PSO based Swarm Intelligence Technique for Multi-Objective Classification Rule Mining

Anil Kumar Mishra, PhD
Assistant Professor
Department of IT, UTM
Shillong, Meghalaya

ABSTRACT
Today's real world faces different kinds of complex optimization problems. The existing methodologies can't cope of with such complex problems. This paper presents classification rule mining as a multi-objective problem rather than a single objective one. Multi-Objective optimization is a challenging area and focus for research. Here two modern domains of research are discussed one is swarm intelligence and other is data mining. In this paper PSO is taken as taken as a swarm intelligence algorithm and classification rule mining is taken as the problem domain. In classification rule discovery, classifiers are designed through the following two phases: rule extraction and rule selection. In the rule extraction phase, a large number of classification rules are extracted from training data. This phase is based on two rule evaluation criteria: support (coverage) and confidence. An association rule mining technique is used to extract classification rules satisfying pre-specified threshold values of minimum support (coverage) and confidence. In second phase, a small number of rules are targeted from the extracted rules to design an accurate and compact classifier. In this paper, I used PSO for multiple objective rule selection to maximize the accuracy of the rule sets and minimize their complexity.

General Terms
Swarm Intelligence, Multi-Objective optimization, soft computing.

Keywords
Classification, multi-objective optimization, particle swarm optimization, multi-objective classification problem, pattern recognition, data mining.

1. INTRODUCTION
Soft computing and computational intelligence is the modern face of artificial intelligence. Swarm intelligence is a sub domain of soft computing. Swarm intelligence generally mimics the swarming behavior of group of some living elements such as birds ants, fishes, cats, bees etc.[1,2,3,4].The swarm intelligence is an reliable method because if any of the agents from swarm fails ,then others can achieve the objective. The swarm intelligence technique can be applicable to the problem areas where other traditional optimization technique fails. Data mining is a very lively and fast growing research area in the field of Computer Engineering and Information Technology. The task of data mining is to extract useful knowledge for human users from a large data set. The application areas of data mining includes: medicinal science, market research [14, 15, 16], weather forecasting and prediction, network security [6], life expectancy science [7], and biometric [9], etc.

The application of swarm intelligence to data mining is not always easy due to its heavy computational load especially in the case of a large data set [10, 11] and the complexity of the domain. Particle Swarm Optimization (PSO) is a swarm intelligence technique which is very similar to evolutionary computation methods. However, PSO is considered here because of its plainness and easiness of implementation. Additionally, it needs tuning of less number of parameters and use of a few operators. The relative simplicity of PSO and the population based technique and information sharing mechanism associated with this method have made it a natural candidate to be extended from single to multi-objective optimization. The details of PSO for single and multi-objective optimization problems have been discussed in Section 2.

A critical review of multi-objective optimization in data mining has been presented in [12]. Association rule mining [13] is one of the most fundamental data mining tasks. In its basic form, association rules satisfying pre-specified threshold values of the minimum *support* (coverage) and *confidence* can be extracted form a data set. The application of association rule mining to classification problem is often referred to as classification rule mining or associative classification [14, 15]. Classification rule mining usually consists of two phases: rule extraction and rule selection. In the rule extraction phase, a large number of classification rules are extracted from a data set using an association rule mining technique. Classification rules satisfying pre-specified threshold values of the minimum support (coverage) and confidence are usually extracted from a data set. In rule selection phase, a part of extracted rules are selected or targeted to design an accurate and compact classifier using a binary multi-objective PSO. The accuracy of the designed classifier usually depends on the specification of the minimum support (coverage) and confidence. Their tuning has been discussed for classification rule mining in the literature [16, 17].

Whereas the basic form of association rule mining is to extract association rules satisfying pre-specified threshold values of the minimum support and confidence [13], other rule evaluation measures known as correlation measures have been proposed to qualify the interestingness or goodness of an association rule. Among them are *gain*, *variance*, *Chi-squarred value*, *entropy gain*, *gini*, *Laplace*, *lift* and *conviction*.

A measure is null-invariant if its value is free from the influence of null-transactions (i.e., transactions that do not contain any of the item sets being examined). Because large databases typically have numerous null transactions, a null-invariant correlation measure should be used, such as all-confidence or cosine. When interpreting correlation measure values, it is important to understand their implications and limitations.

It is shown in [19] that the best rule according to any of the above-mentioned measures is a Pareto-optimal rule with respect to support (coverage) and confidence. The use of an

evolutionary multi-objective (EMO) algorithm was proposed to search for Pareto optimal rules with respect to support (coverage) and confidence for partial classification [20]. Partial classification is the classification of a particular single class from all the other classes. Similar formulation to [20] were used to search for Pareto-optimal association rules [21]. EMO algorithms were also used to search for Pareto-optimal rule sets in classification rule mining [22, 23] where the accuracy of rule sets was maximized and their complexity was minimized.

In this paper, I examine the effect of PSO based multi-objective rule extraction on coverage (support) and confidence from original datasets and PSO based multi-objective rule selection on the accuracy and the complexity of selected rules through computational experiments on some well-known benchmark data sets from the University of California, Irvine (UCI) machine learning repository. I also examine the relation between Pareto-optimal rules and Pareto-optimal rule sets in classifier design. This examination is performed by depicting selected rules together with the candidate rules in the coverage (support) - confidence plane. My objective is to examine whether selected rules in Pareto-optimal rule sets are close to the Pareto front with respect to support (coverage) and confidence.

This paper is organized as follows. In Section 2, I discuss PSO for single and multi-objective optimization problems. I explain some basic concepts in classification rule mining in Section 3. Next i explain two approaches of PSO based multi-objective classification rule mining in Section 4. One approach handles each classification rule as an individual to search for Pareto-optimal rules. The other approach handles each rule set as an individual to search for Pareto-optimal rule sets. Then i explain PSO based multi-objective rule selection as a post-processing procedure in the rule selection phase of classification rule mining in Section 5. Pareto-optimal rule sets are found from a large number of candidate rules, which are extracted from a data set using an association rule mining technique in the rule discovery phase. I report experimental results on some well-known bench mark data sets in Section 6. Experimental results demonstrate the effect of PSO based multi-objective rule selection on the accuracy and the complexity of selected rules. The relation between Pareto-optimal rules and Pareto optimal rule sets is also examined in Section 4. Section 5 concludes the topics by summarizing with some of the promising future research and appended a long list of useful references.

2. APPLICATIONS OF PSO FOR MULTI-OBJECTIVE PROBLEMS

In order to apply the PSO [18, 19, 20] strategy for solving multi-objective problems, the original scheme has to be modified. The algorithm needs to search a set of different solutions (the so-called Pareto front) instead of a single solution (as in single objective optimization). I need to apply Multi-objective Particle Swarm Optimization (MOPSO) to search towards the true Pareto front (non-dominated solutions). Unlike the single objective particle swarm optimization, the algorithm must have a solution pool to store non-dominated solutions found by searching up to stopping criterion (say, upto iteration I_{max}). Any of the solutions in the pool can be used as the global best (gbest) particle to guide other particles in the swarm during the iterated process. The plot of the objective functions whose non-dominated solutions are in the solutions pool would make up for the Pareto front. The pseudocode for general MOPSO is illustrated in algorithm.

Algorithm: General Multi-objective Particle Swarm Optimization Algorithm

```
Begin
    Parameter Settings and initialize Swarm
    Fitness Evaluation and initialize leaders in a leader pool or external archive
    Archive the top best leader from the external archive through calculation of some sort of quality measure for all leaders.
    I = 1                          /* I = Iteration count */
    While (the stopping criterion is not met, say, I < Imax) do
        For each particle
            Select leader in the external archive
            Update velocity
            Update position
            Mutate periodically
        /* optional */
            Fitness Evaluation
            Update pbest
        End for
        Crowding of the leaders
        Update the top best into external archive
    I++
    End While
    Report results in the external archive
End
```

PSO uses an operator that sets the velocity of a particle to a particular direction. This can be seen as a directional mutation operator in which the direction is defined by both the particle's personal best and the global best (of the swarm). If the direction of the personal best is similar to the direction of the global best, the angle of potential directions will be small, whereas a larger angle will provide a larger range of exploration. In contrast, evolutionary algorithms use a mutation operator that can set an individual in any direction (although the relative probabilities for each direction may be different). In fact, the limitations exhibited by the directional mutation of PSO has led to the use of mutation operators (sometimes called turbulence operators) similar to those adopted in evolutionary algorithms.

In the case of multi-objective optimization problems, each particle might have a set of different leaders from which just one can be selected in order to update its position. Such set of leaders is usually stored in a different place from the swarm that is called external archive. This is a repository in which the non-dominated solutions formed so far are stored. The solutions contained in the external archive are used as leaders when the positions of the particles of the swarm have to be updated. Furthermore, the contents of the external archive is also usually reported as the final output of the algorithm.

3. CLASSIFICATION RULE MINING

Let us assume that i have m training patterns (tuples) $\overline{X}_p=(x_{p1}, x_{p2}, \ldots\ldots\ldots, x_{pn})$, p=1, 2, …, m from M classes in an n-dimensional continuous pattern space where x_{pi} is the attribute value of the p^{th} training pattern for the i^{th} attribute. I denote the data set of these m training patterns (or tuples) by D. For our pattern classification problem, i use classification rules of the following type:

Rule R_q: if x_1 is A_{q1} and …… and x_n is A_{qn} then Class C_q with CF_q (1)

where R_q is the label of the q^{th} rule, $\overline{x}=(x_1, x_2, \ldots x_n)$ is an n-dimensional pattern vector, A_{qi} is an antecedent interval for the i^{th} attribute, C_q is a class label, and CF_q is a rule weight (i.e. certainity grade).

I denote the classification rule R_q in (1) as "$\overline{A}_q \Rightarrow C_q$" where $\overline{A}_q=(A_{q1}, A_{q2}, \ldots, A_{qn})$. Each antecedent condition "x_i is A_{qi}" in (1) means the inclusion relation "$x_i \in A_{qi}$". It should be noted that classification rules of the form in (1) do not always have n antecedent conditions. Some rules may have only a few conditions while others may have many conditions. That is, A_{qi} in (1) can be a don't care condition.

In the field of association rule mining, two rule evaluation measures called support and confidence have often been used [13]. Rule support and confidence are two measures of rule interestingness. They respectively reflect the usefulness and certainty of extracted rules.

For example, following is an association rule mined from a data set D, shown with its confidence and support :

(age = young) $\wedge$ (credit=OK) => buys_computer=yes [support=20%, confidence=93%] (2)

where "$\wedge$" represents a logical "AND". More formally, let D be a dataset of tuples. Each tuple in D is described by n attributes, A_1, A_2, ……, A_n and a class label attribute, A_{class}. An item, p, is an attribute- value pair of the form (A_i, v) where A_i is an attribute taking a value, v. A data tuple $\overline{x}=(x_1, x_2, \ldots\ldots x_n)$ satisfies an item, $p=(A_i, v)$, iff $x_i=v$, where x is the value of the i^{th} attribute of $\overline{x}$. Association rules can have any number of items in the rule antecedent (left-hand side) and any number of items in the rule consequent (right-hand side). However, when mining association rules for use in classification, we are only interested in association rules of the form $p_1 \wedge p_2 \wedge \ldots. p_l \Rightarrow A_{class} = C$ where the rule antecedent is a conjunction of items, $p_1, p_2, \ldots\ldots\ldots, p_l$ ($l \leq n$), associated with a class label, C. For a given rule, R, the percentage of tuples in D satisfying the rule antecedent that also have the class label C is called the confidence of R. From a classification point of view, this is akin to rule accuracy. In the above example, a confidence of 93% for Association Rule (2) means that 93% of the customers in D who are young and have an OK credit rating belong to the class buys_computer=yes. The percentage of tuples in D satisfying the rule antecedent and having class label C is called the support of R. A support of 20% for Association Rule (2) means that 20% of the customers in D are young, have an OK credit rating and belong to the class buys_computer=yes.

The occurrence frequency of a rule in data set D is the number of tuples that contain the rule (i.e. compatible in both antecedent part and consequent class). This is also known as support count of that rule.

Let us denote the support count of the classification rule $\overline{A}_q \Rightarrow C_q$ by $SUP(\overline{A}_q \Rightarrow C_q)$, which is the number of patterns compatible with both the antecedent part A_q and the consequent class C_q. $SUP(A_q)$ and $SUP(C_q)$ are also defined in the same manner, which is the number of patterns compatible with Aq and C_q, respectively. The support of the classification rule $\overline{A}_q \Rightarrow C_q$ is defined as

$$\text{Support}\left(\overline{A}_q \Rightarrow C_q\right)=\frac{SUP\left(\overline{A}_q \Rightarrow C_q\right)}{|D|} \quad (3)$$

Where |D| is the cardinality of the data set D (i.e. |D|=m). On the other hand, the confidence of $\overline{A}_q \Rightarrow C_q$ is defined as

$$\text{Confidence}\left(\overline{A}_q \Rightarrow C_q\right)=\frac{SUP\left(\overline{A}_q \Rightarrow C_q\right)}{SUP\left(A_q\right)} \quad (4)$$

The confidence is directly used as the rule weight in this paper.

In partial classification [13, 14, 15], the following measure called coverage is often used instead of the support:

$$\text{Coverage}\left(\overline{A}_q \Rightarrow C_q\right)=\frac{SUP\left(\overline{A}_q \Rightarrow C_q\right)}{SUP\left(C_q\right)} \quad (5)$$

Since the consequent class is fixed in partial classification (i.e., since the denominator of (5) is constant) the maximization of the coverage is the same as that of the support.

In classification rule mining[14, 15], an association rule mining technique such as Apriori [13] is used in the rule extraction phase to efficiently extract classification rules satisfying pre-specified threshold values of the minimum support and confidence. These two parameters are assumed to be pre-specified by users. A part of extracted rules are selected to design an accurate and compact classifier in the rule selection phase.

Let S be a set of selected classification rules. That is, S is a classifier. When a new pattern $\overline{x}_p$ is to be classified by S, I choose a single winner rule with the maximum rule weight among compatible rules with $\overline{x}_p$ in S. The consequent class of the winner rule is assigned to $\overline{x}_p$. When multiple compatible rules with different consequent classes have the same maximum rule weight, the classification of $\overline{x}_p$ is rejected in PSO based multi-objective rule selection in this chapter. Only when the accuracy of the finally obtained rule set is to be evaluated, I use random tiebreak among those classes with the same maximum rule weight in computational experiments.

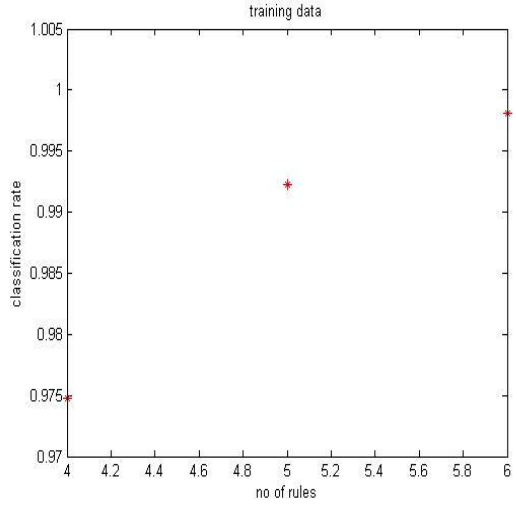
4. EXPERIMENTAL RESULTS

PSO-based multi-objective rule selection was made using 50% of each data set as test pattern. Experimental results are shown in Figure 1-5 while i observed very similar trade-off relations between the accuracy on training patterns and the number of selected rules for all the five data sets in Figure (1)(a), I obtained different results on test patterns as shown in Figure (1)(b).

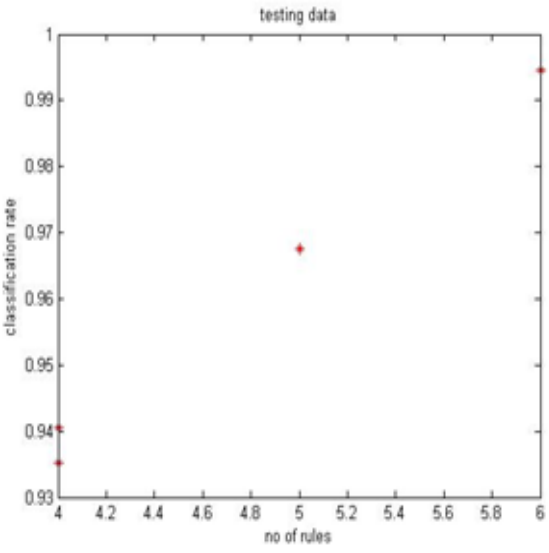
Finally I examine the relation between Pareto optimal rules and Pareto optimal rule sets. More specifically, we examine

whether Pareto optimal rules were selected in each rule at in Figures 1-5 by depicting the selected rules in the support-confidence plane. In Figures, it has shown candidate rules and Pareto optimal rules for the (breastw) data set and the (Cleveland heart) data set. Pareto-optimal rules are shown by asterisk in Figure 4 and 5. Since Pareto-optimal rules are defined by support maximization and confidence maximization, they locate along the upper-right limit of candidate rules in the support-confidence plane.

In Figures 2 and 3 we show the location of the selected rules in three rule sets for the (breastw) data set. Experimental results on the (cleveland heart disease) data set are shown in Figures in the same manner. I can see from these Figures that dominated rules were often selected in Pareto optimal rule sets. Whereas selected rules are often forming Pareto optimal rules in Figures 4 and 5, they are usually very close to class-wise Pareto optimal rules. As shown in this Figure, selected rules are class-wise Pareto optimal or very close to class wise Pareto optimal rules in many cases.



(a) Training Data



(b)Testing Data

Figure 1: Results by PSO-based multi objective rule selection (breastw)

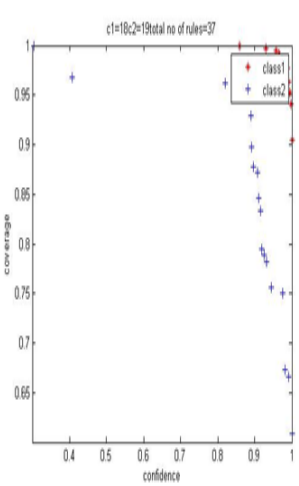


Figure 2: Pareto optimal rules (breastw)

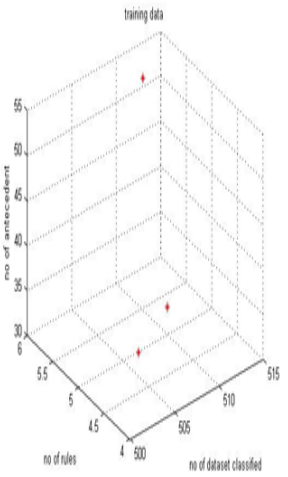


Figure 3: Pareto optimal rules set (breastw)

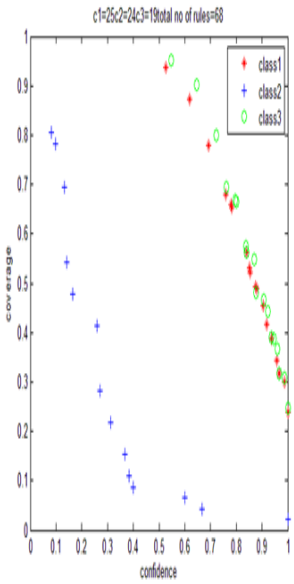


Figure 4: Pareto optimal rules set (balance Scale)

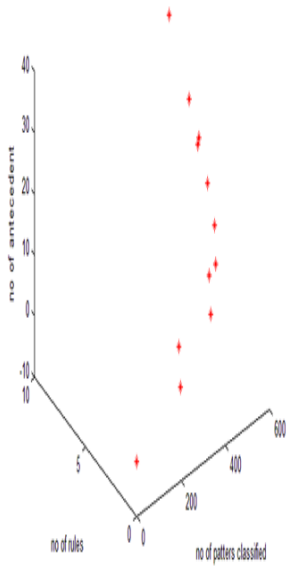


Figure 5: Pareto optimal rules set (balance scale)

5. CONCLUSION AND FUTURE WORKS

In this paper, PSO for multi-objective classification rule mining is explained. Finally, the relation between Pareto optimal rules and Pareto optimal rule sets are examined. It was shown that dominated rules were often selected in Pareto optimal rule sets. They were, however, very close to class wise Pareto optimal rules in many cases. The PSO based swarm intelligence technique can be applied for other multi objective problems such as numerical optimization, cloud computing and software testing. It can be hybridized with other evolutionary techniques or swarm intelligence technique to give better results. When the publications are assembled.

6. REFERENCES

[1] Yong Rui, Thomas S. Huang, Michael Ortega and Sharad Mehrotra, "Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval", IEEE Transactions on Circuits and Video Technology, Vol. 8, No. 5, pp. 644-655, 1998

[2] Mishra Anil Kumar, Das Madhabananda and Panda T. C., "A Hybrid Swarm Intelligence Optimization for Benchmark Models by Blending PSO with ABC", International Review on Modelling & Simulations, Vol. 6, No., pp. 291, 2013.

[3] Anil Kumar Mishra, Madhabananda Das and T. C. Panda, "Hybrid Swarm Intelligence Technique for CBIR Systems" IJCSI International Journal of Computer Science Issues, Vol. 10, No 2, pp. 6-11, March 2013.

[4] Anil Kumar Mishra, Artificial Bee Colony Based Swarm Optimization Technique for Content-Based Image Retrieval System, KIIT University, (2014), Bhubaneswar, Odisha.

[5] Sasisekharan, R., Seshadri, V., & Weiss, S. M. (1996). Data mining and forecasting in large-scale telecommunication networks. *IEEE Expert: Intelligent Systems and Their Applications , 11* (1), 37-43.

[6] Shao, L. S., & Fu, G. X. (2008). Disaster prediction of coal mine gas based on data mining. *Journal of Coal Science and Engineering , 14* (3), 458-463.

[7] Pei, J., Upadhyaya, S. J., Farooq, F., & Govidaraju, V. (2004). Data mining for intrusion detection: Techniques, applications, and systems. *Proceedings of the 20th International Conference on Data Engineering.* IEEE Computer Society.

[8] Wong, S., & Li, C. S. (Eds.). (2006). *Life Science Data Mining.* World Scientific.

[9] *SAS Institute White Paper*. (2001). Retrieved from Data Mining in the Insurance Industry: http:// www.sas.com/products/miner/index.html.

[10] Gutierrez, F. J., Lerma-Rascon, M. M., Salgado-Garza, L. R., & Cantu, F. J. (2002). Biometrics and Data Mining: Comparision of data mining-based keystroke dynamics methods for identity verification. *Proceedings of the Second Mexican International Conference on Artificial Intelligence, Lecture Notes in Computer Science: Advances in Artificial Intelligence,. 2313*, pp. 460-469. London, UK: Springer-Verlag.

[11] Cano, J. R., Herrera, F., & Lozano, M. (2006). On the combination of evolutionary algorithms and stratified strategies for training set selection in data mining. *Applied Soft Computing , 6*, 323-332.

[12] Cano, J. R., Herrera, F., & Lozano, M. (2005). Stratification for scaling up evolutionary protoype selection. *Pattern Recognition Letters , 26*, 953-963.

[13] Coenen, F., & Leng, P. (2005). Obtaining best parameter values for accurate classification. *Proceedings of 5th IEEE Conference on Data Mining*, (pp. 549-552).

[14] Coenen, F., Leng, P., & Zhang, L. (2005). Threhold tuning for improved classification association rule mining. *Lecture Notes in Artificial Intelligence, Advances in Knowledge Discovery and Data Mining – PAKDD 2005, 3518*, pp. 216-225. Berlin.

[15] Bayardo Jr., R. J., & Agrawal, R. Mining the most interesting rules. *Proceedings of 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 145-153).

[16] de la Iglesia, B., Reynolds, A., & Rayward-Smith, V. J. (2005). Developments on a Multi-Objective Metaheuristic (MOMH) Algorithm for Finding Interesting Sets of Classification Rules. *Evolutionary Multi Criterion Optimization EMO 2005,Lecture Notes in Computer Science. 3410*, pp. 826-840. Berlin: springer.

[17] Ishibuchi, H., & Nojima, Y. (2005). Accuracy-Complexity Tradeoff Analysis by Multiobjective Rule Seleciton. *Workshop on Computational Intelligence in Data Mining* (pp. 39-48). ICDM.

[18] Bergh, F. V., & Engelbrecht, A. P. (2004). A Cooperative Approach to Particle Swarm Optimization. *IEEE Transaction on Evolutionary Computation , 8* (3), 225-239.

[19] Kennedy, J., & Eberhart, R. C. (1995). Particle Swarm Optimization. *Proceedings of the IEEE International Conference on Neural Networks*, (pp. 1942-1948). Perth, Australia.

[20] Coello, C. A., Pulido, T., & Lechuga, M. S. (2004). Handling Multiple Objectives with Particle Swarm Optimization. *IEEE Transactions on Evolutionary Computation , 8* (3), 256-279.

[21] Deb, K. (2001). *Multi-Objective Optimization Using Evolutionary Algorithm.* Chichester: John Willey & Sons.

[22] Ghosh, A., & Nath, B. T. (2004). Muli-Objective Rule Mining using Genetic Algorithms. *Information Sciences , 163*, 123-133.

[23] Kaya, M. (2006). Multi-Objective Genetic Algorithm based Approaches for Mining Optimized Fuzzy Association Rules. *Soft Computing , 10*, 578-586.