Multi Label Spatial Semi Supervised Classification using Spatial Associative Rule Mining and Evolutionary Algorithms

J.Arunadevi[1] and Dr.V.Rajamani[2]

[1]Assistant Professor, Thiagarajar School of Management, Madurai, India
[2]Indra Ganesan College of Engineering, Tirchirapalli, India

Abstract

Multi-label spatial classification based on association rules with multi objective genetic algorithms (MOGA) enriched by semi supervised learning is proposed in this paper. It is to deal with multiple class labels problem. In this paper 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. Semi supervised learning is done through the process of rule cover clustering. Finally associative classifier is built with a sorting mechanism. The algorithm is simulated and the results are compared with MOGA based associative classifier, which out performs the existing.

Keywords:

Multi label Classification, Associative Classification, MOGA, HEA, Rule Cover cluster

1. Introduction

Land is very restricted resource subsequently, it is important to recognize it's prospective and optimize its use. Cohon and Jared L [1] said due to the complex needs and a large number of criteria such as environmental, economic, sociological and natural factors, decision-makers need to use techniques of multi-objective planning and multi-criteria analysis in many social activities related to land, especially in the field of planning of spatial organization. The congregation of satellite metaphors from in-flight photographs has motivated the craving of the technical population to use this massive size data for studies.

Chakhar, S. and Mousseau, V [2] discussed spatial decision problems as those problems in which the decision implies the selection among several potential actions or alternatives that are

David Bracewell, et al. (Eds): AIAA 2011,CS & IT 03, pp. 197–210 , 2011.
© CS & IT-CSCP 2011
DOI : 10.5121/csit.2011.1318

associated with some specific locations in space. One of the many aspects of this problem is cost effective and efficient accessing of a set of services or infrastructural facilities by a group of demand points or clients. Because of the large number of specific objectives, which are needed to be considered in this decision making, 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.

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 Richard Frank et. al [3]. Koperski, K [4] 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.

Stuart Ness [5] said traditional methods for classification have been based on finding accurate results to the domain classification and have not been focused on speed, efficiency, or scalability. He also said the drawbacks of supervised and unsupervised learning. Supervised Classification requires significant time from a domain scientist, it is has significant limitations in dealing with large data sets such as spatial data. For unsupervised classification the downside is that with large data sets, computation time is lengthy and results in an inefficient means of creating a classifier.

Semi supervised classification also has the open problems which are unique to the spatial realm, such as dealing with neighborhood techniques that consider near by neighbors to be considered. Open problems within the traditional realm, such as dealing with random sampling of data, as well as finding the global maximum are to be concentrated. Another area of interest with the increasing need for continual processing is to find more efficient ways of creating the classifiers in order to improve speed.

We propose a semi supervised classifier which uses association and clustering optimized with the evolutionary algorithms. This is proposed to deal with the problem which is hard to settle by existing methods. This algorithm is used to mine single-label rules optimized by the Hybrid Evolutionary Algorithm (HEA), and then combines the single labels to disjoint sets by rule cover clustering optimized by GA. Multi label classification enabled by MOGA is applied to each cluster and produces the Multi label Classifier. Given a new test occurrence, the algorithm first finds the nearest cluster and then uses the respective cluster to classify it. Thus, the computational complexity caused by the high dimensional attributes decreases while the performance and efficiency increases. This approach also addresses the problem of near by neighbors a unique problem of spatial area by association, random sampling by MOGA , global maximum by ACO and construct an efficient way to increase the speed of classification through the filtering process.

The paper is organized as follows: Section 2 deals with the background concepts of semi supervised spatial classification, SAR, Rule cover clustering, MOGA and the ACO applied for the optimization of the rule generation. Section 3 deals with approach followed in this paper, Section 4 explains the comparison metrics. Section 5 discusses the results obtained and Section 7 discusses the application of the paper under the area of consideration and Section 8 gives conclusion of the paper

2. Background Study

This section is divided into three parts. Section 2.1 discusses the concepts of semi supervised spatial classification, section 2.2 discusses the use of associative classification, section 2.3 says about the need for the evolutionary algorithms for the semi supervised spatial classification and section 2.4 gives the review on clustering for classification with association rules.

2.1 Semi supervised spatial classification

Zhu, Xiaojin [6] said semi supervised classification is a special form of classification. It was derived from the use of non-labeled samples to assist with the supervised learning method. The main goal of this method is to use both labeled and unlabeled data to build better classifiers. Because semi-supervised learning requires less human effort and gives higher accuracy, it is of great interest both in theory and in practice. Decision directed methods for Semi supervised classification have been referred under various names by different communities in [7-12]. Expectation-maximization (EM) is a well known class of iterative algorithms for maximum-likelihood or maximum a posteriori estimation in problems with incomplete data [13] [14]. A co-training approach to semi-supervised classification was proposed by Blum and Mitchell [15]. Several authors reported experimental results which show the effectiveness of co training [16-18].

Foli [19] reported that the fundamental issue about the conditions under which, and the extent at which, the use of unlabeled data with co-training can increase classification accuracy is basically unsolved . These traditional algorithms can be used for the spatial domain with an additional constraint of dealing with neighborhood techniques that consider near by neighbors to be considered. The other open issues to be addressed are dealing with random sampling of data, as well as finding the global maximum and to find more efficient ways of creating the classifiers in order to improve speed, where in many algorithms concentration is on accuracy.

2.2 Associative Classification

Associative Classification (AC) is a branch of a larger area of scientific study known as data mining. Fayyad et al.[20] defined 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 F. et al [21] uses classification is a special case of association rule mining, in which the antecedent of the rule is the label attribute. W. Li, J. Han, and J. Pei [22] presented associative classification algorithm that selects and analyses the correlation between high confidence rules. Yin, X. and Han, J [23] presented a greedy associative classification algorithm called Classification based on Predictive Association Rules (CPAR).

2.3 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. exist in [24] and [25] 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. 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. A. L. Corcoran and S. Sen in [26] 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 Colorni, Dorigo and Maniezzo [27]-[28] and the first Ant System (AS) was developed by Dorigo [29] 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 [30]-[32].

2.4 Clustering for Classification

Zeng et.al [33] said clustering followed by classification, can be viewed as a conceptual approach. Another strategy of combining clustering and classification is through iterative reinforcement. The clustering-based multi-label classification (CBMLC) framework was discussed by Nasierding et. al in [34]. Kyriakopoulou and Kalamboukis said [35], an ideal situation would be for the classifier to have information about the distribution of the testing examples before it classifies them. In this paper we propose clustering of the association rules to address the correlation between the neighborhoods in spatial data and then followed by classification, which turns out to a semi supervised classifier.

3 Proposed Methodology

In this paper we propose a multi label classifier based on the HEA, 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. Final stage 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. 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.

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

$$f(x) = \frac{[\,(w_1 * \text{Comprehensibility}) + (w_2 * \text{J-Measure}) + (w_3 * \text{Confidence})\,]}{[\ w_1+w_2+w_3\ ]}$$

where w_1, w_2, and w_3 are used defined weights.

Pseudo code for clustering the rules generated

Input : set of rules generated by the HEA $\mathbf{R}_y$={ X_i -> 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 $\mathbf{R}_y$ 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. $\mathbf{r}_y = \mathbf{r}_y \cup \mathbf{r}$
5. Delete the highest rule cover from the cluster cover
6. End While

Output : the representative rule set.

Apply GA 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.

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. Our proposed algorithm uses MOGA to generate the multi label rules from the rules obtained in the previous stage for each cluster. The objectives we have under consideration is high predictive accuracy and high comprehensibility, which are conflicting objectives.

The rules are of the form A1^ A2…An -> C. The antecedent part of the rule is a conjunction of conditions say A (conjunction of A1, A2…An). Predictive Accuracy is defined by Dehuri, S., Mall, R.[36] as

$$PA = (|A\&C| - ½)/ |A| \qquad (1)$$

where |A| is the number of examples satisfying all the conditions in the antecedent A and |A&C| is 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 ½ is subtracted to penalize the rules covering few training examples.

Xian-Jun Shi and Hong Lei [37] discussed the standard way of measuring comprehensibility 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

$$Cp = (L-n)/ (L-1). \qquad (2)$$

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

$$\text{Fitness} = [\, (y_1 * PA + y_2 * Cp) / (y_1 + y_2) \,]$$

where y_1, y_2 are the weights defined by the user.

Repeated learning has been done using the different generations of the MOGA and until no more frequent item sets 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 item sets 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 multi label rules are prioritized based on the sorting procedure. The sorting procedure uses the support, confidence and J measure of the rules. We also take 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

$$SM = \frac{((z_1 * \text{Support}) + (z_2 * \text{J-Measure}) + (z_3 * \text{Confidence}))}{[(z_1 + z_2 + z_3)]}$$

Where, z_1,z_2 and z_3 are used defined weights.

4. 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 [38] 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.

5. RESULTS AND DISCUSSIONS

We have used the synthesized dataset, which has been collected in and around the city Madurai. This data has been collected based on the geographic, demographic, psychographic and behavioralistic of the customer characteristics for our research. The target group are students, professionals, working women, home makers and the senior citizens.

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.

We have followed the procedure adopted by by Xinqi Zheng and Lu Zhao [39], where we take advantage of import wizard in Matlab to accomplish the import of data file.

Table 1 : Environmental parameters for GA

Population size	100
Crossover rate (C)	0.8
Mutation rate (M)	0.1
Stopping criteria	100 generations

We compute our 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.

Table 2: Comparison of the algorithms based on Hamming Loss

	MOGA based AC	**Semi supervised MOGA based AC**
Iteration 1	0.068	0.062
Iteration 2	0.069	0.061
Iteration 3	0.064	0.06
Iteration 4	0.066	0.062
Iteration 5	0.065	0.062
Iteration 6	0.066	0.063
Iteration 7	0.066	0.062
Iteration 8	0.067	0.062
Iteration 9	0.067	0.062
Iteration 10	0.068	0.063
Mean	0.0666	0.0619

The results in the Table 2 indicates that Semi supervised MOGA based AC outperforms MOGA based AC. The hamming loss in our proposed approach is reduced around 7% compared to its counterpart.

Percentage of correct labels not predicted and incorrect labels predicted has been reduced by the proposed approach. This is because of the associative classifier learned through the optimized semi supervised algorithm. The clustering process splits the original training data into smaller parts of similar labels. So the percentage of non prediction and incorrect labels has been reduced since only small volume is tested compared to the full dataset.

Table 3: Comparison of the algorithms based on Accuracy

	MOGA based AC	**Semi supervised MOGA based AC**
Iteration 1	0.698	0.723
Iteration 2	0.694	0.724
Iteration 3	0.694	0.729
Iteration 4	0.699	0.731
Iteration 5	0.692	0.722
Iteration 6	0.692	0.722
Iteration 7	0.694	0.724
Iteration 8	0.695	0.72
Iteration 9	0.695	0.722
Iteration 10	0.695	0.722
Mean	0.6948	0.7239

The results in the Table 3 indicates that Semi supervised MOGA based AC outperforms MOGA based AC. The accuracy in the proposed approach is increased around 4.01%. This is due to the fact that the semi supervised learning provides the environment for the improved accuracy by searching in the limited and related labels.

Table 4: Comparison of the algorithms based on Precision

	MOGA based AC	**Semi supervised MOGA based AC**
Iteration 1	0.889	0.901
Iteration 2	0.885	0.897
Iteration 3	0.885	0.903
Iteration 4	0.886	0.902
Iteration 5	0.888	0.903
Iteration 6	0.885	0.904
Iteration 7	0.886	0.904
Iteration 8	0.885	0.899
Iteration 9	0.887	0.899
Iteration 10	0.887	0.897
Mean	0.8863	0.9009

The results in the Table 4 indicates that Semi supervised MOGA based AC outperforms MOGA based AC. The precision in the proposed approach is increased around 1.62% compared to the non supervised associative classifier

Table 5: Comparison of the algorithms based on recall

	MOGA based AC	Semi supervised MOGA based AC
Iteration 1	0.684	0.693
Iteration 2	0.685	0.695
Iteration 3	0.683	0.69
Iteration 4	0.683	0.694
Iteration 5	0.683	0.694
Iteration 6	0.686	0.695
Iteration 7	0.686	0.695
Iteration 8	0.683	0.696
Iteration 9	0.689	0.695
Iteration 10	0.689	0.698
Mean	0.6851	0.6945

The results in the table 5 indicates that Semi supervised MOGA based AC outperforms MOGA based AC. The recall in the proposed approach is increased around 1.3% compared to the non supervised associative classifier.

Percentage of correct labels that were predicted has been increased by the proposed approach. This is because of the associative classifier undergone a semi supervised approach optimized by the genetic algorithm with the objectives as the predictive accuracy and comprehensibility. The first phase of the association rule generation is also optimized so that the prediction percentage of the correct labels is increased. The number of labels is limited, to search so the percentage of correct labels predicted is increased.

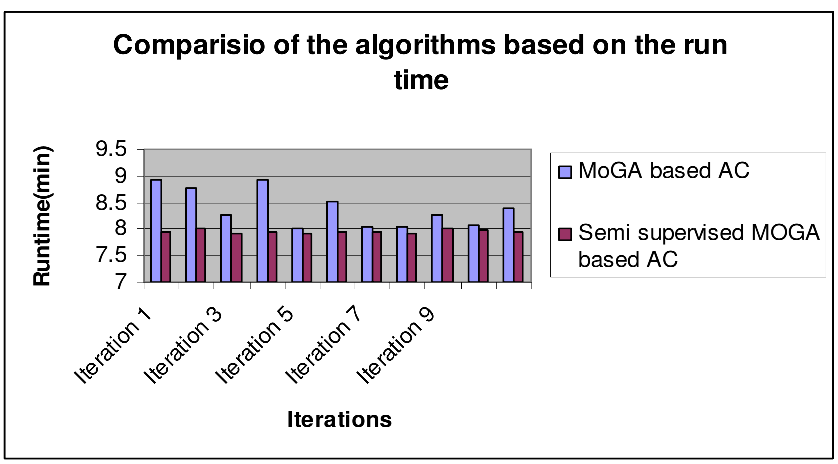

Fig 1: Comparison of the proposed approach with other benchmark algorithm based on the run time

Table 6: Comparison of the algorithms based on mean run time

	MOGA based AC	Semi supervised MOGA based AC	Improvement in the execution time
Time (min)	8.4	7.957	5.1 %

The result for the mean run time of the classifier for an instance is given in Table 6. In comparing the times for the execution of the classifiers Semi supervised MOGA based AC outperforms the other. This is due to the fact that the reduction in the number of instances in the relative label to be searched. Due to the knowledge from the clustering phase the classifier works faster.

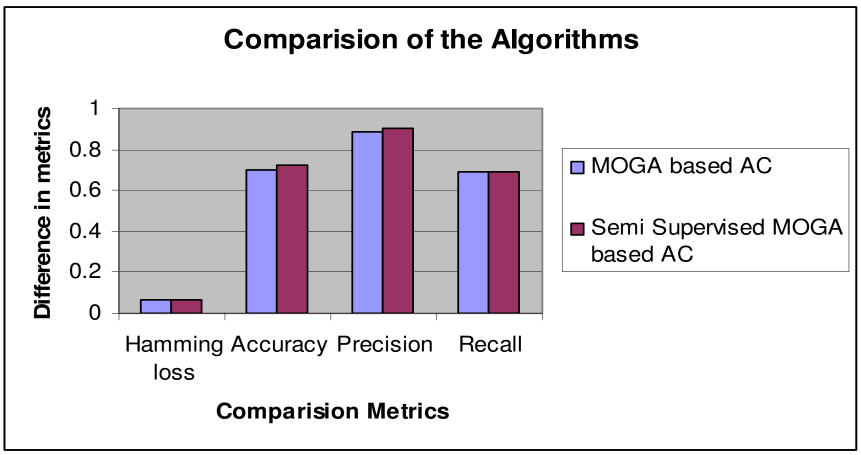

Fig 2: Comparison of the proposed approach with other benchmark algorithm based on the four measures

The consolidated report has been depicted in the Fig 2, it shows that the proposed approach shows remarkable improvement over the other algorithm for multi label classification based on the four metrics considered.

The improved performance exhibited by the proposed approach is due to the fact that, we have used

(1) Semi supervised learning helps us to minimize the number of labels to be searched.
(2) Problem of random sampling of the data is encountered by using GA for rearranging the rules in various orders based on the fitness preferred by the user.
(3) This also helps the clustering for not suffering from the order of the input.
(4) GA is applied for retaining nearest neighbors in common cluster, so that the searching labels in the relative cluster induces increase the speed of classification.

The collective combination of the HEA and MOGA in the various process of Multi label prediction is very effective. The effectiveness of semi supervised learning is demonstrated using the tables and graphs generated from the results obtained.

6. 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 work can be extended to the incremental learning of the training.

Reference:

[1] Cohon, Jared L.: Multiobjective Programming and Planning, Dover Publications Inc., Mineola, New York 2004

[2] Chakhar, S., and Mousseau, V. Towards a typology of spatial decision problems, Annales du LAMSADE**, No. 2,** 2004, pp 125-154.

[3] Richard Frank , Martin Ester , Arno Knobbe, A multi-relational approach to spatial classification, Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, Paris, France ,June 28-July 01, 2009,

[4] Koperski, K.: Progressive Refinement Approach to Spatial Data Mining,Ph.D. thesis,Computing Science, Simon Fraser University, 1999.

[5] Stuart Ness, A Brief Survey: Spatial Semi-Supervised Image Classification**,** http://www users.cs.umn.edu/~sness/csci8701/hw/hw6/Revised-Survey_Spatial-Semi-Supervised_Learning.pdf

[6] Zhu, Xiaojin. Semi-Supervised Learning Literature Survey. Computer Sciences, Univeristy of Wisconsin Madison. Technical Report 1530, University of Wisconsin – Madison (http://www.cs.wisc.edu/~jerryzhu/pub/ssl_survey.pdf.)

[7] Seeger M., Learning with labeled and unlabeled data, Technical Report, University of Edinburgh, Institute for Adaptive and Neural Computation, Dec. 2002, pp. 1-62.

[8] Kemp T., Waibel A., Unsupervised Training of a Speech Recognizer: Recent Experiments, Proc. Eurospeech, vol. 6, 1999, pp. 2725- 2728.

[9] Nagy G., Shelton G.L., Self-corrective character recognition systems, IEEE Trans. On Information Theory, IT-12, no. 2, April 1966, pp. 215-222.

[10] G. Nagy, Classifiers that improve with use, Proc. Conference on Pattern Recognition and Multimedia, IEICE, Tokyo, February 2004, pp. 79-86.

[11] Inoue M., Ueda N., Exploitation of unlabeled sequences in hidden markov models, IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 25, No. 12, Dec. 2003, pp. 1570-1581.

[12] Young T.Y., Farjo A., On decision directed estimation and stochastic approximation, IEEE Trans. On Information Theory, Sep -1972, pp. 671-673.

[13] Dempster A.P., Laird N.M., Rubin D.B., Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society, Series B, 39(1), 1977, pp. 1-38.

[14] Nigam K., McCallum A.K., Thrun S., Mitchell T., Text classification from labeled and unlabeled documents using EM, Machine Learning, 39, 2000, pp. 103-134.

[15] Blum A., Mitchell T., Combining labeled and unlabeled data with co-training, Proc. of the Workshop on Computational Learning Theory, 1998, pp. 92-100.

[16] Seeger M., Learning with labeled and unlabeled data, Technical Report, University of Edinburgh, Institute for Adaptive and Neural Computation, Dec. 2002, pp. 1-62.

[17] Nigam K, Ghani R., Analyzing the effectiveness and applicability of co-training, Proc. 5th Int. Conf. On Information and Knowledge Management, 2000, pp. 86-93.

[18] Goldman S., Zhou Y., Enhancing supervised learning with unlabeled data, Proc. Of the 7th Int. Con. On Machine Learning, ICML 2000, 2000, pp. 327-334.

[19] F. Roli, Semi-Supervised Multiple Classifier Systems: Background and Research Directions, *Multiple Classifier Systems*, Springer Verlag, Vol. 3541/2005, 2005 pp 674-676

[20] Fayyad, U., Piatetsky-Shapiro, G., Smith, G. & Uthurusamy, R. Advances in Knowledge Discovery and Data Mining. Menlo Park, CA: AAAI Press, 1998

[21] Thabtah F., Cowling P., and Peng Y. (2004): Multi-label Classification Learning, Proceedings of the IEEE 2004 International Conference on Advances in Intelligent Systems (AISTA ’04). Luxembourg, Luxembourg, Nov. 2004, pp. 207-213.

[22] W. Li, J. Han, and J. Pei. CMAR: Accurate and efficient classification based on multiple-class association rule. In Proceeding of the First IEEE International Conference on Data Mining (ICDM’01), San Jose, CA, Nov. 2001, pp. 369- 376.

[23] Yin, X. and Han, J.: CPAR: Classification Based on Predictive Association Rules. Proc SIAM Int Conf on Data Mining (SDM'03), 2003,331-335.

[24] S. Dehuri et al, “Application of Elitist Multi-objective Genetic Algorithms for Classification Rule Generations,” Applied Soft Computing, vol. 8, no. 1, 2008,pp.477-487.

[25] S. Dehuri, S. Ghosh, and A. Ghosh, “Genetic Algorithms for Optimization of Multiple Objectives in Knowledge Discovery from large Databases,” A. Ghosh (Eds.), Multi-objective Evolutionary Algorithms for Knowledge Discovery from Databases, 2008, pp. 1-22.

[26] A. L. Corcoran and S. Sen, “Using Real-valued Genetic Algorithms to Evolve Rule Sets for Classification,” Proc. 1st IEEE Conf. Evolutionary Computation, June 1994,pp. 120–124.

[27] A . Colorni, M. Dorigo, and V. Maniezzo. Positive feedback as a search strategy. Technical Report No. 91-016, Politecnico di Milano, Italy,1991.

[28] A. Colorni, M. Dorigo, and V. Maniezzo . The ant system: an autocatatlytic process. Technical Report No. 91- 016, Politecnico di Milano, Italy, 1991.

[29] M. Dorigo. Optimization, Learning and Natural Algorithms. Ph.D. Thesis, Politecnico di Milano, Italy,1992

[30] 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.

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

[32] B. Holldobler and E.O. Wilson. The Ants. Springer-Verlag: Berlin, 1990.

[33] Zeng, H.J., Wang, X.H., Chen, Z., Lu, H., Ma, W.Y.: CBC: Clustering based text classification requiring minimal labeled data. Proceedings of the 3rd IEEE International Conference on Data Mining, Melbourne, Florida, USA 2003

[34] Nasierding, G., Tsoumakas, G., Kouzani, A.Z.: Clustering Based Multi-Label Classification for Image Annotation and Retrieval. In: IEEE International Conference on Systems, Man, and Cybernetics, 2009, pp. 4627-4632

[35] A. Kyriakopoulou and T. Kalamboukis. Text classification using clustering. In ECML-PKDD Discovery Challenge Workshop Proceedings, 2006.

[36] Dehuri, S., Mall, R. "Predictive and Comprehensible Rule Discovery Using A Multi-Objective Genetic Algorithm", Knowledge Based System, volume 19, 2006, pp: 413-421.

[37] Xian-Jun Shi, Hong Lei , " A Genetic Algorithm-Based Approach for Classification Rule Discovery". International Conference on Information Management, Innovation Management and Industrial Engineering, 2008, Volume: 1, 2008, pp 175-178.

[38] Tsoumakas, G., Katakis, I., Vlahavas, I.: "Mining Multi-label Data.",Maimon, O., Rokach, L. (eds.) Data Mining and Knowledge Discovery Handbook, 2nd edn. Springer, Heidelberg ,2009.

[39] Zheng X.Q., Zhao L., *Association rule analysis of spatial data mining based on matlab a case of ancheng township in china.* Proceedings-1st International Workshop on Knowledge Discovery and Data Mining, WKDD, 2008, pp 76-80.