

An Evolutionary Multi Label Classification Using Associative Rule Mining

¹J. Arunadevi and ²V. Rajamani

¹Department of MCA, Thiagarajar School of Management, Madurai, Tamil Nadu, India

²Indra Ganesan College of Engineering, Tirchirapalli, Tamil Nadu, India

Abstract: Multi-label spatial classification based on association rules with Multi Objective Genetic Algorithms (MOGA) is proposed to deal with multiple class labels problem which is hard to settle by existing methods. In this study we adapt problem transformation for the Multi label classification. We use Hybrid evolutionary algorithm for the optimization in the generation of spatial association rules which addresses single label. MOGA is used to combine the single labels into multi labels with the conflicting objectives predictive accuracy and comprehensibility. Finally we built the classifier with a sorting mechanism. The algorithm is executed and the results are compared with Decision trees and Neural network based classifiers, the proposed method out performs the existing.

Key words: Multi label classification, associative classification, MOGA, HEA, single label, India

INTRODUCTION

The phenomenal growth of spatial data increases the importance of the spatial data mining which is used to mine fascinating and constructive but inherent knowledge. Classification of spatial stuff is a primary task in spatial data mining. Spatial classification is defined as the task of learning models to predict class labels based on the features of entities as well as the spatial relationships to other entities and their features by (Frank *et al.*, 2009; Koperski, 1999) said the goal of spatial classification is to learn the concept associated with each class on the basis of the interaction of two or more spatially-referenced objects or space-dependent attributes, according to a particular spacing or set of arrangements.

Rapid urbanization process amplifies the need for space planning purposes with emphasis on user preferences on the construction of urban infrastructure for housing, work and a variety of supporting activities of the population. Because of the large number of specific objectives which are needed to be considered in this planning, the application of spatial classification with multi objectives can have a significant impact on the quality, speed and cost of the planning. Classification is the important task which can be employed for the space planning with multi objectives.

Classification is used to predict the class attributes and the association rule is used to discover correlations among the attributes. We can use association rule mining to find the rules for classification and it is called as

Associative Classification (AC). Li *et al.* (2001), Liu *et al.* (1998), Thabtah *et al.* (2004a) and Yin and Han (2003) shows proof that AC algorithms outperforms the traditional classification. But there are few challenges in the associative classification to be addressed so that it can be widely used. Thabtah (2006) says incremental learning, noise in test data sets and the extraction of multi label rules are the major limitations to be concentrated.

Researchers propose a spatial associative classification algorithm optimized with the evolutionary algorithms for the classification for space planning to meet the user preferences. Researchers consider classification problem as the multi objective one since if it is considered as the single objective, Dehuri and Cho (2008) says the main demerit associated is that the generated rules are often more complex than necessary and not easy to comprehend. The reason behind is that the local, greedy search performed by traditional algorithms selects only one feature at a time and therefore, the feature space is approximated by a set of hypercubes. In real-world applications, the feature space is often very complex and a large set of such hypercubes might be needed to approximate the class boundaries among different classes.

Multi-label classification based on association rules with Multi Objective Genetic Algorithms (MOGA) is proposed to deal with multiple class labels problem which is hard to settle by existing methods, this algorithm decomposes multi-label data to mine single-label rules optimized by the Hybrid Evolutionary Algorithm (HEA) then combines labels with the same attributes to generate

multi-label rules with the help of MOGA. It extracts partial dataset features filtered by MOGA to build the initial classifier through assembling and conducts classification prediction by assembling the classifiers. Thus, the computational complexity caused by the high dimensional attributes decreases while the performance and efficiency increases.

Multi objective view on spatial classification: Mennis and Guo (2009) said classification is about grouping data items into classes (categories) according to their properties (attribute values). Ester *et al.* (1997), Koperski *et al.* (1998) said 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. Ester *et al.* (1997) proposed a neighborhood graph based extension of decision trees that considers both non-spatial attributes of the classified objects and relations with neighboring objects. However, the proposed method does not take into account hierarchical relations defined on spatial objects as well as non-spatial attributes. Malerba *et al.* (2003) proposed to exploit the expressive power of predicate logic to represent both spatial relations and background knowledge such as spatial hierarchies. Ghamrawi and McCallum (2005) said single-label classification assigns an object to exactly one class when there are two or more classes. Multi-label classification is the task of assigning an object simultaneously to one or multiple classes.

Chen *et al.* (2010) said Multi-label classification problem is an extension of traditional multi-class classification problem in which the classes are not mutually exclusive and each sample may belong to several classes simultaneously. Tsoumakas *et al.* (2009) said existing methods for multi-label classification fall into two main categories problem transformation and algorithm adaptation. Problem transformation maps the multi-label learning problem into one or more single-label problems. The most widely-used problem transformation method considers the prediction of each label as an independent binary classification task. Brinker *et al.* (2006) transform the multi label classification task into one or more single-label classification, regression or label ranking tasks. Algorithm adaptation extends specific learning algorithms in order to handle multi-label data directly, it modify standard single-label learning algorithm for multi-label classification. Methods proposed by both Zhang and Zhou (2007) MLKNN, Cheng and Hullermeier (2009) IBLR are considered algorithm adaptation state-of-the-art multi label classification algorithms that exploit

instance-based learning. Researchers introduce the problem of Multi label classification as the Multi objective classification with the help of genetic algorithms and the association rule mining.

Associative classification: AC is a branch of a larger area of scientific study known as data mining. Fayyad *et al.* (1998) define data mining as one of the main phases in knowledge discovery from databases which extracts useful patterns from data. AC integrates two known data mining tasks, association rule discovery and classification to build a model (classifier) for the purpose of prediction. Classification and association rule discovery are similar tasks in data mining with the exception that the main aim of classification is the prediction of class labels while association rule discovery describes correlations between items in a transactional database. Thabtah *et al.* (2004b) uses classification is a special case of association rule mining, in which the antecedent of the rule is the label attribute. Li *et al.* (2001) presented associative classification algorithm that selects and analyses the correlation between high confidence rules. Yin and Han (2003) proposed a greedy associative classification algorithm called CPAR.

Need for the evolutionary algorithms: Over the past decade, population-based Evolutionary Algorithms (EAs) 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. Multi-Objective Evolutionary Algorithms (MOEAs) are a popular approach to confronting these types of problem. A lot many research contributions by Dehuri *et al.* (2008a, b) as a problem solving tools of rule mining. The use of EAs as a tool of preference is due to such problems being typically complex with both a large number of parameters to be adjusted and several objectives to be optimized. EAs which can maintain a population of solutions are in addition able to explore several parts of the Pareto front simultaneously. The robustness and domain-independent capabilities of EAs attracts researchers to evolve a set of classification rules.

Genetic Algorithm (GA) based classifier systems usually fall into two basic categories, the Michigan approach and the Pittsburgh approach. The main difference between these two stems from the chromosome encoding schemes in the population of individuals. In the Michigan approach each individual with fixed length encodes a single prediction rule. In this approach there are at least two possibilities for discovering a set of rules. The first one is let each run of the GA discover a single rule (the best chromosome produced in all generations)

and simply run the GA multiple times to discover a set of rules. But the disadvantage of this strategy is that it is computationally expensive, requiring many GA runs. The second possibility is to design a more elaborate GA where a set of individuals-possibly the whole population corresponds to a set of rules. Whereas in the Pittsburgh approach, each individual is represented by a variable-length string and encodes a complete set of rules. Corcoran and Sen (1994) said the Pittsburgh approach is better suited for static domains and batch-mode learning in which all training samples are available before the learning process starts and the Michigan approach is more flexible to handle incremental-mode learning in which training samples arrive over time and dynamically changing domains.

ACO is a paradigm for designing meta heuristic algorithms for combinatorial optimization problems. The ACO algorithm was first introduced by Dorigo *et al.* (1991a, b) and the first Ant System (AS) was proposed by Dorigo (1992) in his Ph.D. thesis. The ACO is a meta-heuristic algorithm which utilizes the inspiration from real ant colonies behaviours to find a shortest path from a food source to the nest without using visual cues by exploiting pheromone information (Beckers *et al.*, 1992; Goss *et al.*, 1989; Holldobler and Wilson, 1990).

MATERIALS AND METHODS

In this study we propose a three stage multi label classifier based on the HEA, the MOGA and Association rule mining. The first stage generates the optimized spatial association rules by the use of the HEA. In the second stage the multi label rules are generated by the MOGA. Final stage the multi label classifier is built with a sorting mechanism applied to the rules generated.

Application of HEA for optimized spatial association rule mining: SAR uses apriori algorithm for the generation of the rules. Here the rules generated by apriori using the hybrid evolutionary algorithm. The MOGA is used to achieve the multi objective by with a Pareto based multiple-objective genetic algorithm. The possible rules are represented as chromosomes and a suitable encoding/decoding scheme has been defined, it also provides the diversity of associations among the rules generated by elitism. We follow the Michigan approach for the optimization. To increase the efficiency of the MOGA, we are using the ACO which limits the algorithm from falling to the local optimal solution.

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. Then the rules are clustered using the rule cover based on the consequent information.

String representation: Chromosomes are encoded as real numbers the number of genes in each chromosome is equal to the number of item sets considered. Each gene will have 4 digits for vector index. A sample chromosome may look like as follows:

0001 0102 0204 0302 0401 0500 0601
0702 0802 0901 1002 1101 1201

Here, the first two numbers in each gene represents the attribute and the next two denotes the value, fourth gene has the value 0302 where 03 refers to the age group and 02 refers to the third age group ranges from 23-25. Like wise all the gene has been encoded, once the initial population is generated now we are ready to apply genetic operators.

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

$$f(x) = \frac{(w1 * \text{Comprehensibility}) + (w2 * \text{J-Measure}) + (w3 * \text{Confidence})}{w1 + w2 + w3}$$

Where, w1, w2, w3 are used defined weights.

Reproduction (selection): The selection process selects chromosomes from the mating pool directed by the survival of the fittest concept of natural genetic systems. In the proportional selection strategy adopted in this study, a chromosome is assigned a number of copies which is proportional to its fitness in the population, go into the mating pool for further genetic operations. Roulette wheel selection is used for the proportional selection strategy.

Crossover: Crossover is a probabilistic process that exchanges information between two parent chromosomes for generating two child chromosomes. In this study, single point crossover with a fixed crossover probability of C is used. For chromosomes of length l, a random integer called the crossover point is generated in the range (1, l-1). The portions of the chromosomes lying to the right of the crossover point are exchanged to produce two offspring.

Mutation: Each chromosome undergoes mutation with a fixed probability M. For binary representation of chromosomes, a bit position (or gene) is mutated by

simply flipping its value. Since we are considering real numbers in this study, a random position is chosen in the chromosome and replace by a random number between 0-9.

Pseudo code 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.
    
```

Generation of multi-label rules by moga: For a given training dataset D, traditional associative classification algorithms produce only one single-label rules set and form a default label for the remaining unclassified instances. The proposed algorithm uses MOGA to generate the multi label rules from the rules obtained in the previous stage. The objectives we have under consideration is high predictive accuracy and high comprehensibility which are conflicting objectives. The rules are of the form $A_1 \wedge A_2 \dots A_n \rightarrow C$. The antecedent part of the rule is a conjunction of conditions say A (conjunction of $A_1, A_2 \dots A_n$). Predictive Accuracy is defined by Dehuri and Mall (2006) as:

$$PA = (|A \text{ and } C| - \frac{1}{2}) / |A|$$

Where:

- $|A|$ = The number of examples satisfying all the conditions in the antecedent A
- $|A \text{ and } C|$ = The number of examples that satisfy both the antecedent A and the consequent C

Intuitively, this metric measures Predictive Accuracy in terms of how many cases both antecedent and consequent hold out of all cases where the antecedent holds. The term 1/2 is subtracted to penalize the rules covering few training examples

Shi and Lei (2008) proposed the standard way of measuring comprehensibility (Brinker *et al.*, 2006) is to count the number of condition in the rule. If a rule has at most L condition, the comprehensibility of the rule (or individual) p can be defined as:

$$C_p = (L-n)/(L-1)$$

where, n is the length of the rule (or individual) p.

The fitness function is computed as the arithmetic weighted mean of Comprehensibility and Predictive Accuracy. The fitness function is given by:

$$Fitness = (W1*PA+W2*C_p)/(W1+W2)$$

Where W1 is the weight defined by the user. Repeated learning has been done using the different generations of the MOGA and until no more frequent itemsets can be discovered. At this stage, any remaining unclassified instance form a default label. This process results in learning from several subsets of the original training data and generates few rules sets.

Multi label classifier: When the learning process is finished and no further frequent itemsets are found, a merging of the rules sets produced from each training data is performed to obtain a multi-label classifier. When merging the rules sets, the multilabel rules are prioritized based on the sorting procedure. The sorting procedure uses the support, confidence and J measure of the rules and by taking into account the fact that the highest priority rules are those that have been derived from the original dataset during the first iteration then those generated in the second iteration and so on. The sorting has been done by the weighted average of the above three measures. The weights of the measures are defined by the user. The Sorting Measure (SM) is defined as:

$$SM = ((w1 * Support)+(w2* J-Measure)+(w3 * Confidence))/(w1 +w2+w3)$$

Comparison metrics: Multi-label evaluation metrics fall into two main categories: prediction-based and ranking based. Prediction-based metrics evaluate how well the algorithm predicts the actual set of correct labels for each instance. Ranking-based metrics evaluate how well the algorithm ranks the labels relative to one another. Tsoumakas *et al.* (2009) use the following standard multi-label prediction-based evaluation metrics.

Hamming loss is the percentage of correct labels not predicted and incorrect labels predicted. Accuracy is the percentage of true positives out of the total true positives, false positives and false negatives. Precision is the percentage of predicted labels that were correct. Recall is the percentage of correct labels that were predicted.

RESULTS AND DISCUSSION

The environmental setup are Population size = 100, Mutation rate (M) = 0.5, Crossover rate (C) = 0.8,

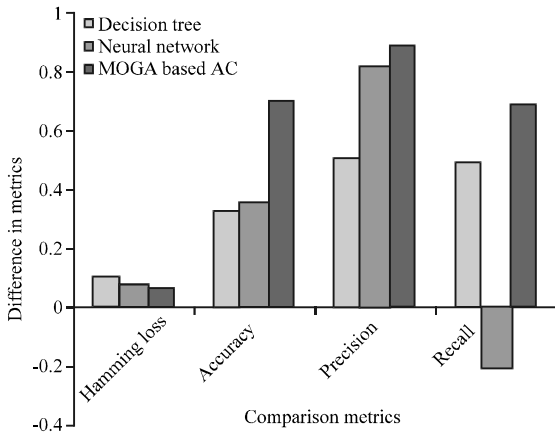


Fig.1: Comparison of the algorithms

Table 1: Mean run time of the classifiers

Mean run time	Decision tree	Neural Network	MOGA based AC
Time (min)	9.02	15	8.4

Elitism = 0.5. The stopping criterion used is the non evolution of the archive during 100 generations, once the minimal number of generations has been over passed.

We have used the synthesized dataset for the research. The general procedure of data mining is:

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

The specific procedure is as following:

- As done by Zheng and Zhao (2008) we 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:
- Run algorithm to generate the single labelrules
- Run algorithm to generate the Multi label rules
- Run algorithm for building the classifier (Fig. 1)

We compute the results using ten-fold cross-validation for each method over each data set. It is defined as break data into 10 sets of size $n/10$, train on 9 datasets and test on 1. Repeat 10 times and compute mean. The result for the mean run time of the classifiers are shown in Table 1.

In comparing the times for the execution of the classifiers MOGA based AC outperforms the other two methods. This is due to the fact that the reduction in the number of rules in the first step.

CONCLUSION

This study proposed a methodology for the Multi label spatial classification optimized by the MOGA and the SAR using the Hybrid Evolutionary algorithm. 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 research can be extended to the incremental learning of the training.

REFERENCES

- Beckers, R., J.L. Deneubourg and S. Goss, 1992. Trails and U-turns in the selection of a path by the ant *Lasius niger*. *J. Theor. Biol.*, 159: 397-415.
- Brinker, K., J. Furnkranz and E. Hullermeier, 2006. A unified model for multilabel classification and ranking. *Proceedings of the 17th European Conference on Artificial Intelligence*, Aug. 29-Sept. 1, Italy, pp: 489-493.
- Chen, B., L. Ma and J. Hu, 2010. An improved multi-label classification method based on svm with delicate decision boundary. *Int. J. Innovative Comp. Inform. Control*, 6: 1605-1614.
- Cheng, W. and E. Hullermeier, 2009. Combining instance-based learning and logistic regression for multilabel classification. *Mach. Learn.*, 76: 211-225.
- Corcoran, L. and S. Sen, 1994. Using real-valued genetic algorithms to evolve rule sets for classification. *Proceedings of the 1st IEEE Conference on Evolutionary Computation*, June 1994, Orlando, FL, USA., pp: 120-124.
- Dehuri, S. and R. Mall, 2006. Predictive and comprehensible rule discovery using a multi-objective genetic algorithm. *Knowledge Based Syst.*, 19: 413-421.
- Dehuri, S. and S.B. Cho, 2008. Multi-objective classification rule mining using gene expression programming. *Proceedings of the 3rd International Conference on Convergence and Hybrid Information Technology*, Nov. 11-13, Busan, pp: 754-760.
- Dehuri, S., S. Ghosh and A. Ghosh, 2008a. Genetic Algorithms for Optimization of Multiple Objectives in Knowledge Discovery from Large Databases. In: *Multi-Objective Evolutionary Algorithms for Knowledge Discovery from Databases*, Ghosh, A. (Eds.). Springer, Heidelberg, pp: 1-22.

- Dehuri, S., S. Patnaik, A. Ghosh and R. Mall, 2008b. Application of elitist multi-objective genetic algorithm for classification rule generation. *Applied Soft Comput.*, 8: 477-487.
- Dorigo, M., 1992. Optimization, learning and natural algorithms. Ph.D Thesis, Politecnico de Milano, Italy.
- Dorigo, M., V. Maniezzo and A. Colormi, 1991a. Positive feedback as a search strategy. Technical Report No. 91-016, Politecnico di Milano, Italy. <http://www.citeulike.org/user/Bc91/article/1372086>.
- Dorigo, M., V. Maniezzo and A. Colormi, 1991b. The ant system: An autocatalytic optimizing process. Technical Report TR91-016, Politecnico di Milano.
- Ester, M., H.P. Kriegel and J. Sander, 1997. Spatial data mining: A database approach. Proceedings of the 5th International Symposium on Advances in Spatial Databases, (ASD'97), Springer-Verlag, Berlin, pp: 47-66.
- Fayyad, U., G. Piatesky-Shapiro, G. Smith and R. Uthurusamy, 1998. *Advances in Knowledge Discovery and Data Mining*. AAAI Press, Menlo Park, California.
- Frank, R., M. Ester and A. Knobbe, 2009. A multi-relational approach to spatial classification. Proceedings of the 15th ACM SIGKDD international conference on Knowledge Discovery and Data Mining, June 28-July 01, Paris, France, pp: 309-318.
- Ghamrawi, N. and A. McCallum, 2005. Collective multi-label classification. Proceedings of the 14th ACM International Conference on Information and Knowledge management, Oct. 31-Nov. 5, Bremen, Germany, pp: 195-200.
- Goss, S., S. Aron, J.L. Deneubourg and J.M. Pasteels, 1989. Self-organized shortcuts in the argentine ant. *Naturwissenschaften*, 76: 579-581.
- Holldobler, B. and E.O. Wilson, 1990. *The Ants*. 1st Edn., Springer-Verlag, Berlin, Heidelberg, ISBN: 978-3-540-52092-4, pp: 10-55.
- Koperski, K., 1999. Progressive refinement approach to spatial data mining. Ph.D. Thesis, Computing Science, Simon Fraser University.
- Koperski, K., J. Han and N. Stefanovic, 1998. An efficient two-step method for classification of spatial data. Proceedings of the International Symposium on Spatial Data Handling (SDH'98), Vancouver, BC Canada, pp: 45-54.
- Li, W., J. Han and J. Pei, 2001. CMAR: Accurate and efficient classification based on multiple class-association rules. Proceedings of the 2001 IEEE International Conference on Data Mining, Nov. 29-Dec. 2, San Jose, CA., pp: 369-376.
- Liu, B., W. Hsu and Y. Ma, 1998. Integrating classification and association rule mining. Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining, Aug. 27-31, AAAI Press, New York, pp: 80-86.
- Malerba, D., F. Esposito, A. Lanza, F.A. Lisi and A. Appice, 2003. Empowering a GIS with inductive learning capabilities: The case of INGENS. *Comp. Environ. Urban Syst.*, 27: 265-281.
- Mennis, J. and D. Guo, 2009. Spatial data mining and geographic knowledge discovery: An introduction. *Comp. Environ. Urban Syst.*, 33: 403-408.
- Shi, X.J. and H. Lei, 2008. A genetic algorithm-based approach for classification rule discovery. Proceedings of the International Conference on Information Management, Innovation Management and Industrial Engineering, Dec. 19-21, Taipei, pp: 175-178.
- Thabtah, F., 2006. Challenges and interesting research directions in associative classification. Proceedings of the 6th IEEE International Conference on Data Mining Workshops, Dec. 18-22, Hong Kong, pp: 785-792.
- Thabtah, F., P. Cowling and Y. Peng, 2004a. MMAC: A new multi-class, multi-label associative classification approach. Proceedings of the 4th IEEE International Conference on Data Mining, (ICDM'04), Brighton, UK., pp: 217-224.
- Thabtah, F., P. Cowling and Y. Peng, 2004b. Multi-label classification learning. Proceedings of the IEEE International Conference on Advances in Intelligent Systems, November 2004, Luxembourg, Luxembourg, pp: 207-213.
- Tsoumakas, G., I. Katakis and I. Vlahavas, 2009. Mining Multi-Label Data. In: *Data Mining and Knowledge Discovery Handbook*, Maimon, O. and L. Rokach (Eds.). 2nd Edn., Springer, Heidelberg.
- Yin, X. and J. Han, 2003. CPAR: Classification based on predictive association rules. Proceeding of the International Conference on Data Mining, (ICDM'03), San Francisco, CA., pp: 331-335.
- Zhang, M.L. and Z.H. Zhou, 2007. ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recognit.*, 40: 2038-2048.
- Zheng, X. and L. Zhao, 2008. Association rule analysis of spatial data mining based on matlab-a case of ancheng township in China. Proceedings of the 1st International Workshop on Knowledge Discovery and Data Mining, Jan. 23-24, Adelaide, South Australia, pp: 76-80.