm is that its performance on the large database is inefficient. The present paper proposed a new algorithm by extracting maximum frequent itemsets based on spatial multidimensional quantitative dataset. Algorithms for mining spatial association rules are similar to association rule mining except consideration of special data, the predicates generation and rule generation processes are based on Apriori. The proposed method (SAS) Scaled Aprori on Spatial multidimensional quantitative dataset in the paper reduces the number of itemsets generated and also improves the execution time of the algorithm.**

Keywords- ***association rules; spatial dataset; X tree.***

I. INTRODUCTION

Data mining and knowledge discovery have become popular fields of research. A significant subset of this research is looking at the particular semantics of space and time and the manner in which they can be sensibly accommodated into data mining algorithms [12].

Scaling is considered an important aspect in Data Mining. The problem of scalability in Data mining is not only how to process such large sets of data, but how to do it within a useful timeframe. Scalability means that as a system gets larger, its performance improves correspondingly. The main purpose of data mining techniques is to find hidden information and unknown relations within an amount of data.

Spatio-Temporal applications like climate change modeling and analysis [1], transportation systems, forest monitoring[2], diseases spreading[3], temporal geographic information systems[4] and environmental systems[5] process spatial, temporal and attribute data elements for knowledge discovery. The spatial objects are characterized by their position, shape and spatial attributes. Spatial attributes are properties of space and spatial objects located in specific positions inherit these attributes. Spatial attributes refer to the whole space and can be represented as layers, each layer represent one theme. The temporal objects are characterized by two models of time that are used to record facts and information about spatial objects.

Association rules mining are one of the major techniques of data mining and it is perhaps the most common form of local-pattern discovery in unsupervised learning systems. The technique is likely to be very practical in applications which use the similarity in customer buying behavior in order to make peer recommendations.

Association rule-based approaches focus on the creation of transactions over space so that an apriori like algorithm [28] can be used. Transactions over space can use a reference-feature centric [29] approach or a data-partition [30] approach. The reference feature centric model is based on the choice of a reference spatial feature and is relevant to application domains focusing on a specific Boolean spatial feature, e.g., incidence of cancer. Domain scientists are interested indicating theco-locations of other task relevant features (e.g., asbestos) to the reference future. Transactions are created around instances of one user specified reference spatial feature. The association rules are derived using the apriori [28] algorithm. The rules found are all related to the reference feature.

In this 21st century the developments in spatial data acquisition, mass storage and network interconnection, volume of spatial data has been increasing dramatically. Vast data satisfied potential demands of exploring the earth’s resource and environment by human being, widening exploitable information source, but the processing approaches of spatial data lag behind severely, and are unable to discover relation and rules in large amount of data efficiently and make full use of existing data to predict development trend.

This work is the extension of Aprori-UB which uses multidimensional access method, UB-tree to generate Better association rules with high support and confidence. In multidimensional databases, objects are indexed according to several or many independent attributes. However, this task cannot be effectively realized using many standalone indices and thus special indexing structures have been developed is last two decades. Indexing and querying high dimensional databases.

This paper has the following sections. Section 2 represents the previous work done in the same field .Section 3 gives the conceptual details used in the proposed algorithm. Section 4 highlights the proposed Aprori-UB method .Section 5 gives the implementation details. Section 6 discusses the conclusion and future scope.

II. Related Work

Spatial datasets need to be preprocessed to construct the transaction database before mining spatial association rules according to the main idea of mining spatial association rules at present. Imam Mukhlash and Benhard Sitohang put forward the framework of spatial data preprocessing, including feature (spatial and non-spatial) selection based on spatial parameters, performing dimension reduction and selection of non-spatial attributes, performing data categorization based on non-spatial data parameters, performing join operations for spatial objects based on spatial parameters and transforming into output form [16].

The preparation and preprocessing of spatial datasets: The spatial datasets in the case included the elevation, the slope and the aspect with the spatial resolution of 100m and the land cover map. A spatial database is defined as a collection of inter-related geodspatial data that can handle and maintain a large amount of data which is shareable between different GIS applications.

The consistency with little or no redundancy and maintenance of data quality including updating. The self-descriptive analysis with metadata and high performance by database management system with database including security access control mechanism.

Spatial Data Mining (SDM) is a process of spatial support decision, which aims at extracting the implicit, unknown, potential, useful spatial and non-spatial knowledge from spatial data, including general geometry rules, spatial characteristics rules, spatial classification rules, spatial clustering rules, spatial association rules and so on [1]. Spatial association rule, termed as spatial association location pattern [2], is one of the most important branches in the SDM, which means a rule indicating certain association relationships among a set of spatial and nonspatial attributes of geographical objects. Because of the complexity of spatial data, the main idea of extracting spatial association rules is to mine spatial association rules in the transaction database categorized from spatial data using some mining algorithms.

A spatial database is a collection of spatially referenced data that acts as a model of reality. This database model represents a selected set or approximation of phenomena which are deemed important enough to represent the digital representation might be for some past, present or future time period .

The Apriori algorithm [3] is one of the most commonly used algorithms in mining association rules at present, and its typical application was market basket analysis to discover customer shopping patterns [4]. Apriori Algorithm can be used to generate all frequent itemset. A Frequent itemset is an itemset whose support is greater than some user-specified minimum support (denoted L, where k is the size of the itemset). A Candidate itemset is a potentially frequent itemset (denoted C , where k is the size of the itemset).

A. Generate the candidate itemsets in N1. Save the frequent itemsets in N2.

B. Steps

1) Generate the candidate itemsets in C from the frequent itemsets in L,k-1

2) Join L k-1 p with L q, as follows: insert nto C select p.item from L k-1 k p, L q ,where, p.itemk-1, p.itemk-1 , . . . , p.item = q.item

3) Generate all (k-1)-subsets from the candidate itemsets in Ck

4) Prune all candidate itemsets from C• where, some (k-1)-subset of the candidate itemset is not in the frequent itemset L k

5) Scan the transaction database to determine the support for each candidate itemset in Nk.

6) Save the frequent itemsets in L k.

III. Concept Used

X-tree has data nodes and normal directory nodes. A data node contains minimum bounding rectangles (MBRs) together with pointers to the actual data objects: (MBR, p) while the directory node consists of MBRs together with pointers to sub-MBRs. In addition, the X-tree introduces another type of nodes: super nodes. A super-node is large directory node of variable size (a multiple of the usual block size). Figure 7 shows an example of an X-tree structure with three kinds of nodes: directory node, leaf node, and super node.

Since the original X-tree was proposed [15], there have been several implementations of X-trees. One of them is the X+-tree [19]. The X+-tree allows the increase of the size of super-nodes in the X-tree to some degree. Technically, in order to avoid overlap, which is bad for performance, a super node might grow during the insertion. However, the linear scan of a large super node can be a problem. In the X-tree, the size of a super-node can be many times larger than size of a normal node. In the X+-tree, the size of super-node is at most the size of a normal node multiplied by a given user parameter MAX_X_SNODE. When the super-node becomes larger than the upper limit, the super-node has to be split into two new nodes.

The X-tree (eXtended node tree) is a new index structure supporting efficient query processing of high-dimensional data. The goal is to support not only point data but also extended spatial data and therefore, the X-tree uses the concept of overlapping regions. The X-tree therefore avoids overlap whenever it is possible without allowing the tree to degenerate; otherwise, the X-tree uses extended variable size directory nodes, so-called supernodes. In addition to providing a directory organization which is suitable for high-dimensional data, the X-tree uses the available main memory more efficiently

The X-tree may be seen as a hybrid of a linear array-like and a hierarchical R-tree-like directory. It is well established that in low dimensions the most efficient organization of the directory is a hierarchical organization.

The reason is that the selectivity in the directory is very high which means that, e.g. for point queries, the number of required page accesses directly corresponds to the height of the tree. This, however, is only true if there is no overlap between directory rectangles which is the case for a low dimensionality. It is also reasonable, that for very high dimensionality a linear organization of the directory is more efficient.

The reason is that due to the high overlap, most of the directory if not the whole directory has to be searched anyway. If the whole directory has to be searched, a linearly organized directory needs less space and may be read much faster from disk than a block-wise reading of the directory. For medium dimensionality, an efficient organization of the directory would probably be partially hierarchical and partially linear.

The problem is to dynamically organize the tree such that portions of the data which would produce high overlap are organized linearly and those which can be organized hierarchically without too much overlap are dynamically organized in a hierarchical form. The algorithms used in the X-tree are designed to automatically organize the directory as hierarchical as possible, resulting in a very efficient hybrid organization of the directory.

A. *Structure of the X-tree*

The overall structure of the X-tree is presented in Figure 1. The data nodes of the X-tree contain rectilinear minimum bounding rectangles (MBRs) together with pointers to the actual data objects, and the directory nodes contain MBRs together with pointers to sub-MBRs (cf. Figure 1).

The X-tree consists of three nodes: data nodes,normal directory nodes, and supernodes. Supernodes are large directory nodes of variable size (a multiple of the usualblock size).

The basic goal of supernodes is to avoid splits in the directory that would result in an inefficient directory structure. The alternative to using larger node sizes is highly overlapping directory nodes which would require accessing most of the son nodes during the search process. This, however, is more inefficient than linearly scanning the larger supernode.

The X-tree is completely different from an R-tree with a larger block size since the X-tree only consists of larger nodes where actually necessary. As a result, the structure of the X-tree may be rather heterogeneous as indicated in Figure 1[7]. Due to the fact that the overlap is increasing with the dimension, the internal structure of the X-tree is also changing with increasing dimension.

In Figure 1, three examples of X-trees containing data of different dimensionality are shown. As expected, the number and size of supernodes increases with the dimension. For generating the examples, the block size has been artificially reduced to obtain a drawable fan-out.

Due to the increasing number and size of supernodes, the height of the X-tree which corresponds to the number of page accesses necessary for point queries is decreasing with increasing dimension [7].

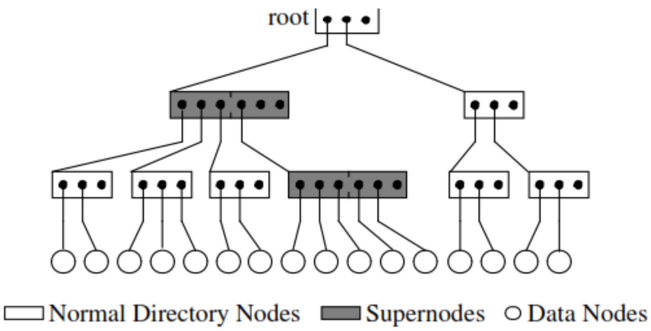


Figure1: Structure of a X-tree.

IV. Proposed Work

The SAS algorithm puesodocode of the algorithm is:

a) Identify the correlated data in the spatial dataset.

b) Find all frequent item sets.

c) Generate scaled association rules from the frequent item sets

d) Identify the quantitative elements.

e) Sorting the item sets based on the frequency and quantitative elements.

f) Use Xtree to create a multidimensional spatial dataset.

g) Discard the infrequent item value pairs

h) Iterate the steps c to f till the required mining results are achieved.

Let I = {i1, i2 … i n items} be a set of items, and T a set of transactions, each a subset of I. An association rule is an

Implication of the form A=>B, where A and B are non-intersecting. The support of A=>B is the percentage of

The transactions that contain both A and B:

X tree psedocode :

Input: A set of M current model tree nodes M A set of current support tree nodes X.

Output: A list Z of feasible sets of points

```
1) $X \leftarrow \{\}$ and $X_{curr} \leftarrow X$
2) IF we cannot prune based on the mutual compatibility of M:
3) FOR each $s \in X_{curr}$
4) IF s is compatible with M:
5) IF s is "too wide":
6) Add s's left and right child to the end of $X_{curr}$
7) ELSE
8) Add s to X.
9) IF we have enough valid support points:
10) IF all of $v \in M$ are leaves:
11) Test all combinations of points owned by the model nodes, using the support nodes' points as potential support.
Add valid sets to Z.
12) ELSE
13) Let $m*$ be the non-leaf model tree node that owns the most points.
```

14) Search using m∗'s left child in place of m∗and S'instead of S.

15) Search using m∗'s right child in place of m∗and Sinstead of S.

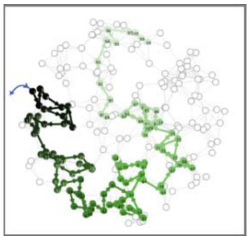


Figure 2 : spatial attributes

The main idea of the X-tree is to avoid overlap of bounding boxes in the directory by using a new organization of the directory which is optimized for high dimensional space. The X-tree avoids splits which would result in a high degree of overlap in the directory. Instead of allowing splits that introduce high overlaps, directory nodes are extended over the usual block size, resulting in so-called supernodes. The supernodes may become large and the linear scan of the large supernodes might seem to be a problem.

Additionally, oversize shelves are organized as chains of disk pages which cannot be read sequentially.We implemented the X-tree index structure and performed a detailed performance evaluation using very large [6]. The X-tree also provides much faster insertion times (about 4 times faster than ub-tree).

V. Implementation and Future Work:

In order to show the performance of the proposed algorithm, we applied the algorithm to Diabetes Data Set which was obtained from UCI Machine Learning Repository[16].This dataset is multivariate, TimeSeries and has 20 attributes. After discovering rules, they have to be presented in understandable form to the user. Java programs since Java byte-codes are compiled or interpreted by the Java Virtual Machine resulting in performance penalty. The core of the X+-tree implementation in [15] is reused, with changes and additions made to data structure and functions.

Spatial Trend Detection: Spatial trends describe a regular change of non-spatial attributes when moving away from a start object o. The existence of a global trend for a start object o indicates that if considering all objects on all paths starting from the values for the specified attribute(s) in general tend to increase (decrease) with increasing distance. Our algorithm detects regions showing a certain global trend, and algorithm local-trends then finds within these regions some paths having the inverse trend (see figure 3).

The algorithm mine the frequent itemsets by using a divideand-conquer strategy as follows: SAS first compresses the database representing frequent itemset into a frequent-pattern tree, or X-tree, which retains the itemset association information as well. The next step is to divide a compressed database into set of spatial databases (a special kind of projected database), each associated with one frequent item. Finally, mine each such database separately, particularly, the construction of X-tree and the mining of X-tree (figure 6).

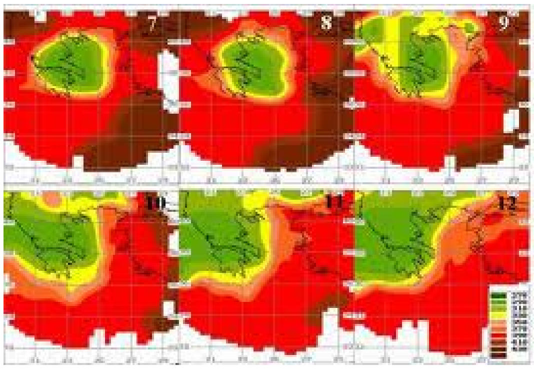
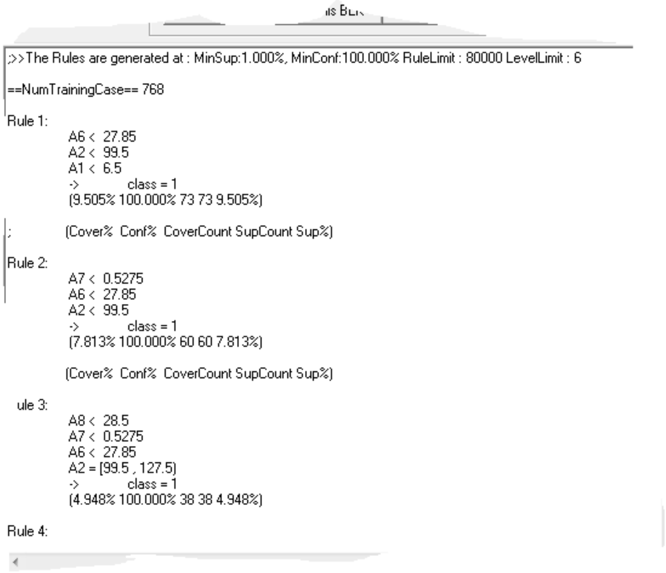
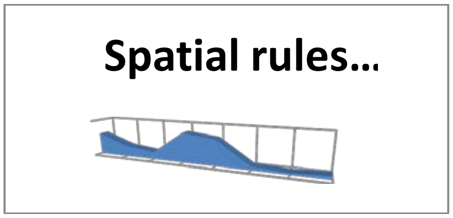


Figure 3 : trend detection phases.



```
;>>The Rules are generated at : MinSup:1.000%, MinConf:100.000% RuleLimit : 80000 LevelLimit : 6

==NumTrainingCase== 768

Rule 1:
         A6 <  27.85
         A2 <  99.5
         A1 <  6.5
         ->          class = 1
         (9.505% 100.000% 73 73 9.505%)

;       (Cover%  Conf%  CoverCount SupCount Sup%)

Rule 2:
         A7 <  0.5275
         A6 <  27.85
         A2 <  99.5
         ->          class = 1
         (7.813% 100.000% 60 60 7.813%)

        (Cover%  Conf%  CoverCount SupCount Sup%)

 ule 3:
         A8 <  28.5
         A7 <  0.5275
         A6 <  27.85
         A2 = [99.5 , 127.5)
         ->          class = 1
         (4.948% 100.000% 38 38 4.948%)

Rule 4:
```

Figure 4: Support and confidence of dataset with rules .



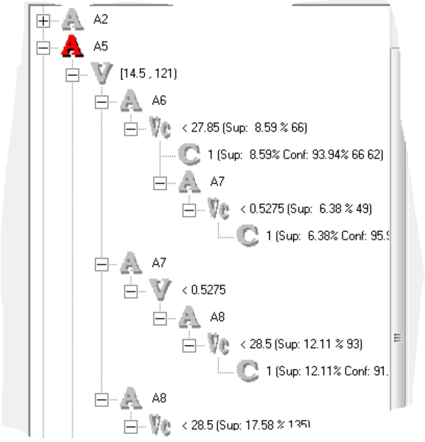
Spatial rules…

Figure 5:Rules generation

The experiment focused on evaluating aprori, Aprori-ub and SAS algorithms. Since we were interested in seeing the best performance, we used diabetes data set samples. The minsupp, minconfidence level and average rule error was compared in figure 8. The evaluation shows that our proposed SAS generated strong association rule with less rule generation error on spatial multidimensional dataset.

VI. Conclusion

In this paper we proposed a practical to find frequent patterns using X Tree. X-tree compresses both dense and sparse datasets by using numerical value representation. In this method we consider Fibonacci number characteristics to find CFP (closed frequent pattern) and then the MFP (maximal frequent pattern) our approach is efficient on both dense and sparse database.



```
⊞ A A2
⊟ A A5
   ⊟ V [14.5 , 121)
      ⊟ A A6
         ⊟ Vc < 27.85 (Sup:  8.59 % 66)
            C 1 (Sup:  8.59% Conf: 93.94% 66 62)
            ⊟ A A7
               ⊟ Vc < 0.5275 (Sup:  6.38 % 49)
                  C 1 (Sup:  6.38% Conf: 95.
      ⊟ A A7
         ⊟ V < 0.5275
            ⊟ A A8
               ⊟ Vc < 28.5 (Sup: 12.11 % 93)
                  C 1 (Sup: 12.11% Conf: 91.
      ⊟ A A8
         ⊟ Vc < 28.5 (Sup: 17.58 % 135)
```

Figure 6: X tree function calculation.

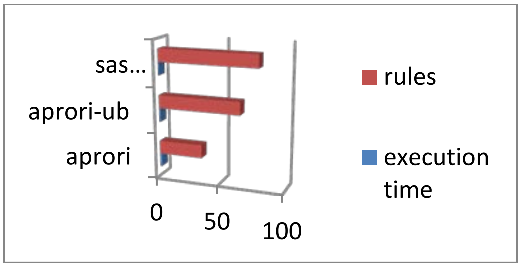


Figure 7 : Comparison of SAS with other algorithms

The algorithm will positively enhance the efficiency of judgment the relationship between spatial objects and further can be used in association analysis. The creation of maximal frequent patterns is done by intersecting the ordered list (OL) of similar type which reduces the search space.

Spatially, association is a relationship between spatial objects. Association analysis is one of the most widely research topics in data mining. The main focus of association rule mining is to generate hypothesis rather than to test them as is commonly achieved using statistical techniques (15). The concept of association rule, introduced by Agrawal, was used for analyzing market basket data to mine customer shopping patterns.

Spatial association algorithms find the frequent sets in spatial and non-spatial databases,and inter-relationship between different variables that are not explicitly stored in the spatial database. In many situations there is a need to discover spatial association rules, rules that associate one or more spatial objects. To confine the number of rules, the concept of minimum support and minimum confidence are used. The intuition behind this is that in large databases,there may exist a large number of associations between objects but most of them will be applicable to only a small number of objects, or the confidence of the rule may be low. However, a strong rule is a rule with large support, i.e., no less than the minimum support threshold, and a large confidence, i.e., no less than the minimum confidence threshold e.g. is_a (X, city) within (X,maharastra) adjacent_to (X,water) close_to(X, Karnataka)…(92%). The rule states that 92% of the cities within Maharastra and adjacent to water are close to Karnataka, which associates predicates is a, within, and adjacent_to with spatial predicate close_to. The quality of the rule is measured in the terms of the surprise associated with it. To calculate the surprise or interestingness associated with the mined rule the correlation and chi-square test technique is adopted.

We have generated scaled association rules with high support and confidence. Other future work in this field includes discovery algorithms with dynamic changes of µ level, improved performance strategies and new measures for rule management.

REFERENCES

[1] Auroop R Ganguly and Karsten Steinhaeuser, (2008) ”Data Mining for climate change and impacts”,IEEE international conference on data mining workshops,ICDMW,15-19,Dec,2008,Italy.

[2] T. Cheng and J. Wang, (2006) “Applications of spatio-temporal data mining and knowledge discovery (STDMKD) for forest fire prevention”, ISPRS Commission VII Mid-term Symposium "Remote Sensing: From Pixels to Processes, Enschede, the Netherlands, 8-11 May 2006 .

[3] Diego Ruiz-Moreno, Mercedes Pascual, Michael Emch, Mohammad Yunus, (2010) “Spatial clustering in the spatio-temporal dynamics of endemic cholera“, BMC Infectious Diseases, Volume 10, 2010.

[4] Yang ping , Tang Xinming , Wang Shengxiao, (2008) “Dynamic cartographic representation of SpatioTemporal data“, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVII. Part B2. Beijing 2008 .

[5] Z. Obradovic , D. Das, V.Radosavljevic, K.Ristovski, S.Vucetic, (2010) “Spatio-Temporal characterization of aerosols through active use of data from multiple sensors, “ISPRS TC VII Symposium – 100 Years ISPRS, Vienna, Austria, July 5–7, 2010

[6] Günther O., Noltemeier H.: ‘Spatial Database Indices For Large Extended Objects’,Proc. 7 th Int. Conf. on Data Engineering, 1991, pp. 520-527.

[7] Stefan Berchtold,Daniel A. Keim,Hans-Peter Kriegel., The X-tree:An Index Structure for High-Dimensional Data,Proceedings of the Twenty-second International Conference on Very Large Data-bases,Mumbai ,India .

[8] J.F. Roddick, K. Hornsby, and M. Spiliopoulou.YABTSSTDMR - yet another bibliography of temporal,spatial and spatio-temporal data mining research.In K.P. Unnikrishnan and R. Uthurusamy,eds, SIGKDD Temporal Data Mining Workshop,pages 167–175, San Francisco, CA, 2001. ACM.

[9] [Online] Available: http://www. http://www.cs.rpi.edu/~zaki/ software/

[10] [Online] Available: http://www. http://www.csc.liv.ac.uk/~frans/KDD/Software/FPgrowth/fpGrowth.html

[11] [Online] Available: http://www. http://www.csc.liv. ac.uk/~frans/KDD/Software/FPgrowth/FPtree.java

[12] [Online] Available: http://www. http://www.sigkdd.org/ kddcup/index.php?section=1998&method=data

[13] [Online] Available: http://www. http://www.kdnuggets.com/ software/associations.html

[14] [Online] Available: http://www. ttp://hen. wikipedia.org /wiki/Weka_machine_learning

[15] [Online] Available: http://www. http://en. wikipedia.org/ wiki/Weka_machine_learning#ARFF_file

[16] [Online] Available: http://www. http://www.cs.waikato. ac.nz/ml/weka/

[17] [Online] Available: http://www. http://sourceforge.net/ projects/weka/files/weka-3-7-windows-jre/3.7.4/weka-3-74jre.exe/download

[18] [Online] Available: http://www. www.sigkdd.org/kddcup/

[19] [Online] Available: http://www. http://kdd.ics.uci.edu/

[20] [Online] Available: http://www. http://www.kdnuggets.com/ datasets/

[21] Sujni Paul, (2010) “An Optimized Distributed Association rule mining algorithm in Parallel and distributed data mining with XML data for improved response time”, International Journal of Computer

Science and Information Technology, Volume 2, Number 2, April 2010.

[22] Yangming JIANG and Siwen BI, (2008) ”Dynamic Object-Oriented Model and its Applications for Digital Earth”, Digital Earth Summit on Geoinformatics, Nov,12-14,2008,Germany.

[23] ZHANG Ruiju et al, (2005) “An Object Oriented Spatio-Temporal Data Model”, Proceedings of International Symposium on Spatio-Temporal Modeling, Spatial Reasoning, Analysis, Data Mining and Data Fusion, 27-29, Aug,2005, peking University, China.

[24] S. Nadi and M.R.Delavar, (2005) “Toward a General Spatio-Temporal Database Structure for GIS applications”, Proceedings of International Symposium on Spatio-Temporal Modeling, Spatial Reasoning, Analysis, Data Mining and Data Fusion, 27-29, Aug,2005, peking University, China.

[25] Souheil Khaddaj,Abdul Adamu and Munir Morad, (2005) “Construction of an Integrated Object Oriented System for Temporal GIS”, American Journal of Applied Sciences 2(12), 2005, pp.1584-1594,ISSN 1546-9239.

[26] Salvatore Rinzivillo and Franco Turini, (2005) ”Extracting spatial association rules from spatial transactions”, Proceedings of GIS’05 13 annual ACM international workshop on geographic information systems,Germany.

[27] Schluter T and Conrad S, (2010) ”Mining Several kinds of Temporal association rules enhanced by Tree structures”, Second international conference on Information,Process and Knowledge Management,2010, Saint Maarten.

[28] R. Agrawal and R. Srikant. “Fast algorithms for mining association rules in large databases. “,Research Report RJ 9839, IBM Almaden Research Center, San Jose, California, June 1994.

[29] Koperski, K., and J. Han. 1995. Discovery of spatial association rules in geographic information databases. In Proceedings of 4th International Symposium on Large Spatial Databases, SSD95, Maine: 47-66.

[30] Iwaki, Hideki, Masaaki Kijima & Yuji Morimoto (2001). “An economic premium principle in a multiperiod economy,” Insurance: Mathematics and Economics 28,325-339.

AUTHORS PROFILE

Dr. M.N. Doja is Professor in the Department of Computer Engineering, Faculty of Engineering and Technology, Jamia Millia Islamia. He received his B.Sc. (Engg), M.Tech. and Ph.D. degrees from B.I.T, I.I.T. Delhi and Jamia Millia Islamia, New Delhi respectively. His areas of research are Software Engineering, Networks, Security, Simulation, Operating System and Soft Computing. of. Doja is a member of Academic Council, Board of Research Studies and Board of Studies of a number of universities including Ambedkar University Lucknow, NSIT New Delhi, A.M.U. Aligarh, Guru Gobind Singh Indraprastha University Delhi, NIT Jalandhar, Hamdard University New Delhi etc. He has been a member of a number of committee for various universities in various capacities. He has been expert member/member of various committee constituted by UGC and AICTE.

Sapna Jain is a Phd Fellow in the Jamia Hamdard University who has obtained her MCA (Masters of Computer Application) degree from Maharishi Dayanand University, ndia. Her area of research is Scalability of data mining algorithms.

Dr. M Afshar Alam is professor in Department of Computer Science, Jamia Hamdard,New Delhi.He has teaching experience of more than 17 years.He has authored 8 books and guided PhD research works.He has more than 30 publications in international/national/journal/conference proceddings. He has delivered special lectures as a resource person at various academic institutions and conferences He is a member of expert committees of UGC,AICTE and other national and international bodies.His research areas include software re-engineering,data miningbioinformatics and fuzzy databases.