How to cite this article:

Basir, M. A., Yusof, Y., & Hussin, M. S. (2018). Optimization of attribute selection model using bio-inspired algorithms. *Journal of Information and Communication Technology, 18*(1), 35-55.

OPTIMIZATION OF ATTRIBUTE SELECTION MODEL USING BIO-INSPIRED ALGORITHMS

[1] Mohammad Aizat Basir, [2] Yuhanis Yusof & [1] Mohamed Saifullah Hussin

[1]School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu, Malaysia

[2]School of Computing, Universiti Utara Malaysia, Malaysia

aizat@umt.edu.my; yuhanis@uum.edu.my; saifullah@umt.edu.my

ABSTRACT

Attribute selection which is also known as feature selection is an essential process that is relevant to predictive analysis. To date, various feature selection algorithms have been introduced, nevertheless they all work independently. Hence, reducing the consistency of the accuracy rate. The aim of this paper is to investigate the use of bio-inspired search algorithms in producing optimal attribute set. This is achieved in two stages; 1) create attribute selection models by combining search method and feature selection algorithms, and 2) determine an optimized attribute set by employing bio-inspired algorithms. Classification performance of the produced attribute set is analyzed based on accuracy and number of selected attributes. Experimental results conducted on six (6) public real datasets reveal that the feature selection model with the implementation of bio-inspired search algorithm consistently performs good classification (i.e higher accuracy with fewer numbers of attributes) on the selected data set. Such a finding indicates that bio-inspired algorithms can contribute in identifying the few most important features to be used in data mining model construction.

Received: 15 July 2018 Accepted: 4 October 2018 Published: 11 December 2018

Keywords: Feature selection, bio-inspired algorithms, data classification, data mining.

INTRODUCTION

Real world data set usually consists of a large number of attributes. It is common that some of those attributes are irrelevant and consequently affects the data mining model. In situations where a rule has too many conditions, having large size of attributes, makes the rule becomes less interpretable. Based on this understanding, it is important to reduce the number of features to be used in data mining model construction. In practical situations, it is recommended to remove the irrelevant and redundant dimensions for less processing time and labor cost. Jensen and Shen (2003) stated that the data set with large number of attributes is known as a data set with high dimensionality. The high dimensionality data set leads to a phenomenon known as the curse of dimensionality where computation time is an exponential function of the number of the dimensions. There are also cases where model contains redundant rules and/or attributes. When faced with difficulties resulting from the high dimension of space, the ideal approach is to reduce this dimension, without losing relevant information in the data. If there are a large number of rules and/or attributes in each rules, it becomes more and more vague for the user to understand and difficult to exercise and utilize. Rule redundancy and/or attribute complexity can be overcome by reducing the number of attributes in a data set and removing irrelevant or less significant rules. This can reduce the computation time and storage space.

The main drawback of rule/attribute complexity reduction is the possibility of information loss. It is important to point out two critical aspects of attribute reduction problems, which are the degree of attribute optimality (in terms of subset size and corresponding dependency degree) and time required to achieve this attribute optimality. For example, existing methods such as Quick Reduct and Entropy-Based Reduction (EBR) methods which was created by Jensen and Shen (2001) performed reduction quickly but does not guarantee a minimal size of subset in many cases (Jensen & Shen, 2001; Suguna & Thanushkodi, 2010; Yue, Yao, Abraham, & Liu, 2007) whereas hybrid methods which combine Rough Set-Based Attribute Reduction (RSAR) and swarm algorithms such as GenRSAR proposed by Jensen and Shen (2003) and AntRSAR, Particle Swarm Optimization (PSORSAR) and BeeRSAR developed by Suguna and Thanushkodi (2010) improved the accuracy but requires large processing time (Suguna, Thanushkodi, & Nadu, 2011).

Feature selection, also known as attribute selection is the process of selecting a subset of relevant features (attributes) to be used in model construction. It is the process of choosing a subset of important features so that

the feature space is optimally reduced to evaluation criterion. Feature selection can reduce both the data and the computational complexity. In general, it can be viewed as a search problem where each state in the search space represents a subset of possible features. For example, if the search space is small, analyzing all subsets in any order and search will complete in a short time. However, the search space is usually not small, 2N where the number of dimensions N in typical data-mining application is large (N>20). Regarding this issue, the search strategy is very important to find near-optimal subsets of features that further improve the quality of the data mining process. Chandrashekar and Sahin (2014) claimed that although feature selection is a well-developed research area with various methods, researchers still try to find better methods to produce efficient classifiers.

RELATED WORK

There are several common feature selection search methods applied in solving problem such as Best First Width Search in classical planning problem by Lipovetzky and Geffner (2017), Genetic Search in fleet routing problem by Borthen, Loennechen, Wang, Fagerholt and Vidal (2017) and Greedy Stepwise Search in fisheries by Zarkami, Moradi, Pasvisheh, Bani and Abbasi (2018). Another method was proposed by Hamdani, Won, Alimi and Karray (2011) and it is based on Genetic Algorithm (GA) with bi-coded chromosome representation and a new evaluation function. The method used a hierarchical algorithm with homogeneous and heterogeneous population to minimize the computational cost and speed up the convergence time. They claimed that the heterogeneous GA performed global search among solutions with different sizes and the best solutions are passed to homogeneous GAs to locally optimize the solutions. Due to the parallel nature of their proposed method, the method showed good performance when compared against heuristic algorithms and simple GA.

Scatter search has been used in credit scoring as implemented by Wang, Hedar, Wang and Ma (2012) to perform a search through the feature subset space to identify important features. It starts with a population of many significant and diverse feature subsets, and stops when the assessment criteria is higher than a given threshold or does not have improve any longer. Several global and local search algorithms have been deployed for optimization purposes. New filter-wrapper hybrid based were invented by Adair, Brownlee and Ochoa (2018) and Rodriguez-Galiano, Luque-Espinar, Chica-Olmo and Mendes (2018) . The proposed method has provided more reliable solutions, where the solutions are more able to generalize unseen data. Similarly, improved algorithm based on monarch butterfly was invented by Faris, Aljarah and Mirjalili (2018) which experiment results shown high efficiency at global optimization.

New hybrid Binary Particle Swarm Optimization (BPSO) and Evolutionary Algorithm (EA) based feature selection method was created by Zhang, Xiong, Zhong and Thompson (2018). Inspired by the concept of BPSO, new particle's position updating process was designed in binary search space. Experimental results demonstrates the proposed method produced better classification accuracy and outperform Extended Nearest Neighbor, k-Nearest Neighbor, Naïve Bayes, and Linear Discriminant Analysis on the eight datasets selected from University of California (UCI) Machine Learning Repository (Asuncion & Newman, 2017).

In addition, the Ant Colony Optimization (ACO) algorithm was applied to find the optimum features for breast cancer diagnosis of Raman-based cancer (Fallahzadeh, Dehghani-Bidgoli & Assarian, 2018). The result shows that ACO feature selection improves the diagnostic accuracy of Raman-based diagnostic models. Likewise, a hybrid approach for feature subset selection using ACO and multi-classifier ensemble was proposed by Shahzad, Ellahi, Naseer and Waseem Shahzad (2018). In the research, ACO was used to enhance the predictive accuracy of filters method. Extensive experimentation indicates that the use of ACO has the ability to generate small subsets and attained higher classification accuracy (Alwan & Ku-Mahamud, 2017).

Independent RSAR hybrid with Artificial Bee Colony (ABC) algorithm has been introduced by Suguna, Thanushkodi and Nadu (2011). They grouped the instances based on decision attributes. Then, they applied Quick Reduct Algorithm (Chouchoulas & Shen, 2001) to find the reduced feature set for each class. To this set of reducts, they utilized ABC algorithm to select a random number of attributes from each set, based on the RSAR model, to find the final subset of attributes. An experiment was carried out on five different datasets from the UCI machine learning (Ibrahim, Abdullah & Saripan, 2009) and compared with six different algorithms which are general RSAR, Entity-based Reduct by Jensen and Shen (2001), GenRSAR and AntRSAR by Jensen and Shen (2003) and Particle Swarm Optimization based RSAR and BeeRSAR by Suguna and Thanushkodi (2010). They claimed the proposed method can find very minimal reduct than existing methods.

New nature-inspired feature selection technique based on bats behavior has been proposed by Nakamura et al. (2012). The technique implemented wrapper approach that combines the power of exploration of the bats together with the speed of the Optimum-Path Forest classifier (Papa, Falcão, Albuquerque & Tavares, 2012). Nakamura et al. (2012) claimed that the proposed technique can find the set of features that maximizes the accuracy in a validating set. Their experiment employed five public datasets to accomplish this task, in which Bat Algorithm has been compared against Binary Firefly Algorithm (Falcon, Almeida & Nayak, 2011) and Binary Gravitational Search Algorithm (Rashedi, Nezamabadi-Pour & Saryazdi, 2010). They claimed the proposed

algorithm out-performed the compared techniques in 3 out of 5 datasets, being the second best in the remaining two datasets.

Cuckoo Search Algorithm which as introduced by Yang and Deb (2009) and has also been used to solve feature selection problem (Shehab, Khader & Laouchedi, 2018). For instance, modified cuckoo search algorithm with rough sets has been proposed by Aziz and Hassanien (2018). This modified cuckoo search algorithm imitates the obligate brood parasitic behavior of some cuckoo species in combination with the Levy flight behavior of some birds. The proposed algorithm shows the capability to reduce the number of features in reduct set while considering the classification quality into account. Also, Usman, Yusof, Naim and Naim (2018) proposed a prediction method based on Cuckoo Search algorithm. Two algorithms namely Cuckoo Search Algorithm and Cuckoo Optimization Algorithm were used during subset generation and results shows that both algorithms significantly selected fewer number of features as well as improved prediction accuracy on selected datasets.

In 2010, Yang created the Firefly Algorithm and it was employed in many area and this includes feature selection application. For example, Firefly Algorithm Based Wrapper-Penalty Feature Selection method for cancer diagnosis has been developed by Sawhney, Mathur and Shankar (2018). The developed method explored the inclusion of a penalty function to the existing fitness function promoting the Binary Firefly Algorithm. This reduces the feature set to an optimal subset while increasing the classification accuracy. In addition, Marie-Sainte and Alalyani (2018) proposes feature selection in Arabic text classification based on Firefly Algorithm. The proposed algorithm has been successfully applied in different combinatorial problems and obtained high precision value in improving Arabic text classification.

Similar to existing work in feature selection, this paper aims to present a model for obtaining the optimal number of attributes for the employed datasets. The model consists of the combination of search methods and reduction algorithms. Different reduction algorithm methods are experimented together with various attribute selection search methods to produce the reduction set model. Thereafter, optimization method using bio-inspired search algorithms been applied to obtain the best reduction set of the attribute and further tested with 5 various data sets.

METHODS

The methodology is shown in Figure 1. It consists of eight activities in two phases: (1) Data Collection; (2) Data pre-processing; (3) Dimensionality Reduction; (4) Model Training & Testing (attribute reduction); (5) Model Training (optimized attribute set); (6) Model Testing (optimized attribute set);

(7) Model Evaluation. The expected output from phase 1 is a model which consists of combination of search methods and reduction algorithms.

Step 1 (Data Collection): Arrhythmia data set was selected from UCI Machine Learning Repository. Arrhythmia data set was selected due to its large size of features that make it challenging to explore (Namsrai et al., 2013). The Arrhythmia data set was used to distinguish between the presence and absence of cardiac arrhythmia and classify it in one of the 16 groups. Details of Arrhythmia dataset can be referred in Table 1.

Step 2 (Data Pre-processing): Data set that has missing values has been pre-processed in order to make sure that data set is ready to be experimented. In this step, data set that has missing value (denoted as '?' in original dataset) can be replaced either with 0 or mean value. Both approaches have been tested and results show no different in terms of performance. This research opts to replace missing value with "0". More than 400 missing values were identified in the data set especially in attributes with nominal values.

Step 3 (Dimensionality Reduction): Seven (7) search methods and three (3) reduction algorithms have been used in order to search for the optimal attributes (i.e local and global search). With these two search, Montazeri et al. (2013) explained that the exploration and exploitation will be balanced, hence solution space is searched effectively The employed search methods and reduction algorithms are the ones commonly used in data mining studies by Aggarwal (2013), Hall and Holmes (2003), Hall et al. (2009) and Masilamani et al. (2010).

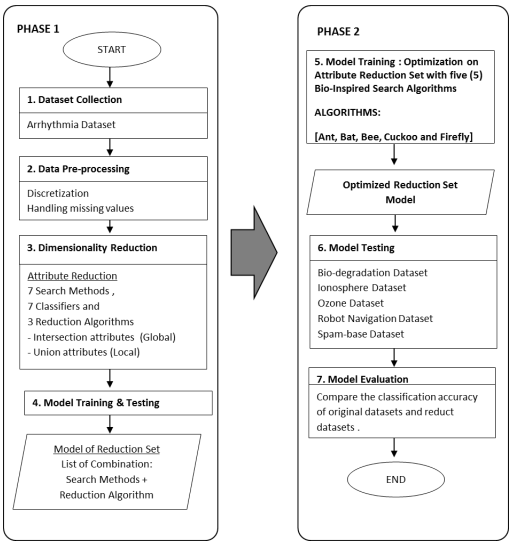

Figure 1. Phases in methodology.

In this step, the intersection of attributes (global search) and union attributes (local search) were identified. For global search, the intersection of attributes was identified from the results of each possible combination of search methods and reduction algorithms. Regarding the intersection results, the next stage is to produce local search where union of the attributes was classified. Consider the example below:

Original Data set ARHYTHMIA ={A1,A2,A3,A4,A5,A6,A7,A8,A9,A10,A11,A12,A13}
Combination Reduction algorithm A + Search method A = Reduct Data set {A1, A2, A3}
Combination Reduction algorithm A + Search method B= Reduct Data set {A2, A3, A6}
Combination Reduction algorithm A + Search method C = Reduct Data set {A3, A2, A9}
Combination Reduction algorithm B + Search method A= Reduct Data set {A10, A2, A13}
Combination Reduction algorithm B + Search method B= Reduct Data set {A4, A13, A10}
Combination Reduction algorithm B + Search method C = Reduct Data set {A7, A10, A13}
The attribute intersection of Data set ARHYTHMIA for Reduction algorithm A with Search method A, B and C is {A2, A3}.
The attribute intersection of Data set ARHYTHMIA for Reduction algorithm B with Search method A, B and C is {A10, A13}.

Based on the example*, the union of the attributes in the two intersections will be {A2, A3, A10, and A13}.*

Step 4 Model Training & Testing: In this step, the selected attributes obtained from previous step were further tested to produce a model that consist the best combination of search method and reduction algorithm. Various combinations of search methods and reduction algorithms were tested with selected attributes to achieve the best combination list for the model. The best combination list was determined by the less number of reductions with good classification accuracy. The output of this step is the model of reduction set with best combination list of search methods and reduction algorithms.

Step 5 Model Training (Optimization on Attribute Reduction Set): Next, the model of reduction set produced in phase 1 was further tested in phase 2 in order to obtain the optimize reduction set. In this step, five (5) bio-inspired search algorithms namely Ant, Bat, Bee, Cuckoo and Firefly have been applied with the aim of reducing the number of attributes for an optimal result by utilizing the capabilities of swarm intelligence searching methods. The output of this step is the optimized reduction set model.

Step 6 Model Testing (Test Model with Various Data set): In this step, the optimized reduction set model obtained from previous step was further tested with various UCI benchmark data sets (Asuncion & Newman, 2017) namely Bio-degradation, Ionosphere, Ozone, Robot Navigation and Spam-base to confirm the performance of the model. These data sets include discrete and continuous attributes and represent various domains. The reason for choosing these data set is to produce a generalize model. Information of the data sets is shown in Table 1.

Step 7 Model Evaluation (Model with good accuracy): In this step, the classification accuracy for all employed data sets were analyzed and compared. The final outcome is the classification accuracy with optimal number of attributes.

Table 1

Data Set Characteristics

Data set	**# of Attributes**	**# of Instances**	**# of Classes**
Arrhythmia	279	452	16
Bio-degradation	41	1055	2
Ionosphere	34	351	2
Ozone	72	2536	2
Robot Navigation	24	5456	4
Spam-base	57	4601	2

All the six (6) data sets were tested using seven (7) search methods, 7 classifiers and 3 reduction algorithms in phase 1. In this research, WEKA was used to perform classification tasks and the measure to be used is the classification accuracy. In this research, accuracy rate is important to indicate how well the model performs in classifying the data. The higher the accuracy the better the model is. The parameters setting for all five bio-inspired search algorithms have been summarized in Table 2. The default setting has been employed for the population size, number of iteration, mutation probability with other specific setting for each search algorithms in WEKA Software developed by Hall et al. (2009).

Table 2

Parameters setting for all Bio-Inspired Algorithms

Search Algorithm	Population Size	Iteration	Mutation Probability	Others
ANT	20	20	0.01	Evaporation rate: 0.9 Pheromone rate: 2.0 Heuristic rate: 0.7
BAT	20	20	0.01	Frequency: 0.5 Loudness: 0.5
BEE	30	20	0.01	Radius Damp: 0.98 Radius Mutation: 0.80
CUCKOO	20	20	0.01	Pa rate: 0.25 Sigma rate: 0.70
FIREFLY	20	20	0.01	Beta zero: 0.33 Absorption Coefficient: 0.001

RESULTS AND DISCUSSION

The outputs of each phase are presented in this section. The results are presented as percentage of accuracy for each list of combination search methods and reduction algorithm. The optimized numbers of attributes with the classification accuracy are also presented. Table 3 shows the results of a list of combination search methods and reduction algorithm that have been obtained in phase 1. This is the attribute reduction model that has been used to perform the first reduction of the datasets before further optimize reduction in phase 2. The first step of attribute reduction is very crucial stage in order to acquire good starting point of attribute selection before optimization step. The purpose of this step is to achieve better attribute reduction at the early stage of attribute selection where the search space been explored before further optimization search being extended in phase 2. In Table 3, the selected combination of search algorithm with best reduction algorithm has produced good reduction set with significant percentage of classification accuracy after comprehensive combination trial in step 3 (dimensionality reduction). Thus, this combination list is considered reliable to be used to perform as an early first stage of attribute reduction.

Table 3

Summary of the Attribute Reduction Model in Phase 1 (Arrhythmia Data Set)

Search Algorithm	Best Reduction Algorithm	#ATTR	#Sel Attr	% Acc
BestFirstSearch	WrapperSubsetEval + Bayes Net (Classifier)	279	19	80.29
GeneticSearch	WrapperSubsetEval + Bayes Net (Classifier)	279	127	79.58
GreedyStepwise	WrapperSubsetEval + Bayes Net (Classifier)	279	20	79.81
LineForwardSelection	WrapperSubsetEval + Naïve Bayes (Classifier)	279	18	79.14
ScatterSearch	CFSSubsetEval / FilteredSubsetEval	279	20	80.15
RaceSearch	ClassifierSubsetEval + Bayes Net (Classifier)	279	26	78.82

The results of phase 2 are shown in Table 4 until Table 8. Table 4 shows the performance of the optimized model with Bio-degradation data set. Significant reduction can be seen (refer RaceSearch row) where almost half number of attributes (from 7 to 4) been reduced by applying all five bio-inspired search algorithms. All five (5) algorithms performed better classification accuracy from the previous reduction (79.39% to 81.04%).

Nevertheless, bio-inspired search algorithms performed as much as previous reductions in term of number of reductions and classification accuracy for almost all rows. This event related to optimum capacity of search been made in phase 1 where no more area of extraction can be discover with optimization algorithms.

Table 4

Performance of the Optimized Attribute Reduction Model on Bio-Degradation Data Set

PHASE 1					PHASE 2		
Search Algorithm	**Best Reduction Algorithm**	**#ATTR**	**#Sel Attr**	**% Acc**	**Search Algorithm (Bio-Inspired)**	**#Sel Attr**	**% Acc**
BestFirst Search	Wrapper Subset Eval + Bayes Net (Classifier)	41	10	82.17	ACO	10	82.17
					BAT	10	82.17
					BEE	10	82.17
					CUCKOO	10	82.17
					FIREFLY	10	82.17

(continued)

PHASE 1					PHASE 2		
Search Algorithm	**Best Reduction Algorithm**	**#ATTR**	**#Sel Attr**	**% Acc**	**Search Algorithm (Bio-Inspired)**	**#Sel Attr**	**% Acc**
GeneticSearch	Wrapper Subset Eval + Bayes Net (Classifier)	41	23	87.47	ACO	13	86.63
					BAT	**16**	**87.74**
					BEE	**18**	**87.74**
					CUCKOO	17	87.18
					FIREFLY	18	87.18
GreedyStepwise	Wrapper Subset Eval + Bayes Net (Classifier)	41	10	82.17	ACO	10	82.17
					BAT	10	82.17
					BEE	10	82.17
					CUCKOO	10	82.17
					FIREFLY	10	82.17
LineForward Selection	Wrapper Subset Eval + Naïve Bayes (Classifier)	41	6	81.89	ACO	6	81.89
					BAT	6	81.89
					BEE	6	81.89
					CUCKOO	6	81.89
					FIREFLY	6	81.89
ScatterSearch	CFS Subset Eval / Filtered Subset Eval	41	14	82.45	ACO	14	82.45
					BAT	13	82.45
					BEE	14	82.45
					CUCKOO	14	82.45
					FIREFLY	14	82.45
RaceSearch	Classifier SubsetEval + Bayes Net (Classifier)	41	7	79.39	**ACO**	**4**	**81.04**
					BAT	**4**	**82.75**
					BEE	**4**	**81.04**
					CUCKOO	**4**	**81.04**
					FIREFLY	**4**	**81.04**
SubsetSize Forward Selection	CFSSubsetEval / Filtered Subset Eval	41	14	82.45	ACO	14	82.09
					BAT	13	82.56
					BEE	14	82.09
					CUCKOO	14	82.09
					FIREFLY	14	82.09

Table 5 shows the performance of the optimized model with Ionosphere data set. It can be seen that all bio-inspired search algorithm managed to reduce number of attributes (refer GeneticSearch and LineForwardSelection row). Improvement on the classification accuracy can be observed where all five bio-inspired search algorithms achieved more than 2% of accuracy with half attributes were reduced (refer to GeneticSearch row). Even though there is a slight decrement in the classification accuracy (from 89.92% to 88.24% with 1 % accuracy reduced) (refer RaceSearch row), however, this situation does not influence the reliability of the 4 (four) optimal attributes that have been selected since all 5 (five) bio–inspired algorithms selected the same 4 (four) attributes out of 5 (five) attributes in phase 1.

Table 5

Performance of the Optimized Attribute Reduction Model on Ionosphere Data Set

PHASE 1					**PHASE 2**		
Search Algorithm	**Best Reduction Algorithm**	**#ATTR**	**#Sel Attr**	**% Acc**	**Search Algorithm (Bio-Inspired)**	**#Sel Attr**	**% Acc**
BestFirstSearch	Wrapper Subset Eval + Bayes Net (Classifier)	34	8	94.12	ACO	8	94.12
					BAT	8	94.12
					BEE	8	94.12
					CUCKOO	8	94.12
					FIREFLY	8	94.12
GeneticSearch	Wrapper Subset Eval + Bayes Net (Classifier)	34	13	90.76	**ACO**	**7**	**93.28**
					BAT	**6**	**94.12**
					BEE	**6**	**92.44**
					CUCKOO	**6**	**93.28**
					FIREFLY	**6**	**93.28**
GreedyStepwise	Wrapper Subset Eval + Bayes Net (Classifier)	34	8	94.12	ACO	8	94.12
					BAT	8	94.12
					BEE	8	94.12
					CUCKOO	8	94.12
					FIREFLY	8	94.12
LineForward Selection	Wrapper Subset Eval + Naïve Bayes (Classifier)	34	9	93.28	ACO	6	90.76
					BAT	9	93.28
					BEE	5	92.44
					CUCKOO	6	90.76
					FIREFLY	5	92.44
ScatterSearch	CFS Subset Eval / Filtered Subset Eval	34	13	90.76	ACO	13	90.76
					BAT	13	90.76
					BEE	13	90.76
					CUCKOO	13	90.76
					FIREFLY	13	90.76
RaceSearch	Classifier Subset Eval + Bayes Net (Classifier)	34	5	89.92	**ACO**	**4**	**88.24**
					BAT	**4**	**88.24**
					BEE	**4**	**88.24**
					CUCKOO	**4**	**88.24**
					FIREFLY	**4**	**88.24**
SubsetSize Forward Selection	CFS Subset Eval / Filtered SubsetEval	34	13	90.76	ACO	13	90.76
					BAT	13	90.76
					BEE	13	90.76
					CUCKOO	13	90.76
					FIREFLY	13	90.76

Table 6 shows the performance of the optimized model with Ozone data set. In GeneticSearch row, five (5) bio-inspired algorithms performed well by

diminishing 3 (three) attributes to only 1 (one) attribute remain. It can be considered an extreme case where only a single attribute is selected to represent the model. Interestingly, with 1 (one) attribute, the same level of classification accuracy (93.85%) has been achieved.

The same exciting case (refer RaceSearch row) can be seen when implementing BAT search algorithm where it managed to obtain higher classification accuracy (91.3% to 93.62%) with a single attribute. This situation could be related to the advantage of global search features in bat algorithm to acquire an optimal reduction set even though with small number of attributes. However, 4 (four) other bio-inspired search algorithms fail to mine the optimal attributes due to small search space (number of attributes) to be explored.

Table 6

Performance of the Optimized Attribute Reduction Model on Ozone Data Set

PHASE 1					**PHASE 2**		
Search Algorithm	**Best Reduction Algorithm**	**#ATTR**	**#Sel Attr**	**% Acc**	**Search Algorithm (Bio-Inspired)**	**#Sel Attr**	**% Acc**
BestFirstSearch	Wrapper Subset Eval + Bayes Net (Classifier)	72	5	93.97	ACO	5	93.97
					BAT	5	93.97
					BEE	5	93.97
					CUCKOO	5	93.97
					FIREFLY	5	93.97
GeneticSearch	Wrapper Subset Eval + Bayes Net (Classifier)	72	4	93.85	**ACO**	**1**	**93.85**
					BAT	**1**	**93.85**
					BEE	**1**	**93.85**
					CUCKOO	**1**	**93.85**
					FIREFLY	**1**	**93.85**
GreedyStepwise	Wrapper Subset Eval + Bayes Net (Classifier)	72	5	93.97	ACO	5	93.97
					BAT	5	93.97
					BEE	5	93.97
					CUCKOO	5	93.97
					FIREFLY	5	93.97
LineForward Selection	Wrapper Subset Eval + Naïve Bayes (Classifier)	72	4	93.97	ACO	4	93.97
					BAT	4	93.97
					BEE	4	93.97
					CUCKOO	4	93.97
					FIREFLY	4	93.97
ScatterSearch	CFS Subset Eval / Filtered Subset Eval	72	12	83.99	**ACO**	**3**	**93.85**
					BAT	**3**	**93.85**
					BEE	**3**	**93.85**
					CUCKOO	**3**	**93.85**
					FIREFLY	**3**	**93.85**

(continued)

PHASE 1					PHASE 2		
Search Algorithm	**Best Reduction Algorithm**	**#ATTR**	**#Sel Attr**	**% Acc**	**Search Algorithm (Bio-Inspired)**	**#Sel Attr**	**% Acc**
RaceSearch	ClassifierSubsetEval + Bayes Net (Classifier)	72	2	91.3	ACO	0	0
					BAT	**1**	**93.62**
					BEE	0	0
					CUCKOO	0	0
					FIREFLY	0	0
SubsetSize Forward Selection	CFSSubsetEval / FilteredSubsetEval	72	15	82.25	ACO	15	82.25
					BAT	15	82.25
					BEE	15	82.25
					CUCKOO	15	82.25
					FIREFLY	15	82.25

Table 7 shows the performance of the optimized model with Robot Navigation data set. Interesting pattern can be examined where no reduction of attributes for all 5 (five) bio-inspired search algorithms except for GeneticSearch row. When a certain number of selected attributes have been obtained, no more attribute reduction can be made even though bio-inspired algorithm is employed. However, when GeneticSearch was employed in Phase 1, it generates a feature set of 8 (eight) attributes, and the bio-inspired algorithms have reduced it to 5 (five) attributes. Yet, the outcome is a higher accuracy.

Table 7

Performance of the Optimized Attribute Reduction Model on Robot Navigation Data Set

PHASE 1					PHASE 2		
Search Algorithm	**Best Reduction Algorithm**	**#ATTR**	**#Sel Attr**	**% Acc**	**Search Algorithm (Bio-Inspired)**	**#Sel Attr**	**% Acc**
BestFirstSearch	Wrapper Subset Eval + Bayes Net (Classifier)	24	6	97.09	ACO	6	97.09
					BAT	6	97.09
					BEE	6	97.09
					CUCKOO	6	97.09
					FIREFLY	6	97.09
GeneticSearch	Wrapper Subset Eval + Bayes Net (Classifier)	24	8	96.28	**ACO**	**5**	**96.93**
					BAT	**5**	**96.93**
					BEE	**5**	**96.93**
					CUCKOO	**5**	**96.93**
					FIREFLY	**5**	**96.93**

(continued)

PHASE 1					PHASE 2		
Search Algorithm	**Best Reduction Algorithm**	**#ATTR**	**#Sel Attr**	**% Acc**	**Search Algorithm (Bio-Inspired)**	**#Sel Attr**	**% Acc**
GreedyStepwise	Wrapper Subset Eval + Bayes Net (Classifier)	24	6	97.09	ACO	6	97.09
					BAT	6	97.09
					BEE	6	97.09
					CUCKOO	6	97.09
					FIREFLY	6	97.09
LineForward Selection	Wrapper Subset Eval + Naïve Bayes (Classifier)	24	6	97.04	ACO	6	97.04
					BAT	6	97.04
					BEE	6	97.04
					CUCKOO	6	97.04
					FIREFLY	6	97.04
ScatterSearch	CFS Subset Eval / Filtered Subset Eval	24	6	96.71	ACO	6	96.71
					BAT	6	96.71
					BEE	6	96.71
					CUCKOO	6	96.71
					FIREFLY	6	96.71
RaceSearch	Classifier SubsetE val + Bayes Net (Classifier)	24	5	95.42	ACO	5	95.42
					BAT	5	95.42
					BEE	5	95.42
					CUCKOO	5	95.42
					FIREFLY	5	95.42
SubsetSize Forward Selection	CFS Subset Eval / Filtered Subset Eval	24	6	96.71	ACO	6	96.71
					BAT	6	96.71
					BEE	6	96.71
					CUCKOO	6	96.71
					FIREFLY	6	96.71

The result of Spam-base data set is depicted in Table 8 where bio-inspired search algorithms performed well to reduce the feature set produced in phase 1. However, when the feature set is reduced to have less than 15 attributes (for example refer to BestFirstSearch[CUCKOO], GreedyStepwise[CUCKOO], RaceSearch[ANT, BAT, BEE, CUCKOO and FIREFLY]), the classification accuracy started to decrease (around 1%) from the previous reduction set. Such an event is related to the issues of the algorithms in phase 1 (one) that occupied almost all of the search space where the optimization algorithms (bio-inspired) has very limited space to explore and exploit.

Table 8

Performance of the Optimized Attribute Reduction Model on Spam-base Data Set

PHASE 1					**PHASE 2**		
Search Algorithm	**Best Reduction Algorithm**	**#ATTR**	**#Sel Attr**	**% Acc**	**Search Algorithm (Bio-Inspired)**	**#Sel Attr**	**% Acc**
BestFirstSearch	Wrapper Subset Eval + Bayes Net (Classifier)	57	18	92.65	ACO	18	92.65
					BAT	18	92.65
					BEE	**16**	**92.84**
					CUCKOO	13	91.63
					FIREFLY	**16**	**92.84**
GeneticSearch	Wrapper Subset Eval + Bayes Net (Classifier)	57	34	93.16	**ACO**	**28**	**93.35**
					BAT	24	92.9
					BEE	**23**	**93.8**
					CUCKOO	**23**	**93.73**
					FIREFLY	**22**	**93.54**
GreedyStepwise	Wrapper Subset Eval + Bayes Net (Classifier)	57	18	92.65	ACO	18	92.65
					BAT	18	92.65
					BEE	16	92.84
					CUCKOO	13	91.63
					FIREFLY	**16**	**92.84**
LineForward Selection	Wrapper Subset Eval + Naïve Bayes (Classifier)	57	17	92.65	ACO	15	92.46
					BAT	**15**	**92.71**
					BEE	15	92.33
					CUCKOO	15	92.33
					FIREFLY	15	92.58
ScatterSearch	CFS Subset Eval / Filtered Subset Eval	57	15	92.33	ACO	15	92.33
					BAT	15	92.33
					BEE	15	92.33
					CUCKOO	15	92.33
					FIREFLY	15	92.33
RaceSearch	Classifier Subset Eval + Bayes Net (Classifier)	57	8	88.68	ACO	6	87.92
					BAT	6	87.92
					BEE	6	87.92
					CUCKOO	6	87.92
					FIREFLY	6	87.92
SubsetSize Forward Selection	CFS Subset Eval / Filtered Subset Eval	57	15	92.33	ACO	15	92.33
					BAT	15	92.33
					BEE	15	92.33
					CUCKOO	15	92.33
					FIREFLY	15	92.33

Similarities of attributes between the selected final attributes in phase 1 and phase 2 are summarized in Table 9. It can be seen that BAT algorithm shows good generalization as it produced significant reduction attributes with acceptable classification accuracy, especially in the Ozone dataset. The other four (4) bio-inspired algorithms also performed well in reducing the size of the feature set while improving classification accuracy. This can be seen in the Ionosphere, Ozone, and Robot-navigation datasets.

Table 9

The best bio-inspired algorithm with percentage of similarity selected attribute and classification accuracy improvement

Dataset	Search Algorithm	Reduction Algorithm	Bio-Inspired Algorithm	Number of Attribute Selected		% Similarity	% classification accuracy improvement
				Phase 1	Phase 2		
Bio-degradation	RaceSearch	Classifier Subset Eval + Bayes Net (Classifier)	BAT	7	4	100%	3.36%
Ionosphere	GeneticSearch	Wrapper Subset Eval + Bayes Net (Classifier)	BAT	13	6	100%	3.36%
Ozone	ScatterSearch	CFS Subset Eval / Filtered Subset Eval	BAT	12	3	100%	9.86%
Robot Navigation	GeneticSearch	Wrapper Subset Eval + Bayes Net (Classifier)	BAT	8	5	100%	0%
Spam-base	GeneticSearch	Wrapper Subset Eval + Bayes Net (Classifier)	FIREFLY	34	22	100%	0.38%

CONCLUSION

The contribution of this paper in the area of data is significant as it provides insight of manipulation of bio-inspired algorithms in exploration and exploitation of the search space in reducing size of feature set. This paper investigates on feature selection method where seven (7) attribute selection search methods, seven (7) classifiers and three (3) reduction algorithms were compared and

tested on six (6) data sets. The initial obtained model (combination of search methods and reduction algorithm) has generated good classification accuracy with relevant features. The models were further optimized using bio-inspired search algorithms to obtain the best reduction set model of the attribute. The experimental results demonstrates that the attribute reduction model with the implementation of bio-inspired search algorithm consistently perform good classification task on the selected data set while using a smaller set of features. Manipulation on parameter setting of the bio-inspired search algorithms can be considered for future work in order to determine the best setup to acquire more promising results.

ACKNOWLEDGMENT

The authors gratefully acknowledge use of service and facilities of Universiti Malaysia Terengganu (UMT), Universiti Utara Malaysia (UUM) and Kementerian Pendidikan Malaysia (KPM).

REFERENCES

Adair, J., Brownlee, A. E. I., & Ochoa, G. (2018). Mutual information iterated local search: A wrapper-filter hybrid for feature selection in brain computer interfaces. *Lecture Notes in Computer Science*, *10784*, 63–77. doi.org/10.1007/978-3-319-77538-8_5

Aggarwal, M. (2013). Performance analysis of different feature selection methods in intrusion detection. *International Journal of Scientific & Technology Research*, *2*(6), 225–231. Retrieved from www.ijstr.org

Alwan, H. B., & Ku-Mahamud, K. R. (2017). Mixed-variable ant colony optimization algorithm for feature subset selection and tuning SVM parameter. *International Journal of Bio-Inspired Computation, 9*(1), 53-63.

Asuncion, A., & Newman, D. J. (2017). UCI machine learning repository. Retrieved from http://archive.ics.uci.edu/ml

Aziz, M. A. El, & Hassanien, A. E. (2018). Modified cuckoo search algorithm with rough sets for feature selection. *Neural Computing and Applications*, *29*(4), 925–934. doi.org/10.1007/s00521-016-2473-7

Borthen, T., Loennechen, H., Wang, X., Fagerholt, K., & Vidal, T. (2017). A genetic search-based heuristic for a fleet size and periodic routing problem with application to offshore supply planning. *EURO Journal on Transportation and Logistics*, *7*(2), 121–150. doi.org/10.1007/s13676-017-0111-x

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers and Electrical Engineering*, *40*(1), 16–28. doi.org/10.1016/j.compeleceng.2013.11.024

Chouchoulas, A., & Shen, Q. (2001). Rough set-aided keyword reduction for text categorization. *Applied Artificial Intelligence*, *15*(9), 843–873. https://doi.org/10.1080/088395101753210773

Falcon, R., Almeida, M., & Nayak, A. (2011). Fault identification with binary adaptive fireflies in parallel and distributed systems. In *2011 IEEE Congress of Evolutionary Computation, CEC 2011* (pp. 1359–1366). IEEE. doi.org/10.1109/CEC.2011.5949774

Fallahzadeh, O., Dehghani-Bidgoli, Z., & Assarian, M. (2018). Raman spectral feature selection using ant colony optimization for breast cancer diagnosis. *Lasers in Medical Science*, 1–8. doi.org/10.1007/s10103-018-2544-3

Faris, H., Aljarah, I., & Mirjalili, S. (2018). Improved monarch butterfly optimization for unconstrained global search and neural network training. *Applied Intelligence*, *48*(2), 445–464. doi.org/10.1007/s10489-017-0967-3

Hall, M. A., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter*, *11*(1), 10–18. doi.org/10.1145/1656274.1656278

Hall, M. A., & Holmes, G. (2003). Benchmarking attribute selection techniques for discrete class data mining. *IEEE Transactions on Knowledge and Data Engineering*, *15*(6), 1437–1447. doi.org/10.1109/TKDE.2003.1245283

Hamdani, T. M., Won, J. M., Alimi, A. M., & Karray, F. (2011). Hierarchical genetic algorithm with new evaluation function and bi-coded representation for the selection of features considering their confidence rate. *Applied Soft Computing*, *11*(2), 2501–2509. doi.org/10.1016/j.asoc.2010.08.020

Ibrahim, N. K., Abdullah, R. S. A. R., Saripan, M. I., Ibrahim, N. K., Abdullah, R. S. A. R., & Saripan, M. I. (2009). Artificial neural network approach in radar target classification. *Journal of Computer Science*, *5*(1), 23–32. doi.org/10.3844/jcssp.2009.23.32

Jensen, R., & Shen, Q. (2001). A rough set-aided system for sorting WWW bookmarks. *Lecture Note in Computer Science*, 95–105. doi.org/https://doi.org/10.1007/3-540-45490-X_10

Jensen, R., & Shen, Q. (2003). Finding rough set reducts with ant colony optimization. In *Proceedings of the 2003 UK workshop on Computational Intelligence* (Vol. 1, pp. 15–22). Retrieved from http://users.aber.ac.uk/rkj/pubs/papers/antRoughSets.pdf

Larabi Marie-Sainte, S., & Alalyani, N. (2018). Firefly algorithm based feature selection for arabic text classification. *Journal of King Saud University - Computer and Information Sciences*. doi.org/10.1016/j.jksuci.2018.06.004

Lipovetzky, N., & Geffner, H. (2017). Best-first width search : exploration and exploitation in classical planning. In *Proceedings of the 31th Conference on Artificial Intelligence (AAAI 2017)* (pp. 3590–3596).

Masilamani, A., Anbarasi, M., & Anupriya, E. (2010). Enhanced prediction of heart disease with feature subset selection using genetic algorithm. *International Journal of Engineering Science and Technology*, *2*(10), 5370–5376. Retrieved from https://www.researchgate.net/publication/50361284

Montazeri, M., Montazeri, M., Naji, H. R., & Faraahi, A. (2013). A novel memetic feature selection algorithm. In *Proceedings of the 5th Conference on Information and Knowledge Technology* (pp. 295–300). IEEE. doi.org/10.1109/IKT.2013.6620082

Nakamura, R. Y. M., Pereira, L. a. M., Costa, K. a., Rodrigues, D., Papa, J. P., & Yang, X.-S. (2012). BBA: a binary bat algorithm for feature selection. In *Proceedings of the 25th SIBGRAPI Conference on Graphics, Patterns and Images*, 291–297. doi.org/10.1109/SIBGRAPI.2012.47

Namsrai, E., Munkhdalai, T., Li, M., Shin, J.-H., Namsrai, O.-E., & Ryu, K. H. (2013). A feature selection-based ensemble method for arrhythmia classification. *Journal of Information Processing Systems*, *9*(1), 31–40. doi.org/10.3745/JIPS.2013.9.1.031

Papa, J. P., Falcão, A. X., De Albuquerque, V. H. C., & Tavares, J. M. R. S. (2012). Efficient supervised optimum-path forest classification for large datasets. *Pattern Recognition*, *45*(1), 512–520. doi.org/10.1016/j.patcog.2011.07.013

Rashedi, E., Nezamabadi-Pour, H., & Saryazdi, S. (2010). BGSA: binary gravitational search algorithm. *Natural Computing*, *9*(3), 727–745. doi.org/10.1007/s11047-009-9175-3

Rodriguez-Galiano, V. F., Luque-Espinar, J. A., Chica-Olmo, M., & Mendes, M. P. (2018). Feature selection approaches for predictive modelling of groundwater nitrate pollution: An evaluation of filters, embedded and wrapper methods. *Science of the Total Environment*, *624*, 661–672. doi.org/10.1016/j.scitotenv.2017.12.152

Sawhney, R., Mathur, P., & Shankar, R. (2018). A firefly algorithm based wrapper-penalty feature selection method for cancer diagnosis. *Lecture Note in Computer Science*, 438–449. doi.org/10.1007/978-3-319-95162-1_30

Shahzad, W., Ellahi, A., Naseer, A., & Waseem Shahzad, P. (2018). A hybrid approach for feature subset selection using ant colony optimization and multi-classifier ensemble. *International Journal of Advanced Computer Science and Applications*, *9*(1), 306–313. doi.org/10.14569/IJACSA.2018.090142

Shehab, M., Khader, A. T., & Laouchedi, M. (2018). A hybrid method based on Cuckoo algorithm for global optimization problems. *Journal of Information and Communication Technology*, *17*(3), 469-491.

Suguna, N., & Thanushkodi, K. (2010). A novel rough set reduct algorithm for medical domain based on bee colony. *Journal of Computing*, *2*(6), 49–54.

Suguna, N., Thanushkodi, K. G., & Nadu, T. (2011). An independent rough set approach hybrid with artificial bee colony algorithm for dimensionality reduction. *American Journal of Applied Sciences*, *8*(3), 261–266. doi.org/10.3844/ajassp.2011.261.266

Usman, A. M., Yusof, U. K., Naim, S., & Naim, S. (2018). Cuckoo inspired algorithms for feature selection in heart disease prediction. *International Journal of Advances in Intelligent Informatics*, *4*(2), 95. doi.org/10.26555/ijain.v4i2.245

Wang, J., Hedar, A.-R., Wang, S., & Ma, J. (2012). Rough set and scatter search metaheuristic based feature selection for credit scoring. *Expert Systems with Applications*, *39*(May 2012), 6123–6128. doi.org/10.1016/j.eswa.2011.11.011

Yang, X.-S., & Deb, S. (2009). Cuckoo search via lévy flights. In *Proceedings of World Congress on Nature & Biologically Inspired Computing* (pp. 210–214). India: IEEE. https://doi.org/10.1109/NABIC.2009.5393690

Yang, X. (2010). *Nature-Inspired Metaheuristic Algorithms: Second Edition*. *University of Cambridge, United Kingdom*. doi.org/10.1016/B978-0-12-416743-8.00005-1

Yue, B., Yao, W., Abraham, A., & Liu, H. (2007). A new rough set reduct algorithm based on particle swarm optimization. In *Second International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2007* (pp. 397–406). Springer.

Zarkami, R., Moradi, M., Pasvisheh, R. S., Bani, A., & Abbasi, K. (2018). Input variable selection with greedy stepwise search algorithm for analysing the probability of fish occurrence: A case study for Alburnoides mossulensis in the Gamasiab River, Iran. *Ecological Engineering*, *118*(1 August 2018), 104–110. doi.org/10.1016/j.ecoleng.2018.04.011

Zhang, N., Xiong, J., Zhong, J., & Thompson, L. (2018). Feature selection method using BPSO-EA with ENN classifier. In *Proceedings of the Eighth International Conference on Information Science and Technology (ICIST)* (pp. 364–369). IEEE. doi.org/10.1109/ICIST.2018.8426154