International Journal of Computer Science and Electronics Engineering (IJCSEE) Volume 4, Issue 2 (2016) ISSN 2320–4028 (Online)

Attribute Selection-based Ensemble Method for Dataset Classification

Mohammad Aizat bin Basir, and Faudziah binti Ahmad

Abstract—**Attribute reduction and classification task are an essential process in dealing with large data sets that comprise numerous number of input attributes. There are many search methods and classifiers that have been used to find the optimal number of attributes. The aim of this paper is to find the optimal set of attributes and improve the classification accuracy by adopting ensemble classifiers method. Research process involves 2 phases; finding the optimal set of attributes and ensemble classifiers method for classification task. Results are in terms of percentage of accuracy and number of selected attributes. 6 datasets were used for the experiment. The final output is an optimal set of attributes with ensemble classifiers method. The experimental results conducted on public real dataset demonstrate that the ensemble classifiers methods consistently show improve classification accuracy on the selected dataset. Significant improvement in accuracy and optimal set of attribute selected is achieved by adopting ensemble classifiers method.**

Keywords—**Attribute reduction, Classification, Reduction algorithm, Ensemble Classifier.**

I. Introduction

Real world dataset usually consists of a large number of attributes. It is very common for 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 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 rule, 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 that two critical aspects of attribute reduction problem 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 find reduct 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].

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.

However, the use of a subset of a feature set may disregard important information contained in other subsets. Consequently, classification performance is reduced. Therefore, this paper aims to find the optimal set of attributes and improve the classification accuracy by adopting ensemble classifiers method. Firstly, an optimal set of attributes subsets are extracted by applying various search method and reduction algorithm to the original dataset. Then an optimal set of attributes is further classified by adopting a classification ensemble approach. In the experiment, 6 various datasets were used. The experiment results showed that the performance of the ensemble classifier has improved the classification accuracy of the dataset. This paper is organized as follows: in Section II, related work is discussed. The proposed method is presented in Section III. In Section IV, the experimental results are given. Finally, conclusions presented in Section V.

Manuscript received April. 8, 2016. This work was supported by UUM, UMT and KPM, Malaysia.

Mohammad Aizat bin Basir is currently pursuing doctoral degree program in computer science in Universiti Utara Malaysia, Malaysia.

Faudziah binti Ahmad is currently Assoc. Prof. in computer science in Universiti Utara Malaysia, Malaysia.

II. RELATED WORKS

There many researches in feature selection methods for constructing an ensemble of classifiers. The ensemble feature selection method is a set of the classifiers, each of which solve the same original task, are joined in order to obtain a better global combination classifier, with more accurate and reliable estimates or decisions than can be obtained from using a single classifier. The aim of designing and using the ensemble method is to achieve a more accurate classification by combining many weak learners.

Previous studies show that methods like bagging improve generalization by decreasing variance. In contrast, methods similar to boosting achieve this by decreasing the bias [4]. [5] demonstrated a technique for building ensembles from simple Bayes classifiers in random feature subsets.

[6] explored tree based ensembles for feature selection. It uses the approximately optimal feature selection method and classifiers constructed with all variables from the TIED dataset. [7] presented the genetic ensemble feature selection strategy, which uses a genetic search for an ensemble feature selection method. It starts with creating an initial population of classifiers where each classifier is generated by randomly selecting a different subset of features. The final ensemble is composed of the most fitted classifiers.

[8] suggested a nested ensemble technique for real time arrhythmia classification. A classifier model was built for each 33 training sets with enhanced majority voting technique. The nested ensembles can alleviate the unlikelihood problem of a classifier being generated when learning the classifier by an old dataset and limited input features. One of the reasons that make the ensemble method popular is that ensemble methods tend to solve dataset problems.

III. METHODOLOGY

The methodology is shown in Figure 1. It consists of five (5) steps: (1) data collection; (2) data pre-processing; (3) dimensionality reduction; (4) classify optimal set of attributes by using ensemble classifiers method; (5) Result-improved classification accuracy: ensemble classifier methods has been compared with datasets that do not use the ensemble classifier method. The output of phase 1 (step 1 – 3) is the optimal set of attributes. For phase 2 (step 4 – 5), the output is the improved classification accuracy by adopting ensemble classifiers method for the classification task. The details of steps involved are described below:-

Step 1 (Data Collection): Six (6) different datasets were selected from UCI Machine Learning Repository. Arrhythmia datasets is one of the dataset selected due to its many features that make it challenging to explore [9]. Other five (5) datasets also were taken from different domain in order to confirm the suitability of the ensemble classifiers.

Step 2 (Data Pre-processing): Dataset that has missing values has been pre-processed in order to make sure that dataset is ready for experimentation. All datasets were discretized since it has numeric data but need to use classifier that handles only nominal values.

Step 3 (Dimensionality Reduction): 8 search methods and 10 reduction algorithms have been used in order to get the optimal set of attributes. The output of this step is the optimal set of attributes.

Step 4 (classify optimal set of attributes by using ensemble classifiers method): In this step, the optimal sets of attributes obtained from the previous step were classified by adopting ensemble classifier method.

Step 5 (Model with good accuracy): In this step, the performance (% classification accuracy) of the dataset that used ensemble classifier methods has been compared with datasets that do not use the ensemble classifier method. The output of this step is the improved classification accuracy with optimal number of attributes.

Standard six datasets namely Arrhythmia, Bio-degradation, Ionosphere, Ozone, Robot Navigation and Spam-base from the UCI [10] were used in the experiments. These datasets include discrete and continuous attributes and represent various field of data. The reason for choosing this dataset is to confirm the ensemble classifier is suited to all field of data. The information on the datasets is shown in Table 2.1.

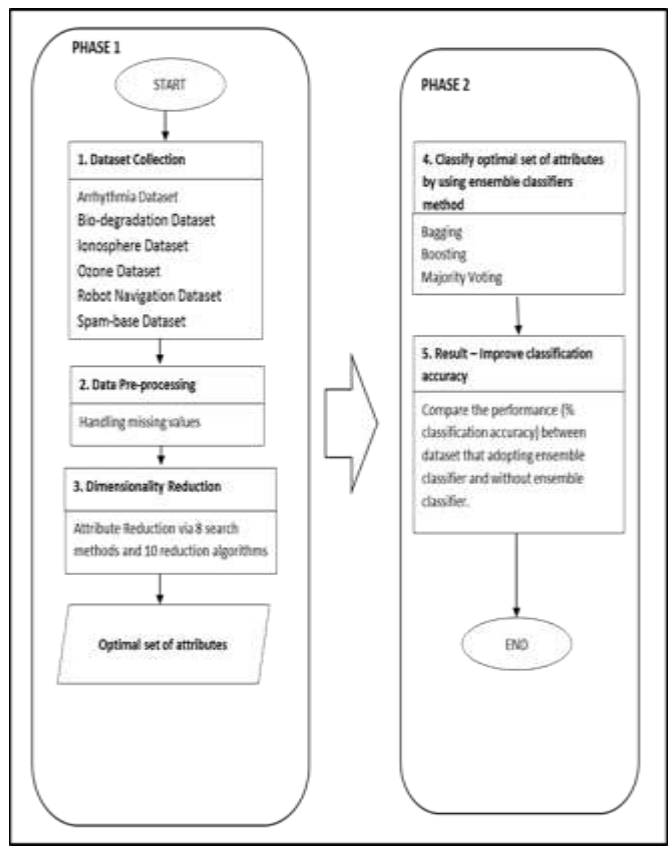

Fig. 1. Methodology

IV. RESULTS AND DISCUSSION

The outputs for phase 1 and phase 2 are presented in section 3.1 and 3.2. The performance results are presented as the percentage of classification accuracy with the optimal set of attributes.

A. *Phase 1 (Step 1-3)*

TABLE I: LIST OF AN OPTIMAL SET OF ATTRIBUTES SELECTED

Dataset	Search Method	Reduction Algorithm	# Attr	# SelAttr
Arrhythmia	Best First Search	WrapperSubsetEval	279	19
Bio-degradation	Best First Search	WrapperSubsetEval	41	10
Ionosphere	Greedy Stepwise	WrapperSubsetEval	34	8
Ozone	Race Search	ClassifierSubsetEval	72	5
Robot Navigation	SubsetSizeForwardSelection	CFSSubsetEval	24	6
Spam-base	Genetic Search	WrapperSubsetEval	57	18

Table I shows the results of an optimal set of attributes selected by using various search method and reduction algorithm. In phase one (1), eight (8) search methods namely Best First Search, Genetic Search, Exhaustive Search, Greedy Stepwise Search, Linear Forward Selection Search, Scatter Search, Subset Size Forward Selection Search and Ranker Search were applied. In addition, ten (10) reduction algorithms that is CfsSubsetEval, ClassifierSubsetEval, ConsistencySubsetEval,FilteredSubsetEval, ChisquaredAttributeEval, FilteredAttributeEval, GainRatioAttributeEval, InfoGainAttributeEval, PrincipalComponent and WrapperSubsetEval were adopted. It can be seen that Arrhythmia and Ozone dataset produced a massive attribute reduction which is more than 90% reduction.

Best first search (BSF) was used with WrapperSubsetEval for Arrhythmia dataset since BFS is a robust search [11] and better for dataset studied [12]. The rest of the dataset achieved more than 60% attribute reduction. Wrapper method (WrapperSubsetEval) performed better for 4 out of 6 datasets selected with combination of various search method. These experiments confirmed that significant attribute reduction can be accomplished by combining the right search method and reduction algorithm.

B. *Phase 2 (Step 4-5)*

In phase 2, each selected set of attributes for the six (6) various dataset namely Arrhythmia, Bio-degradation, Ionosphere, Ozone, Robot Navigation and Spam-base were classified using ensemble classifier methods of boosting, bagging and voting. In this phase, classifiers like Naïve Bayes, the Decision Tree, SVM, and the Bayes Network were evaluated with ensemble method. 70% of the dataset being used as training and the remaining 30% was used for testing data. Classification was performed using WEKA [13]. The results are shown in Table II through Table VI.

TABLE II: CLASSIFICATION RESULT OF USING BAYESNET AND BAYESNET WITH ENSEMBLE CLASSIFIER METHOD

Dataset	Without Ensemble Classifier	Ensemble Classifer	
	BayesNet	Boosting + BayesNet	Bagging + BayesNet
	Acc (%)	Acc (%)	Acc (%)
Arrhythmia	81.62	81.62	80.88
Bio-degradation	82.59	84.22	82.91
Ionosphere	94.29	95.04	94.29
Ozone	94.08	93.88	94.08
Robot Navigation	97.08	97.85	97.12
Spam-base	92.68	93.13	92.53

Table II shows the classification result of using BayesNet and BayesNet with ensemble method. BayesNet with boosting method improves the classification accuracy of 4 datasets namely Bio-degradation, Ionosphere, Robot Navigation and Spam-base. These results are in line with those of previous studies that an ensemble can be more accurate only if the individual classifiers disagree with each other [14]. 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.

TABLE III: CLASSIFICATION RESULT OF USING SUPPORT VECTOR MACHINE AND SUPPORT VECTOR MACHINE WITH ENSEMBLE CLASSIFIER METHOD

Dataset	Without Ensemble Classifier	Ensemble Classifer	
	SVM	Boosting + SVM	Bagging + SVM
	Acc (%)	Acc (%)	Acc (%)
Arrhythmia	76.9	76.31	77.78
Bio-degradation	81.41	82.39	81.46
Ionosphere	93.89	93.86	93.86
Ozone	93.9	93.91	93.9
Robot Navigation	98.95	98.78	98.99
Spam-base	93.12	93.41	93.3

Table III shows the classification result of using BayesNet and BayesNet with ensemble method. SMO classifier with bagging method increased the classification accuracy of Arrhythmia dataset. Many studies have shown that aggregating the prediction of multiple classifiers can improve the performance achieved by a single classifier [15]. In this case, Bagging known as a “bootstrap” ensemble method that creates individuals for its ensemble by training each classifier on a random redistribution of the training set. These results seem to be consistent with other research as found in [16] that shown bagging with SMO is a promising and practical scheme to classify cancer dataset that can make a highly reliable prognostication of patients possibly suffering from cancer. In contrast, Boosting method with SMO classifier performed with better accuracy for Bio-degradation dataset. In this case, these results are consistent with data obtained in [17] which claimed that boosting with SMO performs better than other approaches of using component classifiers such as Decision Trees and Neural Networks. Besides these, they stated that boosting with SMO demonstrates good performance on imbalanced classification problems.

TABLE IV: CLASSIFICATION RESULT OF USING DECISION TREE AND DECISION TREE WITH ENSEMBLE CLASSIFIER METHOD

Dataset	Without Ensemble Classifier	Ensemble Classifer	
	DT	Boosting + DT	Bagging + DT
	Acc (%)	Acc (%)	Acc (%)
Arrhythmia	72.8	73.64	75.06
Bio-degradation	83.48	83.81	83.89
Ionosphere	92.25	93.51	92.53
Ozone	93.8	93.8	93.86
Robot Navigation	95.12	97.46	95.91
Spam-base	92.54	93.49	92.77

Table IV shows the Classification result of using Decision Tree and Decision Tree with ensemble classifier method. Decision Tree classifier with bagging method performed well to enhance the accuracy for Arrhythmia and Bio-degradation datasets. In addition, boosting method with decision tree Decision Tree Classifier produced better accuracy result for Ionosphere and Robot Navigation datasets. These results are consistent with data obtained in [18] which shows that the efficiency of the Decision Tree increases with every ensemble method.

TABLE V: CLASSFICATION RESULT OF USING NAÏVE BAYES AND NAÏVE BAYES WITH ENSEMBLE CLASSIFIER METHOD

Dataset	Without Ensemble Classifier	Ensemble Classifer	
	NB	Boosting + NB	Bagging + NB
	Acc (%)	Acc (%)	Acc (%)
Arrhythmia	79.57	79.57	79.55
Bio-degradation	84.36	84.46	84.32
Ionosphere	94.56	94.53	94.46
Ozone	94.03	93.81	94.03
Robot Navigation	96.71	97.78	96.71
Spam-base	93.37	93.19	93.36

Table V shows the classification result of using Naïve Bayes and Naïve Bayes with ensemble classifier method. Boosting Method with Naïve Bayes classifier performed slightly better with accuracy for Robot Navigation dataset. In this case, it is similar to BayesNet with boosting method in [14] in which the advantage of boosting with Bayesian classifier has the ability to improve the classification performance.

In summary, results have shown significant improvement in term of classification accuracy when using ensemble classifier method.

V. CONCLUSION

In this paper, eight (8) search methods with ten (10) reduction algorithms were tested with 6 datasets. Experimental results benchmark dataset demonstrate that the ensemble classifiers method namely bagging and boosting significantly perform better than other approaches of not using ensemble method. Beside these, it is found that right combination of search methods and reduction algorithms shown good performance on extracting optimal number of attributes. For future research, method of finding the suitable match between search method, reduction algorithm and ensemble classifiers can be developed to get a better view of the datasets.

ACKNOWLEDGMENT

The authors wish to thank Universiti Utara Malaysia (UUM), Universiti Malaysia Terengganu (UMT) and Kementerian Pendidikan Malaysia (KPM). This work was supported by UUM, UMT and KPM, Malaysia.

REFERENCES

[1] R. Jensen and Q. Shen, "Finding rough set reducts with ant colony optimization," *Proc. 2003 UK Work.*, vol. 1, no. 2, pp. 15–22, 2003.

[2] N. Suguna and K. Thanushkodi, "A Novel Rough Set Reduct Algorithm for Medical Domain Based on Bee Colony," vol. 2, no. 6, pp. 49–54, 2010.

[3] B. Yue, W. Yao, A. Abraham, and H. Liu, "A New Rough Set Reduct Algorithm Based on Particle Swarm Optimization," pp. 397–406, 2007. http://dx.doi.org/10.1007/978-3-540-73053-8_40

[4] R. E. Schapire, Y. Freund, P. Bartlett, and W. S. Lee, "Boosting the margin: A new explanation for the effectiveness of voting methods," *Ann. Stat.*, vol. 26, no. 5, pp. 1651–1686, 1998. http://dx.doi.org/10.1214/aos/1024691352

[5] A. Tsymbal, S. Puuronen, and D. W. Patterson, "Ensemble feature selection with the simple Bayesian classification," *Inf. Fusion*, vol. 4, no. 2, pp. 87–100, 2003. http://dx.doi.org/10.1016/S1566-2535(03)00004-6

[6] E. Tuv, "Feature Selection with Ensembles , Artificial Variables , and Redundancy Elimination," *J. Mach. Learn. Res.*, vol. 10, pp. 1341–1366, 2009.

[7] D. W. Opitz, "Feature selection for ensembles," *Proc. Natl. Conf. Artif. Intell.*, pp. 379–384, 1999.

[8] M. E. A. Bashir, M. Akasha, D. G. Lee, G. Yi, K. H. Ryu, E. J. Cha, J.-W. Bae, M.-C. Cho, and C. W. Yoo, "Nested Ensemble Technique for Excellence Real Time Cardiac Health Monitoring.," in *International Conference on Bioinformatics & Computational Biology, BIOCOMP 2010, July 12-15, 2010, Las Vegas Nevada, USA, 2 Volumes*, 2010, pp. 519–525.

[9] E. Namsrai, T. Munkhdalai, M. Li, J.-H. Shin, O.-E. Namsrai, and K. H. Ryu, "A Feature Selection-based Ensemble Method for Arrhythmia Classification," *J Inf Process Syst*, vol. 9, no. 1, pp. 31–40, 2013. http://dx.doi.org/10.3745/JIPS.2013.9.1.031

[10] D. Aha, P. Murphy, C. Merz, E. Keogh, C. Blake, S. Hettich, and D. Newman, "UCI machine learning repository," *University of Massachusetts Amherst*, 1987. .

[11] M. GINSBERG, *Essentials of Artificial Intelligence*. Elsevier, 1993.

[12] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, no. 1–2, pp. 273–324, 1997. http://dx.doi.org/10.1016/S0004-3702(97)00043-X

[13] I. . H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*. 2005.

[14] L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Trans. Pattern Anal. Mach.*, vol. 12, no. 10, pp. 993–1001, 1990. http://dx.doi.org/10.1109/34.58871

[15] C. D. Sutton, "Classification and Regression Trees, Bagging, and Boosting," *Handbook of Statistics*, vol. 24. pp. 303–329, 2004. http://dx.doi.org/10.1016/S0169-7161(04)24011-1

[16] B. Zhang, T. D. Pham, and Y. Zhang, "Bagging Support Vector Machine for Classification of SELDI-ToF Mass Spectra of Ovarian Cancer Serum Samples – Bookmetrix Analysis," in *AI 2007: Advances in Artificial Intelligence*, 2007, pp. 820–826.

[17] X. Li, L. Wang, and E. Sung, "AdaBoost with SVM-based component classifiers," *Eng. Appl. Artif. Intell.*, vol. 21, no. 5, pp. 785–795, Aug.

2008.
http://dx.doi.org/10.1016/j.engappai.2007.07.001

[18] "Improving classification of J48 algorithm using bagging,boosting and blending ensemble methods on SONAR dataset using WEKA." [Online]. Available: https://www.erpublication.org/admin/vol_issue1/upload Image/IJETR022513.pdf. [Accessed: 03-Apr-2016].