

An Ensemble of Classifiers with Genetic Algorithm Based Feature Selection

Zili Zhang and Pengyi Yang

Abstract—Different data classification algorithms have been developed and applied in various areas to analyze and extract valuable information and patterns from large datasets with noise and missing values. However, none of them could consistently perform well over all datasets. To this end, ensemble methods have been suggested as the promising measures. This paper proposes a novel hybrid algorithm, which is the combination of a multi-objective Genetic Algorithm (GA) and an ensemble classifier. While the ensemble classifier, which consists of a decision tree classifier, an Artificial Neural Network (ANN) classifier, and a Support Vector Machine (SVM) classifier, is used as the classification committee, the multi-objective Genetic Algorithm is employed as the feature selector to facilitate the ensemble classifier to improve the overall sample classification accuracy while also identifying the most important features in the dataset of interest. The proposed GA-Ensemble method is tested on three benchmark datasets, and compared with each individual classifier as well as the methods based on mutual information theory, bagging and boosting. The results suggest that this GA-Ensemble method outperform other algorithms in comparison, and be a useful method for classification and feature selection problems.

Index Terms—Ensemble Classifiers, Multi-objective Genetic Algorithms, Decision Tree, Artificial Neural Networks, Support Vector Machines.

I. INTRODUCTION

MACHINE learning algorithms have been widely used in various fields to analyze and extract valuable information and patterns from large datasets with noise and missing values [1], [2], [3], [4]. One fundamental task of those learning algorithms is sample classification which is heavily relied on feature selection or extraction.

Learning algorithms are usually divided into two different categories: supervised learning, unsupervised learning [5]. In this work, we will focus on supervised learning.

The learning algorithms used in the classification process are refereed as classifiers, and several types of classifiers have been developed including decision trees, various types of artificial neural networks (ANN), support vector machines (SVM), and so on. Each of these classifiers uses different learning strategies. A common method used in supervised learning to improve classification accuracy and decrease computation complexity is feature selection [6]. In many applications, feature selection is essential as it can also help to identify

important and meaningful traits [7], [8], [9]. Many feature selection approaches are available [6], [8], [10], [11], [12], [13], which can be categorized as deterministic or stochastic feature selection. However, deterministic feature selection results in high-dimensional datasets are often local optimal, while stochastic feature selection results are usually unstable [11].

A growing body of studies indicates that every single learning strategy has its own shortcomings and none of them could consistently perform well over all datasets. To overcome the shortcomings of individual methods, ensemble methods have been suggested as the promising measures [1], [14], [15], [16], [17]. For instance, the empirical study of ensemble system for data classification by Chandra and Yao [18] suggest that the ensemble systems tend to achieve higher accuracy and generalize better than single method. An ensemble of classifiers is a collection of classifiers that individual decisions are combined typically by means of weighed or un-weighted voting [15]. Some applications of different ensemble methods in real world datasets have demonstrated their power [20], [21], [22], [23].

The necessary and sufficient condition for an ensemble classifier to outperform its individual members is that the combined classifiers are accurate and diverse [16], [24]. In addition, previous studies have illustrated that the key requirements to successful ensemble methods are:

- the individual classifiers used to form the ensemble must have error rates less than 0.5 when classifying data, and
- the errors of those are uncorrelated at least in some extent [15].

In our previous work, we explored different hybrid algorithms—the combination of GA with decision tree (GADT), the combination of GA with artificial neural network (GANN), and the combination of GA with support vector machine (GASVM). They are used to analyze microarray data and SNP genotype data [7], [8]. All three algorithms have been proved powerful in sample classification and trait related feature selection.

In this study, a novel hybrid algorithm is proposed, which is the combination of a multi-objective Genetic Algorithm and an ensemble classifier. While the ensemble classifier, which consists of a decision tree classifier, an Artificial Neural Network (ANN) classifier, and a Support Vector Machine (SVM) classifier, is used as the classification committee, the multi-objective Genetic Algorithm is employed as the Random Subspacing (RS) method and the feature selector to facilitate data classification. This GA-Ensemble algorithm is essentially the combination of our previous algorithms. Nevertheless, the

Zili Zhang is with School of Engineering and Information Technology, Deakin University, Geelong, Victoria, Australia, 3217.
E-mail: zzhang@deakin.edu.au

Pengyi Yang is with Intelligent Software and Software Engineering Laboratory, Southwest University, Chongqing, China, 400715.
E-mail: mikesilver7@msn.com

objective of the use of ensemble classifiers and the combination with multi-objective GA feature selector is to further improve the overall data classification accuracy and feature selection reproducibility.

Since each of the three classifiers uses its own yet different learning strategies to classify the data, a diversely aggregated ensemble classifier can be obtained given the effective integration method was employed. This ensemble method differ itself from bagging and boosting strategies [25] because the diversity between classifiers is inherent in the inductive algorithms themselves other than manipulating the training dataset. The classification results over three benchmark datasets [28] are compared to see if this GA-Ensemble algorithm outperforms the individual ones. Furthermore, the results are also compared with those obtained by methods based on mutual information theory [13] and those obtained with bagging and boosting of decision tree [26].

The rest of the paper is structured as follows: Section II outlines the GA-Ensemble algorithm. The GA feature selector and ensemble classifiers in the GA-Ensemble algorithm are detailed in Sections III and IV, respectively. Evaluation is presented in Section V. Section VI concludes the paper.

II. OUTLINE OF THE GA-ENSEMBLE ALGORITHM

The GA-Ensemble algorithm proposed in this study is the combination of a multi-objective GA and an ensemble classifier consisting of a decision tree classifier, a standard multiple layer perceptron back propagation ANN classifier, and a support vector machine (SVM) classifier. A multi-objective evolutionary algorithm (which is similar to multi-objective GA) has been firstly employed in ensemble classifier construction by Chandra and Yao [27]. However, different from their application which optimize the diversity and the accuracy of the base classifiers explicitly, we incorporate these two optimization goals implicitly. Figure 1. illustrates the structure of the proposed system.

The learning steps of this algorithm can be described as following:

- 1) Initially, the global multi-objective GA randomly creates a set of chromosomes representing various feature sets.
- 2) Using all chromosomes in the set as the inputs of the classifiers. After classifiers evaluate certain feature set, they return the evaluation accuracies of this set to GA. GA then calculates the mean score and the consensus of this feature set.
- 3) After the whole population has been evaluated, GA selects favorite chromosomes with high fitness scores.
- 4) The crossover and mutation operations are then conducted on selected chromosomes with a predefined p_c (probability of crossover) and p_m (probability of mutation), respectively; and the next generation begins.
- 5) Repeat steps 2-4 until terminating generation is reached and the final chromosomes are printed out as the near optimal set of features for classification.

III. THE MULTI-OBJECTIVE GA FEATURE SELECTOR

The proposed ensemble approach utilized three classifiers, each will assess data and features with their own learning

strategies. Thus, a multi-objective GA is employed to balance their assessments and facilitate their diversity. The fitness function of this multi-objective GA is defined as follows:

$$fitness_1(s) = \frac{\sum_{j=1}^n accuracy_j(s)}{n} \quad (1)$$

$$fitness_2(s) = consensus(s) \quad (2)$$

$$fitness(s) = \frac{fitness_1(s) + fitness_2(s)}{2} \quad (3)$$

Where $accuracy_j(s)$ specify the classification accuracy of the j th classifier upon the s th feature subset, while $consensus(s)$ specify the classification accuracy using consensus upon the s th feature subset.

The first part of the fitness function tries to optimize the target feature set into a subset which has superior power on accurate sample classification with not only one specific classifier but the whole classification committee. This part of the function improves the generalization ability of the resulting feature set [22]. As to the second part of the fitness function, it tries to optimize the target features set into a superior set in producing high consensus classification. This part of the function promotes the selected features in creating diverse classifiers implicitly, which in turn leads to the high sample classification accuracy [19].

The use of GA in this algorithm is two-fold. On the one hand, GA works as a RS method and a feature selector to select and rank different features based on their importance. This is extremely useful when the features in the given dataset are large and redundant, while the number of the samples are small. With the help of GA, informative features can be selected and uninformative ones will be removed. Otherwise, the uninformative features will increase the complexity of computation and introduce noisy and redundant data to the process [29]. By doing so, over-fitting can also be avoided in some extent. Moreover, the selected features can be further studied to find their special association with data.

On the other hand, when analyzing large datasets, feature selection is critical in improving the classification accuracy of the classifiers. It is widely acknowledged that the classification accuracy of ANN and SVM is affected by the size of the datasets. This is especially phenomenal when the number of the features is large. Moreover, it is both hard and unnecessary to use all data features as the inputs [7], [8] because it not only adds more computational expenses but also decreases the classification power of classifiers. By using GA, one can scale down the number of the inputs while also maintain or improve the classification accuracy of ANN and SVM. The need of combining GA with decision tree lies in that the decision tree algorithm is deterministic and it always uses the highest ranked feature – the feature with highest gain value – to split the dataset every time. This results in only one tree being created and it may be a locally optimal classifier, while an alternative one with a different splitting point can perform better [30]. This shortcoming is also more severe when the number of

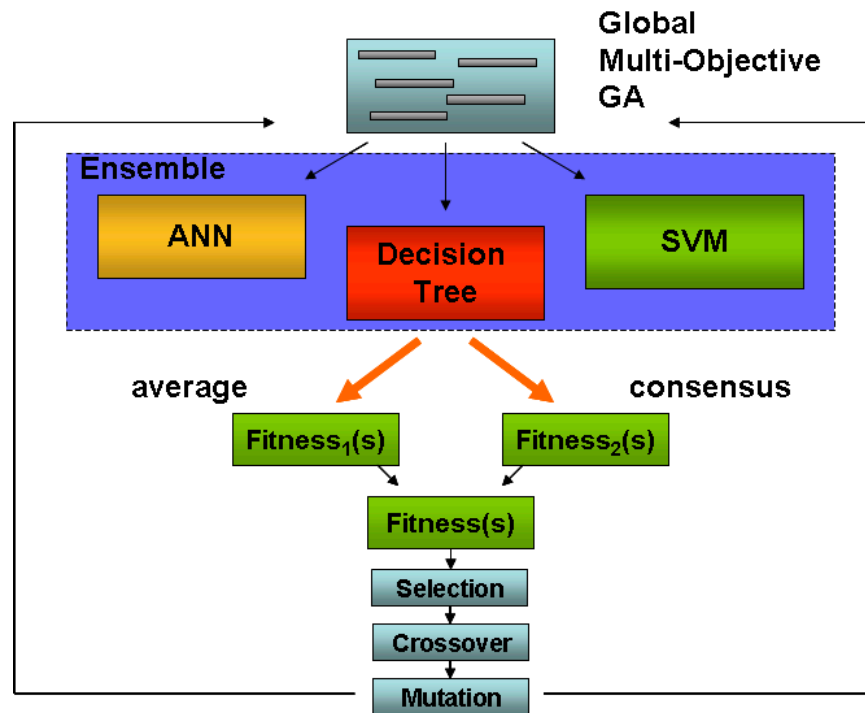


Fig. 1. The Architecture of the GA-Ensemble Algorithm

features been considered is large. When using GA to create different subsets, different decision trees can be produced and the favorite ones will be selected by GA for later iterations. This can help the decision tree to overcome the pitfall of the local optimal classification as well as identify important features.

IV. THE ENSEMBLE CLASSIFIER

Majority voting is one of the simplest strategies in implementing combination classifiers. Yet, the power of this strategy is comparable to other complex methods [31]. In n classifiers majority voting, consensus is made by k classifiers where

$$k = \begin{cases} n/2 + 1 & \text{if } n \text{ is even} \\ (n + 1)/2 & \text{if } n \text{ is odd} \end{cases} \quad (4)$$

The three classifiers namely decision tree, ANN and SVM are integrated as a consensus committee. In the feature selection phase, each candidate feature combination produced by multi-objective GA will be fed into the ensemble classifier. After a feature combination has been input into the ensemble classifier, the three classifiers contained in this ensemble classifier will use the input features to learn and classify data sample, separately. When each classifier returns its classification accuracy by using certain feature combination, the multi-objective GA will calculate the consensus, using majority voting. Then the consensus score with the average score of three classifiers will be used as the fitness score of this feature combination.

In the evaluation phase, the best feature combination selected by GA-Ensemble is used to make sample classification with the test set. When a querying sample is input, three

different classifiers will give their own prediction of which class this sample belongs to, and the majority voting is conducted to decide the final class it should be.

V. EVALUATION

This section presents the experimental results we conducted.

A. Datasets

Three benchmark datasets, all obtained from UCI Repository [28], have been used to evaluate the proposed method. The first dataset is called Sonar dataset. The task is to discriminate between sonar signals bounced off a metal cylinder at different angles under various conditions and those bounced off a roughly cylindrical rock. The first class contains 111 samples and the second class contains 97 samples obtained from rocks under similar conditions. Each sample has 60 features representing the energy within a particular frequency band, integrated over a certain period of time [12]. The second dataset, named Ionosphere, contains 351 samples collected from radar signals, and 225 samples from it belong to class “good” while other 126 samples belong to class “bad”. Each sample has 34 features. The last dataset is called Soybean (large) which contains 307 samples and 35 features. This dataset is different from the first two datasets in that the feature of the dataset is characterized as “categorical” instead of real numbers, and the number of the class is 19 instead of 2. Table 1 is the summary of the datasets used in evaluation.

All three datasets have long been utilized as evaluation datasets in many classification and feature selection studies [12], [13], [32], [33], [34] because they contain random noise, redundant features and the samples are linear inseparable.

TABLE I
DATASETS DESCRIPTION.

Dataset	Num. of Feature	Num. of Sample	Class Num.
Sonar [28]	60	208	2
Ionosphere [28]	34	351	2
Soybean [28]	35	683	19

B. GA-Ensemble Algorithm Implementation

The ANN adopted in this study is a three layers fully connected neural network. The number of the neuron in the first layer corresponds to the number of the input data features, and the number of the hidden neuron is the ceiling of the half of the input ones. Only one neuron is used in the output layer. The learning strategy of this ANN is consistent with that proposed by Brierley and Batty [35]. The learning rate from input layer to hidden layer is set to 0.4 and the learning rate 0.03 is set for the hidden layer to output layer. 1000 training epochs are used to train the ANN.

An easy to use yet very powerful SVM classifier package, SVM-Torch II [36], is employed to construct SVM for the ensemble classifier. The default parameters are used when performing learning and classification. The kernel of the SVM is set as polynomial kernel with exponent of 2.

As for the decision tree, one of the most popular decision tree algorithm package C5.0, an improvement of C4.5 [37], is used to build decision tree and carry out simple classification.

Starting population size of GA is set to 100. The probability of crossover p_c and the probability of mutation p_m are 0.7 and 0.03, respectively. The single point mutation and crossover are used in genetic operation parts, and the binary tournament selection method [38] is adopted to select favorite gene combinations. The termination condition is that GA reaches the 50th generation, and the program terminates after the selected genes of the 50th generation are printed out.

C. Cross Validation

5-fold cross validation is conducted to evaluate the overall accuracy of all utilized methods. Both Sonar and Ionosphere datasets are randomly divided into five separate subsets, and while the four folds are used to train the algorithms, the remaining one fold is used to evaluate the classification accuracy of each method. The validation process repeats five times until all data in the sets are tested and the average classification accuracy is then calculated.

D. Z-score Calculation

In order to evaluate stability and reproducibility of each method, an independent re-run of every method is conducted. The feature combinations in last 10 generations of GA (after 40th generation) are extracted and the top 30 feature combinations with highest classification accuracy from the last 10 generations are then selected to compare in two independent runs to evaluate the stability and reproducibility.

A z-score or standard score is a measure of how many standard deviation units an individual raw score away from the mean of the distribution [39]. We employ this statistic method

to calculate selection frequency of each feature. A high z-score of a feature indicates it's frequently selected. All evaluation results are z-score transformed as following [11]:

$$Z = [F_i - E(F_i)]/\sigma \quad (5)$$

where the $E(F_i)$ and σ are then calculated as following:

$$E(F_i) = P(\text{feature}_i) \cdot A \quad (6)$$

$$\sigma = \sqrt{P(\text{feature}_i) \cdot [1 - P(\text{feature}_i)] \cdot A} \quad (7)$$

In the above formulas, $A = 30$, which is the number of the top 30 feature combinations. $P(\text{feature}_i) = d/T$, where d is the feature combination length and T is the total feature number. F_i is the number of times feature_i is selected.

E. Results

Previous studies show that the mean errors are relatively low and the classification accuracies are likely to be high when the combination size of the feature is small [12], [13]. Thus, in this study, feature sets with size of 6 and 12 for Sonar dataset and size of 5 and 10 for Ionosphere dataset as well as Soybean dataset are used to test our method. Firstly, each individual methods are tested separately, then the proposed GA-Ensemble approach is tested to compare with individual methods.

Tables 2 to 4 provide detail information of the results obtained with Sonar dataset, Ionosphere dataset and Soybean dataset for GADT, GANN, GASVM, and GA-Ensemble algorithms, respectively. All classification accuracies are calculated by averaging the 5-fold cross validation results with the best combination from each algorithm five times.

As shown in Tables 2, 3 and 4, GA-Ensemble method achieved the best classification accuracies, with 80.02% and 83.95% using 6-feature and 12-feature combinations in Sonar dataset, with 92.22% and 93.54% using 5-feature and 10-feature combinations in Ionosphere dataset, and with 94.97% and 95.37% using 5-feature and 10-feature combinations in Soybean dataset.

Table 5 provides the classification results obtained by using bagging and boosting of C4.5 algorithms. As can be seen, the results obtained by GA-Ensemble are comparable or better. The results of the first two datasets are also compared with those obtained with the method based on mutual information theory reported in [13]. In [13], several kinds of classification and feature selection methods are studied, which are all based on mutual information theory. The highest classification accuracies of Sonar dataset with 6 features and 12 features were obtained by 'TMFS with MIFS-U' and 'MIFS-U', with

TABLE II
CLASSIFICATION ACCURACY WITH SONAR DATASET

Methods	6 selected feature	12 selected feature
GADT	(F_{11} F_{25} F_{36} F_{38} F_{39} F_{45}) 79.72%	(F_{11} F_{14} F_{15} F_{21} F_{23} F_{33} F_{37} F_{42} F_{45} F_{47} F_{52} F_{60}) 80.63%
GANN	(F_{11} F_{17} F_{20} F_{27} F_{36} F_{46}) 78.61%	(F_1 F_4 F_7 F_{11} F_{12} F_{17} F_{20} F_{22} F_{49} F_{54} F_{57} F_{58}) 82.01%
GASVM	(F_9 F_{11} F_{12} F_{28} F_{37} F_{46}) 79.22%	(F_2 F_6 F_{11} F_{15} F_{20} F_{22} F_{31} F_{35} F_{36} F_{45} F_{46} F_{48}) 79.82%
GA-Ensemble	(F_4 F_9 F_{12} F_{36} F_{46} F_{48}) 80.02%	(F_3 F_{11} F_{16} F_{18} F_{19} F_{23} F_{32} F_{35} F_{37} F_{39} F_{45} F_{47}) 83.95%

TABLE III
CLASSIFICATION ACCURACY WITH IONOSPHERE DATASET

Methods	5 selected feature	10 selected feature
GADT	(F_3 F_4 F_5 F_{15} F_{27}) 91.47%	(F_2 F_3 F_4 F_5 F_6 F_8 F_{13} F_{23} F_{27} F_{32}) 90.90%
GANN	(F_1 F_5 F_{11} F_{25} F_{27}) 87.85%	(F_1 F_2 F_5 F_8 F_{18} F_{22} F_{24} F_{25} F_{27} F_{32}) 88.22%
GASVM	(F_1 F_5 F_8 F_{27} F_{29}) 91.45%	(F_2 F_3 F_4 F_5 F_6 F_7 F_{12} F_{16} F_{24} F_{27}) 93.16%
GA-Ensemble	(F_1 F_5 F_7 F_8 F_{27}) 92.22%	(F_1 F_5 F_8 F_9 F_{10} F_{20} F_{24} F_{26} F_{27} F_{32}) 93.54%

TABLE IV
CLASSIFICATION ACCURACY WITH SOYBEAN DATASET

Methods	5 selected feature	10 selected feature
GADT	(F_1 F_9 F_{15} F_{17} F_{35}) 92.18%	(F_1 F_{10} F_{12} F_{17} F_{18} F_{23} F_{29} F_{31} F_{32} F_{35}) 93.41%
GANN	(F_1 F_3 F_{15} F_{17} F_{32}) 94.65%	(F_3 F_5 F_{12} F_{15} F_{18} F_{19} F_{21} F_{22} F_{32} F_{35}) 94.85%
GASVM	(F_3 F_{15} F_{17} F_{26} F_{29}) 94.68%	(F_1 F_3 F_6 F_{14} F_{15} F_{22} F_{26} F_{29} F_{31} F_{35}) 94.29%
GA-Ensemble	(F_1 F_3 F_{17} F_{32} F_{35}) 94.97%	(F_1 F_3 F_{15} F_{17} F_{18} F_{22} F_{29} F_{31} F_{32} F_{35}) 95.37%

TABLE V
CLASSIFICATION ACCURACY WITH OTHER ENSEMBLE METHODS

Methods	Sonar Data	Ionosphere Data	Soybean
Bagging C4.5	75.48%	92.02%	92.83%
AdaBoosting C4.5	80.29%	91.74%	93.27%

79.31% and 81.51%, respectively. Those results are 0.5%-2% lower than those achieved by GA-Ensemble method. For Ionosphere datasets, the best classification accuracy is achieved by 'MIFS-U' ($\beta = 1.0$), which is 91.18% for 5 features and 92.02% for 10 features, respectively. For the proposed method, the classification accuracies are again 1%-2.5% better off.

Figure 2 illustrates the z-score of the features selected from Sonar dataset and Ionosphere dataset, respectively. Two independent runs of each method are drawn on the same sub-graph to show the reproducibility and the stability. As can be seen from the diagram, the two independent runs using GA-Ensemble method are better overlapped compared with those using single classifier with single objective GA, in every case. These results demonstrate that GA-Ensemble method is comparatively more stable in feature selection. In addition, frequently selected features are calculated with high z-score. For Sonar dataset, F_{11} , F_{36} , F_{45} and F_{46} are the most

frequently selected features. For Ionosphere dataset, features F_1 , F_3 , F_5 , F_7 and F_{27} are the favorite ones in the selected results. As for Soybean dataset, the favorite features are F_1 , F_3 , F_{15} , F_{17} , F_{32} and F_{35} . It is worth noting that the selection of the smaller feature sets are generally more stable than bigger ones, with GA selector.

VI. CONCLUSION

In this study, we demonstrated that the proposed GA-Ensemble method outperforms the GADT, GANN and GASVM algorithms in both classification accuracy and stability of feature selection with three benchmark datasets. The classification accuracy with GA-Ensemble method is also generally higher than that obtained by the methods based on mutual information theory, bagging and boosting of C4.5. The GA-Ensemble employs different classifiers to select features and use majority voting to make sample classification. The idea is that different classifiers will use their own learning

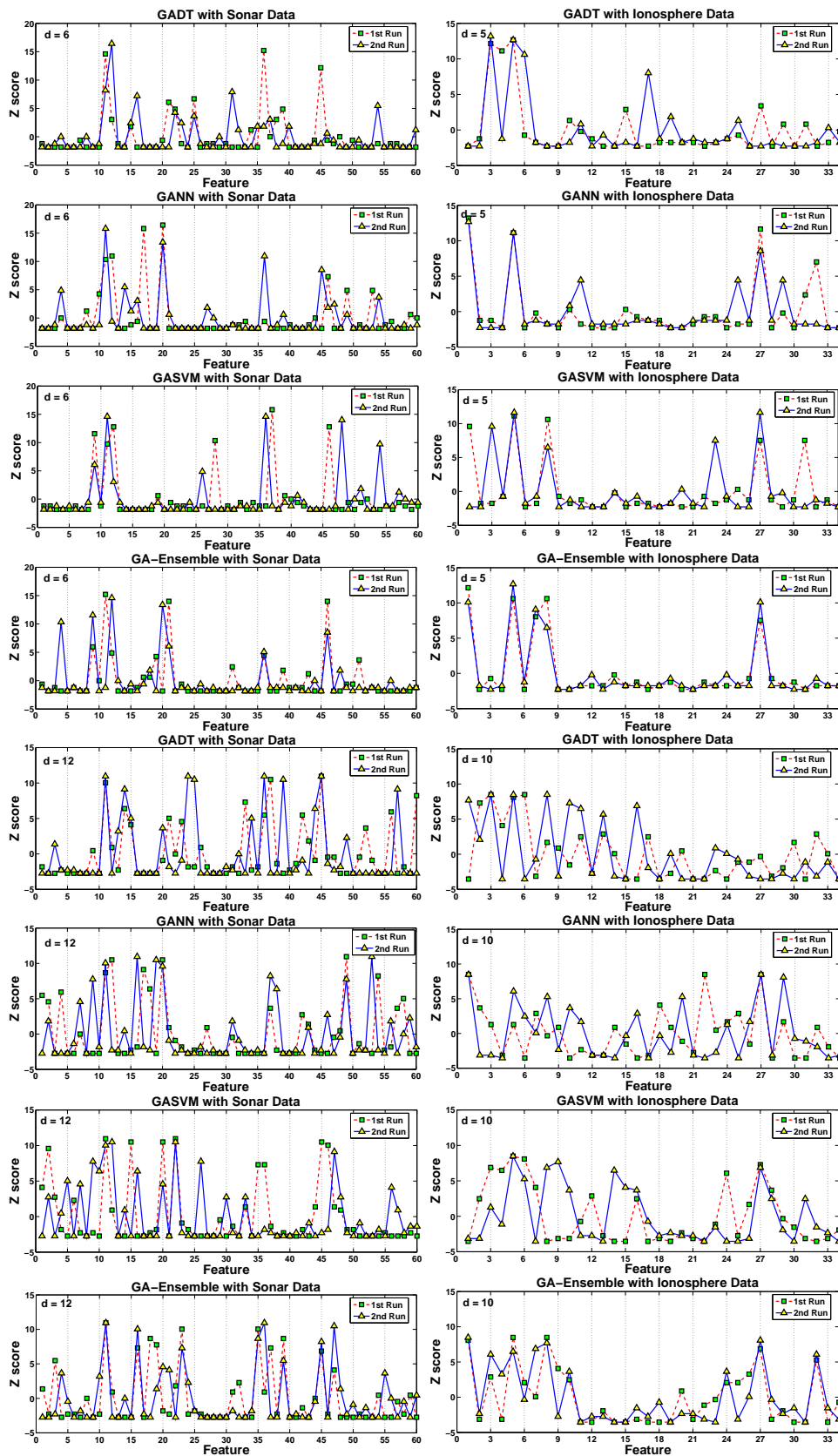


Fig. 2. Feature selection with Sonar dataset and Ionosphere dataset. Features selected by GADT, GANN, GASVM and GA-Ensemble methods. Z-score test is conducted to indicate the selected features. Note d is the feature combination length.

strategies to generate data classification hypothesis, and taking more hypotheses into consideration can improve data classi-

fication accuracy as well as generalization ability. The results suggest that the GA-Ensemble algorithm be a promising feature selection and sample classification algorithm.

REFERENCES

- [1] L. Xu, A. Krzyzak, and C. Suen, "Methods of combining multiple classifiers and their applications to handwritten numerals," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, pp. 418-435, 1992.
- [2] C. Suen, C. Nadal, T. Mai, R. Legault, and L. Lam, "Computer recognition of unconstrained handwritten numerals," *Proc. IEEE*, vol. 80, pp. 1162-1180, 1992.
- [3] A. Tan and D. Gilbert, "An empirical comparison of supervised machine learning techniques in bioinformatics," *Proceedings of the First Asia Pacific Bioinformatics Conference*, vol. 19, pp. 219-222, 2003.
- [4] S. Cho and H. Won, "Machine Learning in DNA Microarray Analysis for Cancer Classification," *Proceedings of the First Asia Pacific Bioinformatics Conference*, vol. 19, pp. 189-198, 2003.
- [5] S. Theodoridis and K. Koutroubas, *Pattern Recognition (Third Edition)*. Elsevier Press, 2006.
- [6] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, pp. 1226-1238, 2005.
- [7] P. Yang and Z. Zhang, "A Hybrid Approach to Selecting Susceptible Single Nucleotide Polymorphisms for Complex Disease Analysis," submitted to BMEI08 Conference.
- [8] P. Yang and Z. Zhang, "Hybrid methods to select informative gene sets in microarray data classification," *Proceedings of AI 2007, LNAI 4830*, Springer, pp. 811-815, 2007.
- [9] C. Ding and H. Peng, "Minimum Redundancy Feature Selection From Microarray Gene Expression Data," *Journal of Bioinformatics and Computational Biology*, vol. 3, no. 2, pp. 185-205, 2005.
- [10] K. Srinivasa, K. Venugopal, and L. Patnaik, "Feature Extraction using Fuzzy C-Means Clustering for Data Mining Systems," *International Journal of Computer Science and Network Security*, vol. 6, no.3A, 2006.
- [11] L. Li, C. Weinberg, T. Darden, and L. Pedersen, "Gene selection for sample classification based on gene expression data: study of sensitivity to choice of parameters of the GA/KNN method," *Bioinformatics*, vol. 17, pp. 1131-1142, 2001.
- [12] M. Skurichina and R. Duin, "Combining Feature Subsets in Feature Selection," *MCS 2005 LNCS 3541*, Springer, pp. 165-175, 2005.
- [13] N. Kwak and C.-H. Choi, "Input Feature Selection by Mutual Information Based on Parzen Windows," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1667-1671, Dec. 2002.
- [14] A. Tsymbal, M. Pechenizkiy, and P. Cunningham, "Diversity in search for ensemble feature selection," *Information Fusion*, vol. 6, pp. 83-98, 2005.
- [15] T. Dietterich, "Ensemble Methods in Machine Learning," *Proceedings of the First International Workshop on MCS. LNCS*, Springer, vol. 1857, pp. 1-15, 2000.
- [16] G. Giacinto and F. Roli, "An approach to the automatic design of multiple classifier systems," *Pattern Recognition Letters*, vol. 22, no. 1, pp. 25-33, 2001.
- [17] D. Miller and L. Yan, "Critic-driven Ensemble Classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 10, pp. 2833-2844, 1999.
- [18] A. Chandra, and X. Yao, "Evolving hybrid ensembles of learning machines for better generalisation," *Neurocomputing*, vol. 69, pp. 686-700, 2006.
- [19] D. Ruta and B. Gabrys, "Application of the Evolutionary algorithms for Classifier Selection in Multiple Classifier Systems with Majority Voting," *Proceedings of MCS 2001, LNCS 2096*, Springer, pp. 399-408, 2001.
- [20] X. Li, S. Rao, Y. Wang, and B. Gong, "Gene Mining: A Novel and Powerful Ensemble Decision Approach to Hunting for Disease Genes Using Microarray Expression Profiling," *Nucleic Acids Research*, vol. 32, no. 9, pp. 2685-2694, 2004.
- [21] S. Cho and P. Chanho, "Speciated GA for Optimal Ensemble Classifiers in DNA Microarray Classification," *Evolutionary Computation, 2004. CEC2004. Congress on*, vol. 1, pp. 590- 597, 2004.
- [22] G. Bontempi, "A Blocking Strategy to Improve Gene Selection for Classification of Gene Expression Data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 4, no. 2, pp. 293-300, 2007.
- [23] Á. Blanco, M. Martín-Merino, and J. de las Rivas, "Ensemble of Dissimilarity Based Classifiers for Cancerous Samples Classification," *PRIB 2007, LNBI 4774*, Springer, pp. 178-188, 2007.
- [24] L. Hansen and P. Salamon, "Neural Network Ensembles," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 12, pp. 993-1001, 1990.
- [25] E. Bauer, R. Kohavi, "An empirical comparison of voting classification algorithms: bagging, boosting, and variants," *Machine Learning*, vol. 36, no. 1-2, pp. 105-139, 1999.
- [26] T.G. Dietterich, "An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization," *Machine Learning*, vol. 40, no. 2, pp. 139-157, 2000.
- [27] A. Chandra, and X. Yao, "Ensemble learning using multi-objective evolutionary algorithm," *Journal of Mathematical Modelling and Algorithms*, vol. 5, no. 4, pp. 417-445, 2006.
- [28] C. Blake and C. Merz, UCI repository of machine learning databases, <http://www.ics.uci.edu/MLRepository.html>, 2007.
- [29] A. Ben-Dor, L. Bruhn, N. Friedman, I. Nachman, M. Schummer, and N. Yakhini, "Tissue Classification with Gene Expression Profiles," *Journal of Computational Biology*, vol. 7, pp. 559-584, 2000.
- [30] E. Keedwell and A. Narayanan, *Intelligent Bioinformatics*. John Wiley & Sons, Ltd, pp. 159, 2005.
- [31] L. Lam and Y. Suen, "Application of Majority Voting to Pattern Recognition: An Analysis of its Behaviour and Performance," *IEEE Transactions on Systems, Man, and Cybernetics* vol. 27, no. 5, pp. 553-568, 1997.
- [32] R. Battiti, "Using Mutual Information for Selecting Features in Supervised Neural Net Learning," *IEEE Trans. on Neural Networks*, vol. 5, no. 4, pp. 537-550, July 1994.
- [33] M. Tesmer and P. Estevez, "AMIFS: Adaptive Feature Selection by Using Mutual Information," *Proceedings of the IEEE International Joint Conference on Neural Networks IJCNN2004, Budapest, Hungary, July 26-29*, pp. 303-308, 2004.
- [34] M. Tan, and L. Eshelman, "Using weighted networks to represent classification knowledge in noisy domains," *Proceedings of the Fifth International Conference on Machine learning*, pp. 121-134, 1988.
- [35] P. Brierley and B. Batty, *Data Mining with Neural Networks - an Applied Example in Understanding Electricity Consumption Patterns*. Knowledge Discovery and Data Mining (ed Max Bramer). pp. 240-303, IEE, 1999.
- [36] R. Collobert and S. Bengio, "SVM-Torch: Support vector machines for large-scale regression problems," *Journal of Machine Learning Research*, vol. 1, pp. 143-160, 2001.
- [37] J. Quinlan, *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.
- [38] D. Goldberg and K. Deb, "A Comparative Analysis of Selection Schemes used in Genetic Algorithms," *Foundations of Genetic Algorithms*. San Mateo, CA: Morgan Kaufmann pp. 69-93, 1991.
- [39] S.L. Jackson, *Research Methods and Statistics: A Critical Thinking Approach (2 edition)*. Wadsworth Publishing, 2005.