Journal of Computer Science 3 (11): 882-886, 2007
ISSN 1549-3636

Exploring Spatial ARM (Spatial Association Rule Mining) for Geo-Decision Support System

[1]Ranjana Vyas, [2]Lokesh Kumar Sharma and [3]U.S. Tiwary
[1]School of Studies in Computer Science, Pt. Ravishankar Shukla University Raipur-492010, India
[2]Fraunhofer Institut Intelligente Analyse und Informationssysteme,
Schloss Birlinghoven, 53754 Sankt Augustin, Germany
[3]Indian Institute of Information Technology Allahabad, India

Abstract: Geographical Decision Support System (Geo-DSS) is a demanding field, since enormous amount of spatial data have been collected in various applications, ranging form Remote Sensing to GIS, Computer Cartography, Environmental Assessment and Planning. Although some efforts were made to combine spatial mining with Spatial Decision Support System but mostly researchers for spatial database are using a popular data mining approach-Apriori based association rule mining. There are two major limitations in existing approaches; the biggest being, that in a typical Apriori based spatial association the same records are required to be scanned again and again to find out the frequent sets. This becomes cumbersome, as spatial data is already known to be large in size. As far as *sparse data* is concerned, an Apriori based spatial association rule may even be considered but when there is *dense data* there were other approaches giving better performance. Researchers discuss only the positive spatial association rules; they have not considered the spatial negative association rules. Negative association rules are very useful in some spatial problems and are capable of extracting some useful and previously unknown hidden information. As this approach makes computation faster, it is thus better candidate for integration into Geo-DSS architectural framework. We have tried to design a particular Decision support system using spatial positive and negative association rule with efficient P-Tree and T-Tree.

Key words: Association Rule Mining (ARM), p-tree, t-tree, Decision Support System (DSS), Spatial Association Rule Mining (SPARM)

INTRODUCTION

Decision Support Systems have been under discussion, development and investigation by Information Systems researchers for more than 35 years. In 1970s many vendors, practitioners and academicians advocated DSS and much optimism about DSS applications were created. Despite the hyperbole, the success rate of DSS implementations has been less than anticipated. Recent development in database technology and a paradigm shift in Information Management with the emergence of OLAP, Data Warehousing and Data Mining during 1990's have shown definite promise to rejuvenate the DSS concepts, technologies and applications. A new generation of techniques and tools is emerging to intelligently assist humans in analyzing mountains of data and finding critical nuggets of useful knowledge and in some cases to perform the analyses automatically. Decision Support System is most widely used for making decision in complex systems (e.g. management and organizational operations, industrial processes or investment portfolios, the command and control of military units or the control of nuclear power plants). There is empirical evidence of human intuitive judgment and decision making which can be far from optimal and it deteriorates even further with complexity and stress. In many situations the quality of decision is important, one such important aspect can be the Geographical Decision support system which is one of the demanding fields, as huge amount of spatial data have been collected in various applications, ranging from remote sensing to GIS, computer cartography, environment assessment and planning. Many researchers had tried to combine spatial mining with spatial decision support system but mostly researchers used popular data mining Approach i.e. Apriori based ARM. The basic approach of Apriori has shown some

Corresponding Author: Ranjana Vyas, School of Studies in Computer Science, Pt. Ravishankar Shukla University Raipur, India-492010

major limitations and the biggest disadvantage of Apriori based spatial Association rule is that same datasets are required to be scanned again and again in order to find out the frequent sets. This approach is very cumber some in general and it becomes more expensive with spatial data, which is known to be large sized. When the nature of data is sparse, Apriori based spatial Association rule may even be considered but in case of dense data there were other approaches giving better performance. It is also been noted that most of the existing approaches mainly focused on mining positive spatial Association Rules whereas negative association rules are useful in some spatial problems and are capable of extracting some useful and previously unknown hidden information.

In this study we have proposed a novel approach of mining spatial positive and spatial negative association rule mining using P tree and T tree which are very useful in some spatial problems and capable of extracting some useful and previously unknown hidden information.

Spatial ARM (Spatial Association Rule Mining): Researchers mainly focus in reflecting structures of spatial objects and spatial/or Non spatial relationships that contain spatial predicates e.g. adjacent_to, near_by, inside, close_to, intersecting, etc Spatial association rules can represent object/predicate relationships containing spatial predicate. For ex the following rules are spatial association rules.

Non spatial consequent with spatial antecedent (s): Is_a (X, town)∧intersects (X, highway)→adjacent_to (X, water)…(80%).

Spatial consequent with non-spatial/spatial antecedent (s): Is_a (X, gas_station)→close_to (X, highway)..(75%).

Various kinds of spatial predicate can be involved in spatial association rules. They may represent topological relationships between spatial objects such as disjoin, intersects, inside/outside, adjacent_to, covers/covered_by equal, etc. They may also represent spatial orientation or ordering, such as left, right, north, east, etc, or contain some distance information, such as close_to, far_away, etc. For systematic study of the mining of spatial association rules, some preliminary concepts are discussed in[8].

Association rule discovery seeks rules of the form $P \rightarrow Q$ with support and confidence greater than or equal to, user specified support minimum support (ms) and minimum confidence (mc) thresholds respectively. This is referred to as the support-confidence framework[1] and the rule $P \rightarrow Q$ is an interesting positive association rule. An item set that meets the user specified minimum support is called frequent item set. Accordingly an infrequent item set can be defined as an item set that does not meet the user specified minimum support. Like positive rule, a negative rule $P \rightarrow \neg Q$ also has measure of its strength, confidence, defined as the ratio supp $(P \cup \neg Q)$/supp (P) where supp $(\neg Q)$ can be measured by 1- supp (Q).

Example Let supp (c) = 0.6, supp (t) = 0.4, supp $(t \cup c)$ = 0.05 and mc = 0.52. The confidence of $t \rightarrow c$ is supp $(t \cup c)$/supp (t) = 0.05/0.4 = 0.125<mc (= 0.52) and supp $(t \cup c)$ = 0.05 is low. This indicates that $t \cup c$ is an infrequent item set and that $t \rightarrow c$ cannot be extracted as rule in support confidence framework. However, supp $(t \cup \neg c)$ = supp (t)-supp $(t \cup c)$ = 0.4-0.05 = 0.35 is high and the confidence of $t \rightarrow \neg c$ is the ratio supp $(t \cup \neg c)$/supp (t) = 0.35/04 = 0.875>mc. Therefore $t \rightarrow \neg c$ is a valid rule.

By extending the definition in[8,9,10] negative spatial association rule discovery is proposed to be defined as follows:

The support of a conjunction of predicate, $P = P_1 \Lambda \ldots \Lambda P_m$, in a set S denoted as supp (P/S), is the number of objects in S which satisfy P versus the cardinality of S. The confidence of rule $P \rightarrow \neg Q$ is the ratio of supp $(P \wedge \neg Q/S)$ versus supp (P/S) i.e. the possibility that a member of S does not satisfy Q when the same member of S satisfies P. A single predicate is called 1-predicate. A conjunction of k single predicates is called a k-predicate

INCORPORATING INTERESTING ITEM SET

The proposed approach of mining both positive and negative association rule may have an exponential number of predicates in a database and only some of them are useful for mining association rule of interest. Therefore it is also an important issue to efficiently search the interesting itemset. In this study we have used a pruning strategy[17] to find out potentially interesting itemset. An interesting function[16,17], interest (X, Y) = $\left|$ supp $(X \cup Y)$-supp (X) supp (Y) $\right|$ and a threshold mi (minimum interestingness) are used. Using this approach, we can establish an effective pruning strategy for efficiently identifying all frequent itemsets of potential interest in a database[12].

Integrating this interest (X,Y) mechanism into the support-confidence framework, for both positive and negative rule discovery, our search is constrained to seeking interesting rules on certain measures and pruning is the removal of all uninteresting branches that

cannot lead to an interesting rule that would satisfy those constraints. This concept of Interestingness becomes prudent in our approach of Spatial Association Rule Mining (SPARM) as it addresses the issues arising out of mining negative association rules as well.

We have earlier explored[12].The method of mining spatial association rule with example of thematic data of Chhattisgarh state-India, in which we have examined the method of posing a data mining query. In this approach firstly a set of relevant data is retrieved by and then it is generalized close_to relationship between towns and the other four classes.

Of entities is computed at a relatively coarse resolution level using a less expensive spatial algorithm such as the MBR data.

We can have large K-predicate sets, at the first level (for 50 towns in Chhattisgarh) spatial Association Rules can be extracted directly from this table then similarly large K-predicate sets can be extracted at the sec level and then again the spatial Association rule can be extracted from it.

AN ALGORITHM FOR MINING SPATIAL ASSOCIATION RULE

Algorithm: Mining the spatial association rules defined in a large spatial database.

Input: The input consists of spatial database, a mining query and two thresholds.

Output: Strong multiple level spatial association rules for the relevant sets of objects and relations.

Explanation of the detailed steps of the algorithm

Step 1: It is accomplished by the execution of a spatial query. All the task relevant objects are collected into one database.

Step 2: In this step we execute some efficient spatial algorithm at a coarse resolution level. For example, R-tree or fast MBR technique and plane-sweep algorithm[3,11] can be applied to extract the objects, which are approximately close to each other, corresponding to computing g_close_to for the task relevant data.

Step 3: In this step, we use the concept of partial support counting using the P-tree (Partial support tree). The idea is to copy the input data (in one pass) into a data structure, that maintains all the relevant aspects of the input and then this maintains all the relevant aspects of the input and then mine this structure. A P-tree is a set enumerated tree structure in which to store partial counts for item sets. The top, single attribute, level comprises an array of references to structures of the form shown to the right, one for each column. Each of these top-level structures is then the root of a sub-tree of the overall P-tree[2]. The advantages offered by the P-tree table are:

- Reduced storage requirements (particularly where the data set contained duplicate rows)
- Faster run times because the desired total support counts had already been partially calculated

Step 4: In this step first, we examine the P-tree and create T-tree[2]. The T-tree is generated in an Apriori manner. There are a number of features of the P-tree Table that enhance the efficiency of this process:

- The first pass of the P-tree will be to calculate supports for singletons and thus the entire P-tree must be traversed. However, on the sec pass when calculating the support for doubles we can ignore the top level in the P tree, i.e. we can start processing from index 2. Further, at the end of the previous pass we can delete the top level (cardinality = 1) part of the table. Consequently as the T-tree grows in size the P-tree table shrinks
- To prevent double counting, on the first pass of the P-tree, we update only those elements in the top-level array of the T-tree that correspond to the column numbers in node codes (not parent codes). On the sec pass, for each P-tree table record found, we consider only those branches in the T-tree that emanate from a top level element corresponding to a column number represented by the node code (not the parent code). Once the appropriate branch has been located we proceed down to level 2 and update those elements that correspond to the column numbers in the *union* of the parent and node codes. We then repeat this process for all subsequent levels until there are no more levels in the T-tree to consider

IMPLEMENTATION

The Algorithm explained here was implemented taking thematic map data of Chhattisgarh state of India and using programming language JAVA. The experiment was performed on a Pentium IV having 128 MB RAM. The performances of various Spatial A. R. M. were observed while increasing the Number of Objects in the spatial database and comparing their performance in terms of execution time.

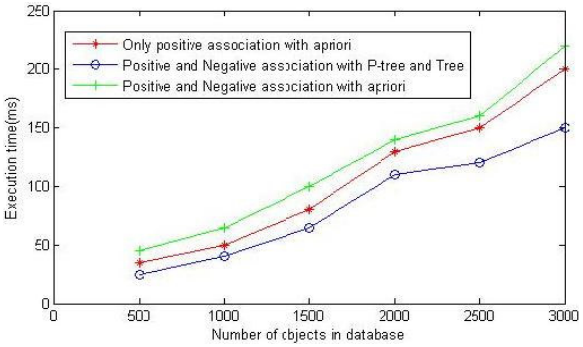

Fig. 1: Graph showing performance of SPARM algorithms generating multilevel positive and negative Associationships

Implementation of the algorithms generated at multilevel positive and negative Associationships. Figure 1 shows the performance of the various algorithms for generating spatial association rules and it is evident from comparative graphs that the algorithm with both Positive and Negative Association rules having P-tree and T-tree outperforms the other approaches.

INTEGRATION OF SPARM WITH P-TREE and T-TREE INTO GEO-DSS

Integration of Data Mining Technologies and DSS though has been discussed in some literature but there were limitation in perception of DMT, as only Classifiers were conceived for integration with DSS. Also a Decision Support System was mostly discussed for simple data types, but a special DSS for Complex data types like Multimedia data, Image files and thematic data was not considered for such a integration. Our proposed approach also uses the descriptive modeling approach such a Association Rule Mining for Spatial data. We propose Spatial Association Rule Mining using P-tree and T-tree to be integrated into a Geo-DSS (Fig. 2).

DSS can be either data or model oriented. Data oriented DS tools involve no models, but enable good understanding through segmentation, slicing, dicing, drilling down, rolling-up and other operations. The proposed architectural framework of our Geo-DSS does not migrate from the general theoretical approach to decision-making, which follows the Simon's model of decision-making but at the same time it takes into account the suggestions of Frank Kriwaczek et al.[15]. While one of the significant DSS architecture proposal by E.G. Mallach[11] consists of a database, a model base, possibly a knowledge base and a user interface.

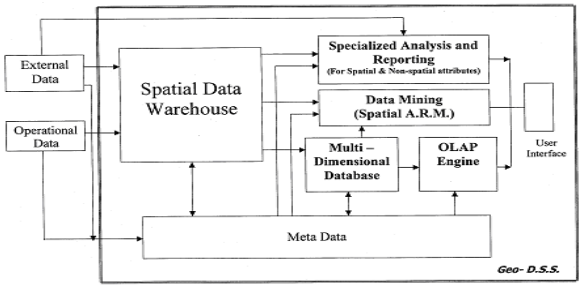

Fig. 2: Architectural integration of Spatial ARM into a Geo-DSS

The architecture proposed for Geo-DSS has Spatial Data Warehouse getting feed from both the External data source and Operation data source as well. The Spatial Data Warehouse with Meta Data may be integrated with the specialized Association Rule Mining, aptly designed for spatial database. The provisions for Specialized Analysis and Reporting for Spatial and Non Spatial attributes enables the customized presentation. The quality of many decisions in business and politics concerning e.g., site planning, marketing, distribution and spatial development directly depends on how easily spatial factors can be taken into account on all levels.

With our proposed services we enable the potential users to turn spatio-temporal information into a ubiquitous high quality component of their business processes and services. Our proposed Geo-DSS will let everyone explore spatially referenced factors intuitively by manipulating interactive thematic maps. Maps combined with data mining, modeling and simulation methods allow synthesizing all available data, from CRM, data warehouses into compact and comprehensive foundations for decisions, control and automated prognosis.

CONCLUSION

A modern and effective Spatial DSS having all the requisite support technologies like OLAP, Specialized Analysis and Reporting along with Spatial Association Rule Mining mechanism is required for future planners and decision makers; new and efficient methods are needed to integrate the related Information Technologies to discover knowledge from large spatial databases. Geo-DSS, serving such purposes is fast becoming an essential software tool for government officials and city planners- for decisions, analysis, planning and management of census, voting and mining and also for developing, systems for consultation and integration.

Although there were some attempts of using Association Rule Mining Algorithm for Spatial data but the approach for the same was having severe limitations. The present study not only proposes a novel and more efficient approach for ARM but also gives an architectural framework of a Geo-DSS incorporating this new approach. Basically, the algorithm presented in this study discusses efficient mining procedures for positive and negative spatial association rules, which explores techniques at multiple approximation and abstraction levels. There are several advantages of the T- tree based approaches over others, as it constructs a highly compact P-tree which is usually substantially smaller than the original database and thus save the cost of subsequent mining process. This method also deals with the cases where there exist multiple concept hierarchies.

The proposed architecture though discusses only Spatial Association Rule Mining, that may be a limiting factor but the same approach may include other efficient spatial data mining approaches, while implementing the P-tree and T-tree methods coupled with both Positive and Negative Association Rule Mining.

REFERENCES

1. Agrawal, R. and R. Srikant, 1994. Fast algorithm for mining association rule. In: Proc. Int. Conf. VLDB Santiago, Chile, pp: 487-499.
2. Francs, C., P. Leng and S. Ahmed, 2004. Data structure for association rule mining: T-tree and P-tree: IEEE Transaction on Knowledge Discovery and Data Eng., 16 (6): 774-778.
3. Dunham, M.H., 2003: Data Mining Introductory and Advance Topics. Pearson Education Inc.
4. Guetting, R.H., 1994. An Introduction to Spatial Database Systems: Special Issue on Spatial Database System of the VLDB J., 3 (4): 357-399.
5. Han, J. and Y. Fu, 1995. Discovery of Multiple Level association rules from large database. In: Proc. Int. Conf. VLDB, pp: 420-431.
6. Han, J., J. Pei, Y. Yin and R. Mao, 2004. Mining Frequent Patterns with out Candidate Generation: A Frequent Pattern Tree Approach: Data Mining and Knowledge 7. Discovery: An Int. J., Kluwer Academic Publishers, 8 (1): 53-87.
8. Han, J., J. Pei and Y. Yin, 2000. Mining Frequent Patterns without Candidate Generation: ACM-SIGMOD, pp: 1-12.
9. Koperski, K. and J. Han, 1995. Discovery of Spatial Association Rules in Geographic Information Database: Springer-Verlag LNCS, 951, pp: 47-66.
10. Malerba, D. and F.A. Lisi, 2001. An ILP method for spatial association rule mining. Working notes of the first workshop on Multi Relational Data mining, Freiburg, Germany, pp: 18-29.
11. Malerba, D., F.A. Lisi and A.A. Francesco, 2003. Mining Spatial Association Rules in Census Data: A Relational Approach: Intell. Data Anal., 7 (6): 541-566.
12. Mallach, E.G., 1994. Understanding Decision Support Systems and Expert Systems, Irwin.
13. Sharma, L.K., O.P. Vyas, U.S. Tiwary and R. Vyas, 2005. A Novel Approach of Multilevel Positive and Negative Association Rule Mining for Spatial Databases: Springer Verlag LNAI 3587, pp: 620-629.
14. Sharma, L.K., U.S. Tiwary and O.P. Vyas, 2004. An efficient approach to Spatial Association rule mining. Int. Conf. On ISPR IIIT Allahabad, India.
15. Shekhar, S., S. Chawla, S. Ravadam, X. Liu and C. Lu, 1999. Spatial databases-accomplishments and research needs: IEEE Transactions on Knowledge and Data Eng.,. 11 (1): 45-55.
16. Smith, G.B. and S.M. Bridge, 2002. Fuzzy Spatial Data Mining: IEEE NAFIPS, pp: 184-189.
17. Vyas, R., L.K. Sharma and O.P. Vyas, 2004. A novel approach to spatial association rule mining with multilevel multidimensionality: National conference on IORPM 2004.
18. Wu, X., C. Zhang and S. Zhang, 2004. Efficient mining of both positive and negative association rule: ACM Transaction on Inform. Syst., 22 (3): 381-405.