Providing a Combination Classification (Honeybee Clooney and Decision Tree) Based on Developmental Learning

Seyed Ahad Zolfagharifar[1], Faramarz Karamizadeh[2] & Hamid Parvin[3]

[1] Department of Computer Engineering, Yasouj Branch, Islamic Azad University, Yasouj, Iran

[2] Department of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran

[3] Department of Computer Engineering, Yasouj Branch, Islamic Azad University, Yasouj, Iran

Correspondence: Faramarz Karamizadeh, Department of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran. Tel: 98-938-668-1700. E-mail: f.karamizadeh@gmail.com

Abstract

The aim of this study is to provide a combination classification based on developmental learning in the proposed method using algorithms inspired by nature (honeybee Clooney) and decision tree, by using algorithm classifier consensus is proposed that this method, at first classifier once implemented and based on the detection rate of input data agreement in the final consensus which is an innovation in this research. To implement the proposed algorithms used MATLAB software. Note that, this is an increase compared to the classifiers ensemble, it have accuracy and fix. This shows that this method of making the ensemble by helping bee Clooney algorithm, when appropriate and effective which the number of data collection records is high or the number of study characteristics is high. In this study, we proposed algorithm on 8 samples tested. However, training time of this method compared with simple ensemble is a slower process but this method compared with simple ensemble method has higher accuracy, this shows, if we want a higher accuracy, we should be spent more time.

In general, if the accuracy of the process have a large importance for us, this method can be a good option to get the results that almost optimal.

Keywords: classifiers consensus, honeybee Clooney, combination, decision tree, classifiers ensemble

1. Introduction

Classification specifically means the discriminant function so that this function n dimension space to the decision region of classes mapped. Today, for diagnosis and problem solving classifiers used. The use of classification learning is an effective approach in the machine learning that in recent years researchers have paid it. In this type of learning in order to improve learning accuracy, the results of classification (such as MLP[1], SVM (Note 1) etc.) are combined together, to this point say the first phase or the classifiers production. Neural networks are the most common choice for basic classifiers. The theory results and experimental show that, when collective learning is better than learning of the basic classifiers which basis classifiers has the acceptable performance and different in the error.

Two classification when have verity in the error which patterns classified incorrectly that are different. Difference in the error cases of basis classifiers, causes covers each other errors and therefore, diversity in error is basics point in the success of a combinatorial classifiers system.

Now the second stage or the combination of classifiers which is done after the first step, at this step, we get classifiers conclude used in the previous step as input to one or more of other classification to base on the output of the first step decision-making.

There are several modes for production and the combination of the first step (classifiers production) and the second step (combined classification). We intend in this article to compare the production different scenarios and combined classifiers and compare them and display output of them by simulating pay, as well as how combine them in the best form. in all the results, we have used standard data sets, see the table, a significant increase in accuracy either on experimental data sets as well as on the validation data set is clear.

We're going to study the effect of classifiers on the basis classifiers in order to increase the combinatorial

classifiers efficiency review.

2. Methodology

2.1 Classifier Combination

Combination of multiple classifiers can be used as a general pattern recognition problem, where inputs are separate classifiers results and outputs are their combination decision considered. This is because the classifiers with different characteristics and or different methodologies can complement each other and cover each other's weaknesses, originated. If several different classifiers with together as vote ensemble their overall error can be significantly reduced. Note that, this title in different places known with different names.

The combination of multiple classifiers, classifier fusion, mixture of experts, committee of neural networks, combinational classification systems, classifier ensemble, divide and conquer classifiers, decision forest and so on. But in general, all methods that benefits from combine the classifier say classifier ensemble while only in the condition usage separate classifier in combination, are applied simultaneously and independently, we have intercourse classifier.

2.2 Hypothesis: Why Classifier Combination Work?

To better understand why classifier combination is effective, attention to the below theorem.

Condorcet Theorem:

If every voter has the possibility of p to correct say and the possibility that the majority of voters, correct comment is m then if $p> 0.5$ it must necessarily $m> p$. In short, m to converge 1, for all $p> 0.5$, provided that the number of voters to the infinite [9].

In general, it can prove that:

•For $p> 0.5$, we have: if L to the infinite then will be m = 1.

•For $p <0.5$, we have: if L to the infinite then will be m = 0.

•For $p = 0.5$, we have: if L to the infinite then will be m = 0.5.

Where L is the number of voters and m majority vote [1].

2.3 Classifier Ensemble Definition

A set of basic classifier which together used to solve the problem of pattern recognition, and their decisions in order to enhance the efficiency of the whole system, are combined with together. Although the composition of the decision how is doing depends on the output of basic classifier in the mix. If their type only the label categories, then used the voting procedures.

2.4 Definition Multiple Classifiers Systems

In general, the combination of classifiers can be done in four levels. The first high-level, means that at the level which the output of several classifiers combine with a variety of methods. Here, we have at least several basic classifiers which their decisions with method for example majority combined with together. We named this level, combination level. The next level is the classifier level. At this level, the different basis classifier can use to build multiple classifiers systems of compound. The next level is the level of features where it is assumed that we have a number of features and for each classifiers a subset of the features choose. The last level is the level of data (Figure 1).
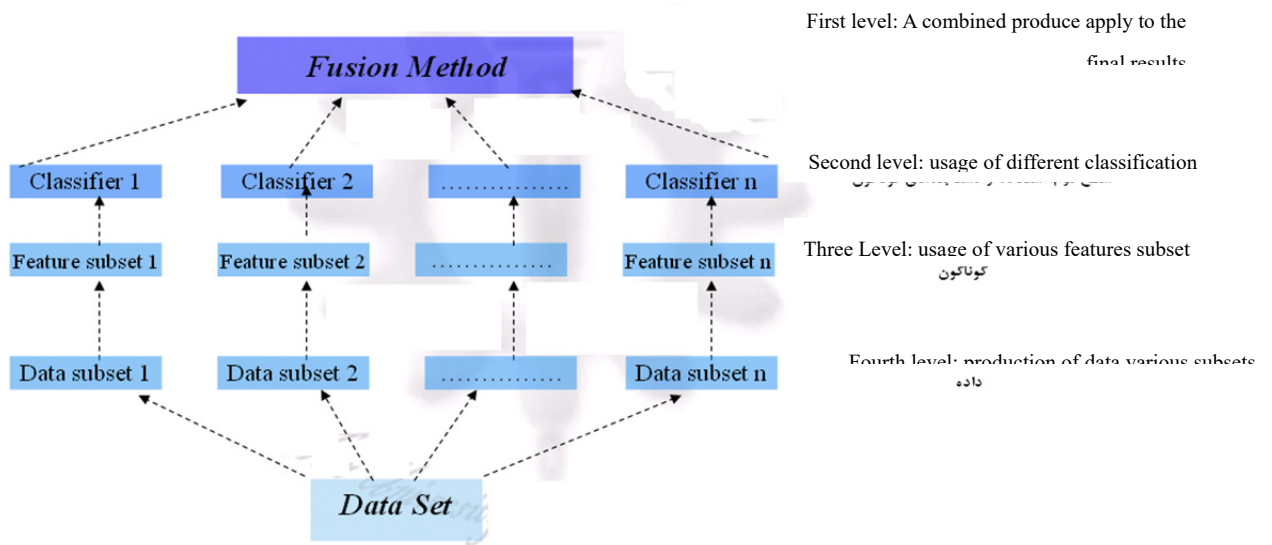
Figure 1. The different levels of classifiers ensemble production

2.5 Classifiers Output Combination

What a way use to different classifiers output combination depend on the output of existing classifiers in the ensemble that we look at kind of them here.

The first type: the output of each classifiers D_i for a given sample pattern as is right and wrong, so that for a given set Z, classifiers Di output vector y_i produces. y_{ij} equal to one if classifiers D_i pattern of z_j classification correct and otherwise is zero.

The second type data: each classifiers D_i produces a label δ_i for a test pattern X, so that for a pattern a vector $\delta = [\delta_1, \delta_2 \dots \delta_n]$ occurs which L is the number of classifiers. Here, we do not have any information about the reliability labels and even we do not have replacement label to label a pattern.

The third type data: Here's each classifiers D_i a priority of the correct label to the wrong label.

The fourth type data: Here's each classifiers D_i produces a vector of c dimensional $[d_{i,1}, d_{i,2}, \dots, d_{i,n}]$ value of $d_{i,j}$ generates the validity of the hypothesis that, "sample X to category ω_j has member "display. In most cases $d_{i,j}$ is a number between zero and one. Now, check several methods of output combining of various classifiers pay.

A) Majority vote: suppose that classifiers outputs as binary vectors of c dimensional $[d_{i,1}, d_{i,2}, \dots, d_{i,n}]$ that $d_{i,j} = 1$, if classifiers D_i sample X placed in the category ω_j and otherwise $d_{i,j} = 0$. Vote of plurality mood for category ω_k for calculated following formula:

$$\sum_{i=1}^{L} d_{i,k} = \max_{j=1}^{c} \sum_{i=1}^{L} d_{i,j}$$

(1)

If in case, for example two class (c = 2) simple majority (50 percent of the vote plus one) to a class k this vote called majority vote. One of the weaknesses of the majority vote does not guarantee the performance boost, with three classifiers with 60% accuracy at a minimum mood by majority vote to reach 40% efficiency. The best performance which is obtainable from a combination method of the pattern of success and worst possible performance called pattern of failure. Although the pattern of success for example, the mood top is 90 percent however, there is not guarantee to the success event. Attention to the table below

Table 1. Summary of output moods of the three classifiers for ten pattern

No.	111 a	101 b	011 c	001 d	110 e	100 f	010 g	000 h	P_{maj}	$P_{maj} - p$
Pattern of success										
1	0	3	3	0	3	0	0	1	0.9	0.3
2	2	2	2	0	2	0	0	2	0.8	0.2
3	1	2	2	1	3	0	0	1	0.8	0.2
4	0	2	3	1	3	1	0	0	0.8	0.2
5	0	2	2	2	4	0	0	0	0.8	0.2
6	4	1	1	0	1	0	0	3	0.7	0.1
7	3	1	1	1	2	0	0	2	0.7	0.1
8	2	1	2	1	2	1	0	1	0.7	0.1
9	2	1	1	2	3	0	0	1	0.7	0.1
10	1	2	2	1	2	1	1	0	0.7	0.1
11	1	1	2	2	3	1	0	0	0.7	0.1
12	1	1	1	3	4	0	0	0	0.7	0.1
Identical classifiers										
13	6	0	0	0	0	0	0	4	0.6	0.0
14	5	0	0	1	1	0	0	3	0.6	0.0
15	4	0	1	1	1	1	0	2	0.6	0.0
16	4	0	0	2	2	0	0	2	0.6	0.0
17	3	1	1	1	1	1	1	1	0.6	0.0
18	3	0	1	2	2	1	0	1	0.6	0.0
19	3	0	0	3	3	0	0	1	0.6	0.0
20	2	1	1	2	2	1	1	0	0.6	0.0
21	2	0	2	2	2	2	0	0	0.6	0.0
22	2	0	1	3	3	1	0	0	0.6	0.0
23	2	0	0	4	4	0	0	0	0.6	0.0
24	5	0	0	1	0	1	1	2	0.5	−0.1
25	4	0	0	2	1	1	1	1	0.5	−0.1
26	3	0	1	2	1	2	1	0	0.5	−0.1
27	3	0	0	3	2	1	1	0	0.5	−0.1
Pattern of failure										
28	4	0	0	2	0	2	2	0	0.4	−0.2

Table 1 mood a (111) means that the number of patterns that classifiers of D_1, D_2 and D_3, the classification action for them doing right, and mood b (101) means that the number of patterns that classifiers of D_1 and D_3 classification action for them doing right, and classifiers of D_2 classification them wrong.

B) Weighted majority vote: if existing classifiers in a combination ensemble do not have the same precision, it is reasonable that closer classifiers have a greater impact on the final decision. However, $d_{i,j}$ defined as follows:

$$d_{i,j} = \begin{cases} 1, & if \quad D_i \quad labels \quad X \quad in \quad \omega_j \\ 0, & otherwise \end{cases}$$

(2)

Then separation function, to ω_j class such obtain:

$$g_i(X) = \sum_{i=1}^{L} b_i d_{i,j}$$

(3)

That b_i is a coefficient for classifiers of D_i. For example, assume that three classifiers of D_1 and D_2 and D_3 with accuracy of 0.6, 0.6 and 0.7. By combining majority vote method, $p_{maj} = 0.696$ but with the weighted majority vote combination method with weights $b_1 = b_2 = 0$ and $b_3 = 1$ (which practically we remove vote of classifiers of D_1 and D_2) $p_{maj} = p_3 = 0.7$ will be. One way to weighting to the classifiers for achieve maximum majority vote accuracy through the following formula:

$$b_i \propto \log \frac{p_i}{1 - p_i}$$

(4)

That pi is the vote accuracy of D_i classifier, provided that L classifier of D_i independently of one another [1].

2.6 The Proposed Method

Overall the classifier combination may be done in the four levels or procedures. The combination level or integration, classification level, features level and data level.

In this study, used the combination level or fusion method. The combination level is the level that the outputs of multiple classifier combined with a variety of methods. Here, there is at least some basic classifier that their decisions by way of example, the majority vote combined with together. At this level, to any classifier within each class weighted and the participation rate of each class in the final classification determined. Each classifier how percentage involved in the consensus.

In the proposed method, first using decision tree algorithm divided data collection in different classifiers and each classifier divided the smaller classifier then the classifiers training and outcome of education and testing of learning amount of decision trees classifier stored, the results of the previous step which represents the amount of learning each classifier in the class shows to the defined performance function sent, using bee colony algorithm, the efficiency function improved. Then amount of performance and the percentage recognition of each class, obtained in the classifier and each classifier based on percentage recognition of subset class participate in consensus. To better understand the example.

Examples:

Suppose that we have three device of separators fruit or classifier (A, B, C) we want to use this device separated apples and pears, it means that each classifier had two mood or class and in total six mood arise. To each class of the classifiers gets a random value, actually have a value between zero and one which represents the recognition of apple from pear by any device, as the initial value such as Table 2 we arrived.

Table 2. Apply random weight to the chromosomes

type of mode	Classifier A		Classifier B		Classifier C	
	Class1(pear)	Class2 (apple)	Class1(pear)	Class2 (apple)	Class1(pear)	Class2 (apple)
Random value (chromosomes random)	0.9	0.4	0	0.1	0.4	0.2

After classification, identification of classes and choose the initial state that the accuracy of each classifier implies to the device (Classifier) get some fruit as example (data collection) and in fact solutions are opinion of classifier to a class save.

In this example as shown in Figure 2, half of input data are apples and the other half are pears.
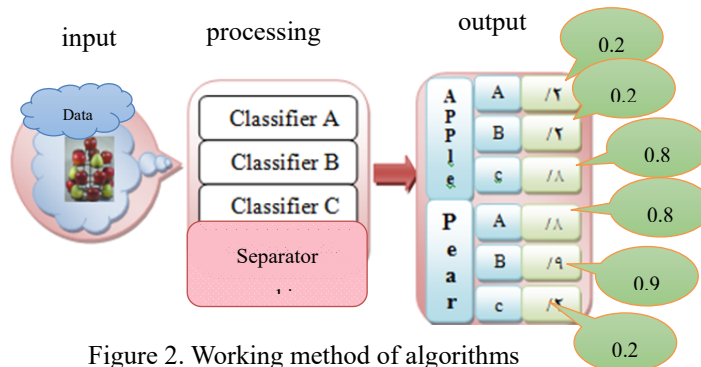

Figure 2. Working method of algorithms

Table 3. Calculate the new weight of Clooney

type of mode	Classifier A		Classifier B		Classifier C	
	Class1	Class2	Class1	Class2	Class1	Class2
Classifier	0.2	0.8	0.2	0.9	0.8	0.2

After determining the vote of each classifier for example Table 3 about the input data, now should earn consensus for each class.

$$W_{Apple} = \frac{\sum_{i=1}^{k}(R_i^{apple} \times D_i^{apple})}{\sum_{i=1}^{k}(R_i^{apple})} \times P(Apple) \tag{5}$$

In the formula R_i^{apple} initial values of classifier to class II (apple), D_i^{apple}, the values obtained in the first run of machine for the second class, K number of classifiers and P is percent of the class data rather than all data that in this example, the data half   is apple and the other half is pear considered. For all classes is calculated.

In this example and according to data values and consensus votes each input data classifier, pears or class 1 has been detected.

Then, because of classifier accuracy percent is low, we will update the initial values and re-run the algorithm we do this several times until the accuracy of the algorithm close to the desired value, means in fact, with training data collection teach algorithm and at the end between all outputs choose input data related to best accuracy. Finally, from the experimental data sets to test the algorithms use.

2.7 The Proposed Algorithm

2.7.1 Decision Tree Algorithm

The decision tree algorithm works on a specific goal analyzes several features and provide conditions for forecasts and targeted of sales. Decision trees used for predict category variables, classification trees are called because the samples set in category or classes. Decision trees that are used for prediction of continuous variables called regression trees [2]. Most learning algorithms, decision tree based on a greedy searching action top to bottom in the space of the existing trees.

In the decision tree ID3 used a statistical value that called the Gain information to determine that how much can be a feature training examples according to their classification separate.

Entropy:

The purity rate (disorder or lack of purity) defines a set of examples. If the set S includes examples of positive and negative from a sense of purpose, the entropy S with respect to the classification of Boolean defined as follows.

$$E(s) = -(p \log_2 p + \bar{p} \log_2 \bar{p}) \tag{6}$$

Information Gain:

Information Gain of a feature is the amount of entropy reduction that by isolating example through this feature achieved.

In other words, the information Gain (Gain (S, A)) for a feature such as A to set of examples S defined as follows:

$$Information gain = Entro~(s) - \Sigma sv|s|v \in Values~(A) * \text{Entropy}(s) \tag{7}$$

Which Values (A) is the set of all the features of A and VS is a subset of S which A has V value. In the above definition, first term is the data entropy and the second term is the amount of entropy expected after the separation of data [2].

2.7.2 Bee Clooney Algorithm

In the bee Clooney algorithm (ABC) Bees include in three groups:

In the algorithm ABC, for the first half of the population of bees are worker bees and the other half are browser bee. For each food source, there is only a worker bee. In other words, the number of worker bees equal to the

number of food sources around the hive. Worker bees have been exhausted at working in food sources will be leading browser bees.

The main steps of the algorithm is given below:

- Initialization.
- Repeat.

(A) Place of the worker bees in food source in the memory:

(B) Place of the browser bees in food source in the memory:

(C) Send the leading bees to search for new food sources:

- to (achieve desired situation.)

2.8 Evaluation Criteria

To evaluate the performance of proposed algorithms have been used the two criteria. One error variance or error of error and other statistical analysis. These two criteria are accurate criteria and at work as the main criteria of quality classifier which they have been used [1].

2.9 Evaluation of Different Methods

To evaluate different algorithms and compare algorithms with the proposed method at first the algorithms and the proposed method on several data sets of table 4 run (on each running ten times) then the percent of accurate and error each method earned in each time running and will be entered in an Excel file and gave mean, standard deviation of accuracy and error of each method on every collection.

Table 4. Collection of used data

row	set	Explanation
1	Breast	This collection of data related to risk factors for breast cancer in women of America.
2	Bupa	The data collection analysis liver disorders
3	Galaxy	Collect classified from information to radar returns the ionosphere
4	Glass	The data collection related to glass, the glass produced.
5	ionosphere	This data set for eleven information categories radar ionosphere signal to send or receive it.
6	SAHeart	Data collection for men who have been diagnosed with heart disease.
7	wine	The data collection related to recognition of consumed.
8	Iris	The data collection for the identification of lily flowers

The average achieved order in decreasing until algorithm based on accuracy sorted and determine which method is the best accuracy in the data collection.

2.10 Accuracy in Different Method

For doing this study, we have 20 samples of the classified algorithm according to the Table 5 selected randomly and with the available data sets, they've performed 10 times and the results were saved. To run the sample algorithms used the Weka software. This box diverse tool and comprehensive through a common interface available, so that the user can compare different methods with together and methods that are more appropriate to consider the issues, recognize.

Table 5. Table of classification algorithms

row	1	2	3	4	5	6	7	8	9	10
Name of input classifier	MLP	IBK or KNN	SMO or SVM	Naive Bayes	random forest	random tree	stacking	adabootm1	bagging	multy scheme
Basic classifier(output)	-	-	-	-	-	-	bagging	Decision stump	decision stump	decision stump
row	11	12	13	14	15	16	17	18	19	20
Name of input	vote	staking	multy	staking	stacking	vote	vote	Vote	stacking	vote

| classifier | | | scheme | | | | | | | |
| Basic classifier(output) | decision stump | desionstump | filtered classifier | j48 | MLP | MLP | rbfclassifair | reptree | tree_j48 | zeroR |

2.11 Results of Comparative

If the average of the existing classifier results put together in the past part and the results regular based on data tables 6 show better result.

Attachment (6): Results of a variety methods of classifier assembly on the different sets (thick values are the best results in each dataset)

Name of input classifier	breast	bupa	galaxy	glass	ionosphere	SAHeart	wine	iris
before my ABC	95.80	64.20	67.42	66.62	92.71	68.82	92.08	96.00
after my ABC	95.78	66.52	68.66	66.31	92.71	69.57	94.34	96.67
MLP	92.34	68.32	68.76	67.06	91.51	68.48	97.19	96.47
IBK or KNN	95.67	62.96	79.60	69.44	86.75	64.81	94.83	95.40
SMO or SVM	96.93	58.09	25.33	57.57	88.03	72.36	98.82	96.00
Naive Bayes	96.31	55.68	57.86	48.22	82.48	70.15	97.36	95.53
ranodom tree	93.79	64.00	70.31	69.95	88.77	62.36	92.19	92.80
random forest	65.01	57.97	24.77	35.51	64.10	65.37	39.89	33.33
adabootm1-Decision stump	94.58	66.00	26.07	44.81	91.11	70.56	89.27	95.67
bagging-decision stump	93.22	60.75	31.67	46.45	83.05	66.32	85.34	72.20
staking-decision stump	65.01	57.97	24.77	35.00	64.10	65.37	35.45	33.33
staking- j48	93.75	66.78	58.33	65.19	89.37	70.58	88.76	89.53
staking- MLP	95.99	68.99	69.81	65.93	90.94	67.58	97.42	96.67
stacking-tree_j48	65.01	57.97	24.77	35.28	64.10	65.37	39.89	33.33
stacking-bagging	65.01	57.97	24.77	34.95	64.10	65.37	39.89	33.33
vote- zeroR	65.01	57.97	24.77	35.51	64.10	65.37	39.89	33.33
vote -rbfclassifair	92.80	59.97	26.01	44.81	82.48	64.72	58.60	66.67
vote- decision stump	92.49	59.54	26.04	44.86	82.54	65.19	59.10	66.67
vote - reptree	94.41	64.55	66.69	65.37	89.37	68.31	90.28	93.40
vote- MLP	96.15	69.04	68.58	66.78	91.17	68.74	97.19	96.67
multy scheme-filtered classifier	95.14	57.62	56.87	71.17	90.03	70.00	90.95	93.53
multy scheme-decision stump	92.50	60.55	26.01	44.81	82.45	65.15	58.20	66.67
mean	88.54	62.49	48.29	54.68	83.29	67.52	77.80	77.77
Standard deviation	12.41	4.44	21.56	13.63	10.61	2.58	23.61	25.43

2.12 Observation Charts

For better analysis of results in this section with several number of classifier status graphs image down.
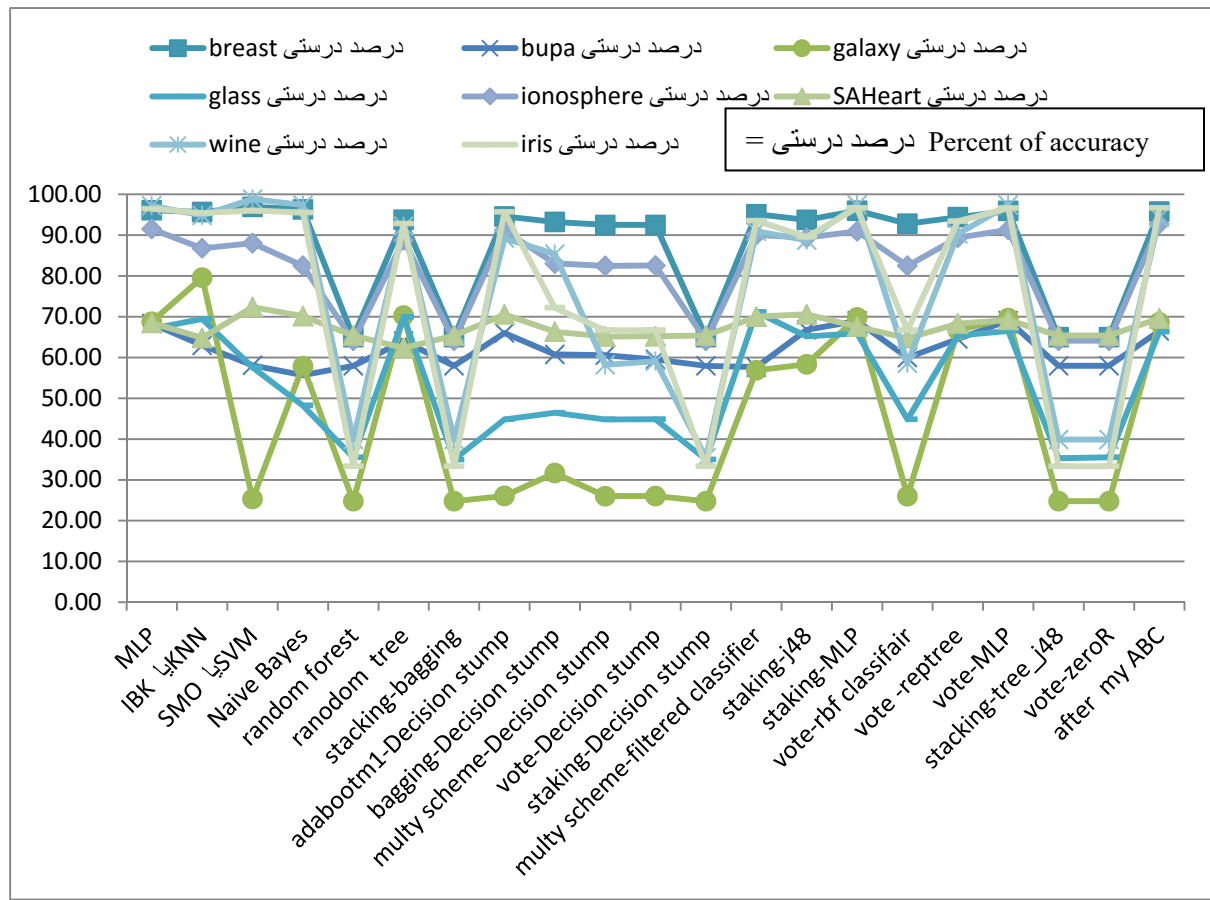
Figure 1. Classifier performance mean

In Figure 1, we can see the difference between mean. In this graphs clearly show there is a significant difference between the mean of data for example data set of galaxy which display withe green color, consider, you will see that output of datasets from 20 percent to 80 percent related to KNN classifier   that has just been diagnosed.

To view the mean and standard deviation of classifier accuracy percent criteria than the data set, Figure 2 is continued. As you can see, this chart shows the average of the red line, the above square of red line and Low Square shows the standard deviation of the results.
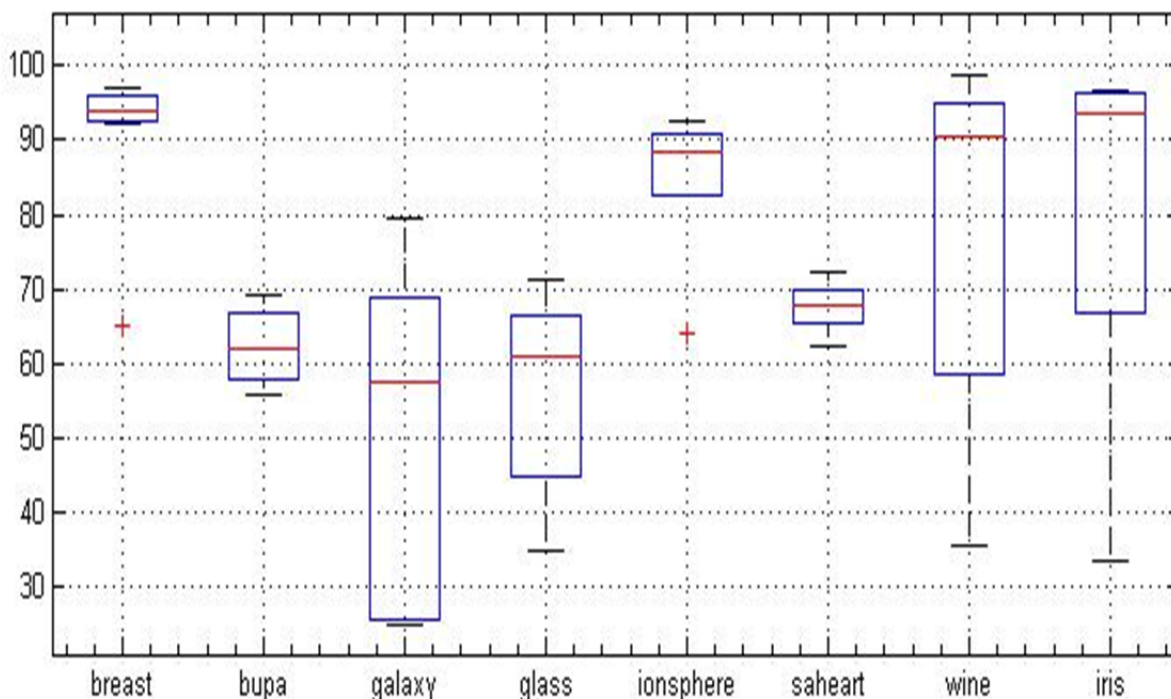
Figure 2. Mean and SD of standard deviation of the classifier implementation results

3. Conclusion

Due to growing demand in recent decades to efficient intelligent computing providing a classification algorithm with better performance is very justified and valuable. Studies have shown that simple issues solve by common classification algorithm but difficult issues with high complexity, their use is inefficient. Even resolution issues while our knowledge of the relations between classes is poor, can be very difficult. To solve this growing problem have been proposed learning techniques under areas.

In this paper, we tried to compared a variety of ways and to date of combined classification and came to the conclusion that the proposed combined classification method which inspired methods from the nature of use, in cases where the number of data or number of great features, it might be the best method. However, to continue this work can be more criteria for mapping data collection and the number of sub-categories for each numerical algorithm consider that can be adhered reviewed.

In this research tried to a general method for classification in environments prone to error raised. Although the report of mining data cited that numerical methods how that can be imposed on a non-numerical data, but it is suggested the output of this method for data collection that benefits numerical features to be employed.

Reference

Akbani, R., Kwek, S., & Jakowicz, N. (2004). Applying support vector machinesto imbalanced datasets. In Proceedings of European Conference on MachineLearning, pages 39-50, Pisa, Italy.

Fukunaga, K. (1972). Introduction to Statistical Pattern Recognition. Academic Press, Inc., Orlando, FL, 1972.

Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematicstudy. *Intelligent Data Analysis Journal, 6*(5), 429-450.

Kohonen, T. (1990). Improved versions of learning vector quantization. In Proc. InternationalJoint Conference on Neural Networks, San Diego, CA, I545–I550.

Kuncheva, L. I. (2005). Combining Pattern Classifiers, Methods and Algorithms. New York: Wiley.

Marquis, J. A. (1784). Condorcet. Sur les elections par scrutiny. *Histoire de l'Academie Royale desSciences*, 31–34.

Skalak, D. B. (1994). Prototype and feature selection by sampling and random mutation hillclimbing algorithms.In roc. 11th Int. Conf. on Machine Learning, New Brunswick, NJ, 1994, Morgan Kaufmann, Los

Altos, CA, 293–301.

Vapnik, V., & Lerner, A. (1963). Pattern recongintion using generalized portrait method. *Automation and Remont Control, 24,* 774-780.

Wu, G., & Chang, E. Y. (2003). Class-bounday alignment for imbalanced datasetlearning. In Proceedings of the ICML'03 Workshop on Learning from ImbalancedData Sets, Washington, DC, August 2003.

Notes

Note 1.Multi-layer perceptron

Note 2. Super vector machine

Copyrights