Journal of Theoretical and Applied Information Technology
20th March 2013. Vol. 49 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

RECORD-TO-RECORD TRAVEL ALGORITHM FOR ATTRIBUTE REDUCTION IN ROUGH SET THEORY

MAJDI MAFARJA[1,2], SALWANI ABDULLAH[1]

[1]Data Mining and Optimization Research Group (DMO), Center for Artificial Intelligence Technology, Universiti Kebangsaan Malaysia, 43600 UKM, Bangi Selangor, Malaysia.
[2]Department of Computer Science, Faculty of Information Technology
Birzeit University, P.O. Box 14, Birzeit, Palestine
E-mail: mmafarja@birzeit.edu , salwani@ftsm.ukm.my

ABSTRACT

Attribute reduction is the process of selecting a minimal attribute subset from a problem domain while retaining a suitably high accuracy in representing the original attributes. In this work, we propose a new attribute reduction algorithm called record-to-record travel (RRT) algorithm and employ a rough set theory as a mathematical tool to evaluate the quality of the obtained solutions. RRT is an optimization algorithm that is inspired from simulated annealing, which depends on a single parameter called DEVIATION. Experimental results on 13 well known UCI datasets show that the proposed method, coded as RRTAR, is comparable with other rough set-based attribute reduction methods available in the literature.

Keywords: *Rough Set Theory, Attribute Reduction, Record–To–Record Travel Algorithm*

1. .INTRODUCTION

Attribute reduction (AR) which is a NP-hard problem [1] can be defined as the process of finding the most predictive input attributes of a given outcome in many areas such as machine learning, data mining, pattern recognition and signal processing [2]. In these areas, since a huge number of attributes are often involved in datasets, attribute reduction becomes a necessary stage [3-4]. Attribute reduction is concerned in finding a minimum number of attributes N (subset) from the original set with M attributes that is $N < M$.

Attribute reduction is important in improving the performance of the learning algorithms by reducing the problem size and resulting search space by removing the redundant and irrelevant attributes. An attribute is said to be relevant if a decision is depending on it, otherwise it is irrelevant. Whilst, an attribute can be considered as redundant if it is highly correlated with other attributes.

Rough Set Theory [5-7] has been used to find a minimum reducts by generating all possible reducts, and select the one with the lowest cardinality and high dependency. This process is a simple mechanism but it is a time consuming procedure and it is only practical for simple datasets. Given a feature set with N features, the task of attribute reduction can be seen as a search for an optimal feature subset through the competing 2^N candidate subsets. Therefore, heuristic approaches have been considered since it only searches a particular path and find the (minimal) near-optimal subset [3]. Meta-heuristics have been widely used with high dimensional datasets to find better solutions for attribute reduction problems, instead of using the reduction method in the rough set theory.

Meta-heuristic approaches aim to find an acceptable solution within a reasonable computational time. Stochastic methods have been employed to handle attribute reduction problems in rough set theory [8]. For example, Jensen and Shen [2, 9] have studied meta-heuristic approaches to solve the attribute reduction problems. In their works, three methods have been presented i.e. the genetic algorithm (GenRSAR), the ant colony-based method (AntRSAR), and the simulated annealing algorithm (SimRSAR). Hedar et al. [10] considered a memory-based heuristic of tabu search to solve the attribute reduction problem in rough set theory. A great deluge algorithm for attribute reduction was presented by Abdullah and Jaddi [11], followed by Jihad and Abdullah [12] proposal of the composite neighbourhood structure, and Arajy and Abdullah [13] presentation of a hybrid variable neighbourhood search algorithm for the same problem. For the first time, a constructive hyper-heuristics was employed in attribute reduction problems by Abdullah et al. [14]. Ant

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

colony-based approaches were proposed by [15-19]. The firefly algorithm [20], artificial bee colony [21], bee colony optimisation [22], scatter search (SSAR) [23-24] and particle swarm optimisation (PSO) [25-28] were also proposed. Further reading about attribute reduction problems can be found in [8, 29-33].

In this work we proposed a new attribute reduction mechanism that investigates the Record-to-Record Travel algorithm (RRT) for attribute reduction problems in rough set theory. RRT is a single solution-based meta-heuristic algorithm that is originally proposed by Dueck [34]. The details of RRT are discussed in Section 3.

The rest of this paper is organized as follows: Section 2 contains a brief introduction on the rough set theory. The proposed approach is explained and discussed in Section 3, followed by a detailed implementation in Section 4. Numerical results on well-known datasets are reported in Section 5. Finally, a brief conclusion and further scope of the work are stated Section 6.

2. ROUGH SET THEORY

Rough set theory is a mathematical approach to analyse the vagueness and uncertainty in data. The main advantage of using rough set theory for attribute reduction is that the rough set does not require preliminary or additional information about the data. The starting point of the rough set theory is the concept of indiscernibility. A rough set is the approximation of an ambiguous concept [5] (set) by a pair of precise concept [35] known as the lower and upper approximations.

An example dataset, as presented in

Table, shows a two dimensional array; the columns of which are labeled by attributes, rows by the objects of interest and entries of the table are the attribute values. Here, the table consists of four conditional attributes (*a*, *b*, *c*, *d*) and one decision attribute (*e*) and eight objects. The task of attribute reduction is to find the minimal reduct from the conditional attributes so that the resulting reduced dataset remains consistent with respect to the decision attribute.

Table 1: An Example Dataset.

$X \in U$	*A*	*b*	*c*	*d*	$\Rightarrow e$
u0	1	0	2	2	0
u1	0	1	1	1	2
u2	2	0	0	1	1
u3	1	1	0	2	2
u4	1	0	2	0	1
u5	2	2	0	1	1
u6	2	1	1	1	2
u7	0	1	1	0	1

Let an information system be $I=(U, A)$ where U and A are non-empty sets of a finite objects and attributes respectively such that $a: U \rightarrow Va$ for every attribute $a \in A$. Va represents the value of an attribute a. Any subset P of A determines a binary relation *IND (P)* on U, which will be called an indiscernibility relation, and is defined as follows:

$$IND(P) = \{(x, y) \in U^2 | \forall a \in B, a(x) = a(y)\} \qquad (1)$$

The partitioning of U, generated by *IND(P)* will be denoted by *U/IND(P)*, or simple *U/P* and can be calculated as follows:

$$U / IND(P) = \otimes\{a \in P : U / IND(\{a\})\} \qquad (2)$$

The indiscernibility relation will be used next to define the approximations which is the basic concepts of the rough set theory.

Let $X \subseteq U$, is the approximations; $\underline{P}(X)$ and $\overline{P}(X)$ called the P-lower and the P-upper approximation of X respectively. They can be defined as follows:

$$\underline{P}(X) = \{x \in U : P(x) \subseteq X\} \qquad (3)$$

$$\overline{P}(X) = \{x \in U : P(x) \cap X \neq \emptyset\} \qquad (4)$$

To depict above definitions by an example that refers to

Table, if $P = \{b, c\}$, then objects **u1**, **u6** and **u7** are indiscernible as objects **u0** and **u4**. *IND(P)* creates the following partition of U:

$U / IND (P) = U / IND (b) \otimes U / IND (c)$

$=\{\{u0,u2,u4\},\{u1,u3,u6,u7\},\{u5\}\}\otimes\{\{u2,u3,u5\},$

$\{u1,u6,u7\},\{u0,u4\}\}$

$= \{\{u2\},\{u0,u4\},\{u3\},\{u1,u6,u7\},\{u5\}\}$

Journal of Theoretical and Applied Information Technology
20th March 2013. Vol. 49 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Let C and D be an equivalence relation over U, then the positive region can be defined as follows:

$$POS_C(D) = \bigcup_{X \in U/D} \underline{P}\, X \tag{5}$$

The positive region of the partition U/D with respect to P contains all objects of U, that can be uniquely classified to the blocks of the partition U/D, using the knowledge in attributes P. For example: let $P = \{b, c\}$ and $D = \{e\}$, then:

$$POS_C(D) = \bigcup \{\varnothing, \{u2, u5\}, \{u3\}\} = \{u2, u3, u5\}$$

It can be easily shown that when considering attributes b and c, the objects **u1**, **u3** and **u5** can certainly be classified as belong to a class in attribute e.

One of the major issues in the rough set theory is to measure the degree of dependency between attributes. Intuitively, a set of attributes D depends totally on a set of attributes P, denoted as $P \Rightarrow D$, if all values of attributes from D are uniquely determined by values of attributes from P. If there exists a functional dependency between the values of D and P, then D depends totally on P. Dependency can be defined as follows:

for D, $P \subset A$, it is said that D depends on a degree of k ($0 \le k \le 1$) denoted by $P \Rightarrow k$ if:

$$k = \gamma_P(D) = \frac{|POS_P(D)|}{|U|} \tag{6}$$

where $|U|$ denotes the cardinality of set U.

If $k = 1$, we can say that D depends totally on P, whereas if $k < 1$, we can say that D depends partially on P, and if $k = 0$, we will say that D does not depend on P. In the example dataset in Table 1, let $P = \{b, c\}$ and $D = \{e\}$, then the degree of dependency is:

$$\gamma_{\{b,c\}}(\{e\}) = \frac{|POS_{\{b,c\}}(\{e\})|}{|U|}$$

$$= \frac{|\{u2, u3, u5\}|}{|\{u0, u1, u2, u3, u4, u5, u6, u7\}|} = \frac{3}{8}$$

Finding the minimal reducts can be achieved by comparing the degrees of dependency of the generated subsets, where the reduced set has the same degree of dependency of the original set. A reduct is formally defined as a subset R of minimal cardinality of the conditional attribute set C such that $\gamma_R(D) = \gamma_C(D)$ where D is a decision system.

$$R = \{X : X \subseteq C, \gamma_X(D) = \gamma_C(D)\} \tag{7}$$

$$R_{\min} = \{X : X \in R, \forall Y \in R, |X| \le |Y|\} \tag{8}$$

The intersection of all reduced subsets is called the core which contains all those attributes that cannot be removed from the data set without introducing further contradictions.

$$Core(R) = \bigcap_{X \in R} X$$

Using the example shown in Table I, the minimal reduct sets of C are:

$$R = \{\{a, b, c\}, \{a, c, d\}\{b, c, d\}, \{b, d\}, \{c, d\}\}$$

From these sets, the minimal reduct is:

$$R_{\min} = \{\{b, d\}, \{c, d\}\}$$

It is obvious that finding all the possible reducts is a time consuming process, and this is applicable only with small datasets. Calculating all the reducts in aiming to find only the minimal one, but discovering others is pointless. To improve the performance of the above method an alternative strategy is required for large datasets.

3. RECORD-TO-RECORD TRAVEL ALGORITHM

Record-to-Record Travel algorithm (RRT) which is a local search algorithm was originally proposed by Dueck [34]. It differs from simulated annealing algorithm in the mechanism of accepting non-improving solutions. It has the advantage that it depends only on one parameter which is the value of the RECORD-DEVIATION [34]. The algorithm improves an initial solution by searching its neighborhood (generated by randomly flip flop one cell (0 → 1, 1 → 0)) for better solutions based on their evaluation (in this work, it is the degree of dependency). The solution is accepted if its objective value is greater than the RECORD minus the deviation D (RECORD – D). The initial value of the RECORD is equal to the initial objective function. During the search process, the RECORD value is updated with an objective value of the best solution so far. More formally, in the case of maximization, if (Sol_{best}) is the best solution so far and (Sol^*) is the new generated solution, (Sol^*) is accepted if f (Sol^*) is greater than f (Sol_{best}) or lower by a fixed deviation (D). The process is repeated until the stopping condition is satisfied.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Note that in this work, the stopping condition is set as a number of iterations.

Figure. 1 represents the pseudo code of the Record-to-Record Travel algorithm for attribute reduction (RRTAR), which is employed in this work.

3.1. Solution construction and representation

In this work, the initial solution is represented by a one dimensional vector with the dimension equaled to the number of attributes |N| in the original datasets. The initial solution is generated by randomly assigning "1" or '0" to the vector cells, where the cell with the value of one shows that the attribute is selected; otherwise the cell with the value of zero means that the attribute is discarded.

4. EXPERIMENTAL RESULTS

The proposed algorithm was programmed using Java and performed on Intel Pentium 4, 2.33 GHz computer and tested on 13 well-known UCI datasets [2, 9] as shown in Table. For every dataset, the algorithm was executed 20 times.

The results of our approach and the results from the state-of-the art methods are reported in Table 3 and Table 4. The entries in these tables represent the number of attributes in the minimal reducts obtained by each method. The superscripts in parentheses represent the number of runs that achieved the minimal reducts. The number of attribute without superscripts means that the method could obtain this number of attribute for all runs.

```
Record-to-Record Travel Algorithm
Generate a random initial solution Sol
Set Sol_best = Sol
Set Record = f(Sol_best)
while (stopping-criterion is not satisfied)
    generate at random a new solution Sol_trial in the
    neighbour of Sol
    Calculate f(Sol_trial)
    if ( f(Sol_trial)> f(Sol_best))
        Sol ← Sol_trial ; Sol_best ← Sol_trial
        f(Sol) = f(Sol_trial); f(Sol_best) = f(Sol_trial)
    else
        if (f(Sol_trial) == f(Sol_best))
            Calculate | Sol_trial |;
            Calculate | Sol_best |;
        If (|Sol_trial | < |Sol_best|)
            Sol ← Sol_trial ; Sol_best ← Sol_trial
            f(Sol) = f(Sol_trial); f(Sol_best) = f(Sol_trial)
        else
            if ( f(Sol_trial)> Record-D)
            Sol ← Sol_trial ; f(Sol) ← f(Sol_trial)
        end if
    end if

    if (f(Sol_trial)> Record)
        Record = f(Sol_trial);
    end if
Iteration+ +
end while
Calculate cardinality of best solution, | Sol_best |;
Return Best Solution Found
```

Figure 1: The Pseudo Code For RRTAR.

Table 2: List Of The UCI Datasets.

Datasets	No of Attributes	No. of Objects
M-of-N	13	1000
Exactly	13	1000
Exactly2	13	1000
Heart	13	294
Vote	16	300
Credit	20	1000
Mushroom	22	8124
LED	24	2000
Letters	25	26
Derm	34	366
Derm2	34	358
WQ	38	521
Lung	56	32

RRTAR is compared with other rough set attribute reduction methods i.e. Tabu Search (TSAR) by Hedar et al. [10], Ant Colony Optimization (AntRSAR) by Jensen and Shen [2, 9], Genetic Algorithm (GenRSAR) by Jensen and Shen [2, 9], Simulated Annealing (SimRSAR) by Jensen and Shen [2], Ant Colony Optimization (ACOAR) by Liangjun Ke et al. [16], Scatter Search (SSAR) by Wang et al. [23], Great Deluge algorithm (GD-RSAR) by Abdullah and Jaddi [11], Composite Neighbourhood Structure for Attribute Reduction (IS-CNS) by Jihad and Abdullah [12], Hybrid variable neighbourhood search algorithm (HVNS-AR) by Arajy and Abdullah [13], and a Constructive Hyper-Heuristics (CHH_RSAR) by Abdullah et. al [14].

From the results, it can be seen that RRTAR is comparable with the other approaches since it performs better than some approaches in some datasets. It is better than AntRSAR on one datasets (Credit); and better than SSAR on two datasets (ties on three datasets) and also better than GenRSAR in

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

all datasets. RRTAR outperforms TSAR in two datasets. Our method outperforms IS-CNS, HVNS-AR, CHH_RSAR on 1, 1, and 3 instances, respectively.

Here, we are interested in comparing our approach with GD-RSAR and SimRSAR. These two methods are selected because they have the same structure as RRT with the difference only in accepting the worst solutions. The results have shown that our approach is able to obtain better results on all datasets when compared with the GD-RSAR. RRTAR is also comparable with SimRSAR since it is able to obtain one result better than SimRSAR, and ties on 4 datasets.

This clearly proves the RRTAR is comparable with the other meta-heuristic approaches in solving the attribute reduction problem.

As stated in the earlier discussion, RRT shows promising performance when compared with other available methods. We believed that the strength of this method comes from the simplicity of the algorithm since it uses very little information about the structure of the problem, and the number of needed parameters. Hence these make it easier to control the performance of the algorithm. If we compare RRTAR with the other algorithms which have the same structure (i.e. GD-RSAR and SimRSAR); RRTAR needs to set only one parameter (D), while GD-RSAR and SimRSAR need setting for two and three parameters, respectively. RRT differs from SA in that it guarantees that a move which is much worse than the best solution found so far is never be accepted.

Table 3: *Comparisons Of RRTAR With The State-Of-Art Approaches 1.*

Datasets	RRTAR	GD-RSAR	TSAR	SimRSAR	AntRSAR	ACOAR
M-of-N	6	$6^{(10)} 7^{(10)}$	6	6	6	6
Exactly	6	$6^{(7)} 7^{(10)} 8^{(3)}$	6	6	6	6
Exactly2	10	$10^{(14)} 11^{(6)}$	10	10	10	10
Heart	$6^{(9)} 7^{(11)}$	$9^{(4)} 10^{(16)}$	6	$6^{(29)} 7^{(1)}$	$6^{(18)} 7^{(2)}$	6
Vote	$8^{(13)} 9^{(7)}$	$9^{(17)} 10^{(3)}$	8	$8^{(15)} 9^{(15)}$	8	8
Credit	$\mathit{8^{(18)} 9^{(2)}}$	$11^{(11)} 12^{(9)}$	$8^{(13)} 9^{(5)} 10^{(2)}$	$8^{(18)} 9^{(1)} 11^{(1)}$	$8^{(12)} 9^{(4)} 10^{(4)}$	$8^{(16)} 9^{(4)}$
Mushroom	$4^{(6)} 5^{(14)}$	$4^{(8)} 5^{(9)} 6^{(3)}$	$4^{(17)} 5^{(3)}$	4	4	4
LED	$5^{(18)} 6^{(2)}$	$8^{(14)} 9^{(6)}$	5	5	$5^{(12)} 6^{(4)} 7^{(3)}$	5
Letters	8	$8^{(7)} 9^{(13)}$	$8^{(17)} 9^{(3)}$	8	8	8
Derm	$7^{(1)} 8^{(16)} 9^{(3)}$	$12^{(14)} 13^{(6)}$	$6^{(14)} 7^{(6)}$	$6^{(12)} 7^{(8)}$	$6^{(17)} 7^{(3)}$	6
Derm2	$9^{(2)} 10^{(18)}$	$11^{(14)} 12^{(6)}$	$8^{(2)} 9^{(14)} 10^{(4)}$	$8^{(3)} 9^{(7)}$	$8^{(3)} 9^{(17)}$	$8^{(4)} 9^{(16)}$
WQ	$13^{(2)} 14^{(13)} 15^{(5)}$	$15^{(14)} 16^{(6)}$	$12^{(1)} 13^{(13)} 14^{(6)}$	$13^{(16)} 14^{(4)}$	$12^{(2)} 13^{(7)} 14^{(11)}$	$12^{(4)} 13^{(12)} 14^{(4)}$
Lung	$6^{(14)} 7^{(6)}$	$4^{(5)} 5^{(2)} 6^{(13)}$	$4^{(6)} 5^{(13)} 6^{(1)}$	$4^{(7)} 5^{(12)} 6^{(1)}$	4	4

Table 4: *Comparisons Of RRTAR With The State-Of-Art Approaches 2.*

Datasets	RRTAR	IS-CNS	HVNS-AR	GenRSAR	CHH_RSAR	SSAR
M-of-N	6	6	6	$6^{(6)} 7^{(12)}$	$6^{(11)} 7^{(9)}$	6
Exactly	6	6	6	$6^{(10)} 7^{(10)}$	$6^{(13)} 7^{(7)}$	6
Exactly2	10	10	10	$10^{(9)} 11^{(11)}$	10	10
Heart	$6^{(9)} 7^{(11)}$	6	6	$6^{(18)} 7^{(2)}$	6	6
Vote	$8^{(13)} 9^{(7)}$	8	8	$8^{(2)} 9^{(18)}$	8	8
Credit	$\mathit{8^{(18)} 9^{(2)}}$	$8^{(10)} 9^{(9)} \mathit{10^{(1)}}$	$8^{(7)} 9^{(6)} \mathit{10^{(7)}}$	$10^{(6)} 11^{(14)}$	$8^{(10)} 9^{(7)} \mathit{10^{(3)}}$	$8^{(9)} 9^{(8)} 10^{(3)}$
Mushroom	$4^{(6)} 5^{(14)}$	4	4	$5^{(1)} 6^{(5)} 7^{(14)}$	4	$4^{(12)} 5^{(8)}$
LED	$5^{(18)} 6^{(2)}$	5	5	$6^{(1)} 7^{(3)} 8^{(16)}$	5	5
Letters	8	8	8	$8^{(8)} 9^{(12)}$	8	$8^{(5)} 9^{(15)}$
Derm	$7^{(1)} 8^{(16)} 9^{(3)}$	$6^{(18)} 7^{(2)}$	$6^{(16)} 7^{(4)}$	$10^{(6)} 11^{(14)}$	6	6
Derm2	$9^{(2)} 10^{(18)}$	$8^{(4)} 9^{(16)}$	$8^{(5)} 9^{(12)} \mathit{10^{(3)}}$	$10^{(4)} 11^{(16)}$	$8^{(5)} 9^{(5)} \mathit{10^{(10)}}$	$8^{(2)} 9^{(18)}$
WQ	$13^{(2)} 14^{(13)} 15^{(5)}$	$12^{(2)} 13^{(8)} 14^{(10)}$	$12^{(3)} 13^{(6)} 14^{(8)} 15^{(3)}$	16	$12^{(13)} 14^{(7)}$	$13^{(4)} 14^{(16)}$
Lung	$6^{(14)} 7^{(6)}$	$4^{(17)} 5^{(3)}$	$4^{(16)} 5^{(4)}$	$6^{(8)} 7^{(12)}$	$4^{(10)} 5^{(7)} 6^{(3)}$	4

5. CONCLUSION AND FUTURE WORK

The Record-to-Record Travel algorithm for attribute reduction problems in rough set theory has been studied in this paper. The performance of the proposed algorithm is tested on standard benchmark datasets and comparison results have shown that our approach is able to produce good results and comparable with other approaches in the literature. Our future work will concentrate on incorporating the fuzzy logical principle in controlling the parameters in the algorithm, based on certain rules generated from an intelligent fuzzy membership function. This is subject to our future work.

REFERENCES

[1] A. Skowron and C. Rauszer, "The discernibility matrices and functions in information systems," *Intelligent decision support–Handbook of applications and advances of the rough set theory,* pp. 311–362, 1992.

[2] R. Jensen and Q. Shen, "Semantics-Preserving Dimensionality Reduction: Rough and Fuzzy-Rough-Based Approaches," *IEEE Trans. on Knowl. and Data Eng.,* vol. 16, pp. 1457-1471, 2004.

[3] H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*. Boston: Kluwer Academic Publishers 1998.

[4] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *The Journal of Machine Learning Research,* vol. 3, pp. 1157-1182, 2003.

[5] Z. Pawlak, "Rough Sets," *International Journal of Information and Computer Sciences,* vol. 11, pp. 341-356, 1982.

[6] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*: Kluwer Academic Publishers, 1992.

[7] Z. Pawlak, "Rough set approach to knowledge-based decision support," *European journal of operational research,* vol. 99, pp. 48-57, 1997.

[8] J. Bazan*, et al.*, "Rough set algorithms in classification problem," *Rough Set Methods and Applications. Physica Verlag,* pp. 49–88, 2000.

[9] R. Jensen and Q. Shen, "Finding Rough Set Reducts with Ant Colony Optimization," in *Proceedings of the 2003 UK Workshop on Computational Intelligence*, ed, 2003, pp. 15–22

[10] A.-R. Hedar*, et al.*, "Tabu search for attribute reduction in rough set theory," *Soft Comput.,* vol. 12, pp. 909-918, 2006.

[11] S. Abdullah and N. S. Jaddi, "Great Deluge Algorithm for Rough Set Attribute Reduction," in *Database Theory and Application, Bio-Science and Bio-Technology*. vol. 118, Y. Zhang*, et al.*, Eds., ed: Springer Berlin Heidelberg, 2010, pp. 189-197.

[12] S. K. Jihad and S. Abdullah, "Investigating composite neighbourhood structure for attribute reduction in rough set theory," presented at the Intelligent Systems Design and Applications (ISDA), 10th International Conference, 2010.

[13] Y. Z. Arajy and S. Abdullah, "Hybrid variable neighbourhood search algorithm for attribute reduction in Rough Set Theory," presented at the Intelligent Systems Design and Applications (ISDA), 2010.

[14] S. Abdullah*, et al.*, "A constructive hyper-heuristics for rough set attribute reduction," presented at the Intelligent Systems Design and Applications (ISDA), 2010.

[15] J. Wu*, et al.*, "An Approach to Feature Selection Based on Ant Colony Optimization and Rough Set," in *Intelligent Computing and Information Science*. vol. 134, R. Chen, Ed., ed: Springer Berlin Heidelberg, 2011, pp. 466-471.

[16] L. Ke*, et al.*, "An efficient ant colony optimization approach to attribute reduction in rough set theory," *Pattern Recognition Letters,* vol. 29, pp. 1351-1357, 2008.

[17] Y. Chen*, et al.*, "A rough set approach to feature selection based on ant colony optimization," *Pattern Recognition Letters,* vol. 31, pp. 226-233, 2010.

[18] H. Ming, "Feature Selection Based on Ant Colony Optimization and Rough Set Theory," presented at the Computer Science and Computational Technology, 2008. ISCSCT '08. International Symposium on, 2008.

[19] G. Wang*, et al.*, "Study of adaptive parameter control for ant colony optimization applied to feature selection problem.," *Advanced Science Letters.,* vol. in press, 2012.

[20] H. Banati and M. Bajaj, "Fire Fly Based Feature Selection Approach," *International Journal of Computer Science,* vol. 8, pp. 473-480, 2011.

[21] Y. Hu*, et al.*, "A Novel Discrete Artificial Bee Colony Algorithm for Rough Set-based

Feature Selection," *IJACT: International Journal of Advancements in Computing Technology*, vol. 4, pp. 295-305, 2012.

[22] N. Suguna and K. Thanushkodi, "A Novel Rough Set Reduct Algorithm for Medical Domain Based on Bee Colony Optimization," *Journal of Computing,* vol. 2, pp. 49-54, 2010.

[23] W. Jue*, et al.*, "Scatter Search for Rough Set Attribute Reduction," in *Bio-Inspired Computing: Theories and Applications, 2007. BIC-TA 2007. Second International Conference on*, 2007, pp. 236-240.

[24] J. Wang*, et al.*, "Rough set and scatter search metaheuristic based feature selection for credit scoring," *Expert Systems with Applications,* vol. 39, pp. 6123-6128, 2012.

[25] R. Bello*, et al.*, "Two-Step Particle Swarm Optimization to Solve the Feature Selection Problem," presented at the Proceedings of the Seventh International Conference on Intelligent Systems Design and Applications, 2007.

[26] X. Wang*, et al.*, "Feature selection based on rough sets and particle swarm optimization," *Pattern Recognition Letters,* vol. 28, pp. 459-471, 2007.

[27] C. Bae*, et al.*, "Feature selection with Intelligent Dynamic Swarm and Rough Set," *Expert Systems with Applications,* vol. 37, pp. 7026-7032, 2010.

[28] H. Fan and Y. Zhong, "A Rough Set Approach to Feature Selection Based on Wasp Swarm Optimization," *Journal of Computational Information Systems,* vol. 8, pp. 1037-1045, 2012.

[29] J. Dong*, et al.*, "Using Rough Sets with Heuristics for Feature Selection," presented at the Proceedings of the 7th International Workshop on New Directions in Rough Sets, Data Mining, and Granular-Soft Computing, 1999.

[30] R. Jensen and Q. Shen, *Computational Intelligence and Feature Selection: Rough and Fuzzy Approaches*: Wiley-IEEE Press, 2008.

[31] G. H. John*, et al.*, "Irrelevant Features and the Subset Selection Problem ", 1994.

[32] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence* vol. 97, pp. 273-324, 1997.

[33] D. Koller and M. Sahami, "Toward Optimal Feature Selection," presented at the International Conference on Machine Learning 1996.

[34] G. Dueck, "New Optimization Heuristics The Great Deluge Algorithm and the Record-to-Record Travel," *Journal of Computational Physics,*, vol. 104, pp. 86-92, 1993.

[35] R. Slowinski, "Rough set learning of preferential attitude in multi-criteria decision making," *Methodologies for Intelligent Systems,* pp. 642-651, 1993.