

Solving Attribute Reduction Problem using Wrapper Genetic Programming

Mohammed Alweshah Omar A. Alzubi

Jafar A. Alzubi

Sharihan Alaqeel

Al-Balqa Applied University

Summary

Attribute reduction (AR) represents a NP-hard problem, and it is identified as the problematic issue of pinpointing the least (possible) subset of characteristics taken from the reference set. The key issue related to characteristics selectors is the production of a minimal number of reductions representing the reliable meaning of all characteristics. Nevertheless, there is no approach that can ensure optimality in the process of solving this issue. However, some methods are more efficient compared to other ones due to some characteristics of the algorithm. The research of this thesis aims at providing efficient ways that help us find the characteristics which are known as the most informative ones and the least possible features with least possible data loss. This has been done through the combination between wrapper approach and genetic programming algorithm, Wrapper Genetic Programming (WGP). Numerical experiment carried out on 10 real word dataset from the University of California Irvine benchmark data sets (UCI) Repository of Machine Learning Databases has been used and presented in order to show that WGP can give competitive solutions in an efficient manner compared to approaches available in the literature on this issue.

Key words:

Classification, attribute reduction, genetic programming algorithm, wrapper approach.

1. Introduction

Many researchers are focusing on the problem of finding a subset with minimal attributes from an original set of data in an information system [1, 2, 3, 4, 5] Locating such a subset is basically done by using the complete search approach, which locates all possible subsets to find the optimal subsets (i.e., those with a maximum rough set dependency degree). Obviously, a complete search approach is an impractical and complex solution to the problem and is only practical for simple data sets. For high-dimensional data sets, the heuristic search is much faster than the complete search, because it searches according to a particular path in order to find the minimal reduct [6].

A number of attribute reduction approaches exists that can be broadly classified into three categories: the filter approach, In contrast, the wrapper approach requires one predetermined learning model and selects features with the wrapper approach, and the hybrid approach[7] The sole difference between these two approaches is that ,The filter

approach relies on the characteristics of the learning data and selects a subset of features without involving any learning model [8, 9] the aim of improving the generalization performance of that particular learning model [9, 10, 11, 12] .

An alternative way to determine a minimal reduct is to adapt meta-heuristic algorithms. Meta-heuristics can be classified into single based solutions methods or population based methods[13, 14]. Example for single based solution are: local search, tabu search[15], simulated annealing, threshold accepting, variable neighborhood search, iterated local search, guided local search, GRASP and so on. Example of population-based meta-heuristics such as evolutionary algorithms (genetic algorithms, evolution strategies, genetic programming[16], evolutionary programming, estimation of distribution algorithms, differential evolution, and co-evolutionary algorithms), swarm intelligence-based methods (e.g., ant colonies, particle swarm optimization), scatter search, bee colony, artificial immune systems, and so on [17, 18].

Based on the above, the background give motivation for investigate in attribute reduction field exactly in wrapper approach, That is the goal of this paper is to propose a combination between wrapper and population-based algorithm(genetic programming) that can find a minimal subset of attributes and able to deliver good results.

The rest of the paper is organized as follows: Section 2 a general review various search algorithms that have been applied to solve the attribute reduction problem. And Section 3 describes the employment of wrapper approaches to solve the attribute reduction problem, Section 4 presents a discussion of the experimental results. Section 5 concludes the work presented in this paper.

2. Background and Literature

The goal of attribute reduction methods is to reduce the number of attributes in a data set as much as possible, in order to reduce computational complexity while preserving minimal information loss. Attribute reduction is an important pre-processing step in many fields in data mining, because many application problems use data collected from real valued vectors [19].

Attribute reduction has many approaches that have been proposed by different authors who have investigated it from various angles. Many of those approaches are similar in content. Kira & Rendell (1998) idealized approach finds the minimal feature subset that still retains necessary and sufficient properties of the target concept [20]. Narendra & Fukunaga (1977) they involve selecting a subset of M features by classical method that satisfies an established criterion that defines the optimal feature set size from an original set with N features, where $M < N$ [21]. Dash and Liu (1997) divide the range of attribute reduction definitions into four main categories, i.e., classical, idealized, improving prediction accuracy, and approximating original class distribution. In the approximating original class distribution method, the goal of attribute reduction is to select the smallest possible feature set that accurately represents the complete class distribution [22].

a number of meta-heuristic techniques (such as Simulated Annealing, Tabu Search, Genetic Algorithm and Ant Colony Optimization) that have been applied in the past to solve the attribute reduction problem were discussed and the quality of the obtained solutions. Arafat, Elawady et al examine how ant colony optimization applied to this downside for find out lowest rough set reduct Rough Set Theory offers heuristic function to measure the quality of single subset [23]. Also, Tallón-Ballesteros and Riquelme introduces the use of an ant colony optimization (ACO) algorithm, called Ant System, as a search method in two well-known feature subset selection methods based on correlation or consistency measures such as CFS (Correlation-based Feature Selection) and CNS (Consistency-based Feature Selection) [24].

Wang, Zhang et al presented a novel scatter search (SS) based rough set attribute reduction method, shortly SSAR. This method is an artificial-evolutionary-based algorithm and lies among memory-based heuristics [25]. Yongxiong and Li [26] create a hybrid approach using TS and C4.5 ensemble algorithm is proposed and applied to the feature selection for ensemble and relief the unbalanced data problems. Matsui, Katagiri et al formulate the new feature point selection problem as combinatorial optimization one and propose a hybrid metaheuristics through tabu search and memetic algorithm for the formulated feature point selection problem [27].

Unler and Murat propose a modified PSO algorithm for the feature subset selection problem our approach differs in two aspects from the earlier studies using PSO for this problem [28]. Gheyas and Smith SAGA combines the ability to avoid being trapped in a local minimum of SA with a very high 9 rate of convergence of the crossover operator of GA, the strong local search ability of the greedy algorithm and high computational efficiency of generalized regression neural networks GRNN [29].

Moghadasian and Hosseini presented a filter based feature selection method based on Cuckoo search (COA) algorithm and information theory to select a smaller number of features of high dimensional datasets and achieve similar or even better classification performance [30]. Schiezero and Pedrini propose a binary version of the ABC algorithm, where the number of new features to be analyzed in a neighborhood of a food source is determined through a perturbation parameter and each food source stores its quality (fitness), which is given by the accuracy of the classifier using the feature subset indicated by the bit vector [31].

Mafarja and Abdullah memetic algorithm-based approach inside the rough set theory which is a hybridization of genetic algorithm and simulated annealing [32]. Abdullah and Jaddi applied the basic Great Deluge algorithm to solve the attribute reduction problem [33]. The algorithm employs different neighbourhood structures to generate trial solutions and then it searches for the best solution among all valid solutions. Mafarja and Abdullah modified on great deluge they divide the search space into three regions with equal size, referred as interval. Based on this interval value, we form three different level values coded as level1, level2 and level3 and introduce three β values [34]. Emary, Zawbaa et al The objective of this paper was to propose a firefly algorithm (FFA) for feature selection to choose minimal number of features (attributes) and to obtain comparable or even better classification accuracy from utilizing all attributes [35].

The above related work motivates us develop the proposed approach presented in this paper. It is based on modification of wrapper approach by metaheuristic techniques for handling the attribute reduction problem.

3. Proposed Method: Wrapper Genetic Programming (WGP)

As mentioned above, The goal of feature selection is to find the smallest feature subset from data according to the preferred criterion [13]. Feature selection is much harder problem, due to the absence of class labels. Most algorithms for supervised FSS can be classified as filter or wrapper approaches.

On the other hand wrapper algorithms are those that use a classifier in order to assess the quality of a given attribute subset [12].

The accuracy achieved by wrapper algorithms is much greater than those obtained by any other filters. This advantage draw the attention of researchers from the machine learning community to further study the wrapper-based FSS.

The search process result of different search algorithms such as greedy sequential [36], floating [37], best-first search, branch and bound [38], evolutionary algorithms [39],

40] is forwarded to a different classifiers (e.g. KNN and Naive Bayes) in order to evaluate the correctness of the subset obtained by search algorithms. This led to support the belief that wrapper methods results outcome results of other filters.

A generalized wrapper algorithm (the flow chart is presented in figure 1 and its pseudo code algorithm in shown in figure 2) is very similar to the generalized filter algorithm except that it utilizes a predefined mining algorithm A instead of an independent measure M for subset evaluation. For each generated subset S , it evaluates its goodness by applying the mining algorithm to the data with feature subset S and evaluating the quality of mined results.

Therefore, different mining algorithms will produce different feature selection results. Varying the search strategies via the function generates (D) and mining algorithms (A) can result in different wrapper algorithms. Since mining algorithms are used to control the selection of feature subsets, the wrapper model tends to give superior performance as feature subsets found are better suited to the predetermined mining.

Evolutionary algorithms (EAs) operate on a population of individuals that represent possible solutions to a problem.

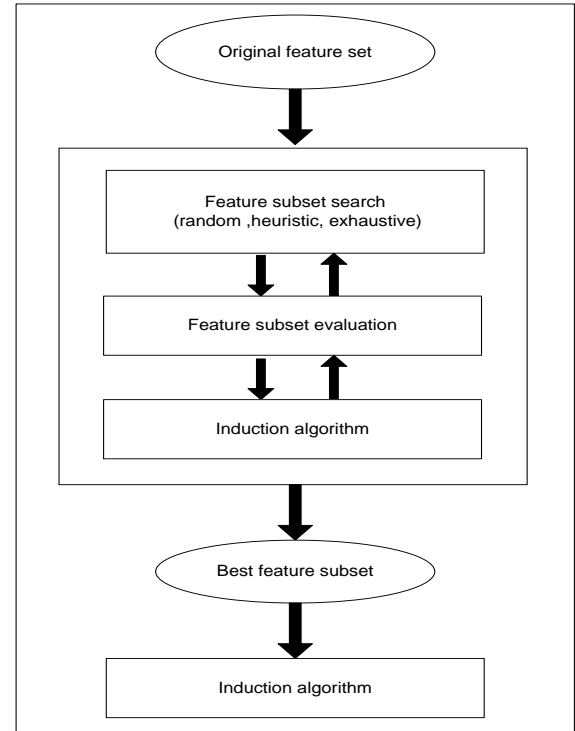


Fig. 1 wrapper approach

wrapper algorithm

input: $D(F_0, F_1, \dots, F_{n-1})$ // a training data set with N features
 S_0 // a subset from which to start the search
 δ // a stopping criterion
output: S_{best} // an optimal subset
begin
 initialize: $S_{best} = S_0$;
 $\gamma_{best} = eval(S_0, D, A)$; // evaluate S_0 by a mining algorithm A
 do begin
 $S = generate(D)$; // generate a subset for evaluation
 $\gamma = eval(S, D, A)$; // evaluate the current subset S by A
 if (γ is better than γ_{best})
 $\gamma_{best} = \gamma$;
 $S_{best} = S$;
 end until (δ is reached);
return S_{best} ;
end;

Fig. 2 Wrapper approach pseudo code

The initial population of individuals may be created entirely at random, or some knowledge about previously known solutions may be used to seed the population. The algorithm evaluates the individuals to determine how well they solve the problem at hand using a user defined

objective function, which is unique to each problem; genetic programming is a kind of (EAs).

The main advantages of evolutionary algorithms are [41]:

1. Being robust to dynamic changes: Traditional methods of optimization are not robust to dynamic changes in the environment and they require a complete restart for providing a solution. In contrary, evolutionary computation can be used to adapt solutions to changing circumstances.
2. Broad applicability: Evolutionary algorithms can be applied to any problems that can be formulated as function optimization problems.
3. Hybridization with other methods: Evolutionary algorithms can be combined with more traditional optimization techniques.
- (4) Solves problems that have no solutions: The advantage of evolutionary algorithms includes the ability to address problems for which there is no human expertise. Even though human expertise should be used when it is needed and available; it often proves less adequate for automated problem-solving routines.

The traditional operators used in both GP and Gas are very similar. However the one used in GP have been is customized in order to work with trees so the offspring do not exceed the maximum depth of the tree.

The operators can be divided into primary genetic operator which includes the reproduction and crossover operators. While the mutation and other novel operators considers secondary ones. The reproduction operator acquire an identical offspring by simply copying the individual. While the crossover operator requires two parents and randomly selects a node in each as a crossover point. The sub-trees rooted at these crossover points are swapped to obtain the offspring, as shown in figure 3.

It is worth mentioning here that there exist many differences in the crossover between GP and Gas. Firstly, in GP the cross over between two identical parents may result in offspring that are completely different from the parent. Secondly, if the root of the parents are crossover points then the crossover degenerates to reproduction. Thirdly, crossover is like point mutation in case if the terminal nodes are two crossover points. In this case it simply changes only single node.

Sub-tree or grow mutation is the most used form of mutation.

Figure 4 illustrates this form where the randomly selected node from the tree which constitute the root of a sub-tree. This sub-tree will be replaced by a new randomly generated sub-tree.

Despite the questionability of the usefulness of mutation in GAs by some researchers, it still plays an important role in relieving the search from stagnation and it re-enter lost alleles into the population. However, mutation is not necessary in GP to avoid stagnation due to the aforementioned fact that an identical parents may result in different offspring through crossover.

The chance for a symbol to disappear from the population is very unlikely due to the following two facts: First, the allele meaning in GP is not related to its position. Second, the number of nodes in the population is relatively very large comparing to the size of both the function and terminal sets.

The main motivation for choosing GA as a random selection approach is its capability to effectively explore large search spaces which is based on the principle of natural selection. This is achieved by evolving a population of individuals, where each individual is a candidate solution to a given problem.

In summary, three operators constitute the genetic algorithm:

- Reproduction which selects good string.
- Crossover which combines good strings in order to create better offspring's.
- Mutation which alters a string locally for the purpose of generating a better string.

The algorithm is terminated at any iteration if the population evaluation and testing satisfy the termination criterion; otherwise the three GA operators will be re-applied on the population and then re-evaluated. This procedure is continued until the termination criterion is met [42]. The above approach is illustrated as a flow chart in figure 5.

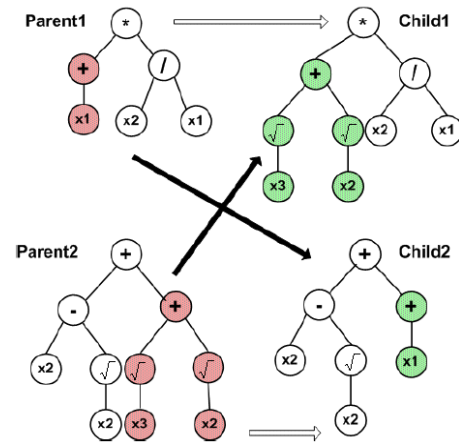


Fig. 3 GP crossover process

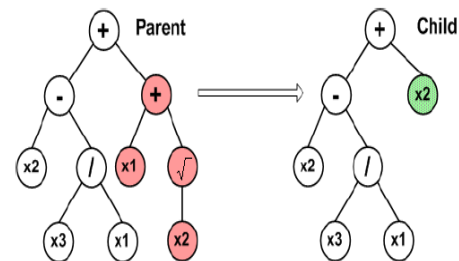


Fig. 4 GP crossover

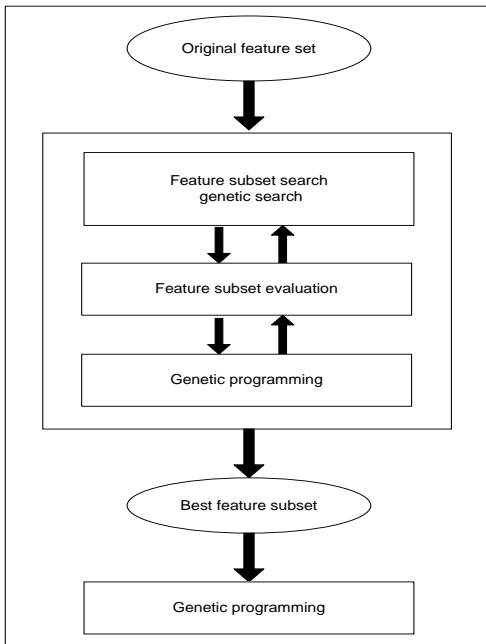


Fig. 5 Proposed WGA method

4. Experiment Results

In this section, the computational results of the proposed methods are presented. The proposed algorithms are programmed using J2EE Java and performed on an Intel Pentium 4, 2.33 GHz computer, and tested on 10 well-known University of California Irvine (UCI) data sets, which have been used by many researchers, are used to test the performance of the proposed algorithms. Table 1 presents a detail description of the data sets used in the experiments. These data sets can be freely downloaded from <http://www.ics.uci.edu/~mlearn/> [43]. The performance of the proposed algorithms is evaluated based on the minimal number of selected attributes, the classification accuracy.

The comparisons in this paper are carried out in terms of the minimal attributes, the classification accuracy. The purpose of this comparison is to evaluate the effectiveness of using a genetic programming classifier with wrapper evaluator in obtaining the minimal attributes, and its effect on the classification accuracy (based on the obtained minimal attributes).

The superscripts in parentheses represent the number of runs that achieved this number of attributes, while the number of attributes without superscripts means that the method could obtain that number of attributes in all of the runs. Further comparison is carried out in terms of the classification accuracy for all data sets table 2 provides

details of the minimal attributes and classification accuracy together that are obtained from WGP.

The results shown in Table 3 indicate that WGP is capable of improving classification accuracy between feature selection' with and without attribute reduction' by WGP for six data sets (Exactly, Derm, LED, Credit, Heart, Vote) as shown in bold and achieved the same accuracy for four data set (M-of-N, Mushroom, Lung, Derm2) with accuracy between 95.35% to 98.44%. The results for the classification accuracy in table 3 show that WGP approach is able to get the best result for five information sets (Heart, Vote, Credit, Derm, Drem2), compared to all methods in the literature.

Table 1: UCI datasets

Datasets	No of Attributes	No. of Objects
M-of-N	13	1000
Exactly	13	1000
Heart	13	294
Vote	16	300
Credit	20	1000
Mushroom	22	8124
LED	24	2000
Drem	34	366
Drem2	34	358
Lung	56	32

Conclusion

In literature, a lot of research questions were once raised to identify robust approach to sort out the attribute discount concern. As a way to answer this research query, two research pursuits have been outlined with the intention of proposing amazing approach to seek out essentially the most informative and minimal attributes with least information loss. This aim has been done through (WGP) approaches.

In this paper, genetic programming algorithm was customize to solving the attribute reduction problem. It was shown that using an intelligent mechanism of genetic programming to produce new programs to increase the probability of finding high-quality reducts. Experimental outcome confirmed that the proposed process (WGP) tremendously outperformed the genetic programming and many other procedures regarding minimal attributes and classification accuracy. The effectiveness of the hybridized wrapper and genetic programming in terms of the classification accuracy is the ability to make a decision on essentially of the most informative attributes to be incorporated in the reductions. The evaluation of the effectiveness of the solution got used to be measured no longer most effective in terms of minimal attributes, but in addition in phrases of classification accuracy.

Table 2 Comparison between feature selection without attribute reduction and feature selection with (WGP) in terms of minimal attributes and classification accuracy.

Datasets	No. of Attr.	Accuracy with genetic programming		Accuracy with wrapper on genetic programming evaluator		No. of Attributes With attribute reduction on max	No. of Attributes With attribute reduction on min
		min	Max	min	max		
Vote	16	61.33	97.67	61.37	97.00	7 : 2,4,5,9,11,12,13	8 : 2,3,4,8,9,11,12,16
M-of-N	13	93.50	100	93.9	100	10:1,2,3,5,7,8,9,10,11,12	9 : 1,2,3,4,5,7,8,9,11
Exactly	13	68.80	91.00	67.7	78.90	8 : 1,3,4,5,7,9,11,12	7 : 2,3,6,8,9,12,13
credit	20	73.40	76.00	72.7	76.10	10:1,2,3,5,6,9,11,15,19,20	11:1,2,3,5,6,9,10,11,15,16,19
Derm	34	96.17	97.81	93.71	98.63	16:1,4,5,7,8,9,14,15,20,22,26,27,28,31,32,33	19:1,2,5,6,7,9,10,13,14,15,18,20,21,22,23,26,28,30,33
Heart	13	82.99	85.37	84.01	93.71	4 : 5,10,11,13	5 : 1,7,10,11,12
LED	24	75.1	98.20	71.7	92.15	8 : 1,2,3,4,5,7,10,15	7 : 1,2,3,4,5,7,15
Derm2	34	96.65	98.88	95.81	99.44	17:3,5,6,7,8,9,13,15,21,22,25,26,28,29,30,33,34	14:3,5,6,8,13,15,20,21,22,24,25,26,28,31
Lung	56	96.88	100	96.87	100	5 : 9,42,43,47,48	27:3,5,6,9,10,11,12,13,14,19,20,21,22,23,24,25,27,28,29,33,39,42,47,48,49,52,55
mushroom	22	91.78	98.92	92.56	98.17	13:1,4,5,7,10,11,12,14,15,18,19,21,22	10:5,7,8,12,13,14,16,17,18,20

Table 3 Comparison with the literature in terms of classification accuracy

Datasets	Fuzzy GD Accruy[34]	Fuzzy RRT	GA-Fuzzy GD Accruy [44]	GA-Fuzzy RRT [45]	GA [45]	WGA
M-of-N	100%	100%	100%	100%	100%	96.75%
Heart	69.17%	31%	14%	0%	100%	94.18%
Mushroom	100%	100%	100%	100%	100%	95.35%
Vote	70%	70%	63%	63%	63%	83.4%
Credit	73%	72.70%	6%	8%	21%	74.7%
Lung	100%	100%	100%	100%	100%	98.44%
LED	100%	100%	100%	100%	100%	87.98%
Derm	49.57%	48.05%	24%	19%	38%	96.99%
Drem2	53.34%	49.72	86%	78%	75%	97.77%
Exactly	100%	100%	100%	100%	64%	79.9%

The comparison of WGP with different on hand approaches in terms of minimal attributes tested the effectiveness of the proposed strategy. However, several questions still need to be answered to enhance the proposed algorithms as a future work; the proposed approaches can be validated and extended to real valued data sets from different domains and real-world problems such as web and text mining, bioinformatics and speech recognition, which may yield interesting results.

References

- [1] J. R. Anaraki and M. Eftekhari, "Improving fuzzy-rough quick reduct for feature selection," in Electrical Engineering (ICEE), 2011 19th Iranian Conference on, 2011, pp. 1-6.
- [2] T. Deng, C. Yang, and X. Wang, "A Reduct Derived From Feature Selection," Pattern Recognition Letters, vol. 33, pp. 1638-1646, 2012.
- [3] M. M. Kabir, M. M. Islam, and K. Murase, "A new wrapper feature selection approach using neural network," Neurocomputing, vol. 73, pp. 3273-3283, 2010.
- [4] N. Al-Najdawi, S. Tedmori, O. A. Alzubi, O. Dorgham, and J. A. Alzubi, "A Frequency Based Hierarchical Fast Search Block Matching Algorithm for Fast Video Communication," International journal of advanced computer science and applications, vol. 7, no. 4, pp. 447-455, 2016.
- [5] J. A. Alzubi, "Diversity-Based Boosting Algorithm," International journal of advanced computer science and applications, vol. 7, no. 5, 2016.
- [6] H. Liu and H. Motoda, Feature extraction, construction and selection: A data mining perspective: Springer Science & Business Media, 1998.
- [7] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," Knowledge and Data Engineering, IEEE Transactions on, vol. 17, pp. 491-502, 2005.
- [8] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in ICML, 2003, pp. 856-863.
- [9] K. Michalak and H. Kwasnicka, "Correlation-based feature selection strategy in neural classification," in Intelligent

- Systems Design and Applications, 2006. ISDA'06. Sixth International Conference on, 2006, pp. 741-746.
- [10] S.-U. Guan, J. Liu, and Y. Qi, "An incremental approach to contribution-based feature selection," *Journal of Intelligent Systems*, vol. 13, pp. 15-42, 2004.
- [11] O. Almomani, M. Al-shugran, J. A. Alzubi, and O. A. Alzubi, "Performance Evaluation of Position-based Routing Protocols using Different Mobility Models in MANET," *International journal of computer application*, vol. 119, pp. 43-48.
- [12] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial intelligence*, vol. 97, pp. 273-324, 1997.
- [13] I. H. Osman and J. P. Kelly, *Meta-heuristics: theory and applications*: Springer Science & Business Media, 2012.
- [14] S. Abdullah, L. Golafshan, and M. Z. A. Nazri, "Re-heat simulated annealing algorithm for rough set attribute reduction," *International Journal of the Physical Sciences*, vol. 6, pp. 2083-2089, 2011.
- [15] H. Zhang and G. Sun, "Feature selection using tabu search method," *Pattern recognition*, vol. 35, pp. 701-711, 2002.
- [16] C.-H. Yang, L.-Y. Chuang, and C. H. Yang, "IG-GA: a hybrid filter/wrapper method for feature selection of microarray data," *Journal of Medical and Biological Engineering*, vol. 30, pp. 23-28, 2010.
- [17] M. Alweshah and S. Abdullah, "Hybridizing firefly algorithms with a probabilistic neural network for solving classification problems," *Applied Soft Computing*, vol. 35, pp. 513-524, 2015.
- [18] M. Alweshah, "Firefly Algorithm with Artificial Neural Network for Time Series Problems," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 7, pp. 3978-3982, 2014.
- [19] J. P. Rushton and A. R. Jensen, "Wanted: More race realism, less moralistic fallacy," *Psychology, Public Policy, and Law*, vol. 11, p. 328, 2005.
- [20] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Proceedings of the ninth international workshop on Machine learning*, 1992, pp. 249-256.
- [21] P. M. Narendra and K. Fukunaga, "A branch and bound algorithm for feature subset selection," *Computers, IEEE Transactions on*, vol. 100, pp. 917-922, 1977.
- [22] M. Dash and H. Liu, "Feature selection for classification," *Intelligent data analysis*, vol. 1, pp. 131-156, 1997.
- [23] H. Arafat, R. M. Elawady, S. Barakat, and N. M. Elrashidy, "Using rough set and ant colony optimization in feature selection," *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, vol. 2, 2013.
- [24] A. J. Tallón-Ballesteros and J. C. Riquelme, "Tackling ant colony optimization meta-heuristic as search method in feature subset selection based on correlation or consistency measures," in *Intelligent Data Engineering and Automated Learning-IDEAL 2014*, ed: Springer, 2014, pp. 386-393.
- [25] J. Wang, Q. Zhang, H. Abdel-Rahman, and M. I. Abdel-Monem, "A rough set approach to feature selection based on scatter search metaheuristic," *Journal of Systems Science and Complexity*, vol. 27, pp. 157-168, 2014.
- [26] W. Yongxiong and K. Li, "Feature and weight selection using Tabu search for improving the recognition rate of duct anomaly," in *Robotics and Biomimetics (ROBIO)*, 2014 IEEE International Conference on, 2014, pp. 2163-2168.
- [27] T. Matsui, Y. Katagiri, H. Katagiri, and K. Kato, "Automatic Feature Point Selection through Hybrid Metaheuristics based on Tabu Search and Memetic Algorithm for Augmented Reality," *Procedia Computer Science*, vol. 60, pp. 1120-1127, 2015.
- [28] A. Unler and A. Murat, "A discrete particle swarm optimization method for feature selection in binary classification problems," *European Journal of Operational Research*, vol. 206, pp. 528-539, 2010.
- [29] I. A. Gheyas and L. S. Smith, "Feature subset selection in large dimensionality domains," *Pattern recognition*, vol. 43, pp. 5-13, 2010.
- [30] M. Moghadasian and S. P. Hosseini, "Binary Cuckoo Optimization Algorithm for Feature Selection in High-Dimensional Datasets," in *International Conference on Innovative Engineering Technologies (ICIET'2014)*, 2014, pp. 18-21.
- [31] M. Schiezero and H. Pedrini, "Data feature selection based on Artificial Bee Colony algorithm," *EURASIP Journal on Image and Video Processing*, vol. 2013, pp. 1-8, 2013.
- [32] M. Mafarja and S. Abdullah, "Investigating memetic algorithm in solving rough set attribute reduction," *International Journal of Computer Applications in Technology*, vol. 48, pp. 195-202, 2013.
- [33] S. Abdullah and N. S. Jaddi, "Great deluge algorithm for rough set attribute reduction," in *Database theory and application, bio-science and bio-technology*, ed: Springer, 2010, pp. 189-197.
- [34] M. Mafarja and S. Abdullah, "Modified great deluge for attribute reduction in rough set theory," in *Fuzzy Systems and Knowledge Discovery (FSKD)*, 2011 Eighth International Conference on, 2011, pp. 1464-1469.
- [35] E. Emary, H. M. Zawbaa, K. K. A. Ghany, A. E. Hassanien, and B. Parv, "Firefly Optimization Algorithm for Feature Selection," in *Proceedings of the 7th Balkan Conference on Informatics Conference*, 2015, p. 26.
- [36] J. Kittler, "Feature set search algorithms," *Pattern recognition and signal processing*, pp. 41-60, 1978.
- [37] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern recognition letters*, vol. 15, pp. 1119-1125, 1994.
- [38] P. Somol, P. Pudil, and J. Kittler, "Fast branch & bound algorithms for optimal feature selection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, pp. 900-912, 2004.
- [39] I. Inza, P. Larrañaga, R. Etxeberria, and B. Sierra, "Feature subset selection by Bayesian network-based optimization," *Artificial intelligence*, vol. 123, pp. 157-184, 2000.
- [40] M. E. ElAlami, "A filter model for feature subset selection based on genetic algorithm," *Knowledge-Based Systems*, vol. 22, pp. 356-362, 2009.
- [41] S. Sivanandam and S. Deepa, *Introduction to genetic algorithms*: Springer Science & Business Media, 2007.
- [42] A. G. Karegowda, M. Jayaram, and A. Manjunath, "Feature subset selection problem using wrapper approach in supervised learning," *International journal of Computer applications*, vol. 1, pp. 13-17, 2010.
- [43] C. Blake and C. J. Merz, "{UCI} Repository of machine learning databases," 1998.
- [44] M. Majdi, S. Abdullah, and N. S. Jaddi, "Fuzzy Population-Based Meta-Heuristic Approaches for Attribute Reduction in

Rough Set Theory," World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering, vol. 9, pp. 2065-2073, 2015.

- [45] M. Mafarja and S. Abdullah, "A fuzzy record-to-record travel algorithm for solving rough set attribute reduction," International Journal of Systems Science, vol. 46, pp. 503-512, 2015.



Mohammed Alweshah received his BS in computer sciences in 1993, from Al-Mustansiriah University, Iraq and his master degree of computer sciences in 2005, from Al-Balqa Applied University, Jordan. Mohammed Alweshah received his Ph.D. degree from the Computer Science Department, School of Information Technology, National University of

Malaysia (UKM), Malaysia in 2013. He was working under the supervision of Prof. Salwani Abdullah. Currently, Alweshah is a assistance professor with the Prince Abdullah Bin Ghazi Faculty of Information Technology, Al-Balqa Applied University, AlSalt-Jordan. His research interests include Heuristics and Meta-heuristics, Evolutionary Algorithms, and data mining problems.



Omar A. Alzubi was born in Jordan. He received his Master degree with distinction in Computer and Network Security from New York Institute of Technology (New York, USA) in 2006. He also holds Ph.D. degree in Computer and Network Security from Swansea University (Swansea, UK) in 2013. He joined Al-Balqa Applied University since 2013 as an assistant

professor in computer and network security.

Dr. Alzubi research interest includes machine learning, network security, cloud security, application of Algebraic-Geometric theory in channel coding, and Elliptic curve cryptosystems. He is also involved in UK-Turkey Higher Education Partnership Program 2011-2013 projects where he proposed a cryptosystem based on Elliptic curves.



Jafar A. Alzubi received a B.Sc (Hons) in Electrical Engineering, majoring Electronics and Communications from the University of Engineering and Technology, Lahore, Pakistan in 2001. In 2005 received M.Sc. (Hons) in Electrical and Computer Engineering from New York Institute of Technology, New York, USA. In 2008, He joined the Wireless Communications Research Laboratory at Swansea University

(Swansea, UK), completing his PhD in Advanced Telecommunications Engineering in June 2012. He is now an Assistant professor at Computer Engineering department, Al-Balqa Applied University; also he is deputy dean of Engineering Faculty. His research interests include Elliptic curves cryptography, classification, using Algebraic-Geometric theory. As part of his research, he designed the first regular and first irregular block turbo codecs using Algebraic Geometry codes and

investigated their performance across various computer and wireless networks.



Sharihan Alaqeel received his BS in computer sciences in 2005, from Al-Hussien University, Jordan and his master degree of computer sciences in 2016, from Al-Balqa Applied University, Jordan. She was working under the supervision of Dr. Mohammed Alweshah. Currently, Alaqeel is a Laboratory supervisor with the Al-Ahliyah Amman University, AlSalt-Jordan. His research interests include Evolutionary Algorithms, and data mining problems.