Journal of Theoretical and Applied Information Technology
10th March 2014. Vol. 61 No.1

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

WATER CYCLE ALGORITHM FOR ATTRIBUTE REDUCTION PROBLEMS IN ROUGH SET THEORY

[1]AHMED JABBAR, [2] SUHAILA ZAINUDIN

[1,2]Data Mining and Optimization Research Group (DMO), Center for Artificial Intelligence Technology, Universiti Kebangsaan Malaysia, 43600 UKM, Bangi Selangor, Malaysia.

E-mail: [1]ahmed.jabbar1986@gmail.com, [2]suhaila@ftsm.ukm.my

ABSTRACT

The attribute reduction is known as the procedure for decreasing the number of features in an information system and its action is a vital phase of data mining processing. In the attribute reduction process, the least subset of attributes is selected (according to rough set theory which is employed as a mathematical tool) from the initial set of attributes with very little loss in information. In this study, a new optimization approach, known as the water cycle algorithm (WCA), has been used for attribute reduction and the rough set theory is employed as a mathematical tool to assess quality of solutions that are produced. The idea of the WC as an optimization algorithm was derived from nature, after examining the whole water cycle process which involves the flow of streams and rivers into the sea in the natural world. The WC-RSAR has been employed in public datasets that are obtainable in UCI. From the findings of the experiments, it has been shown that the suggested method performs equally well or even better than other methods of attribute selection.

Keywords: *Attribute Reduction, Water Cycle Algorithm, Rough Set Theory.*

1. INTRODUCTION

Attribute reduction (AR) is considered as a NP-hard problem [1] and could be described as a process of discovering the most predictable input features of a given result in various fields as signal processing, data mining, pattern recognition and machine learning [2]. Attribute reduction is a requirement in these fields as datasets are often comprised of many attributes [3-4]. Attribute reduction has to do with discovering the least number of attributes, N (subset), from the initial set with M attributes such that N < M. Attribute reduction plays an important role in improving the performance of learning algorithms by diminishing the size of the problem and resultant search space by getting rid of the obsolete and inappropriate attributes. There for, the attribute is deemed to be relevant if an outcome hinges upon it. If not, it is considered to be irrelevant. Meanwhile, a redundant attribute is one that is highly interrelated with other attributes [5].

Pawlak's rough set theory [6-7] had been employed to determine the least reducts by detecting all the conceivable reducts and opting for the one that has the least cardinality and greatest dependency. Although this is considered a simple procedure, it takes a lot of time to be implemented; moreover, it is feasible for small datasets. As an alternative to the use of the reduction approach in rough set theory, a lot of meta-heuristic methods had been tried out on high dimensional datasets in search of the best answers of problems involving the attribute reduction [8].

The major objective of meta-heuristics lies in obtaining a satisfactory solution among a suitable computational period. There are two categories of meta-heuristics – population-based methods and single-based solution methods [9]. The simulated Annealing (SimRSAR) [2], the Tabu Search (TSAR) [11], the Great Deluge algorithm (GD-RSAR) [12], the Investigating Composite Neighbourhood Structure (IS-CNS) [13], the Constructive Hyper-Heuristics (CHH_RSAR) [14], and Hybrid Variable Neighbourhood Search algorithm (HVNS-AR) [15], Modified Great Deluge Algorithm (MGDAR) [16], Record-To-Record Travel Algorithm (RRTAR) [5], Nonlinear Great Deluge Algorithm (NLGDAR) [17]. Are examples of single-based methods, while the Ant Colony Optimisation (AntRSAR) [18], the Genetic Algorithm (GenRSAR) [2, 18], the Ant Colony Optimisation (ACOAR) [20], and the Scatter Search (SSAR) [21], are all population-based methods.

Journal of Theoretical and Applied Information Technology
10th March 2014. Vol. 61 No.1

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

A new approach for attribute reduction problems in rough set theory was put forward in this study. This method was first proposed by Eskandar et al. [22], and it is known as the water cycle algorithm (WCA). 13 typical benchmark datasets are taken from UCI which is available at http://www.ics.uci.edu/~mlearn and tested with the algorithm. The rough set theory was used in order to determine the minimum reduction.

The structure of the paper comes in four sections. Section 2, gives a short introduction on rough set theory. Section 3, we provide a thorough explanation of the application of the water cycle algorithm. Section 4 gives the results of the simulation, and the final section presents the conclusions of the paper.

2. THE ROUGH SET THEORY

The rough set theory [7] is considered an expansion of the standard set theory which favours using estimates in decision making. The characteristics of this theory are similar (in some ways) to the theory of evidence by Dempster-Shafer [23] and the theory of fuzzy set [24]. The rough set by itself is an estimation of an ambiguous concept (set) of two distinct concepts, known as lower and upper approximations, into which the interest domain is grouped into fragmented categories. While the lower approximation describes the objects in the domain which are known for sure as being from the subset of interest, the upper approximation describes those objects which might possibly be from the subset.

Rough Set Attribute Reduction (RSAR) [19] is considered as a filter for extricating knowledge from a domain in a concise manner by saving the subject matter of the information whilst cutting down on the quantity of knowledge that is included. A major benefit of rough set analysis is that it only needs the data that is provided in order for it to operate and no other matters need to be taken into consideration. Its functioning only depends on the data granularity configuration. Upon that lies the main difference between Rough Set Attribute Reduction and the Dempster-Shafer theory, which requires the probability, and the fuzzy set theory, which needs to have the membership values, in order to function.

Table 1 shows an example of a dataset with a two-dimensional array in which the columns have been labelled with attributes, the rows are labelled by the objects with interest, and the entries are comprised the attribute values in the table. In the following example, the table is composed of eight objects, four conditional attributes (*a, b, c, d*) and one decision attribute (*e*). It is the mission of the attribute reduction to locate the smallest reduct among all the conditional attributes in order to unchanged the decision attribute that is produced in the reduced dataset.

Table 1: Dataset Example.

X ∈ U	***A***	***b***	***c***	***d***	**⇒ *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

If I= *(U, A)* represent informational system, consider *U* and *A* are non-empty sets of finite objects and attributes respectively, such that $a: U \rightarrow Va$ for every attribute $a \in A$, with *Va* representing the value of an attribute *a*. Any subset *P* of *A* establishes a binary relation *IND (P)* on *U* that can be known as an *indiscernibility* relation, and can be defined as follows:

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

U/IND(P) or just *U/P* will denote the partitioning of *U*, produced by $IND(P)$ and it can be computed as follows:

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

Next, the *indiscernibility* relation can be employed to identify the approximations, and this forms the main concept of the rough set theory.

Consider $X \subseteq U$ as the approximations with $\underline{P}(X)$ and $\overline{P}(X)$, being the *P*-lower and the *P*-upper approximations of *X* respectively, which could be defined as:

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

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

Journal of Theoretical and Applied Information Technology
10th March 2014. Vol. 61 No.1

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

In order to illustrate the above definitions by an example with reference to the table, consider $P = \{b, c\}$, then the objects **u1**, **u6** and **u7** are indiscernible as the objects **u0** and **u4**. $IND(P))$ founds the next partition of *U*:

$U / IND(P) = U / IND(b) \otimes U / IND(c)$
$=\{\{u0,u2,u4\},\{u1,u3,u6,u7\},\{u5\}\}\otimes\{\{u2,u3,u5\},\{u1,u6,u7\},\{u0,u4\}\}$
$=\{\{u2\},\{u0,u4\},\{u3\},\{u1,u6,u7\},\{u5\}\}$

If *C* and *D* are considered as an equivalence relation over *U*, next the positive region could be defined as:

$$POS_C(D) = \cup_{X \in U/D} \underline{P} X \quad (5)$$

The positive region of the partition *U*/*D* with regard to *P* comprises all objects of *U* that could be exclusively grouped in the blocks of the partition *U*/*D by* using knowledge in attributes *P*. For example, if $P = \{b, c\}$ and $D = \{e\}$, then:

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

When taking into consideration attributes *b* and *c*, it can be clearly shown that the objects **u1**, **u3** and **u5** could be definitely categorized into attribute *e*.

The measurement of the degree of dependency between attributes is considered as one of the key issues in the theory of rough set. A set of attributes *D* is completely dependent automatically on *P* attributes set, denoted as $P \Rightarrow D$, that if all the values of attributes from *D* are determined by the values of attributes from *P*. If the values of *D* and *P* are functionally dependent, then it means that *D* is totally dependent on *P*. Dependency could be defined as:

Regarding $D, P \subset A$, it indicates that *D* depends to an extent on *k* ($0 \le k \le 1$) which is denoted by $P \Rightarrow k$ if:

$$K = \gamma_P(D) = \frac{POS_P(D)}{|U|} \quad (6)$$

Where $|U|$ indicates the cardinality of set *U*. Let $k = 1$, then it can be said that *D* is wholly dependent on *P*. However, if $k < 1$, then it can be said that *D* is partially dependent on *P*, but if $k = 0$, it can then be said that *D* is independent of *P*. In the dataset example given in Table 1, if $P = \{b, c\}$ and $D = \{e\}$, then the degree of dependency is:

$$\gamma_{\{b,c\}}(\{e\}) = \frac{|POS_{\{b,c\}}(\{e\})|}{|U|} = \frac{|\{u2, u3, u5\}|}{|\{u0, u1, u2, u3, u4, u5, u6, u7\}|} = \frac{3}{8}$$

The minimal reducts could be found with contrasting degrees of dependency of subsets that are produced, while the reduced set and the original set have the same degree of dependency. The formal definition for a reduct is that it is a subset, R, with a minimal cardinality of C, the conditional attribute set, such as $\gamma_R(D) = \gamma_C(D)$, where *D* is considered a decision system.

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

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

The core is the area in which all the reduced subsets intersect, and it comprises all those attributes which cannot be eliminated from the dataset without presenting more inconsistencies.

$$Core(R) = \cap_{X \in R} X \quad (9)$$

By employing the example in Table I, we find the minimal reduct sets as:

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

The minimal reduct from these sets is:

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

Clearly, a lot of time is wasted in calculating all the possible reducts as the aim is merely to locate the minimal reduct, and thus this process is only suitable for small datasets. An alternative approach needs to be found so as to increase the performance of the above method to enable it to be applied to large datasets.

3. WATER CYCLE ALGORITHM FOR ATTRIBUTE REDUCTION (WC-RSAR)

3.1. The solution representation and the initial solution generation

In the study, a solution has been given as a one-dimensional vector, as the length of the vector is considered according to the attributes number in the initial dataset. In addition, every value in the vector (cell) is denoted by either “1” or “0”. The value of

Journal of Theoretical and Applied Information Technology
10th March 2014. Vol. 61 No.1

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

"1" indicates that the related attribute is chosen; if not, then the value is set at "0". The subset of the solution is shown in Fig. 1 where 4 attributes have been chosen. i.e. {1, 3, 7, 10}

1	0	1	0	0	0	1	0	0	1	0

Fig. 1: Solution Representation

3.2. The quality measurement and acceptance conditions

The quality of the solution is measured according to the dependency degree, which is denoted as γ. There are two solutions which are the trial solution, *Sol** and the current solution, *Sol*. If the degree of dependency is enhanced such that γ (*Sol**) > γ (*Sol*), then the trial solution *Sol** will be selected but If the degree of dependency is the same for both solutions such that γ (Sol*) = γ (Sol), then the solution of the lesser number of attributes which is (denoted as #) could be selected

.

3.3. The water cycle algorithm

A. Basic concepts

Eskandar et al. put forward the water cycle algorithm (WCA) [22]. The inspiration for the idea of WCA was drawn from observing the nature and came from studying the water cycle and observing the way in which the streams and rivers flow downhill into the sea in natural world.

As water flows down from higher place to lower one, a river or a stream is created. As such, most rivers are formed at the top of mountains where the melting of snow occurs. In turn, the rivers constantly flow down and along this journey they are fed with water from rainfall and from other streams before they subsequently end up in the sea.

A simple diagram depicting part of this water cycle is given in Fig. 2. The water in lakes and rivers start to evaporate. Moreover, during the process of photosynthesis plants either give off or transpire water. Then, the water that is evaporated or transpired goes up into the atmosphere and leads to the formation of clouds that condense in the colder air above. Thus the water is circulated through precipitation and the formation of rain back to the earth again. This process is known as the hydrologic or water cycle [10].
In our natural world, most of the water that comes from the melting of snow or from rainfall seeps into the permeable layer of rock or soil underground and is stored there in large amounts. This aquifer is sometimes referred to as groundwater for more clarification (see percolation arrow in Fig. 2). That water in the aquifer flows in a downward direction underground in the same way that it flows on the surface of the ground. The underground water could be emptied into a lake, swamp or stream. More clouds are formed through the evaporation of water from streams and rivers, together with transpiration from trees and other vegetation, thus causing more rain to fall, and so the cycle goes on. [10].

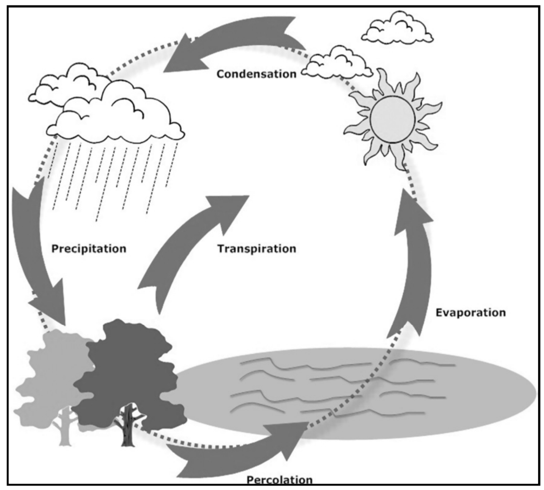

Fig. 2: A Simplified diagram of the hydrologic cycle (water cycle) by Eskandar et al. [22]

B. The proposed WCA

Like all other meta-heuristic algorithms, our suggested method starts with initial population, which can be compared to the raindrops. First, we begin with the assumption that rain or precipitation is available. A sea is selected as the best individual (best raindrop). A number of worth raindrops are selected to represent a river while the remainder of the raindrops are represented streams flowing into the sea and the rivers. Each river takes in water from the streams according to the force of their flow, which will be explained in the subsections that follow. Actually, the quantity of water entering a river and/or sea differs from one stream to another. Furthermore, the flow of the rivers into the sea is as it at the lowest location.

C. Create the initial population

When population-based meta-heuristic methods are employed to resolve an optimization problem, the problem variables values must be structured in form

Journal of Theoretical and Applied Information Technology
10th March 2014. Vol. 61 No.1

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

of an array. This array is named "Chromosome" and "Particle Position" in GA and PSO terminologies, respectively. Hence, in the suggested method, the array for a single solution is appropriately called a "raindrop". A raindrop is an array of $1 \times N_{var}$ in a N_{var}dimensional optimization problem, and then this array can be defined as:

$$Raindrop = [X_1, X_{2,}, X_{3,} \dots, X_N] \quad (10)$$

The raindrop cost could be determined by calculating the function of cost (C) as:

$$C_i = Cost_i = \int (X_1^i, X_2^i, \dots, X_{N_{var}}^i) \quad i = 1,2,3, \dots, N_{pop} \quad (11)$$

Where N_{pop} and N_{vars} are represented the number of raindrops (initial population) and design variables. First, N_{pop} raindrops are created. A number of N_{sr} are chosen as the sea and rivers from the best individuals (minimum values). The raindrop with the least value among the rest is taken as a sea. Actually, N_{sr} represents the total Number of Rivers (user parameter) for a single sea as shown in Eq. (12). The remainder of the population (raindrops that compose the streams that flow down directly into the sea or into the rivers) is determined by using Eq. (13).

$$N_{sr} = Number\ of\ Rivers + \underbrace{1} \quad (12)$$

$$N_{Raindrops} = N_{pop} - N_{sr} \quad (13)$$

The following equation is used to assign raindrops into the sea or the rivers regarding the strength of the flow:

$$NS_n = round\left\{\left|\frac{Cost_n}{\sum_{i=1}^{N_{sr}} Cost_i}\right| \times N_{Raindrops}\right\}, \quad n = 1,2, \dots, N_{sr} \quad (14)$$

N_{sr} , representing a number of streams that flow into certain sea or rivers.

The raindrops together create the streams which are linked with each other to generate new rivers, where some of the streams may directly flow into the sea. All the streams and rivers ultimately end in the sea (best optimal point). The flow of a stream in the direction of a particular river is shown in Fig. 3.

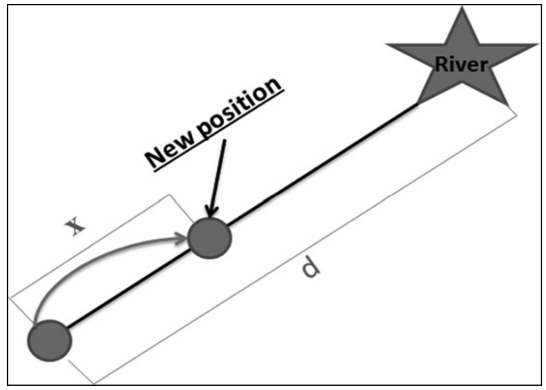

Fig. 3: Schematic View Of Flow Of A Stream To A Particular River (The River And Stream Are Represented By The Star And Circle, Respectively) By Eskandar et al. [22].

This idea can also be applied on rivers that flow into the sea so the new position for the rivers and streams can be given as:

$$X_{Stream}^{i+} = X_{Stream}^i + rand \times C \times (X_{River}^i - X_{Stream}^i) \quad (15)$$

$$X_{River}^{i+} = X_{River}^i + rand \times C \times (X_{Sea}^i - X_{River}^i) \quad (16)$$

where C is represented a value between 1 and 2 (nearer to 2), the best selected value for C is 2.As rand stands for a uniformly distributed random number between 0 and 1, If the solution which is given by a stream is better than its connecting river then the positions of the stream and the river can be exchanged (i.e. the stream becomes the river and vice versa). Similarly, like this exchange may also occur in the position of the sea and the rivers. Fig. 4 illustrates exchange occurs in a stream that is considered the best solution within the other streams and the river.

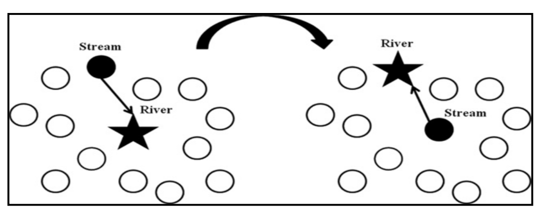

Fig. 4: Exchange In Positions Of The River And The Stream By Eskandar Et Al. [22].

Evaporation is a process where dmax represents small number (closer to zero). If the distance between the sea and the river is less than dmax, it signifies that the river arrived at or linked with the sea. The evaporation process is taken into

Journal of Theoretical and Applied Information Technology
10th March 2014. Vol. 61 No.1

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

consideration in this situation and as can be observed in nature, after ample evaporation has taken place, it will begin to rain or precipitation will occur. A large dmax value will lower the search but a small value will encourage an intensification of the search close to the sea. As such, the intensity of the search close to the sea (the optimum solution) is controlled by the dmax. The value of the dmax adapts accordingly and decreases as:

$$d_{max}^{i+1} = d_{max}^{i} - \frac{d_{max}^{i}}{\max iteration} \quad (17)$$

On completion of the evaporation, the rain process is employed. The raining process involves the formation of streams in various locations by the new raindrops. The following equation is used to specify new locations of the freshly new forming streams:

$$X_{Stream}^{new} = LB + rand \times (UB - LB) \quad (18)$$

Where UB and LB are the upper and lower bounds respectively as identified from the given problem.

Eq. (19) is only used for those streams which flow directly into the sea in order to improve the computational performance of the algorithm and the convergence rate of the controlled problems. The objective of this equation is to foster the creation of the streams that flow straight into the sea in order to increase the search near the sea (the optimum solution) of the feasible area for the controlled problems.

$$X_{Stream}^{new} = X_{sea} + \sqrt{\mu} \times randn(1, N_{var}) \quad (19)$$

Where μ is a coefficient that indicates the range of the search area close to the sea and randn is the normally distributed random number. While the larger value for μ raises the possibility of exiting in the feasible area, the smaller value for μ steers the algorithm to search in a narrow area close to the sea. The suitable value to set for μ is 0.1. From a mathematical perspective, the standard deviation is represented by the term $\sqrt{\mu}$ in Eq. (19) and thus, the concept of variance is accordingly defined as l. By employing these concepts, the individuals that are generated with variance μ are dispersed approximate to the best optimum point which is the (sea) that has been obtained.

D.The flowchart and the Steps of WCA

WCA steps can be summarized as follows:

Step 1: Selecting the WCA initial parameters: $N_{sr}, d_{max}, N_{pop}, max_$iteration.
Step 2: Generating the random initial population and forming the sea, rivers and initial streams (raindrops) by using Equations (12) and (13).
Step 3: Calculating the value (cost) of each raindrop by using Eq. (11).
Step 4: Determining the intensity of the flow for the sea and rivers by using Eq. (14).
Step 5: flowing of the streams into the rivers by using Eq. (15).
Step 6: flowing of the rivers into the sea (the most downhill location) by using Eq. (16).
Step 7: Exchanging the position of the stream with the river in order to obtain the best solution, as illustrated in Fig. 4.
Step 8: like Step 7, whether the river could find a better solution than the sea, exchanging the position of the sea with that of the river as shown in Fig 4.
Step 9: Checking if the conditions of the evaporation are satisfied.
Step 10: checking if the conditions of the evaporation are satisfied, the rain process will occur by using Equations (18) and (19).
Step 11: Reducing the value of d_{max}, which is considered a defined user parameter by using Eq. (17).
Step 12: Checking the criteria of convergence so if the stopping criteria is met, the algorithm will stop, and otherwise it will return to Step 5.

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

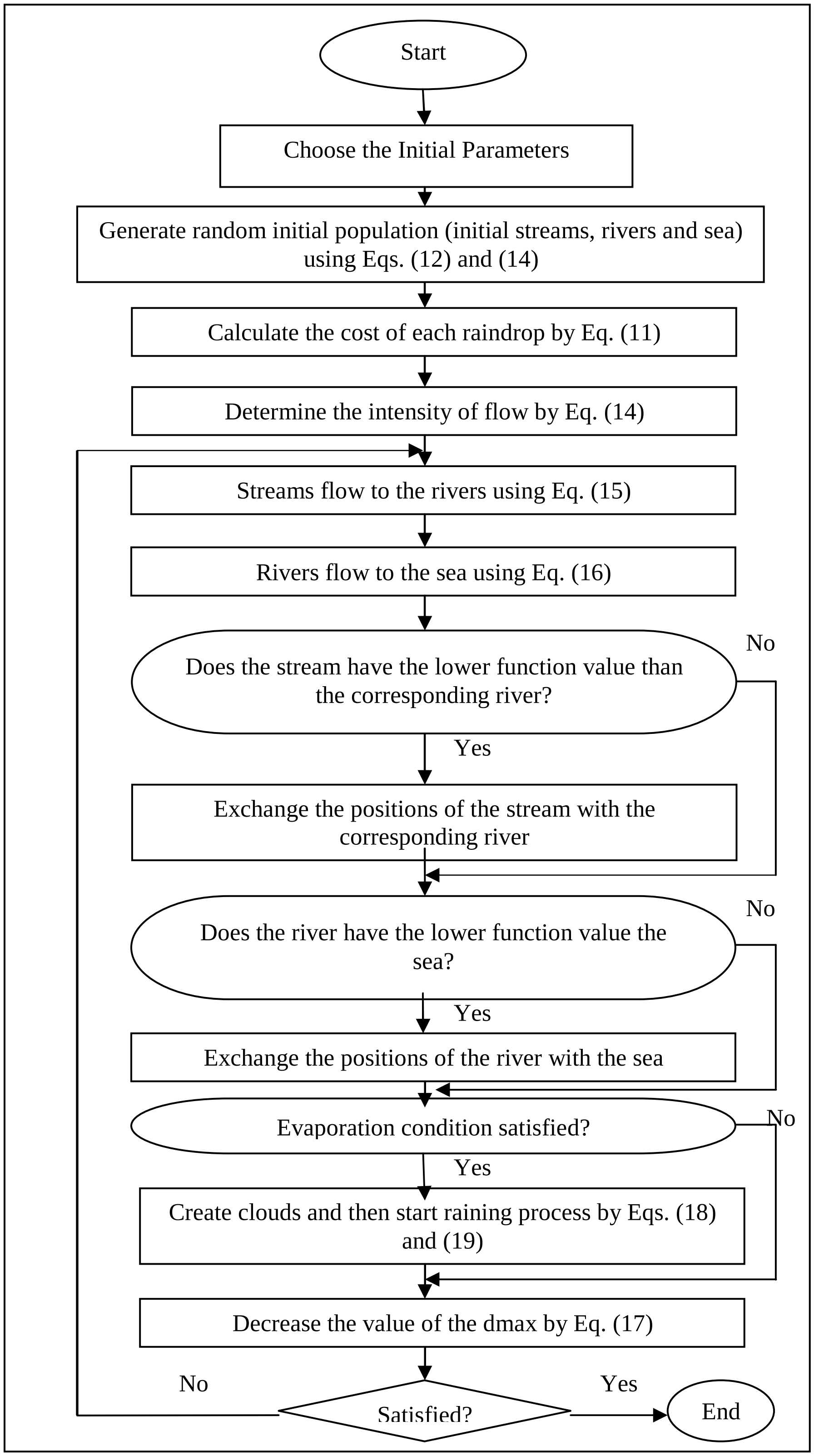

Fig. 5: The flow chart of WCA

Journal of Theoretical and Applied Information Technology
10th March 2014. Vol. 61 No.1

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

4. EXPERIMENTAL RESULTS and DISCUSSION

MATLAB was used to program the proposed algorithm and it was performed on a Core i3 processor, a 1.5 GHz computer and tested on 13 well-known UCI datasets [2, 18] as illustrated in table 2. The algorithm had 20 runs with different initial populations for each dataset as proposed by [2]. The algorithm parameters are pop-size: 20, max-iterations: 25.

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

WCRSAR is contrasted with other attribute reduction methods which have been investigated. The best reduct that is obtained out of 20 runs for each method is recorded, and the number of runs which achieved this reduct has been stated in parentheses. Where a number appears without a superscript it indicates that this method managed to obtain the number of attributes for all the runs.

The contrasted methods are categorised into single-based solution and population-based solution methods. Table 3 and 4, gives a comparison of the results of this study with single-based solution methods (e.g. Simulated Annealing (SimRSAR) [2], Tabu Search (TSAR) [11], Great Deluge algorithm (GD-RSAR) [12], Investigating Composite Neighbourhood Structure (IS-CNS) [13], a Constructive Hyper-Heuristics (CHH_RSAR) [14], Hybrid Variable Neighbourhood Search algorithm (HVNS-AR) [15], Modified Great Deluge Algorithm (MGDAR) [16], Record-To-Record Travel Algorithm (RRTAR) [5], Nonlinear Great Deluge Algorithm (NLGDAR) [17]). Table 5 gives a comparison between the results of this study with population-based solutions methods (e.g. Ant Colony Optimisation (AntRSAR) [18], Genetic Algorithm (GenRSAR) [2, 18], Ant Colony Optimisation (ACOAR) [20], Scatter Search (SSAR) [21]).

Journal of Theoretical and Applied Information Technology
10th March 2014. Vol. 61 No.1

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

Table 3: Comparisons Of WCRSAR With Single-Based Solutions Methods

Datasets	WCRSAR	SimRSAR	TSAR	GD-RSAR	IS-CNS
M-of-N	6	6	6	$6^{(10)}\ 7^{(10)}$	6
Exactly	6	6	6	$6^{(7)}\ 7^{(10)}\ 8^{(3)}$	6
Exactly2	10	10	10	$10^{(14)}\ 11^{(6)}$	10
Heart	$\mathbf{4^{(3)}\ 5^{(17)}}$	$6^{(29)}\ 7^{(1)}$	6	$9^{(4)}\ 10^{(16)}$	6
Vote	$\mathbf{7^{(8)}\ 8^{(12)}}$	$8^{(15)}\ 9^{(15)}$	8	$9^{(17)}\ 10^{(3)}$	8
Credit	**8**	$8^{(18)}\ 9^{(1)}\ 11^{(1)}$	$8^{(13)}\ 9^{(5)}\ 10^{(2)}$	$11^{(11)}\ 12^{(9)}$	$8^{(10)}\ 9^{(9)}\ 10^{(1)}$
Mushroom	$4^{(5)}\ 5^{(15)}$	4	$4^{(17)}\ 5^{(3)}$	$4^{(8)}\ 5^{(9)}\ 6^{(3)}$	4
LED	$5^{(5)}\ 6^{(15)}$	5	5	$8^{(14)}\ 9^{(6)}$	5
Letters	$8^{(13)}\ 9^{(7)}$	8	$8^{(17)}\ 9^{(3)}$	$8^{(7)}\ 9^{(13)}$	8
Derm	$7^{(15)}\ 8^{(5)}$	$6^{(12)}\ 7^{(8)}$	$6^{(14)}\ 7^{(6)}$	$12^{(14)}\ 13^{(6)}$	$6^{(18)}\ 7^{(2)}$
Derm2	$8^{(2)}\ 9^{(14)}\ 10^{(4)}$	$8^{(3)}\ 9^{(7)}$	$8^{(2)}\ 9^{(14)}\ 10^{(4)}$	$11^{(14)}\ 12^{(6)}$	$8^{(4)}\ 9^{(16)}$
WQ	$12^{(1)}\ 13^{(6)}\ 14^{(13)}$	$13^{(16)}\ 14^{(4)}$	$12^{(1)}\ 13^{(13)}\ 14^{(6)}$	$15^{(14)}\ 16^{(6)}$	$12^{(2)}\ 13^{(8)}\ 14^{(10)}$
Lung	$11^{(1)}\ 12^{(19)}$	$4^{(7)}\ 5^{(12)}\ 6^{(1)}$	$4^{(6)}\ 5^{(13)}\ 6^{(1)}$	$4^{(5)}\ 5^{(2)}\ 6^{(13)}$	$4^{(17)}\ 5^{(3)}$

Table 4: Comparisons Of WCRSAR With Singlel-Based Solutions Methods

Datasets	WCRSAR	CHH_RSAR	MGDAR	RRTAR	NLGD-RSAR
M-of-N	6	$6^{(11)}\ 7^{(9)}$	6	6	6
Exactly	6	$6^{(13)}\ 7^{(7)}$	6	6	6
Exactly2	10	10	10	10	10
Heart	$\mathbf{4^{(3)}\ 5^{(17)}}$	6	$6^{(14)}\ 7^{(6)}$	$6^{(9)}\ 7^{(11)}$	9
Vote	$\mathbf{7^{(8)}\ 8^{(12)}}$	8	8	$8^{(13)}\ 9^{(7)}$	$10^{(14)}11^{(6)}$
Credit	**8**	$8^{(10)}\ 9^{(7)}\ 10^{(3)}$	$8^{(13)}\ 9^{(3)}\ 10^{(4)}$	$8^{(18)}\ 9^{(2)}$	11
Mushroom	$4^{(5)}\ 5^{(15)}$	4	$4^{(7)}\ 5^{(13)}$	$4^{(6)}\ 5^{(14)}$	4
LED	$5^{(5)}\ 6^{(15)}$	5	5	$5^{(18)}\ 6^{(2)}$	$7^{(15)}8^{(5)}$
Letters	$8^{(13)}\ 9^{(7)}$	8	$8^{(18)}\ 9^{(2)}$	8	9
Derm	$7^{(15)}\ 8^{(5)}$	6	$6^{(11)}\ 7^{(9)}$	$7^{(1)}8^{(16)}9^{(3)}$	$11^{(17)}12^{(3)}$
Derm2	$8^{(2)}\ 9^{(14)}\ 10^{(4)}$	$8^{(5)}\ 9^{(5)}\ 10^{(10)}$	$8^{(4)}\ 9^{(12)}\ 10^{(4)}$	$9^{(2)}\ 10^{(18)}$	$11^{(15)}12^{(5)}$
WQ	$12^{(1)}\ 13^{(6)}\ 14^{(13)}$	$12^{(13)}\ 14^{(7)}$	$12^{(1)}\ 13^{(11)}\ 14^{(8)}$	$13^{(2)}\ 14^{(13)}15^{(5)}$	$15^{(11)}16^{(9)}$
Lung	$11^{(1)}\ 12^{(19)}$	$4^{(10)}\ 5^{(7)}\ 6^{(3)}$	$4^{(6)}\ 5^{(11)}\ 6^{(3)}$	$6^{(14)}7^{(6)}$	4

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

Table 5: Comparisons Of WCRSAR With Population-Based Solutions Methods

Datasets	WCRSAR	AntRSAR	GenRSAR	ACOAR	SSAR
M-of-N	6	6	$6^{(6)}\ 7^{(12)}$	6	6
Exactly	6	6	$6^{(10)}\ 7^{(10)}$	6	6
Exactly2	10	10	$10^{(9)}\ 11^{(11)}$	10	10
Heart	$\mathbf{4^{(3)}\ 5^{(17)}}$	$6^{(18)}\ 7^{(2)}$	$6^{(18)}\ 7^{(2)}$	6	6
Vote	$\mathbf{7^{(8)}\ 8^{(12)}}$	8	$8^{(2)}\ 9^{(18)}$	8	8
Credit	**8**	$8^{(12)}\ 9^{(4)}\ 10^{(4)}$	$10^{(6)}\ 11^{(14)}$	$8^{(16)}\ 9^{(4)}$	$8^{(9)}\ 9^{(8)}\ 10^{(3)}$
Mushroom	$4^{(5)}\ 5^{(15)}$	4	$5^{(1)}\ 6^{(5)}\ 7^{(14)}$	4	$4^{(12)}\ 5^{(8)}$
LED	$5^{(5)}\ 6^{(15)}$	$5^{(12)}\ 6^{(4)}\ 7^{(3)}$	$6^{(1)}\ 7^{(3)}\ 8^{(16)}$	5	5
Letters	$8^{(13)}\ 9^{(7)}$	8	$8^{(8)}\ 9^{(12)}$	8	$8^{(5)}\ 9^{(15)}$
Derm	$7^{(15)}\ 8^{(5)}$	$6^{(17)}\ 7^{(3)}$	$10^{(6)}\ 11^{(14)}$	6	6
Derm2	$8^{(2)}\ 9^{(14)}\ 10^{(4)}$	$8^{(3)}\ 9^{(17)}$	$10^{(4)}\ 11^{(16)}$	$8^{(4)}\ 9^{(16)}$	$8^{(2)}\ 9^{(18)}$
WQ	$12^{(1)}\ 13^{(6)}\ 14^{(13)}$	$12^{(2)}\ 13^{(7)}\ 14^{(11)}$	16	$12^{(4)}\ 13^{(12)}\ 14^{(4)}$	$13^{(4)}\ 14^{(16)}$
Lung	$11^{(1)}\ 12^{(19)}$	4	$6^{(8)}\ 7^{(12)}$	4	4

From the results given in Table 3 and Table 4, the proposed method is comparable with the available single based methods, it can be seen that WCARSAR produced better results than SimRSAR, TSAR, IS-CNS and MGDAR in 3 datasets, i.e. heart, vote and credit. However WCARSAR has outperforms GD-RSAR in 12 datasets, i.e. m-of-n, exactly, exactly2, heart, vote, credit, mushroom, led, letters, derm, derm2 and wq. WCRSAR outperforms CHH_RSAR in 5 datasets, i.e. m-of-n, exactly, heart, vote, and credit. WCRSAR outperforms RRTAR in 6 datasets i.e. heart, vote, credit, derm, derm2 and wq. WCARSAR obtained better results than NLGD-RSAR in 8 datasets, i.e. heart, vote, credit, led, letters, derm, derm2 and wq.

Table 5 shows the comparison between the proposed method and the available population-based approaches in the literature. From the results it can be seen that the proposed method are ambidextrous to produce better results than the available population-based approaches. It can be said that they are comparable with these approaches since they are able to obtain best results on some datasets. For example, WCARSAR is better than AntRSAR and ACOAR in 3 datasets, i.e. heart, vote and credit. However WCARSAR has outperforms GenRSAR in 12 datasets, i.e. m-of-n, exactly, exactly2, heart, vote, credit, mushroom, led, letters, derm, derm2 and wq. WCRSAR outperforms SSAR in 5 datasets i.e. heart, vote, credit, letters and wq.

5. CONCLUSION AND FUTURE WORK

This paper presented a study on the Water Cycle Algorithm (WCA) for attribute reduction problems in the theory of rough set. The viability of the suggested algorithm had been tested on regular benchmark datasets and from a comparison of the results it has been proven that this method is capable of producing good results and that it is as good as the other methods mentioned in previous studies. Future work should be directed at improving the performance and enhance the good of the WCA by producing a hybrid algorithm of the WCA with a hill climbing algorithm.

REFERENCE

[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.

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

[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] Mafarja, Majdi, and Salwani Abdullah. "Record-To-Record Travel Algorithm For Attribute Reduction In Rough Set Theory." Journal of Theoretical and Applied Information Technology 49.2 (2013).

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

[7] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data: Kluwer Academic Publishers, 1991.

[8] X. Wang, et al., "Feature selection based on rough sets and particle swarm optimization," Pattern Recogn. Lett, vol. 28, pp. 459-471, 2007.

[9] F. Glover and G. Kochenberger, Handbook of Metaheuristics (International Series in Operations Research & Management Science): Springer, 2003.

[10] David S. The water cycle, illustrations by John Yates. New York: Thomson Learning; 1993.

[11] Hedar, Abdel-Rahman, Jue Wang, and Masao Fukushima. "Tabu search for attribute reduction in rough set theory." Soft Computing 12.9 (2008): 909-918.

[12] Abdullah, Salwani, and Najmeh Sadat Jaddi. "Great Deluge Algorithm for Rough Set Attribute Reduction." Database Theory and Application, Bio-Science and Bio-Technology. Springer Berlin Heidelberg, 2010. 189-197.

[13] Jihad, SaifKifah, and S. Abdullah. "Investigating composite neighbourhood structure for attribute reduction in rough set theory." Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on. IEEE, 2010.

[14] Abdullah, Salwani, et al. "A constructive hyper-heuristics for rough set attribute reduction." Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on. IEEE, 2010.

[15] Arajy, Yahya Z., and Salwani Abdullah. "Hybrid variable neighbourhood search algorithm for attribute reduction in Rough Set Theory." Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on. IEEE, 2010.

[16] Mafarja, Majdi, and Salwani Abdullah. "Modified great deluge for attribute reduction in rough set theory." Fuzzy Systems and Knowledge Discovery (FSKD), 2011 Eighth International Conference on. Vol. 3. IEEE, 2011.

[17] Jaddi, Najmeh Sadat, and Salwani Abdullah. "Nonlinear Great Deluge Algorithm for Rough Set Attribute Reduction." Journal of Information Science and Engineering 29.1 (2013): 49-62.

[18] Jensen, Richard, and Qiang Shen. "Finding rough set reducts with ant colony optimization." Proceedings of the 2003 UK workshop on computational intelligence. Vol. 1. No. 2. 2003.

[19] A. Chouchoulas and Q. Shen. Rough set-aided keyword reduction for text categorisation. Applied Artificial Intelligence, 15(9):843-873, 2001.

[20] Ke, Liangjun, Zuren Feng, and Zhigang Ren. "An efficient ant colony optimization approach to attribute reduction in rough set theory." Pattern Recognition Letters 29.9 (2008): 1351-1357.

[21] Wang, Jue, et al. "Scatter search for rough set attribute reduction."Computational Sciences and Optimization, 2009. CSO 2009. International Joint Conference on. Vol. 1. IEEE, 2009.

[22] Eskandar, Hadi, et al. "Water cycle algorithm–A novel metaheuristic optimization method for solving constrained engineering optimization problems." Computers & Structures (2012).

[23] A. Skowron, J.W. Grzymala-Busse. From rough set theory to evidence theory, Advances in the Dempster-Shafer Theory of Evidence,ed. by R. Yager, M. Fedrizzi, and J. Kasprzyk, John Wiley & Sons, Inc. 1994.

[24] Z. Pawlak, A. Skowron. Rough membership functions. In R. Yager, M. Fedrizzi, J. Kacprzyk (eds.), Advances in the Dempster-Shafer Theory of Evidence, Wiley, New York, pp 251-271. 1994.