International Journal of Computer Applications (0975 – 8887)
Volume 51– No.3, August 2012

Estimation of Evolutionary Optimization Algorithm for Association Rule using Spatial Data Mining

N. Naga Saranya
Karpagam University
Coimbatore - 641021

M. Hemalatha
Karpagam University
Coimbatore - 641021

ABSTRACT
The innovative process for spatial data is more risk when compared to relational data. This can be functional for the efficiency and effectiveness of algorithms as well as the difficulty of possible patterns that can be establish in a spatial database. To optimize the rules generated by Association Rule Mining (Apriori method) [1] use hybrid evolutionary algorithm. This research paper present a novel hybrid evolutionary algorithm (HEA) [2] which uses particle swarm optimization for spatial association rule mining with clustering. The proposed HEA algorithm is to enhance the performance of Multi objective genetic algorithm [3][4] by incorporating local search, particle swarm optimization (PSO), for Multi objective association rule mining. Thereafter, particle swarm 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.

Keywords
Spatial Data Mining, Apriori Algorithm, Satellite Data, Hybrid Evolutionary Algorithm, Particle Swarm Optimization

1. INTRODUCTION
Advances in database and data achievement technologies have resulted in huge amount of spatial data, much of which cannot be gladly explored using conventional data analysis techniques. The purpose of spatial data mining is to computerize the mining of exciting and of use patterns that are not clearly represented in spatial datasets.

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) [1][5], 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 work will be on using the other Evolutionary Optimization Algorithms such as PSO (Particle Swarm Optimization) for the rule generation.

2. RELATED STUDY
Qin Ding; Qiang Ding; Perrizo, W [6], to get better association rule from spatial data, the author proposed an efficient approach using Peano count tree (P-tree) structure. This P-tree structure provides a lossless and compressed representation of spatial data. For the rule generation the PARM algorithm is compared with FP-growth and Apriori algorithms and gives best results

Jiangping Chen; Yanan Chen; Jie Yu; Zhaohui Yang; [8], used the apriori for association rule mining and spatial autocorrelation and regression is implemented using spatial data. Here they compared the results between spatial autocorrelation and spatial association rule mining.

Wei Ding Eick, C.F. Jing Wang Xiaojing Yuan,[9], discussed an integrated approach is used for region rule mining and introduced a novel framework to mine regional association rules. The proposed framework is evaluated in a real-world case study that identifies spatial risk patterns from arsenic in the Texas water supply.

Jiangping Chen,[10], discussed the association rule mining for spatial autocorrelation with an cell structure theory. Here it provides an algebra data structure with that rule, then the autocorrelation of the spatial data can be calculated in algebra.

3. PROBLEM DEFINITION
Mining spatial association rules can be defined as below[5]:

Input:
A spatial database (SDB) including geography graph and attribute tables, two series of thresholds

```
SDB(attr, th)
 for every large k-itemset in the spatial database,
{
    minsup[l] and minconf[l] for large 1-itemset
    and
    minsup[k] and minconf[k] for large k-itemset.
}
```

Output:
Output(Some strong spatial association rules)

3.1 Related Definition
Definition 1.
A spatial association rule is a rule in the form shown in Eq.(1).

$P_1 \cap \ldots \cap P_m \Rightarrow Q_1 \cap \ldots \cap Q_n$ (s%; c%_) …….. (1)

where at least one of the predicates $P_1, \ldots P_m$, and $Q1, \ldots Q_n$ is a spatial predicate.
s% is the support of the rule
c% is the confidence of the rule

Definition 2.
The support of a conjunction of predicates shown in Eq.(2) in a set S,

$P = P_1 \cap \ldots \cap P_m$ ……… (2)

It is denoted as ρ(P/S), is the number of objects in S which satisfy P versus the total number of objects of S.
The confidence of a rule P-ÆQ in S, Ψ(P-ÆQ/S), is the possibility that Q is satisfied by a member of S when P is satisfied by the same member of S. A single predicate is

called a 1-predicate. A conjunction of k single predicates is called a
k-predicate. In this paper, the large itemset contains k predicates is called k-itemset, and the set of all the large k-itemset is L_k.

3.2 Apriori Algorithm

Pseudo-code for Apriori Algorithm:

C_k: Candidate itemset of size k
L_k: frequent itemset of size k

L_1 = {frequent items};
for (k= 1; L_k!= ∅; k++) do begin
    C_{k+1} = candidates generated from L_k;
    for each transaction t in database do
    increment the count of all candidates in C_{k+1} that are contained in t
    L_{k+1}= candidates in $C_{k+1}$with min_support
end
return $_kL_k$;

4. ARCHITECTURE OF HYBRID EVOLUTIONARY ALGORITHM

As reported in the literature[3][4][7][5], several techniques and heuristics/meta heuristics have been used to improve the general efficiency of the evolutionary algorithm. Some of most used hybrid architectures are summarized as follows:
1.Hybridization between several Spatial Association Rule mining with clustering
2. Neural network assisted evolutionary algorithms
3. Spatial clustering assisted evolutionary algorithm
4. Particle Swarm Optimization (PSO) assisted evolutionary algorithm
5. Hybridization between evolutionary algorithm and other heuristics (such as local search, tabu search, simulated annealing, hill climbing, dynamic programming, greedy random adaptive search procedure, etc)

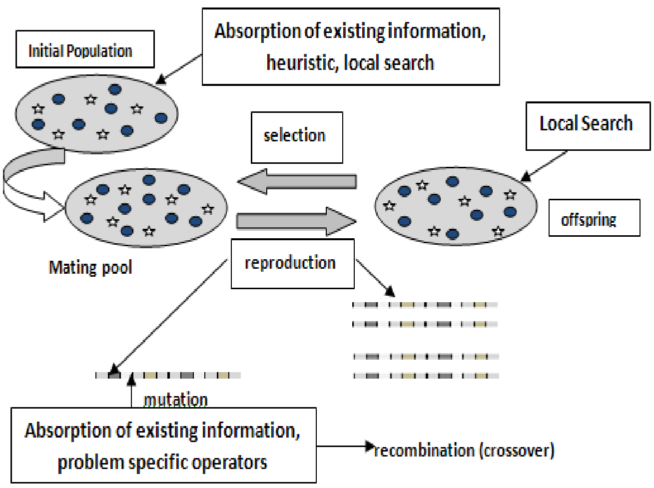

Fig. 1: Hybrid evolutionary algorithm generic architectures

Fig 1, illustrates some of the generic architectures for the various types of hybridization. By problem, we refer to any optimization or even function approximation type problem and intelligent paradigm refers to any computational intelligence technique, local search, optimization algorithms etc.

4.1 Particle Swarm Optimization

In this paper we propose SAR[7] based on the HEA[5][15], the MOGA and Association rule mining with clustering. The first stage generates the optimized spatial association rules by the use of the HEA. In the second stage rule cover is applied to the association rules for clustering optimized with GA. Next stage the Multi label rules are generated by the MOGA[3][4]. Final stage [13][14], the Multi label classifier is built with a sorting mechanism applied to the rules generated.

Pseudo code for optimization of rule generation
1. while (t <= no_of_gen)
2. M_Selection(Population(t))
3. PSO_MetaHeuristic
    while(not_termination)
      generateSolutions()
      pheromoneUpdate()
      daemonActions()
    end while
  end PSO_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.

The fitness function is calculated as the arithmetic weighted average confidence, comprehensibility and J-Measure.

The fitness function is given by
f(x) = [(w1 * Comprehensibility) + (w2 * J-Measure) + (w3 * Confidence)] [w1+w2+w3]
where w1, w2, and w3 are used defined weights.

Pseudo code for clustering the rules generated
Input :
set of rules generated by the HEA Ry={ Xi -> Y | i=1,2,…,n } and the rule cover.
Apply GA for rearranging the rules in various orders based on the fitness preferred by the user.
1. Generate the cluster rule cover
2. count = number of records in the cluster cover
3. while(no of records in the cluster cover > 2% of count)
    Sort all the rules in the Ry in the descending order of the rule cover.
    Take the first rule r with highest rule cover
  If the no of records in the rule cover is <= 2% of count
    Exit while loop
  End if.
4. $r_y = r_y$ U r
5. Delete the highest rule cover from the cluster cover
6. End While

Output :
Output(the representative rule set)

Apply GA[11][12] for retaining nearest neighbours in common cluster.
The optimized representative rule set is used for the segmentation of the consequent. GA is applied at the first stage for the arrangement of the rules based on the fitness; this is to help the clustering for not suffering from the order of the input.

5. FRAMEWORK FOR PROPOSED METHOD

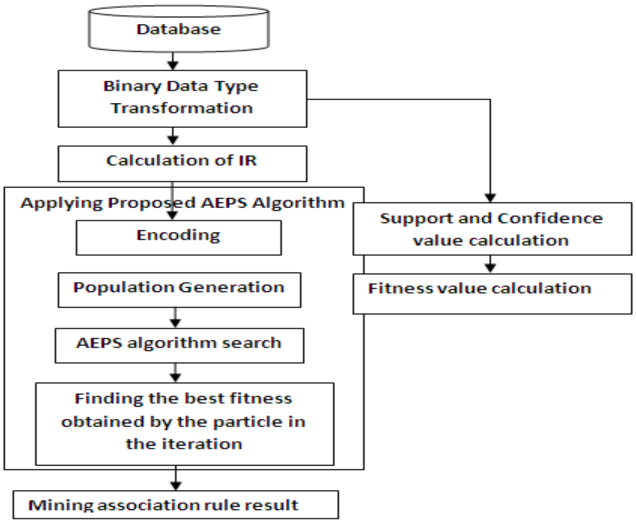

Fig. 2. Framework for Proposed AEPS Algorithm

5.1 Proposed AEPS Association Rule Mining

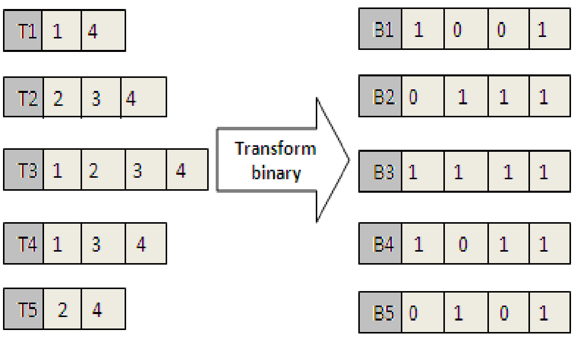

Fig. 3. Data transformation- Binary Data

In Fig. 3, there are five records, say T1 to T5, in the original data. Each of these records is transformed and stored as a binary type. For instance, there are a total of only four different products in the database, so four cells exist for each transaction. Take B4 as an Fig. 3. Data type transformation. example, this transaction only purchased products 2 and 3, so the values of cells 2 and 3 are both "1s," whereas cells 1 and 4 are both "0s."

5.2 IR Value Estimation

This study applies the AEPS algorithm in association rule discovery, as well as in the calculation of IR value which is included in chromosome encoding. The purpose of such an inclusion is to produce more meaningful association rules. Moreover, search efficiency is increased when IR analysis is utilized to decide the rule length generated by chromosomes in particle swarm evolution. IR analysis avoids searching for too many association rules, which are meaningless itemsets in the process of particle swarm evolution. This method addresses the front and back partition points of each chromosome, and the range decided by these two points is called the IR, which is shown in Eq. (3):

IR = [log(mTransNum(m)) + log(nTransNum(n))] [Trans(m, n) /TotalTrans] …. (3)

6. PROPOSED AEPS ALGORITHM

C_k: Candidate itemset of size k
L_k: frequent itemset of size k
L_1= {frequent items};
for (k= 1; L_k!=$\varnothing$; k++) **do begin**
  C_{k+1}= candidates generated from L_k;
  for each transaction t in database do
    increment the count of all candidates in C_{k+1} that are contained in t
  L_{k+1}= candidates in $C_{k+1}$with min_support
end
return $\cup_k L_k$;
for each particle do
  initialize position and velocity of particle
end for
for each particle do
  calculate fitness value
  if fitness value is better than best fitness value in particle history then
    take current value as new
  end if
end for
choose as the particle with best fitness value among all particles in current iteration
for each particle do
  calculate particle velocity
  update particle position
end for

7. IMPLEMENTATION RESULTS

We have used the synthesized dataset, which has been collected from UCI datasets. This data has been collected based on the geographic information from multi spectral satellite landsat image data for our research. The general procedure of data mining is:

- Data preparation (including data selection, data pre treatment and data transformation)
- Data arrangement
- Model building/data mining
- Result evaluation and explanation.

The sample dataset is shown in Fig 4. Apply AEPS algorithm over the datasets to get best rule generation.

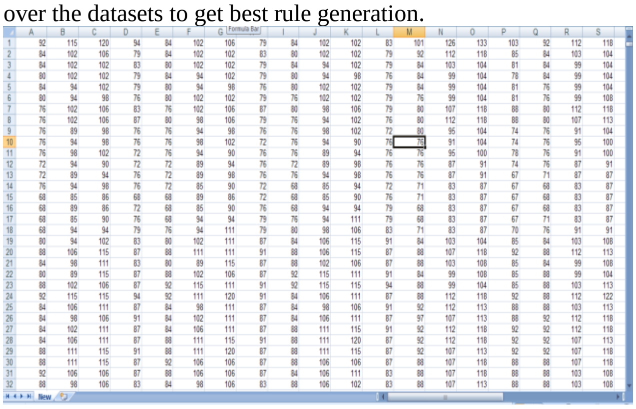

Fig.4. Sample Dataset

Performance of AEPS algorithms rule generated results shown in Table 1.

Table 1: Performance of Proposed AEPS Algorithm

Algorithm	AEPS
Best Fitness	0.881
B.F (iteration)	1
B.F (particle)	4

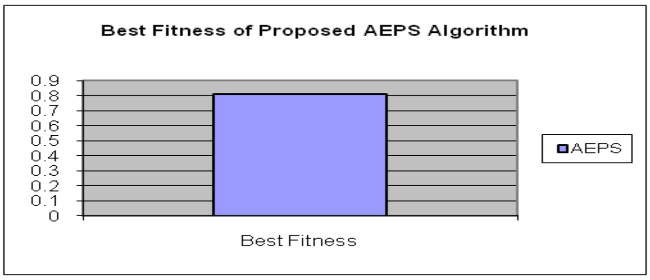

Fig 5: Best Fitness of Proposed AEPS Algorithm

Fig 5, shows the performance of proposed AEPS algorithm in terms of best fitness using spatial data and the best fitness obtained in the iteration , by the particle is shown in Fig 6.

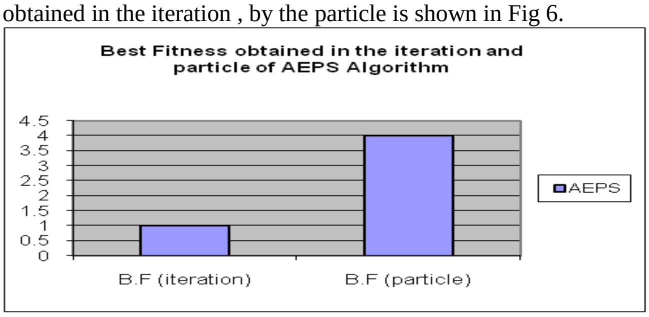

Fig 6: Best Fitness obtained in the iteration and particle of AEPS Algorithm

Comparitive performance in terms of auucuracy of existing and proposed AEPS algorithm shown in Fig 7.

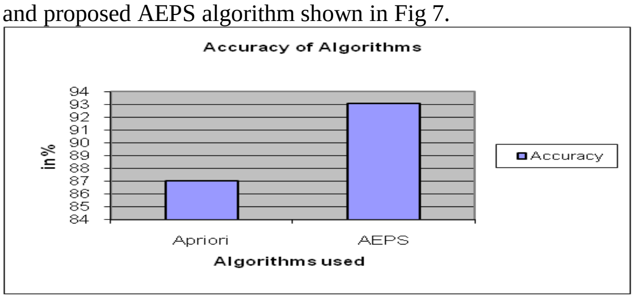

Fig 7: Comparative Accuracy of Algorithms

First set the population rate and while doing the generation part fix any stopping criteria rate shown in Fig 8.

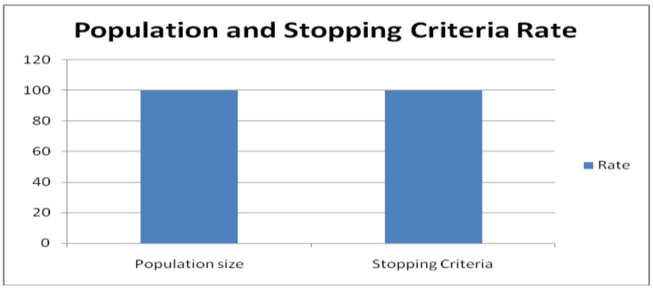

Fig 8. Rate of Population and Stopping Criteria

8. CONCLUSION

This paper proposed a methodology for the Multi label spatial classification optimized by the MOGA and the SAR using the Hybrid Evolutionary Algorithm and the semi supervised learning. The results for the proposed method is promising and also lay a opening for the identification of Multi label which can be further extended to the real world multi label classification, which consider all available classes that pass certain user threshold for each item set. The experimental results indicate that AEPS reached the minimal error rate faster than the other methods, and thus reduces computational cost. The work can be extended to the incremental learning of the training.

9. ACKNOWLEDGMENTS

The authors are grateful to the anonymous referees for their constructive and helpful comments.

10. REFERENCES

[1] Cheng-Hong Yang , Chih-Jen Hsiao and Li-Yeh Chuang , "Accelerated Chaotic Particle Swarm Optimization for Data Clustering", IPCSIT vol.3, 2011, ACSIT Press.

[2] Xueping Zhang; Yixun Liu; Jiayao Wang; Gaofeng Deng; Chuang Zhang, "Hybrid Particle Swarm Optimization with GA Mutation to Solve Spatial Clustering with Obstacles Constraints", Computational Intelligence and Design, 2008, 299 – 302.

[3] J.Arunadevi, V.Rajamani, "An Evolutionary Multi Label Classification using Associative Rule Mining for Spatial Preferences", IJCA Special Issue on "Artificial Intelligence Techniques - Novel Approaches & Practical Applications" AIT, 2011.

[4] Jiang Qing; Lin Hubin; Li Jiaoe; Liu Jing, "The Research on Spatial Data Mining Module Based on Multi-objective Optimization Model for Decision Support System", Intelligent Systems (GCIS), 2010 , 299 – 302.

[5] Imam Mukhlash, Benhard Sitohang, "Spatial Data Preprocessing for Mining Spatial Association Rule with Conventional Association Mining Algorithms", International Conference on Electrical Engineering and Informatics, Indonesia June 17-19, 2007.

[6] Qin Ding; Qiang Ding; Perrizo, W , "PARM—An Efficient Algorithm to Mine Association Rules From Spatial Data" IEEE Transactions on Systems, Man and Cybernetics, 18 November 2008.

[7] Diansheng Guo a, Jeremy Mennis, "Spatial data mining and geographic knowledge discovery—An introduction", Computers, Environment and Urban Systems 33 (2009) 403–408.

[8] Jiangping Chen; Yanan Chen; Jie Yu; Zhaohui Yang; "Comparisons with spatial autocorrelation and spatial association rule mining", Spatial Data Mining and Geographical Knowledge Services (ICSDM), 2011 IEEE International Conference on July 1 2011.

[9] Wei Ding Eick, C.F. Jing Wang Xiaojing Yuan, "A Framework for Regional Association Rule Mining in Spatial Datasets", Data Mining, 2006. ICDM '06. Sixth International Conference on 18-22 Dec 2006.

[10] Jiangping Chen, "An Algorithm About Association Rule Mining Based On Spatial Autocorrelation ",The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B6b. Beijing 2008.

[11] Diansheng Guo a, Jeremy Mennis, "Spatial data mining and geographic knowledge discovery—An introduction",

Computers, Environment and Urban Systems 33 (2009) 403–408.

[12] Xueping Zhang; Yixun Liu; Jiayao Wang; Gaofeng Deng; Chuang Zhang, "Hybrid Particle Swarm Optimization with GA Mutation to Solve Spatial Clustering with Obstacles Constraints", Computational Intelligence and Design, 2008, 299 – 302.

[13] Teng-Sheng Moh, Ameya Sabnis, "Applying Hybrid KEPSO Clustering to Web Pages", ACMSE '10, April 15-17, 2010, Oxford, USA.

[14] Ajith Abraham, Swagatam Das, Amit Konar, "Kernel Based Automatic Clustering Using Modified Particle Swarm Optimization Algorithm", GECCO'07, July 7–11, 2007, London.

[15] Marcelo N. Kapp, Robert Sabourin, Patrick Maupin, "A PSO-Based Framework for Dynamic SVM Model Selection", GECCO'09, July 8–12, 2009, Canada.