

Optimization of Spatial Association Rule Mining using Hybrid Evolutionary algorithm

J.Arunadevi

Lecturer, Department of Computer Applications
Thiagarajar School of Management, Madurai,
India

Dr.V.Rajamani

Principal
Indra Ganesan College of Engineering
Tirchirapalli, India

ABSTRACT

Spatial data refer to any data about objects that occupy real physical space. Attributes within spatial databases usually include spatial information. Spatial data refers to the numerical or categorical values of a function at different spatial locations. Spatial metadata refers to the descriptions of the spatial configuration. Application of classical association rule mining concepts to spatial databases is promising but very challenging. Spatial Association Rule Mining requires new approaches compared to classical association rule mining. Spatial data consists of dependent events compared to transactional data which consist of independent transactions. It is more difficult to classify a discovered spatial association rule as interesting. Instead of much generalized rule more specific rule discovery needs further research.

Spatial Association Rules are association rules about spatial data objects. Either the antecedent or the consequent of the rule must contain some spatial predicates. Spatial association rules are implications of one set of data by another. The main area of concentration in this paper is to optimize the rules generated by Association Rule Mining (Apriori method), using hybrid evolutionary algorithm. The main motivation for using Evolutionary algorithms in the discovery of high-level prediction rules is that they perform a global search and cope better with attribute interaction than the greedy rule induction algorithms often used in data mining. The improvements applied in EAs are reflected in the rule based systems used for classification as described in results and conclusions. The future enhancements will be on using the other Evolutionary Optimization Algorithms such as PSO (Particle Swarm Optimization) for the rule generation.

Keywords

Spatial Association Rule Mining, Evolutionary Optimization Algorithms, Genetic Algorithms, ACO

1. INTRODUCTION

Spatial data refer to any data about objects that occupy real physical space (usually geographic). Attributes within spatial databases usually include spatial information (metadata). Spatial metadata are used to describe a spatial object's geometry and topology. Spatial metadata includes geometric metadata (e.g., shape, size, area, perimeter, location, elevation) and topological metadata (e.g., "neighbor of", "adjacent to", "included in", "includes"). Spatial data have spatial metadata as attributes. The phrase "spatial data" usually refers to both the data and the metadata. More specifically, spatial data refers to the numerical

or categorical values of a function (or measurements) at different spatial locations (e.g., rainfall, land use, population density, property values, elevation, biodiversity, vegetation index, crime rates, ice cover, sea surface temperature). Spatial metadata refers to the descriptions of the spatial configuration (e.g., geometry, topology, spatial relationships).

Spatial data items are naturally linked to neighboring data elements (e.g., contiguous geographic positions). These data elements are not statistically independent. Spatial metadata usually include unique attributes and unique relationships between attributes that are not normally found in other databases. In many cases, spatial data are stored as rasters - rows and columns, with (x, y) location information implied by the placement within the raster (e.g., Remote Sensing images). The (x, y) values are not stored anywhere in the database. Special extraction tools are needed (such as GIS). These differences can pose challenges to standard data mining and database querying algorithms.

Spatial Association Rules are association rules about spatial data objects. Either the antecedent or the consequent of the rule must contain some spatial predicates (such as near) [1]. Spatial association rules are implications of one set of data by another such as the average monthly family income in Madurai for families living near Anna nagar is Rs. 100,000.

Spatial association rule mining is difficult to solve this type of problems directly and always results in exponential complexity. In this paper, we present a novel hybrid evolutionary algorithm (HEA) which uses genetic algorithm with ant colony optimization for spatial association rule mining. The proposed HEA algorithm is to enhance the performance of Multi objective genetic algorithm (MOGA) by incorporating local search, ant colony optimization (ACO), for Multi objective association rule mining. In the proposed HEA algorithm, genetic algorithm is conducted to provide the diversity of associations. Thereafter, ant colony optimization is performed to come out of local optima. From the experiment results, it is shown that the proposed HEA algorithm has superior performance when compared to other existing algorithms.

In general the main motivation for using Genetic Algorithms in the discovery of high-level prediction rules is that they perform a global search and cope better with attribute interaction than the greedy rule induction algorithms often used in data mining.[2].

2. SPATIAL ASSOCIATION RULE MINING

A spatial association rule is of the form $X \Rightarrow Y (c\%)$, where X is called antecedent and Y consequent of the rule. The

antecedent contains a set of predicates from the exploring database, the consequent only represents one predicate, which is not yet included in the antecedent. The rule itself then reflects an existing relationship between predicates in antecedent and consequent. A measure of a rule's strength is confidence ($c\%$), which indicates that c percent of the items satisfying the antecedent also satisfies the consequent. In a large database many association relationships may exist but some may occur rarely or may not hold in most cases. To focus our study to the patterns which are relatively strong, i.e., which occur frequently and hold in most cases, the concepts of minimum support and minimum confidence are introduced. spatial association rules represents object/predicate relationships containing spatial predicates. For example, the following rules are spatial association rules.

- Nonspatial consequent with spatial antecedent(s)

$$\text{is_a}(x,\text{house}) \wedge \text{close_to}(x,\text{beach}) \rightarrow \text{Is_expensive}(x) \quad (90\%)$$
- Spatial consequent with non-spatial /spatial antecedent(s).

$$\text{Is_a}(x,\text{gas_station}) \rightarrow \text{close_to}(x,\text{highway}). \quad (75\%)$$

Various kinds of spatial predicates can be involved in spatial association rules [5].

3. MULTI OBJECTIVE GENETIC ALGORITHM (MOGA)

Multi-objective optimization deals with solving optimization problems which involve multiple objectives. Most real-world search and optimization problems involve multiple objectives (such as minimizing fabrication cost and maximize product reliability, and others) and should be ideally formulated and solved as a multi-objective optimization problem. However, the task of multi-objective optimization is different from that of single-objective optimization in that in multi-objective optimization, there is usually no single solution which is optimum with respect to all objectives. Classical search and optimization methods usually work with a point-by-point principle and thus are required to be applied many times, each time finding one Pareto-optimal solution. Moreover, the efficacy of classical methods largely depends on the shape of the Pareto-optimal region, discreteness of the search space, presence of constraints, and others. Over the past decade, population-based evolutionary algorithms (EAs) (genetic algorithms (GAs) and evolution strategies (ESs)) have been found to be quite useful in solving multi-objective optimization problems, simply because of their ability to find multiple optimal solutions in a single simulation run.

4. ANT COLONY OPTIMIZATION

Ant Colony Optimization (ACO) is a paradigm for designing metaheuristic algorithms for combinatorial optimization problems. The ACO algorithm was first introduced by Colnori, Dorigo and Maniezzo [9] [10] and the first Ant System (AS) was proposed by Dorigo in his Ph.D. thesis [11]. The ACO is a meta-heuristic algorithm, which utilizes the inspiration from real

ant colonies behaviors to find a shortest path from a food source to the nest without using visual cues by exploiting pheromone information [12] [13] [14]. When ant colonies are seeking for food, they leave a kind of chemical compositions, which is called pheromone. The more ants walk through the path, the more pheromone left on the ground. Then, the next ant will choose one path with a probability proportional to the amount of pheromone. Finally this positive feedback process will construct a shortest path from their nest to the food source. The characteristic of ACO algorithms is their explicit use of elements of previous solutions.

Edge Selection:

An ant will move from node i to node j with probability

$$P_{i,j} = \frac{\tau_{i,j}^\alpha \eta_{i,j}^\beta}{\sum \tau_{i,j}^\alpha \eta_{i,j}^\beta}$$

where

$\tau_{i,j}$ is the amount of pheromone on edge ij

α is a parameter to control the influence of $\tau_{i,j}$

$\eta_{i,j}$ is the desirability of edge ij (a priori knowledge, typically $1/d_{i,j}$)

β is a parameter to control the influence of $\eta_{i,j}$

Pheromone Update

$$\tau_{i,j} = (1 - \rho)\tau_{i,j} + \Delta\tau_{i,j}$$

where

$\tau_{i,j}$ is the amount of pheromone on a given edge ij

ρ is the rate of pheromone evaporation

and $\Delta\tau_{i,j}$ is the amount of pheromone deposited, typically given by

$$\Delta T_{i,j}^k = \begin{cases} 1/L_k & \text{if ant } k \text{ travels on edge } i,j \\ 0 & \text{otherwise} \end{cases}$$

where L_k is the cost of the k th ant's tour (typically length).

5. APPLICATION OF HEA FOR SPATIAL ASSOCIATION RULE MINING

During the last two decades, evolutionary algorithms (EAs) have been applied to a wide range of optimization and decision-making problems. Association rule-mining algorithm involves many criteria like comprehensibility, confidence factor and interestingness [7], therefore we treated it as a multiobjective problem rather than single objective one. So we can consider it as a multiobjective problem rather than single objective one. There are two basic approaches to represent the rules, named as Pittsburgh and Michigan. In the Pittsburgh approach each chromosome represents a set of rules and this approach is more suitable for classification rule mining [7]; as we do not have to decode the consequent part and the length of the chromosome limits the number of rules generated. The other approach is called Michigan approach where each chromosome represents a separate rule. In this paper we use the Michigan approach to develop the individuals. The procedures of HEA are as follows. First, MOGA searches the solution space and generates association lists to provide the initial population for ACO. Next, ACO is executed, when ACO terminates, the crossover and

mutation operations of MOGA generate new population. ACO and GA search alternately and cooperatively in the solution space.

Pseudocode for optimization of rule generation

```

1. while (t <= no_of_gen)
2. M_Selection(Population(t))
3. ACO_MetaHeuristic
   while(not_termination)
       generateSolutions()
       pheromoneUpdate()
       daemonActions()
   end while
end ACO_MetaHeuristic
4. M_Recombination_and_Mutation(Population(t))
5. Evaluate Population(t) in each objective.
6. t = t+1
7. end while
8. Decode the individuals obtained from the population with
high fitness function

```

6. RESULTS

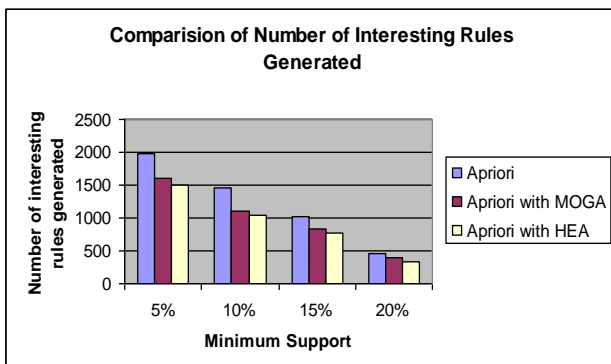
We have used the synthesized dataset for our research. **Data pre-processing.** The general procedure of data mining is: question raise → data preparation (including data selection, data pretreatment and data transformation) → data arrangement → model building/data mining → result evaluation and explanation. Data preparation is the key which determines the success of data mining. The process of spatial data is much more complex[8]. After preprocessing we have transformed the spatial data in term of .xls file

We have implemented the basis of the apriori algorithm of association rule, we programmed to complete the calculation in virtue of M-language in Matlab. The specific procedure is as following.

(1) Take advantage of “import wizard” in Matlab to accomplish the import of data file. Until now, the data fields and character fields are saved separately. For example, the default uses a matrix named “data_num” to keep numerical fields and a matrix named “textdata” to keep character fields.

(2) Run the association rule program to calculate the support and confidence .

The results have been given by the comparison of rules generated with the Apriori association algorithm, the Apriori algorithm with the MOGA and the Apriori algorithm with HEA



7 CONCLUSION

From the graph we can observe that the number of rules generated by using our approach is comparatively less than the number of rules generated by the apriori algorithm and relatively less than MOGA. we demonstrated the effect of HEA-based rule selection as a post-processing procedure in the second phase of association rule mining. Since the results are promising we can extend the experiment using the using the other Evolutionary Optimization Algorithms such as PSO (Particle Swarm Optimization) for the rule generation.

8. REFERENCES

- [1] Margaret H.Dunham and S.Sridhar. “Data Mining Introductory and Advanced Topics”, Pearson Education, 2006.
- [2] Alex A. Freitas, “A Survey of Evolutionary Algorithms for Data Mining and Knowledge Discovery” Postgraduate Program in Computer Science, Pontificia Universidade Catolica do Parana Rua Imaculada Conceicao, 1155. Curitiba - PR. 80215-901. Brazil.
- [3] Dehuri, S., Jagadev, A. K., Ghosh A. And Mall R. 2006. Multi-objective Genetic Algorithm for Association Rule Mining Using a Homogeneous Dedicated Cluster of Workstations. American Journal of Applied Sciences 3 (11): 2086-2095, 2006 ISSN 1546-9239.
- [4] Peter P. Wakabi-Waiswa and Venansius Baryamureeba. Extraction of Interesting Association Rules Using Genetic Algorithms. International Journal of Computing and ICT Research, Vol. 2, No. 1, pp. 26 – 33. <http://www.ijcir.org/volume2-number1/article4.pdf>.
- [5] www.dsi.unive.it/~dm/ssd95.pdf
- [6] Freitas, A.A., 2003. A survey of evolutionary algorithms for data mining and knowledge discovery. In: A. Ghosh, S. Tsutsui (Eds.), Advances in Evolutionary Computing, Springer Verlag, New York, pp: 819–845.
- [7] Dehuri, S. and R. Mall, 2004. Mining predictive and comprehensible rules using a multi-objective genetic algorithm. Advance Computing and Communication (ADCOM), India.
- [8] “Association Rule Analysis of Spatial Data Mining Based on Matlab”, Workshop on Knowledge Discovery and Data Mining, 2008 IEEE DOI 10.1109/WKDD.2008.21
- [9] A . Colorni, M. Dorigo, and V. Maniezzo. Positive feedback as a search strategy. Technical Report No. 91-016, Politecnico di Milano, Italy, 1991.
- [10] A. Colorni, M. Dorigo, and V. Maniezzo . The ant system: an autocatalytic process. Technical Report No. 91- 016, Politecnico di Milano, Italy, 1991.
- [11] R. Beckers, J.L. Deneubourg and S. Goss. Trails and Uturns in the selection of the shortest path by the ant lasius niger. Journal of Theoretical Biology, 159, 1992, pp. 397- 415.
- [12] M. Dorigo. Optimization, Learning and Natural Algorithms. Ph.D. Thesis, Politecnico di Milano, Italy, 1992

[13] S. Goss, S. Aron, J.L. Deneubourg and J.M. Pasteels ().
Self-organized shortcuts in the argentine ant.
Naturwissenschaften, 76, 1989, pp. 579-581.

[14] B. Holldobler and E.O. Wilson (1990). *The Ants*. Springer-
Verlag: Berlin.