Prediction of Facility Location Using Evolutionary Spatial Data Mining

J. Arunadevi*
Dr. V. Rajamani**

Abstract

One of the features that distinguish the Geographical Information Systems (GIS) from other information systems is spatial information function. This function usually provides selection switches and solutions for GIS users. Simultaneously with GIS techniques development, the GIS executive analysis functions have also been developed interestingly. For instance, Facility location allocation is one of the areas of interest in GIS. The problem is evaluated using different methods of spatial data mining, with respect to the association analysis of the spatial data, clustering the associations and classifying them.

Keywords: Facility, location, Allocation, Spatial Data Mining, Multi Objective genetic Algorithm, GIS

Introduction

Land represents an important resource for the economic life of a majority of people in the world. Land is a very restricted resource and it is therefore important to recognize its prospects and optimize its use. Land use however is not only a realm of those directly using it; it is exposed to a part of the wider reality of social and economic development and change. Land use therefore is a highly dynamic process. Due to complex needs and a large number of criteria such as environmental, economical, sociological and natural factors, decision-makers need to use techniques of multi-objective planning and multi-criteria analysis. This is applicable to many social activities related to land, especially in the field of planning of spatial organization.

Spatial decision problems are those problems in which the decision implies the selection among several potential actions or alternatives associated with some specific locations in land. In this kind of spatial allocation problem, facility location is significant, because of the growing needs of the population. The goal of the facility location problem study is to serve a set of

*Assistant Professor, Dept of MCA, Thiagarajar School of Management, Madurai, Tamilnadu, INDIA
**Principal, Indra Ganesan College of Engineering, Tirchirapalli, Tamilnadu, INDIA

Authors are grateful to the anonymous referees for comments and suggestions for improving the text and contents of the paper. Authors alone are responsible for any errors/mistakes still remaining in the paper.

demand points, typically called clients, by opening a set of access points, called facilities. 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. Every enterprise in private and public sectors face the problem of strategically locating facilities to provide services to consumers. These decisions are very important because they require long-term investments in buildings and facilities, as well as a sizable financial outlay.

The solution of a facility location problem is important for the decision makers. They need a decision support tool which will locate the facility based on several criteria. The type of facility, user preference on facility, different services of the facility, facility opening and closing time, facility establishment and relocation cost criteria can also be significant for decision makers to take a decision. The Location allocation problem nowadays not only sets facility in the nearest distance but also tries to add the non distance based criteria to find optimal solution.

It is not very easy to answer a location problem. The reasons are:

i) Uncertainty in future,
ii) Complexity and conflicting factors associated with the site selection problem,
iii) Constraints and limitations of resources to produce a site, etc.

Evolutionary Spatial Data Mining in Facility Location Problem

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. The phenomenal growth of spatial data increases the importance of spatial data mining which is used to mine fascinating and constructive but inherent knowledge. Spatial data mining is a branch of data mining concerned with the discovery of patterns in spatial databases. Because of the large number of specific objectives to be considered in spatial planning, the application of spatial data mining techniques with multi objectives can have a significant impact on the quality, speed and cost of the planning.

The application of traditional data mining techniques to spatial data can result in patterns that are biased or that do not fit the data well. Unfortunately, spatial data mining techniques lack formal mechanisms to help decision-makers for exploring the solution space of their problem with multiple objective. Evolutionary Algorithms (EA) seem particularly suitable to solve multi objective optimization problems because they are able to capture multiple Pareto-optimal solutions in a single simulation run and may exploit similarities of solutions by recombination. The robustness and domain-independent capabilities of EA also can add to the taste to make use of them.

In the spatial facility allocation problem, a large number of specific objectives lead to the complex relationship between the facility's performance and its environmental conditions. Since evolutionary algorithms always have solutions in a population during a search itself, they can provide the best solution accompanied with the spatial data mining techniques. They can be used to provide the solutions to the spatial facility allocations with multi objectives.

Literature Review

Margaret H. Dunham and S. Sridhar (Margaret H. Dunham and S. Sridhar, 2006) advocates Spatial Association Rule mining (SAR) which is about generating 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. Various activities involved in the SAR are computing spatial relationships, generating frequent sets and extracting the association rules. The existing approaches use quantitative reasoning which computes distance relationships during the frequent set generation (Yoo J.S., Shekhar S. and Celik M., 2005)-(Yoo, J.S. and Shekhar S., 2004). These approaches deal only with points, consider only quantitative relationships and do not consider non spatial attributes of geographic data, which may be fundamental importance to knowledge discovery. Qualitative spatial reasoning (Appice, M., Berardi, M., Ceci, M. and Malerba, D, 2005), (Mennis, J. and Liu, J.W., 2005), (Koperski, K. and Han, J., 1995) considers distance and topological relationships between a reference geographic object type and a set of relevant feature types represented by any geometric primitive (e.g. points, lines, and polygons). Bogorny.V (Bogorny, V., 2006) uses the qualitative spatial reasoning approach with prior knowledge and removes well known patterns completely by early pruning the input space and the frequent item sets. Salvatore Orlando (URL: dsi.unive.it) discusses the various kinds of spatial predicates which can be involved in spatial association rules.

Over the past decade, population-based EAs have been found quite useful in solving multi-objective optimization problems simply because of their ability to find multiple optimal solutions in a single simulation run. Alex A. Freitas (Alex A. Freitas, 2003) explains the motivation for using Genetic Algorithms (GA) in the discovery of high-level prediction rules. GA performs a global search and copes better with attribute interaction than the greedy rule induction algorithms often used in data mining. According to Dehuri. S et al. (Dehuri, S., Jagadev, A. K., Ghosh A. And Mall R., 2006) GA for rule discovery can be divided into two broad approaches, the Michigan approach and the Pittsburgh approach. Peter P. Wakabi-Waiswa and Venansius Baryamureeba (Peter P. Wakabi-Waiswa and Venansius Baryamureeba, 2008) say the biggest distinguishing feature between the two is in the mapping of chromosome to the rule.

Ant Colony Optimization (ACO) is a paradigm for designing meta heuristic algorithms for combinatorial optimization problems. The ACO algorithm was first introduced by Colorni, Dorigo and Maniezzo (A. Colorni, M. Dorigo, and V. Maniezzo, 1991) and the first Ant System (AS) was discussed by Dorigo (M. 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 the shortest path from a food source to the nest without using visual cues by exploiting pheromone information (R. Beckers, J.L. Deneubourg and S. Goss, 1992), (S. Goss, S. Aron, J.L. Deneubourg and J.M. Pasteels, 1989) and (B. Holldobler and E.O. Wilson, 1990).

Clustering association rule is one of the meaningful ways of grouping association rules into different clusters. When spatial association rules are generated in order to identify group of targets clustering approach is used. Waler A.Kosters et al. (1999)

select highly ranked (based on confidence) association rules one by one and formed cluster of objects covered by each rule until all the objects in the database are covered. Toivonen.H et al. (1995) forms cluster of rules of the form Xi -> Y, that is, rules with different antecedent but with same consequent Y and they extract representative rules for each cluster as knowledge for the cluster. Pi Dechang and Qin Xiaolin (Pi Dechang and Qin Xiaolin, 2008) form cluster of rules based on structure distance of antecedent. Alipio Jorge (Alipio Jorge, 2004) forms hierarchical clustering of rules based on different distance methods used for rules. G. Li, and H.J.Hamilton (G. Li, and H. J. Hamilton, 2004) discuss different ways of pruning redundant rules including rule cover method. All Associative Classifier (AC) CBA, CMAR are suggested by W. Li et al. (W. Li, J. Han, and J. Pei, 2001). RMR proposed by A. Thabtah, and P. I. Cowling (Thabtah A. and Cowling P.I., 2007) and MCAR proposed by Adriano Veloso et al. (Adriano Veloso, Wagner Meira, Marcos Gonçalves, and Mohammed Zaki, 2007) generate cluster of rules called class-association rule (CAR) with class label as same consequent and they use database (rule) cover to select potential rules to build (AC) classifier model. In most of the association rule mining, the confidence measure is used to rank association rules. Also, other measures such as chi-square, laplace-accuracy is used to select highly ranked rules.

Diansheng Guo and Jeremy Mennis (Diansheng Guo, Jeremy Mennis, 2009) say classification is about grouping data items into classes (categories) according to their properties (attribute values). (Ester, M., Kriegel, H. P., and Sander, J., 1997) and (Koperski, K., Han, J. and Stefanovic, N., 1998) advocates, 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. (Ester, M., Kriegel, H. P. and Sander, J., 1997) discuss a neighborhood graph based extension of decision trees that consider 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. (Malerba, D., Esposito, F., Lanza, A., Lisi, F.A. and Appice, A., 2003) try to exploit the expressive power of predicate logic to represent both spatial relations and background knowledge, such as spatial hierarchies. Nadia Ghamrawi and Andrew McCallum (Nadia Ghamrawi and Andrew McCallum, 2005) say that 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.

Benhui Chen et al. (Benhui Chen, Liangpeng Ma and Jinglu Hu, 2010) say 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. G. Tsoumakas et al. (Tsoumakas, G., Katakis, I. and Vlahavas, I., 2009) formulate existing methods for multi-label classification into two main categories; namely 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, K. et al. (, K., Furnkranz, J. and Hullermeier, E., 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 modifies standard single-label learning algorithm for multi-label classification. Methods adopted by both Min-Ling Zhang and Zhi-Hua Zhou (2007) is Multi-Label k-Nearest Neighbor (MLKNN), Weiwei Cheng and Eyke Hllermeier (2009) is Instance Based Learning by Logistic Regression (IBLR) are state-of-the-art of algorithm adaptation in multi label classification algorithms that exploit instance-based learning.

Associative Classification (AC) is a branch of a larger area of scientific study known as data mining. Fayyad et al. (Fayyad, U., Piatetsky-Shapiro, G., Smith, G. and Uthurusamy, R., 1998) define data mining as one of the main phases in knowledge discovery from databases, which extract 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. (Thabtah F., Cowling P., and Peng Y., 2004) use classification as 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 (2001) present an associative classification algorithm that selects and analyses the correlation between high confidence rules. Yin, X. and Han, J (Yin, X. and Han, J., 2003) develop a greedy associative classification algorithm called Classification based on Predictive Association Rules (CPAR).

Zhu, Xiaojin (Zhu, Xiaojin, 2008) says semi supervised classification is a special form of classification. It is 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 (41-46). Expectation-Maximization (EM) is a well known class of iterative algorithms for maximum-likelihood or maximum a posteriori estimation in problems with incomplete data (47-48). A co-training approach to semi-supervised classification was proposed by Blum and Mitchell (2002). Several authors report experimental results which show the effectiveness of co training (50-52). Foli (53) reports 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.

Problem Formulation

Decisions are often evaluated on the basis of the quality of the processes behind them. It is in this context that spatial data mining is increasingly used to generate alternatives to aid decision-makers in their deliberations. In the process of identifying the best geographic location for a service or production facility some of the factors involved are proximity to source of supply, proximity to customers, proximity to labor, site considerations and quality-of-life issues. The traditional spatial data mining techniques includes association rule mining, clustering and classification. These techniques can be applied to the facility location problem, which generates the knowledge for decision making.

To satisfy various constraints in the problem, multi objectives are to be considered. The application of spatial data mining techniques to solve this type of problems directly always results in exponential complexity. Since no single location may be better than others, the next problem is to identify several locations from which to choose the best one. Because of this nature of the problem, the solution space increases, the size of the database increases and the number of variables involved also increases. The other things to be considered in spatial data are they are multi dimensional and auto correlated.

Population-based 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. Evolutionary algorithms provide the opportunity of promoting the diversity among the solutions which, enhances the probability of identifying the best choice in the multi constraint environment. Global searching capability stimulates the chance of selecting the finest among the given large solution space. Evolutionary algorithms are not affected by the order of the input data. Escaping from the local optimal solutions can be achieved with this said property. It is a preferred tool, as it can efficiently solve problems which are typically complex.

Optimization of the problem can be achieved by customizing evolutionary algorithms when a large number of parameters are to be adjusted and several objectives are to be optimized. Spatial data mining techniques with customized EA can be used to decide the best location for multiple facilities with capacity planning for the maximum output rate.

Frame Work for Proposed Methodology

A novel Hybrid Evolutionary Algorithm (HEA) which uses GA with ACO for spatial association rule mining is presented. 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. The proposed HEA algorithm is to enhance the performance of the MOGA by incorporating local search with 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 the local optima. From the experiment results, it is shown that the proposed HEA algorithm has superior performance when compared to other existing algorithms.

The clustering algorithm groups the rules in the form Xi -> Y for i=1,2,…,n. That is, different rule antecedents Xi's are collected into one group for a same rule consequent Y. The next step is to select a small set of representative rules from each group.

The representative rules are selected based on rule instance cover as follows:

Let Ry={ Xi -> Y | i=1,2,…,n } be a set of n rules for some item-set Y and m(Xi Y) be rule cover, which is the set of tuples/records covered by the rule Xi -> Y in the dataset D. Let Cy be the cluster rule cover for a group or cluster of rules Ry. i.e.,

Cy = m(Ry) = U i=1,2,…n m(Xi Y)

From cluster rule set Ry, find a small set of k rules ry called representative rule set such that m(ry) is almost equal to m(Ry). i.e.,

m(ry) ≈ m(Ry), or U i=1,2,…k m(Xi Y) ≈ i=1,2,…n m(Xi Y), where k<< n.

Multi-label classification based on association rules with multi objective genetic algorithms (MOGA) is proposed to deal with the 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), and then combines labels with the same attributes to generate multi-label rules with the help of the MOGA. It extracts partial dataset features filtered by the 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.

A Semi supervised classifier is proposed, 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 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, and global maximum by ACO and construct an efficient way to increase the speed of classification through the filtering process.

Proposed Algorithm

A multi label classifier based on the HEA, the MOGA and the association rule mining with clustering is proposed. 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. In the next stage the multi label rules are generated by the MOGA. In the final stage the multi label classifier is built with a sorting mechanism applied to the rules generated.

Step 1:
Optimization of the spatial association rule mining through HEA
Input: Dataset D
Output: optimized spatial association rules
Procedure
BEGIN
Application of Apriori for the rule generation;
MOGA with ACO is applied on the rules;
Rule set is generated based on the fitness;
END

Step 2:
Application of clustering to the association rules obtained in step 1
Input: optimized spatial association rules
Output: associative clusters
BEGIN
Application of GA for rearranging the rules in various orders based on the fitness preferred by the user;
Clustering the rules obtained from HEA, processed in step by GA;
Application of GA for retaining nearest neighbors in common cluster;
END

Step 3:
Multi label rules generation with the application of MOGA
Input: associative clusters
Output: Multi label clusters
BEGIN
Step A: Generate the initial population P;
Step B: Subdivide the population into 'm' subpopulation according to the number of objectives (m);
Step C: For each subpopulation 'S' carry out the following steps.
Step C.1: Evaluate the fitness based on the objectives assigned to each subpopulation;
Step C.2: Select the best chromosome 'X' from 'S';
Step C.3: Select two chromosomes 'P1' and 'P2' from 'S';
Step C.4: Apply crossover between 'P1' & 'P2'. Let O1 be the best offspring;
Step C.5: Apply crossover between O1 and 'X'. Let O2 be the best offspring;
Step C.6: if O2 is better than 'X' then replace O2 with 'X';
Step C.7: Iterate steps (C.1 – C.5) until all chromosomes are considered;
Step D: The best chromosome in each subpopulation is compared for best solutions;
Step E: Iterate steps C & D until best Pareto optimal solutions are obtained;
END

Step 4:
Merging of the rule set from multi label clusters to obtain multi label classifier
Input: Multi label clusters
Output: Multi label classifiers
BEGIN
Rules from the clusters are prioritized by MOGA;
The rules are sorted based on the prioritization;
The rules are merged based on the sorting to get multi label classifier;
END

Result and Discussions

The environmental set up is summarized in table 1

Table 1: Environmental parameters for GA

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

Comparison metrics
The metrics used are
1. Hamming loss
2. Accuracy
3. Precision
4. Recall

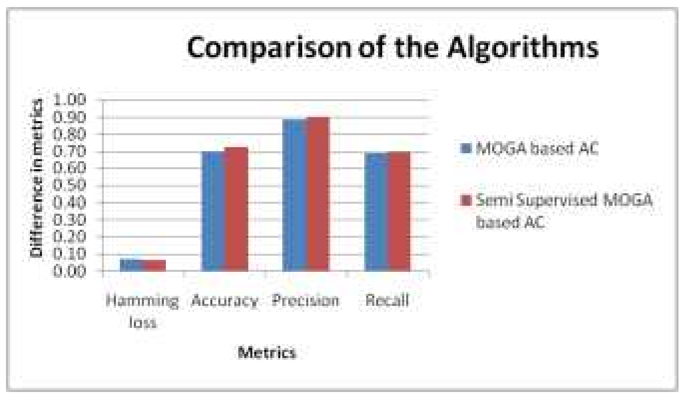

Figure 1: Comparison of the proposed approach with other benchmark algorithm based on the four measures

The consolidated report has been depicted in Figure 1; it shows that the proposed approach holds 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 the proposed approach because:

1) Semi supervised learning helps us to minimize the number of labels to be searched.
2) The 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) GA helps the clustering for not suffering from the order of the input.
4) GA is applied for retaining nearest neighbors in the common cluster, so that the searching labels in the relative cluster induce increase in 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. The region for the study is Annanagar, Madurai, Tamil Nadu, India which is the city taken for the case study and the results based on the various categories are presented.

The primary categorizations are
1. Men
2. Women

The secondary categorizations are
1. Students (Men)
2. Married (Men)
3. Senior citizens (Men)
4. Students (women)
5. Just married (Women)
6. Working women, etc.

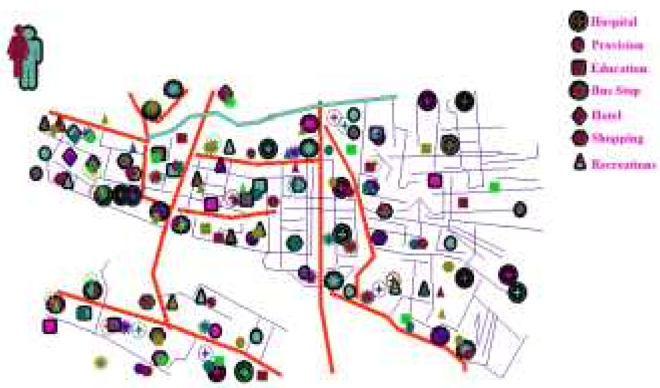

Figure 2: Preference based on Men and Women

The following observations are made
1. The driving distance to the facilities between one to two kilometers from the residence is preferable
2. The same kind of facilities situated nearby is preferable
3. The students and just married people are alike in their expectations
4. Working personals require the facilities of different nature to be situated nearer
5. Hotels and recreations could be nearer
6. All require the accessibility of the facilities within a minimum travel time from one to other

These observations could be extended with the various other parts of the city. The study can be also extended to various types of facilities.

Conclusion

A spatial associative classification algorithm optimized with the evolutionary algorithms for the classification for space planning to meet the user preferences is presented. The classification problem is considered a multi objective one since if it is considered as single objective, the main demerit associated would be that the generated rules would often be more complex than necessary and not easy to comprehend. The reason behind this 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 hyper cubes. In real-world applications, the feature space is often very complex and a large set of such hyper cubes might be needed to approximate the class boundaries among different classes.

A multi-label classification based on association rules with multi objective genetic algorithms is proposed to deal with the 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 and then combines labels with the same attributes to generate multi-label rules with the help of the MOGA. It extracts partial dataset features filtered by the 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 increase.

Future efforts can be pursued further by enhancing the adaptability of the algorithm for multiple cities. The facility allocation for a particular brand of consumable can be planned based on the proposed approach. Decision support system may be developed for the expansion of the business based on the methodology presented.

Expanding the applications of the algorithm into broader GIS topics like optimized path finding and crime detection can be done. Planning for the expansion of the city could be addressed.

The algorithm proposed can be further fine tuned with the concepts of incremental learning and outlier analysis of the classifier which results in the better efficiency of the classifier.

References

1. *Margaret H. Dunham and Sridhar, S. (2006), "Data Mining Introductory and Advanced Topics", Pearson Education.*
2. *Yoo J.S., Shekhar, S. and Celik, M. (2005). "A Join-less Approach for Co-location Pattern Mining: A Summary of Results", In 5th IEEE-ICDM, Houston, p.813-816. IEEE Computer Society.*
3. *Yoo, J.S. and Shekhar S. (2004) "A partial join approach for mining co-location patterns", In 12th ACM-GIS, Washington, p.241-249, ACM Press.*
4. *Appice, M., Berardi, M., Ceci, M. and Malerba, D (2005), "Mining and Filtering Multi-level Spatial Association Rules with ARES", In 15th ISMIS, New York, p.342-353. Springer.*
5. *Mennis, J. and Liu, J.W. (2005), "Mining Association Rules in Spatio-Temporal Data: An Analysis of Urban Socioeconomic and Land Cover Change", Transactions in GIS, v9 (1), January, p. 5-17.*
6. *Koperski, K. and Han, J. (1995), "Discovery of spatial association rules in geographic information databases", In 4th SSD, Portland, p. 47-66. Springer.*
7. *Bogorny, V. (2006), "Enchancing spatial association rule mining in geographic databases". PhD Thesis, Instituto de Informatica da UFRGS, Brazil, October.*
8. *www.dsi.unive.it/~dm/ssd95.pdf*
9. *Alex A. Freitas(2003), "A survey of evolutionary algorithms for data mining and knowledge discovery", Advances in evolutionary*

computing: theory and applications, Springer-Verlag New York, Inc., New York, NY.

10. *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, ISSN 1546-9239.*
11. *Peter P. Wakabi-Waiswa and Venansius Baryamureeba (2008), "Extraction of Interesting Association Rules Using Genetic Algorithms". International Journal of Computing and ICT Research, Vol. 2, No. 1, pp. 26 – 33.*
12. *Colorni, A., Dorigo, M. and Maniezzo, V. (1991). "Positive feedback as a search strategy". Technical Report No. 91-016, Politecnico di Milano, Italy,.*
13. *Colorni, A., Dorigo, M. and Maniezzo, V. (1991), "The ant system: an autocatatlytic process". Technical Report No. 91- 016, Politecnico di Milano, Italy.*
14. *Dorigo, M. (1992), "Optimization, Learning and Natural Algorithms". Ph.D. Thesis, Politecnico di Milano, Italy.*
15. *Beckers, R., Deneubourg J.L and Goss S. (1992), "Trails and Uturns in the selection of the shortest path by the ant lasius niger". Journal of Theoretical Biology, 159, pp. 397- 415.*
16. *Goss, S., Aron, S., Deneubourg, J.L. and Pasteels, J.M. (1989),"Self-organized shortcuts in the argentine ant", Naturwissenschaften, 76, pp. 579-581.*
17. *B. Holldobler and Wilson,E.O. (1990), "The Ants", Springer-Verlag: Berlin,.*
18. *Waler A.Kosters, Elena Marchiori and Ard A.J. Oerlemans (1999), "Mining Clusters with Association Rules", Advances in Intelligent Data Analysis (IDA-99) (D.J.Hand, J.N.Kok and M.R.Berthold, Eds.), Lecture Notes in Computer Science 1642, Springer, pp. 39-50.*
19. *Toivonen, H., Klemettinen, M., Ronkainen, P., Hatonen, K., and Mannila, H. (1995), "Pruning and grouping discovered association rules", In Proceedings ECML-95 Workshop on Statistics, Machine Learning, and Knowledge Discovery in Database, April, pp 47-52.*
20. *Pi Dechang and Qin Xiaolin(2008), "A new Fuzzy Clustering algorithm on Association rules for knowledge management", Information Technology Journal 7(1), pp. 119-124.*
21. *Alipio Jorge(2004), "Hierarchical Clustering for thematic browsing and summarization of large sets of Association Rules", In Proceedings of the 4th SIAM International Conference on Data Mining, Orlando, FL, pp. 178-187.*
22. *Li, G. and Hamilton, H. J. (2004), "Basic association rules", In Proceedings of the 4th SIAM International Conference on Data Mining, Orlando, FL, pp. 166–177.*
23. *Li, W., Han, J. and Pei, J. (2001), "CMAR: accurate and efficient classification based on multiple class-association rules", International Proceedings of IEEE International Conference on Data Mining (ICDM2001), pp. 369–376.*
24. *Thabtah, A., and Cowling, P. I. (2007), "A greedy classification algorithm based on association rule", Applied Soft Computing, Vol. 7, No. 3, pp. 1102-1111, June.*
25. *Adriano Veloso, Wagner Meira, Marcos Gonçalves, and Mohammed Zaki(2007), "Multi-label Lazy Associative Classification", Knowledge Discovery in Databases: PKDD 2007, pp. 605-612.*
26. *Diansheng Guo, Jeremy Mennis (2009), "Spatial data mining and geographic knowledge discovery – An introduction", Computers, Environment and Urban Systems 33, 403–408.*
27. *Ester, M., Kriegel, H. P., & Sander, J. (1997), "Spatial data mining: A database approach", Advances in spatial databases. Berlin: Springer-Verlag, pp. 47–66, Berlin.*
28. *Koperski, K., Han, J., and Stefanovic, N.(1998), " An efficient two-step method for classification of spatial data" , 1998 international symposium on spatial data handling SDH'98, pp. 45–54, Vancouver, BC, Canada.*
29. *Malerba, D., Esposito, F., Lanza, A., Lisi, F.A. and Appice, A. (2003): "Empowering a GIS with Inductive Learning Capabilities: The Case of INGENS". Journal of Computers, Environment and Urban Systems, Elsevier Science, 27. 265-281.*
30. *Nadia Ghamrawi , Andrew McCallum (2005), "Collective multi-label classification", Proceedings of the 14th ACM international conference on Information and knowledge management, Bremen, Germany, October 31-November 05.*
31. *Benhui Chen, Liangpeng Ma and Jinglu Hu (2010), "An improved multi-label classification method based on svm with delicate decision boundary", International Journal of Innovative Computing, Information and Control , Volume 6, Number 4, pp. 1605–1614, April.*
32. *Tsoumakas, G., Katakis, I. and Vlahavas, I. (2009), "Mining Multi-label Data.", Maimon, O., Rokach, L. (eds.) Data Mining and Knowledge Discovery Handbook, 2nd edn. Springer, Heidelberg.*
33. *Brinker, K., Furnkranz, J. and Hullermeier, E. (2006), "A unified model for multilabel classification and ranking", in Proceedings of the 17th European Conference on Artificial Intelligence (ECAI '06), Riva del Garda, Italy, pp. 489-493.*
34. *Min-Ling Zhang and Zhi-Hua Zhou (2007), "Ml-knn: A lazy learning approach to multi-label learning," Pattern Recognition, vol. 40, no. 7, pp. 2038–2048.*
35. *Weiwei Cheng and Eyke Hllermeier (2009), "Combining instance based learning and logistic regression for multilabel classification," Machine Learning, vol. 76, no. 2-3, pp. 211–225, September.*
36. *Fayyad, U., Piatetsky-Shapiro, G., Smith, G. and Uthurusamy (1998), R. Advances in Knowledge Discovery and Data Mining, Menlo Park, CA: AAAI Press.*
37. *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, pp. 207-213, Nov.*
38. *W. Li, J. Han, and J. Pei. (2001), "CMAR: Accurate and efficient classification based on multiple-class association rule", in Proceeding of the First IEEE International Conference on Data Mining (ICDM'01), pp. 369- 376, San Jose, CA, Nov.*
39. *Yin, X. and Han, J. (2003), CPAR: Classification Based on Predictive Association Rules, Proceedings SIAM International Conferene on Data Mining (SDM'03), 331-335.*
40. *Zhu, Xiaojin(2008). "Semi-Supervised Learning Literature Survey", Computer Sciences, Univeristy of Wisconsin-Madison. Technical Report 1530, University of Wisconsin – Madison,.*
41. *Seeger M. (2002), "Learning with labeled and unlabeled data", Technical Report, University of Edinburgh, Institute for Adaptive and Neural Computation, pp. 1-62 ,Dec.*
42. *Kemp T., Waibel A. (1966), "Unsupervised Training of a Speech Recognizer: Recent Experiments", Proc. Eurospeech, vol. 6, pp. 2725- 2728, 1999.Nagy G., Shelton G.L., "Self-corrective character recognition systems", IEEE Trans. On Information Theory, IT-12, no. 2, pp. 215-222, April.*
43. *G. Nagy (2004), "Classifiers that improve with use", Proceedings of Conference on Pattern Recognition and Multimedia, IEICE, Tokyo, pp. 79-86, February.*
44. *Inoue M., Ueda N. (2003), "Exploitation of unlabeled sequences in hidden markov models", IEEE Transactions On Pattern Analysis and Machine Intelligence, Vol. 25, No. 12, pp. 1570-1581, Dec.*

45. *Young T.Y., Farjo A. (1972), "On decision directed estimation and stochastic approximation", IEEE Transactions On Information Theory, pp. 671-673, Sep.*
46. *Dempster A.P., Laird N.M., Rubin D.B.(1977), "Maximum likelihood from incomplete data via the EM algorithm", Journal of the Royal Statistical Society, Series B, 39(1), pp. 1-38,.*
47. *Nigam K., McCallum A.K. (2000), Thrun S., Mitchell T., "Text classification from labeled and unlabeled documents using EM", Machine Learning, 39, pp. 103-134,.*
48. *Blum A., Mitchell T. (1998), "Combining labeled and unlabeled data with co-training", Proceedings of the Workshop on Computational Learning Theory, pp. 92-100.*
49. *Seeger M. (2002), "Learning with labeled and unlabeled data", Technical Report, University of Edinburgh, Institute for Adaptive and Neural Computation, pp. 1-62, Dec.*
50. *Nigam K, Ghani R. (2000), "Analyzing the effectiveness and applicability of co-training", Proceedings 5th International Conference on Information and Knowledge Management, pp. 86-93.*
51. *Goldman S., Zhou Y. (2000), "Enhancing supervised learning with unlabeled data", Proceedings of the 7th International Conference On Machine Learning, ICML 2000, pp. 327-334.*
52. *Roli, F. (2005), "Semi-Supervised Multiple Classifier Systems: Background and Research Directions", Multiple Classifier Systems, Springer Verlag, Vol. 3541/2005, pp 674-676.*