Jurnal
Teknologi

Full Paper

ATTRIBUTE SELECTION MODEL FOR OPTIMAL LOCAL SEARCH AND GLOBAL SEARCH

Mohammad Aizat Basir[a*], Faudziah Ahmad[b]

[a]School of Informatics and Applied Mathematics (PPIMG), Universiti Malaysia Terengganu, 21030 Kuala Terengganu Terengganu, Malaysia
[b]UUM College of Arts and Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia

Article history
Received
8 May 2016
Received in revised form
30 June 2016
Accepted
15 September 2016

*Corresponding author
aizat@umt.edu.my

Abstract

Attribute selection also known as feature selection is an essential process in data sets that comprise numerous numbers of input attributes. However, finding the optimal combination of algorithms for producing a good set of attributes has remained a challenging task. The aim of this paper is to find a list of an optimal combination search methods and reduction algorithm for attribute selection. The research process involves 2 phases: finding a list of an optimal combination search methods and reduction algorithm. The combination is known as model. Results are in terms of percentage of accuracy and number of selected attributes. Six (6) datasets were used for experiment. The final output is a list of optimal combination search methods and reduction algorithm. The experimental results conducted on public real dataset reveals that the model consistently shows the suitability to perform good classification task on the selected dataset. Significant improvement in accuracy and optimal number of attribute selection is achieved with a list of combination algorithms used.

Keywords: Attribute selection, reduction algorithm, search methods, classification

Abstrak

Pemilihan atribut juga dikenali sebagai pemilihan ciri merupakan proses penting dalam set data yang terdiri daripada bilangan yang lebih banyak sifat-sifat input. Walau bagaimanapun, adalah tugas yang mencabar untuk mencari gabungan optimum algoritma untuk menghasilkan satu sifat-sifat set yang baik. Tujuan kertas kajian ini adalah untuk mencari senarai yang kaedah carian gabungan optimum dan algoritma pengurangan untuk pemilihan atribut. Proses penyelidikan melibatkan 2 fasa; mencari senarai yang kaedah carian gabungan optimum dan algoritma pengurangan. Gabungan ini dikenali sebagai model. Keputusan adalah dari segi peratusan ketepatan dan beberapa ciri-ciri yang dipilih. Enam (6) set data telah digunakan untuk eksperimen. Output akhir adalah senarai kaedah carian gabungan optimum dan algoritma pengurangan. Keputusan eksperimen dijalankan ke atas set data awam sebenar menunjukkan bahawa model secara konsisten menunjukkan kesesuaian untuk melaksanakan tugas pengelasan baik pada set data yang dipilih. Peningkatan yang ketara dalam ketepatan dan jumlah optimum atribut yang dipilih dicapai dengan menggunakan senarai algoritma gabungan.

Kata kunci: Pemilihan atribut, algoritma pengurangan, kaedah carian, pengelasan

1.0 INTRODUCTION

Real world dataset usually consist a large number of attributes. It is very common some of those input attributes could be irrelevant and consequently give an impact to the design of a classification model. In situations where a rule has too many conditions, it becomes less interpretable. Based on this understanding, it becomes important to reduce the dimensionality (number of input attributes in the rule) of the rules in the rule set. In practical situations, it is recommended to remove the irrelevant and redundant dimensions for less processing time and labour cost. The amount of data is directly correlated with the number of samples collected and the number of attributes. A dataset with a large number of attributes is known as a dataset with high dimensionality [1]. The high dimensionality of datasets leads to the phenomenon known as the curse of dimensionality where computation time is an exponential function of the number of the dimensions. It is often the case that the model contains redundant rules and/or variables. When faced with difficulties resulting from the high dimension of a space, the ideal approach is to decrease 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 could be overcome by reducing the number of attributes in a dataset and removing irrelevant or less significant rules. This can reduce the computation time, and storage space. Models with simpler and small number of rules are often easier to interpret.

The main drawback of rule/attribute complexity reduction is the possibility of information loss. It is important to point out those two critical aspects of attribute reduction problems 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 performs reduction in less time but could not guarantee a minimal subset [1]–[3] whereas other hybrid methods which combine rough set and swarms algorithm such as GenRSAR, AntRSAR, PSO-RSAR and BeeRSAR methods improve the performance but consume more time [1], [2].

In feature selection, also known as variable selection, attribute selection or variable subset selection is the process of selecting a subset of relevant features (attributes) for use in model construction. It is the process of choosing a subset of original features so that the feature space is optimally reduced to evaluation criterion. Feature selection can reduce both the data and the computational complexity. The raw data collected is usually large, so it is important to select a subset of data by creating feature vectors. Feature subset selection is the process of identifying and removing much of the redundant and irrelevant information possible.

Feature selection in general 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, analysing all subsets in any order and search will get completed in a short time. However, the search space is usually not small, 2^N 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. Although feature selection is a well-developed research area with various methods [4], researchers still try to find better methods to make their classifiers more efficient with possible option which is combinations of generation procedures and evaluation functions [5].

2.0 RELATED WORKS

There are several feature selection search methods. One of them is Best First Search. Best First Search is a feature selection method based on artificial intelligence, which allows backtracking in the search space [6]. This algorithm, similar to the Greedy Hill Climbing algorithm, makes use of local changes in the search space. But in contrast to it when the path for reaching the optimum solution is not hopeful, it is possible to backtrack the search space. Linear search (LS), an extension of Best First search [7] searches the space of feature subsets by Greedy Hill-Climbing augmented with a backtracking facility. Another method by Hamdani *et al*., proposed a new algorithm based on genetic algorithms with bi-coded chromosome representation and new evaluation function [8]. It used a Hierarchical algorithm with homogeneous and heterogeneous population to minimize the computational cost and speed up the convergence time. They claimed, heterogeneous GA performs a global search among the solutions with different sizes and then a number of best solutions are sent to homogeneous GAs to locally optimize the solutions. Due to the parallel nature of their proposed method, the method showed good performance when compared with heuristic algorithms and simple GA.

Genetic search (GS), another feature selection method, is a randomized search method which performs using a simple genetic algorithm [9]. The genetic algorithm finds the feature subset to maximize special output function using techniques inspired by natural evolution. Rank search (RS) uses a feature evaluator (such as gain ratio) to rank all the features. After a feature evaluator is specified, a forward selection search is used to generate a ranking list [10]. Scatter search (SS) [11] has been developed to perform a scatter 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 improvement any longer.

Stepwise search (SWS) is a variation of the Forward search that performs a test to check if a feature can be eliminated without significant reduction in the output function [12]. Unlike other methods, the Tabu search (TS) is proposed for combinatorial optimization problems. It combines a local search with anti-cycling memory-based rules to avoid trapping in local optimal solutions [13]. It performs interactive search by traversing the feature subset space to maximize the target function while taking consideration of the interaction among features. Fast Correlation-Based Filter search (FCBF) [14] has been invented to evaluate features via the relevance and redundancy analysis, and uses the analysis results as guideline to choose features. Similarly, Ranker identifies important features by evaluating each feature individually and ranks the features by the values of their evaluation metrics [15]–[17].

Several global and local search algorithms have been deployed for optimization purposes. Seymour *et al*. [18] performed an experimental comparison of several global (random search, a genetic algorithm, simulated annealing, particle swarm) and local (Nelder-Mead and orthogonal search) optimization algorithms. Similarly, Kisuki *et al*. [19] compared random search, a genetic algorithm, and simulated annealing with pyramid search and window search. In both these studies, the experimental results showed that the random search was more effective than the other algorithms tested. This reason is that in the tuning tasks considered, the number of high-performing parameter configurations is large and hence it is easy to find one of them. While Norris *et al*. [20] implemented the Nelder-Mead simplex method, simulated annealing, and a genetic algorithm in the empirical performance tuning framework Orio, the authors did not conduct an experimental comparison. A number of previous works deploy local search algorithms for empirical performance tuning. Examples include orthogonal search in ATLAS [21], pattern search in loop optimization [22], and a modified Nelder-Mead simplex algorithm in Active Harmony [23], [24].

Ye, Chen, and Liao (2007) have presented a new algorithm for minimum attribute reduction based on Binary Particle Swarm Optimization (BPSO) with Vaccination [25]. Their research started with transformation of the problem of minimum attribute reduction into an unconstrained binary optimization problem. Then they defined suitable fitness function and the equivalence of optimality between the original problems to prove that the transformation has been done. In the next step, they approached to solve the transformed problem using an improved BPSO algorithm combined with some vaccination mechanism. Experimental results on a number of data sets obtained from the UCI machine learning repository show that the proposed algorithm has a higher possibility of finding a minimum reduction and remarkably outperforms some existing algorithms specifically designed for minimum attribute reduction in both quality of solution and computational complexity.

New heuristic approach for solving the minimal attribute reduction problem (MARP) based on the ant colony optimization (ACO) meta-heuristic has been proposed [26]. They developed a new algorithm R-ACO for solving the MARP and the simulation results claimed that their approach can find more minimal attribute reductions more efficiently in most cases. Basically their research improved previous works [1], [27]. The improvement involved reducing the time cost operation by proposed a new model R-Graph to solve the MARP with ACO. With this approach, they solved the problem with increased more reductions especially towards achieving minimal reductions.

Rough Set-based Attribute Reduction (RSAR) namely Independent RSAR hybrid with Artificial Bee Colony (ABC) algorithm has been introduced [28]. They grouped the instances based on decision attributes. Then, they applied Quick Reduct Algorithm [29] 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 at-tributes. An experiment was carried out on five different da-tasets from the UCI machine learning repository. The performance of the reduct is analyzed with Genetic k-Nearest Neighbor (GKNN) classifier and compared with six different algorithms (general RSAR, Entity based Reduct (EBR), Genetic RSAR, Ant RSAR, Particle Swarm Optimization based RSAR (PSORSAR) and with their previous work (BeeRSAR). They claimed the proposed method can find very minimal reduct than the other existing methods.

Zhang *et al*. (2006) has presented the use of AFSA as a new tool which sets up a neural network (NN), adjusts its parameters, and performs feature reduction, all simultaneously. They combined the feature selection and NN architecture problem into an optimization procedure and employs AFSA to resolve it [30]. Also, they performed the feature selection and evolving NN architecture at the same time by AFSA, which is not based on a fixed network. Results showed that their proposed meth-od were able to optimize network architecture to be kept simple, reduce computation and enhance generalization ability of the resulting classifier.

Narendra and Fukunaga (1977) presented an algorithm [31] claimed to be efficient and select the best subset without exhaustive search. A result has shown that the best 12-feature set from a 24-feature set was selected with the computational effort of evaluating only 6000 subsets. Saeys et. al (2007) discussed a basic taxonomy of feature selection techniques, and discussing their use, variety and potential in a number of both common as well as upcoming bioinformatics applications.

This paper aims to present the model for optimal local and global search. The model consists of the best combination of search methods and reduction algorithms. Several attribute selection search methods were explored. Different reduction algorithm methods are experimented together with various attribute selection search methods for finding the best model. Then the model was further tested with 5 datasets.

This paper is organized as follow: - Section 1 briefly on introduction; Section 2 discusses on related works; Section 3 describes on methodology; Section 4 performs results and discussion and Lastly Section 5 present the conclusions.

3.0 METHODOLOGY

The methodology is shown in Figure 1. It consists of eight parts in two phases: (1) data collection; (2) data pre-processing; (3) dimensionality reduction; (4) data training and testing for attribute reduction; (5) model testing & verification; (6) test model with various dataset; (7) model with good accuracy: compare classification accuracy and time to build model for original datasets and reduct datasets. The expected output from phase 1 is the good model which consists of best search methods with best reduction algorithm.

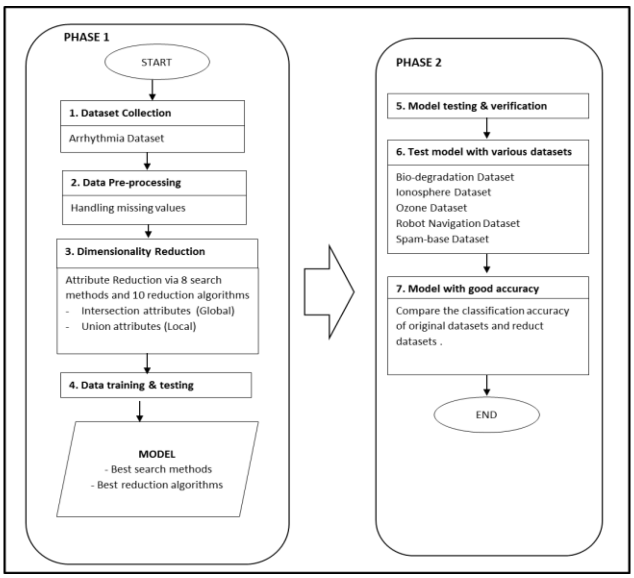

Figure 1 Methodology

Step 1 (Data Collection): Arrhythmia datasets was selected from UCI Machine Learning Repository. Arrhythmia dataset was selected due to its many features that make it challenging to explore [32].

Step 2 (Data Pre-processing): Dataset that has missing values has been pre-processed in order to make sure that dataset is ready to be experimented. In this step, dataset that has missing value (?) replaced with 0 & mean value. This approach has been tested and results show no different in term of performance. This research decided to replace missing value with 0 values.

Step 3 (Dimensionality Reduction) : 8 search methods and 10 reduction algorithms has been used in order to get the best model for finding local & global search of attributes. With these two searches, exploration and exploitation will be balanced, hence solution space is searched effectively [33]. The search methods and reduction algorithms selected are popular algorithms and widely used in data mining analysis. 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. Considering example below:

Original Dataset WEATHER={A1,A2,A3,A4,A5,A6,A7,A8,A9,A10,A11,A12,A13}

Combination Reduction algorithm A + Search method A = Reduct Dataset {A1, A2, A3}

Combination Reduction algorithm A + Search method B= Reduct Dataset {A2, A3, A6}

Combination Reduction algorithm A + Search method C = Reduct Dataset {A3, A2, A9}

Combination Reduction algorithm B + Search method A= Reduct Dataset {A10, A2, A13}

Combination Reduction algorithm B + Search method B= Reduct Dataset {A4, A13, A10}

Combination Reduction algorithm B + Search method C = Reduct Dataset {A7, A10, A13}

The attribute intersection of Dataset WEATHER for Reduction algorithm A with Search method A, B and C is {A2, A3}.

The attribute intersection of Dataset WEATHER for Reduction algorithm B with Search method A, B and C is {A10, A13}.

Based on example given, the union of the attributes can be found between two intersections which is {A2, A3, A10, and A13}.

Step 4 (Data 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.

Step 5 (Model testing & verification): Next, the model produced in phase 1 was further tested in phase 2 in order to confirm the correctness of the model.

Step 6 (Test model with various dataset): In this step, model was tested with various datasets to confirm the correctness of the model.

Step 7 (Model with good accuracy): In this step, the accuracy of original dataset has been compared with reduct datasets. The output of this step is the classification accuracy with optimal number of attributes.

Standard six datasets namely Arrhythmia, Bio-degradation, Ionosphere, Ozone, Robot Navigation and Spam-base from the UCI [34] were used in the experiments. These dataset include discrete and continuous attributes and represent various field of data. The reason for choosing this dataset is to confirm the model suits all field of data. The information on the datasets is shown in Table 1.

Table 1 Dataset characteristics

Dataset	# 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 five (5) datasets were tested using 8 search methods and 10 reduction algorithms.

WEKA (Waikato Environment for Knowledge Analysis) tool was used in this research. WEKA is a popular suite of machine learning software written in Java, developed at the University of Waikato, New Zealand[35]. WEKA contains a collection of visualization tools and algorithms for data analysis and predictive modeling. It provides many different algorithms for data mining and machine learning. In this research, WEKA was used to perform classification tasks which produce an accuracy rate. Accuracy rate is the percentage of test set samples that are correctly classified by the model. For this research accuracy rate is very important in term of determining how well the model perform to classify the data. Higher accuracy rate mean the better the model.

4.0 RESULTS AND DISCUSSION

The outputs for each phase are presented in this section. The performance results are presented as percentage of accuracy for each list of an optimal combination search methods and reduction algorithm.

Table 2 shows the results of a list of an optimal combination search methods and reduction algorithm.

Table 2 Summary of the model

Search Algorithm	Best Reduction Algorithm	#Sel Attr	% Acc
Best First Search	WrapperSubsetEval + Bayes Net (Classifier)	19	80.29
Genetic Search	WrapperSubsetEval + Bayes Net (Classifier)	127	79.58
Genetic Search	WrapperSubsetEval + Bayes Net (Classifier)	127	79.58
Greedy Stepwise	WrapperSubsetEval + Bayes Net (Classifier)	20	79.81
Greedy Stepwise	WrapperSubsetEval + Bayes Net (Classifier)	20	79.81
LineForwardSearch	WrapperSubsetEval + Naïve Bayes (Classifier)	18	79.14
Ranker	LatentSemanticAnalysis	10	94.15
ScatterSearch	CFSSubsetEval / FilteredSubsetEval	20	78.31
RaceSearch	ClassifierSubsetEval + Bayes Net (Classifier)	26	78.82
RaceSearch	ClassifierSubsetEval + Bayes Net (Classifier)	26	78.82
SubsetSizeForwardSelection	CFSSubsetEval / FilteredSubsetEval	18	78.71
SubsetSizeForwardSelection	CFSSubsetEval / FilteredSubsetEval	18	78.71

In phase 2, five (5) various dataset namely Bio-degradation, Ionosphere, Ozone, Robot Navigation and Spam-base were further tested in order to confirm the performance of the model. The results are shown in Table 3 through Table 7.

Table 3 Performance of the model with bio-degradation dataset

Search Algorithm	Best Reduction Algorithm	#Sel Attr	% Acc
Best First Search	WrapperSubsetEval + Bayes Net (Classifier)	10	84.22
Genetic Search	WrapperSubsetEval + Bayes Net (Classifier)	23	86.60
Genetic Search	WrapperSubsetEval + Bayes Net (Classifier)	23	86.68
Greedy Stepwise	WrapperSubsetEval + Bayes Net (Classifier)	10	84.31
Greedy Stepwise	WrapperSubsetEval + Bayes Net (Classifier)	10	84.22
LineForwardSearch	WrapperSubsetEval + Naïve Bayes (Classifier)	6	83.73
ScatterSearch	CFSSubsetEval / FilteredSubsetEval	14	81.85
RaceSearch	ClassifierSubsetEval + Bayes Net (Classifier)	7	80.65
RaceSearch	ClassifierSubsetEval + Bayes Net (Classifier)	7	81.32
SubsetSizeForwardSelection	CFSSubsetEval / FilteredSubsetEval	14	81.75
SubsetSizeForwardSelection	CFSSubsetEval / FilteredSubsetEval	14	83.15

Table 2 shows the performance of the model with Bio-degradation dataset. Genetic search algorithm with WrapperSubsetEval + Bayes Net has a performance of more than 85% accuracy together with 23 selected attribute from 41 attributes of original dataset.

Table 4 Performance of the model with ionosphere dataset

Search Algorithm	Best Reduction Algorithm	#Sel Attr	% Acc
Best First Search	WrapperSubsetEval + Bayes Net (Classifier)	8	95.04
Genetic Search	WrapperSubsetEval + Bayes Net (Classifier)	13	93.55
Genetic Search	WrapperSubsetEval + Bayes Net (Classifier)	13	92.51
Greedy Stepwise	WrapperSubsetEval + Bayes Net (Classifier)	8	95.08
Greedy Stepwise	WrapperSubsetEval + Bayes Net (Classifier)	8	95.04
LineForwardSearch	WrapperSubsetEval + Naïve Bayes (Classifier)	9	94.12
Ranker	LatentSemanticAnalysis	13	92.80
ScatterSearch	CFSSubsetEval / FilteredSubsetEval	5	91.49
RaceSearch	ClassifierSubsetEval + Bayes Net (Classifier)	5	89.69
RaceSearch	ClassifierSubsetEval + Bayes Net (Classifier)	13	92.80
SubsetSizeForwardSelection	CFSSubsetEval / FilteredSubsetEval	13	92.68
SubsetSizeForwardSelection	CFSSubsetEval / FilteredSubsetEval	8	95.04

Table 4 shows the performance of the model with Ionosphere dataset. Greedy Stepwise algorithm with WrapperSubsetEval + Bayes Net performs more than 95% accuracy together with 8 selected attribute from 34 attributes of original dataset.

Table 5 Performance of the model with ozone dataset

Search Algorithm	Best Reduction Algorithm	#Sel Attr	% Acc
Best First Search	WrapperSubsetEval + Bayes Net (Classifier)	5	93.88
Genetic Search	WrapperSubsetEval + Bayes Net (Classifier)	4	93.92
Genetic Search	WrapperSubsetEval + Bayes Net (Classifier)	4	93.92
Greedy Stepwise	WrapperSubsetEval + Bayes Net (Classifier)	5	94.05
Greedy Stepwise	WrapperSubsetEval + Bayes Net (Classifier)	5	93.88
LineForwardSearch	WrapperSubsetEval + Naïve Bayes (Classifier)	4	94.02
Ranker	LatentSemanticAnalysis	12	85.33
ScatterSearch	CFSSubsetEval / FilteredSubsetEval	2	93.49
RaceSearch	ClassifierSubsetEval + Bayes Net (Classifier)	2	93.49
RaceSearch	ClassifierSubsetEval + Bayes Net (Classifier)	15	83.91
SubsetSizeForwardSelection	CFSSubsetEval / FilteredSubsetEval	15	92.62
SubsetSizeForwardSelection	CFSSubsetEval / FilteredSubsetEval	5	93.88

Table 5 shows the performance of the model with Ozone dataset. Greedy Stepwise search algorithm with WrapperSubsetEval + Bayes Net reduction algorithm produced almost minimum number of selected attributes (5 attributes) with highest accuracy of 94%.

Table 6 Performance of the model with robot navigation dataset

Search Algorithm	Best Reduction Algorithm	#Sel Attr	% Acc
Best First Search	WrapperSubsetEval + Bayes Net (Classifier)	5	93.88
Genetic Search	WrapperSubsetEval + Bayes Net (Classifier)	4	93.92
Genetic Search	WrapperSubsetEval + Bayes Net (Classifier)	4	93.92
Greedy Stepwise	WrapperSubsetEval + Bayes Net (Classifier)	5	94.05
Greedy Stepwise	WrapperSubsetEval + Bayes Net (Classifier)	5	93.88
LineForwardSearch	WrapperSubsetEval + Naïve Bayes (Classifier)	4	94.02
Ranker	LatentSemanticAnalysis	12	85.33
ScatterSearch	CFSSubsetEval / FilteredSubsetEval	2	93.49
RaceSearch	ClassifierSubsetEval + Bayes Net (Classifier)	2	93.49
RaceSearch	ClassifierSubsetEval + Bayes Net (Classifier)	15	83.91
SubsetSizeForwardSelection	CFSSubsetEval / FilteredSubsetEval	15	92.62
SubsetSizeForwardSelection	CFSSubsetEval / FilteredSubsetEval	5	93.88

Table 6 shows the performance of the model with Robot Navigation dataset. Subset Size Forward Selection search algorithm with CFSSubsetEval reduction algorithm produced less than 75% from original number of attributes (6 attributes) with the accuracy of 98%.

Table 7 Performance of the model with spam-base dataset

Search Algorithm	Best Reduction Algorithm	#Sel Attr	% Acc
Best First Search	WrapperSubsetEval + Bayes Net (Classifier)	18	93.13
Genetic Search	WrapperSubsetEval + Bayes Net (Classifier)	34	93.48
Genetic Search	WrapperSubsetEval + Bayes Net (Classifier)	34	93.43
Greedy Stepwise	WrapperSubsetEval + Bayes Net (Classifier)	18	93.35
Greedy Stepwise	WrapperSubsetEval + Bayes Net (Classifier)	18	93.13
LineForwardSearch	WrapperSubsetEval + Naïve Bayes (Classifier)	17	93.29
Ranker	LatentSemanticAnalysis	15	92.60
ScatterSearch	CFSSubsetEval / FilteredSubsetEval	8	88.57
RaceSearch	ClassifierSubsetEval + Bayes Net (Classifier)	8	88.62

Search Algorithm	Best Reduction Algorithm	#Sel Attr	% Acc
RaceSearch	ClassifierSubsetEval + Bayes Net (Classifier)	15	92.91
SubsetSizeForwardSelection	CFSSubsetEval / FilteredSubsetEval	15	92.63
SubsetSizeForwardSelection	CFSSubsetEval / FilteredSubsetEval	18	93.13

Table 7 shows the performance of the model with Spam-base dataset. Genetic search algorithm with WrapperSubsetEval + Bayes Net produces 92% accuracy with 34 selected attribute from 57 attributes of original dataset. The strength of the BayesNet that it utilizes the correlation present between the classifiers has the ability to improve the classification performance even if the error rate of individual classifier falls to certain level.

In summary, results shows more than 80% accuracy and large portion of reduction number of attribute achieved with 5 different datasets used. This significance result proved that an optimal list of combination search methods and reduction algorithm can be used in attribute selections for optimal result.

5.0 CONCLUSION

In this paper, 8 attribute selection search methods with 10 reduction algorithms were compared and tested with 6 datasets. The results obtained shows model with an optimal list of combination search methods and reduction algorithm (Table 2). The models were further tested and verified with 5 various datasets to confirm the correctness and validity of the model. The experimental results demonstrates that the model consistently show the suitability to perform good classification task on the selected dataset. However, tradeoffs should be considered when choosing the classification technique to be used. The best model is usually discovered by trial and error on different algorithms. Most of the time researcher must compare or even combinations of available techniques in order to obtain the best possible results.

Acknowledgement

The authors wish to thank Universiti Utara Malaysia (UUM), Universiti Malaysia Terengganu (UMT) and Kementerian Pendidikan Tinggi Malaysia (KPT). This Work supported by UUM, UMT and KPT.

References

[1] Jensen, R., and Shen, Q. 2003. Finding rough set reducts with ant colony optimization. *Proc. 2003 UK Work.* 1(2): 15–22.

[2] Suguna, N., and Thanushkodi, K. 2010. A Novel Rough Set Reduct Algorithm for Medical Domain Based on Bee Colony. 2(6): 49-54.

[3] Yue, B., Yao,W., Abraham,, A., Liu, and H. 2007. A New Rough Set Reduct Algorithm Based on Particle Swarm Optimization. 397-406.

[4] Chandrashekar, G. and Sahin, F. 2014. A Survey On Feature Selection Methods. *Comput. Electr. Eng.* 40(1): 16-28.

[5] Hall , M. a. 1999. Correlation-based Feature Selection for Machine Learning. *Methodology*. 2(1): 1-5.

[6] Hall , M. a. 1999. Correlation-based Feature Selection for Machine Learning. *Methodology*. 2(1): 1-5.

[7] Gütlein, M., Frank, E., Hall, M., and Karwath, A. 2009. Large-scale Attribute Selection Using Wrappers. *IEEE Symposium on Computational Intelligence and Data Mining, CIDM 2009 - Proceedings*. 332-339.

[8] Hamdani, T. M. J,. Won, M. A. M., Alimi, and Karray, F. 2001. Hierarchical Genetic Algorithm With New Evaluation Function And Bi-Coded Representation For The Selection Of Features Considering Their Confidence Rate. *Applied Soft Computing Journal*. 11(2): 2501-2509.

[9] Goldberg, D. E. 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-We.

[10] Battiti, R. 1994. Using Mutual Information For Selecting Features In Supervised Neural Net Learning. *IEEE Trans. Neural Networks*. 5(4): 537-550.

[11] Hedar, A. R., Wang, J., and Fukushima, M. 2008. Tabu Search For Attribute Reduction In Rough Set Theory. *Soft Comput.* 12(9): 909-918.

[12] Kohavi, R. and John, G. 1997. Wrappers For Feature Subset Selection. *Artif. Intell.* 97(1): 273-324.

[13] Hedar, A. R., Wang, J., and Fukushima, M. 2008. Tabu Search For Attribute Reduction In Rough Set Theory. *Soft Comput.* 12(9): 909-918.

[14] Yu, L., and Liu, H. 2003. Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution. *Int. Conf. Mach. Learn.* 1-8.

[15] Kononenko, I. 1994. Estimating attributes: Analysis and extensions of RELIEF. *Machine Learning: ECML-94*. 784: 171-182.

[16] Kira, K., and Rendell, L. A. 1992. *Machine Learning Proceedings*.

[17] Liu, H. L. H., and Setiono, R. 1995. Chi2: Feature Selection And Discretization Of Numeric Attributes. *Proc. 7th IEEE Int. Conf. Tools with Artif. Intell.*

[18] Seymour, K. You, H. and Dongarra, J. 2008. A Comparison Of Search Heuristics For Empirical Code Optimization. *Proceedings - IEEE International Conference on Cluster Computing, ICCC*, Proceeding. 421-429.

[19] Kisuki, T. Knijnenburg, P. M. W. and Boyle, M. F. P. O. 2000. Combined Selection of Tile Sizes and Unroll Factors Using Iterative Compilation. *Proc. 2000 Int. Conf. Parallel Archit. Compil. Tech.* 237.

[20] Norris, B., Hartono, A. and Gropp, W. 2007. Annotations for Productivity and Performance Portability. *Petascale Computing: {A}lgorithms and Applications*. 443-462.

[21] Whaley, R. C. and Dongarra, J. J. 1998. Automatically Tuned Linear Algebra Software. *Proc. ACM/IEEE Conf. Supercomput.* 1-27.

[22] Qasem, K. K. M.-C. J. A., 2006. Automatic Tuning Of Whole Applications Using Direct Search And A Performance-Based Transformation System. *J. Supercomput.* 36(2): 183-196.

[23] Tabatabaee, V. Tiwari, A. and Hollingsworth, J. K. 2005. Parallel Parameter Tuning For Applications With Performance Variability. *Proceedings of the ACM/IEEE 2005 Supercomputing Conference*.

[24] Tiwari, A. Chen, C. Chame, J. Hall, M. and Hollingsworth, J. K. 2009. A Scalable Auto-Tuning Framework For Compiler Optimization. *Proceedings of the 2009 IEEE International Parallel and Distributed Processing Symposium*.

[25] Ye, D. Chen, Z. and Liao, J. 2007. A New Algorithm for Minimum Attribute Reduction Based on Binary Particle

Swarm Optimization with Vaccination. *PAKDD*. 1029-1036.

[26] Yu, H. Wang, G. and Lan, F. 2008. *Rough Sets and Current Trends in Computing*. Berlin, Heidelberg: Springer Berlin Heidelberg.

[27] Jiang, Y. and Liu, Y. 2006. An Attribute Reduction Method Based on Ant Colony Optimization. *6th World Congress on Intelligent Control and Automation*. 1: 3542-3546.

[28] Suguna, N. K. Thanushkodi, G. and Nadu, T. 2011. An Independent Rough Set Approach Hybrid with Artificial Bee Colony Algorithm for Dimensionality Reduction. 8(3): 261-266.

[29] Chouchoulas, A. and Shen, Q. 2001. Rough Set-Aided Keyword Reduction For Text Categorization. *Applied Artificial Intelligence*. 15: 843-873.

[30] Zhang, M. Shao, C. Li, F. Gan, Y. and Sun, J. 2006. Evolving Neural Network Classifiers and Feature Subset Using Artificial Fish Swarm. *2006 IEEE International Conference on Mechatronics and Automation, ICMA*.1598-1602.

[31] Narendra, P. M. and Fukunaga, K. 1977. A Branch and Bound Algorithm for Feature Subset Selection. *IEEE Trans. Comput.* 26(9): 917-922.

[32] Namsrai, E., Munkhdalai, T., Li, M., Shin, J.-H., Namsrai, O.-E., and Ryu, K. H. 2013. A Feature Selection-based Ensemble Method for Arrhythmia Classification. *J Inf Process Syst*. 9(1): 31-40.

[33] Montazeri, M., H. Naji, R., and Faraahi, A. 2013. A novel memetic feature selection algorithm. *The 5th Conference on Information and Knowledge Technology*. 295-300.

[34] Aha, D., Murphy, P., Merz, C., Keogh, E., Blake, C., Hettich, S., and Newman, D. 1987. UCI Machine Learning Repository. University of Massachusetts Amherst.

[35] Witten, I. H. and Frank, E. 2005. *Data Mining: Practical Machine Learning Tools And Techniques*.