A FILTER-BASED FEATURE SELECTION METHOD APPLIED IN DIFFERENT STRUCTURES OF CLASSIFIER ENSEMBLES

Anne Magály de Paula Canuto[1], Karliane Medeiros Ovidio Vale[1], Araken Medeiros Santos[1,2] and Antonino Alves Feitosa Neto[1]

[1]Informatics and Applied Mathematics Department, Federal University of RN, Campus Universitário S/N, Lagoa Nova, CEP 59072-970 Natal, RN – Brazil

[2] Federal Rural University of Semi-Árido - UFERSA, Campus Angicos, CEP 59515-000 Angicos, RN - Brazil

anne@dimap.ufrn.br; karlianev@gmail.com; araken@ufersa.edu.br; antonino_feitosa@yahoo.com.br

Abstract – Diversity is considered as one of the main prerequisites for an efficient use of ensemble systems. One way of increasing diversity is through the use of feature selection methods in ensemble systems. In this paper, a class-based feature selection method for ensemble systems is proposed. The proposed method is inserted into the filter approach of feature selection methods and it chooses only the attributes that are important only for a specific class. An analysis of the performance of the proposed method is also investigated in this paper and it shows that the proposed method has outperformed the standard feature selection method.

Keywords – Ensemble systems, Diversity, Feature Selection

1 Introduction

In the search for efficient pattern recognition systems, it has been often found that no single classifier is entirely satisfactory for a particular task, and hence the idea of combining different classification methods has emerged as potentially very promising [2],[4]. The main example of this idea is the ensemble systems (or committees), which exploit the idea that a pool of different classifiers can offer complementary information about patterns to be classified, improving the effectiveness of the overall recognition process.

In the context of ensemble, diversity is one aspect that has been acknowledged as very important [5]. For example, there is clearly no accuracy gain in a system that is composed of a set of identical base classifiers. One way of increasing diversity is through the use of feature selection or data distribution in ensemble systems. Feature selection methods can be divided into two main approaches, which are: filter or wrapper. This paper presents a filter feature selection method for ensemble systems. Unlike most of the filter methods, the method proposed in this paper uses a class-based filter method, in which attributes that are good only for the corresponding class are chosen to represent this class. The ensembles systems will be composed by classifiers which are expert in answering about the belongingness of an input pattern to a specific class. In this sense, there will be, at least, one classifier per class in the ensemble systems.

In order to analyze the feasibility of the proposed method, it will be compared with a standard filter-based method, which also uses a class-based procedure and they will be applied to four different datasets. In addition, the performance of these systems will be analyzed in different structures (homogeneous and heterogeneous).

2 Feature Selection in Ensembles

Feature selection methods can be defined as the process that chooses the best attributes subset according to a certain criterion, excluding the irrelevant or redundant attributes. In using feature selection methods, it is aimed to improve the quality of the obtained results. In the context of ensembles, feature selection methods aim to reduce redundancy among the attributes of a pattern and to increase the diversity in such systems. Recently, several authors have investigated the use of feature selection methods in ensembles, such as in [3],[6],[7],[9],[10]. There are several feature selection methods that can be used for ensembles, which can be broadly divided into two main approaches, which are: filter and wrapper. In the filter approach, as it can be found in [6],[7],[8], no need for a classification method to be used during the feature selection process. In other words, the feature selection process is independent from the classification method. On the other hand, the wrapper approach, as it can be found in [3],[9],[10], the feature selection process is dependent from the classification method. The feature subset is chosen based on the classification method used. Two different classification methods lead to different feature subset chosen.

The major drawback of the filter approach is the efficiency of these methods, while the major drawback of the wrapper approach is the computational time, since e the filter approach is generally computationally more efficient than the wrapper approach. In the context of ensembles, the major drawback of the wrapper approach is emphasized, since the assessment criterion usually has to take into consideration the accuracy of the whole ensemble system, increasing even further the complexity of this function. In contrast, as the ensemble system used a two-step decision making process (individual classifiers and combination method), the dependency of the chosen subset with the classification methods can be smoothed out, since the classifier accuracy is not the only parameter to define the accuracy of the ensemble system. Because of this, the use of the filter approach to select different subsets of attributes for the individual classifiers in an ensemble has become an interesting choice.

In the filter approach, usually, a ranking procedure is performed in which attributes are assessed for all classes of the problem, called general ranking process. However, it is well known that different classes of a problem can have different particularities and levels of difficulty. When using a general ranking procedure (for all classes), difficulties of one class are distributed among all others. In this sense, classes which are not very difficult to be classified may become more difficult. Moreover, different classes of one problem might need a different number of attributes to be classified. For instance, an attribute can be very important to one class and not very important to other classes. Because of this, the idea of using class-based ranking has emerged.

There are some works in the literature which use class-based feature selection, such as the favorite class method [5]. In these works, the choice of the attributes is based on the importance of the attributes for the analyzed class. However, an attribute can have a similar importance for two or more classes. In this sense, even when using class-based feature selection, this attribute will probably be chosen for both classes. Nonetheless, the choice of this attribute may affect the accuracy of the classifiers, making it confuse patterns of both classes.

2.1 The Class-based Feature Selection Method

Aiming to smooth out the aforementioned problems, a feature selection method is presented. It is a filter method which will search for attributes that are important for one class and not very important for other classes, letting the classification method more secure about the class to be classified (attribute which will not make confusion in the classification method). The main idea is that one classifier, which will be responsible for classifying patterns of one class, will base its decision on attributes which are important only for this class.

In this method, a new ranking strategy is added in the feature selection method. As already mentioned, in a class-based ranking, attributes are ranked, for each class, based on a criterion from the most important to the least one. Then, the first N attributes are picked. In the presented method, a second ranking is executed, in which the position of the attribute in the ranking of the analyzed class, along with the position of this attribute in the other classes are taken into consideration. In this second ranking process, the idea is that if an attribute is highly (in the top positions) ranked in an analyzed class, this means that this attribute is important for the classification process of this class and this will be positively counted for this attribute. Also, if an attribute is also highly ranked on the other classes, this means that this attribute is important for more than one class and the choice of this attribute can affect the accuracy of the classifiers, making it confuse patterns of all classes. This will be negatively counted for the attribute.

The second ranking procedure will be performed based on a RP parameter, in which a parameter is rewarded for this position in the ranking of the analyzed class (the magnitude of this reward depends on the position of the attribute in the ranking) and it is punished by its position in the ranking of the other classes (the magnitude of this punishment depends on the position of the attribute in the ranking). The RP parameter can be described as follows:

$$RP_i = \mathrm{Re}\, w_i - Pun_i \qquad \textbf{(1)}$$

Where:

$$\mathrm{Re}\, w_i = V_{ic} + \frac{NA}{NA + R_{ic}} \qquad \textbf{(2)}$$

$$Pun_i = \sum_{j=1}^{A, j \neq i} V_{jc} + \frac{NA}{NA + R_{jc}} \qquad \textbf{(3)}$$

Where: V_{ic} is the value of attribute i for class c; R_{ic} is the ranking of attribute i in class c, A is the number of attributes and *NA* is the total number of attributes used in the dataset.

Using eq. 1, the second ranking process is made and the first N attributes are picked.

3 Experimental Works

In this paper, an analysis of the presented method is performed, comparing it with the original filter method (only the first ranking). The ensemble systems using feature selection are also compared with ensemble systems with no feature selection. For the ensemble systems using feature selection (original and proposed), two criteria were used as basis (first ranking) for the ranking of the attributes, Spearman Correlation and Variance. The idea of using Spearman Correlation is to have a rank-based correlation measure, while the use of variance is because it is a criterion that does not need the class label vector to calculate its ranking [8].

Four databases are used for the experiments, which are: Gaussian3, Simulated6, St. Jude and Splice. The first two are synthetic databases that simulate microarray data in the gene expression analysis [12]. These two synthetic databases have attributes related to each of their classes. Thus, there is a set of 200 attributes in Gaussian3 that determine each class. This relation between attribute and class is exclusive. The 200 attributes that determine Class 2, for example, do not determine Class 0 or Class 1. Similarly, Simulated6 presents sets of 50 attributes exclusively related to each class. Simulated6 has 6 classes with 50 exclusively related attributes, totaling 300 attributes. As Simulated6 actually has 600 attributes, the remaining 300 are noise.

The last two are real datasets. Splice dataset contains a primate splice-junction gene sequences (DNA) with associated imperfect domain theory, obtained from UCI repository [1]. A total of 3190 Instances using 60 attributes were used. These attributes describe sequences of DNA used in the process of creation of proteins. Finally, St Jude Leukemia is a database obtained from gene expression data of leukemia cells [13]. Its data represents genes of samples of cells with Leukemia. It has 248 samples and each cell has 985 attributes. Each attribute is a number that measures the expression level of a gene found in a cell sample. In this paper, all attributes were normalized to the interval [0],[1].

Four classification methods are used as base classifiers for the ensemble systems, which are: k-NN (nearest neighbor), C4.5 (decision tree), NB (Naïve Bayesian Learning) and MLP (multi-layer Perceptron) neural network. In addition, three fusion-based combination methods will be used, which are: sum, majority vote (MV) and weighted sum (WS).

The ensemble size is defined by the number of classes of the used dataset. In this sense, ensembles of size 3 are used for Gaussian and splice and ensembles of size six are used for Simulated6 and St. Jude. Also, several different configurations are used, which varied from non-hybrid (homogeneous) to non-hybrid (heterogeneous) structures of ensembles. For the heterogeneous structures, ensembles with 2 (HYB 2), 3 (HYB 3) and 4 (HYB 4) different types of classifiers were taken into consideration. As there are several possibilities for each structure, this paper presents the average of the accuracy delivered by all possibilities of the corresponding hybrid structure.

The base components and the ensemble systems were built by using the 10-fold cross-validation methodology. In addition, to compare the impact of the proposed feature selection method, the accuracies of the ensembles when using the proposed feature selection method were compared with the original feature selection methods. To do this comparison, a statistical test was applied, which is called hypothesis test (t-test) [5], with a confidence level of 95% ($\alpha = 0.05$).

4 Results and Discussion

Table 1 shows accuracy and standard deviation of the ensemble systems when using no feature selection. As it can be observed from Table 1, the HYB 4 structure was not applied to the Gaussian and splice datasets. It is because these datasets have only three classes and it was used one classifier per class. In this sense, at most, three different classifiers (HYB 3) were used in this dataset.

Table 1: Accuracy (Acc) ± Standard Deviation (SD) of the ensemble systems using no feature selection

	Simulated			**St Jude Leukemia**		
	MV	SUM	W-SUM	MV	SUM	W-SUM
NH	86.70±10.00	87.50±11.10	85.80±11.40	96.70±3.20	96.70±3.10	96.50±3.50
H2	86.90±10.50	87.00±11.60	87.20±10.90	96.80±3.60	96.90±3.40	97.10±3.30
H3	90.50±10.60	90.40±10.70	89.60±10.40	98.30±2.80	98.40±2.70	98.30±2.80
H4	91.10±10.30	92.10±10.00	89.90±10.50	98.40±2.80	98.30±2.80	98.30±2.81
	Gaussian3			**Splice**		
NH	81.60±10.80	81.70±10.80	68.70±12.90	86.90±2.00	87.00±1.90	87.30±1.88
H2	84.20±8.70	79.20±9.10	82.50±10.20	89.30±1.90	91.80±1.90	92.30±1.82
H3	94.70±6.10	90.40±10.40	92.90±5.50	92.80±1.70	93.60±1.80	93.90±1.82

As it can be seen from Table 1, the accuracy of the ensemble systems, in most of the cases, increased when increased the number of different types of classifiers. In all datasets, the highest accuracies were obtained, for all combination methods, when using complete hybrid structures (HYB 4 for simulated and St. Jude and HYB 3 for Gaussian and splice).

4.1 The Synthetic Datasets

Table 2 shows accuracy and standard deviation of the ensemble systems when using both feature selection methods and, in both methods, a fixed number of attributes was allocated for each classifiers

(200 for Gaussian3 and 100 for Simulated6). The choice of the number of attributes was done in order to have all the attributes allocated for one classifier. For instance, in Simulated6 dataset, the choice of 100 attributes for 6 classifiers means a total of 600 attributes (the number of attributes of one pattern).

Table 2: Accuracy and Standard Deviation of the ensemble systems using feature selection methods for the synthetic datasets with a fixed number of attributes for each classifier

		Gaussian3				
		SO			VO	
	MV	SUM	W-SUM	MV	SUM	W-SUM
NH	86.66±8.90	83.33±14.20	87.08±7.90	89.58±7.30	90.00±7.20	90.41±7.10
H2	87.87±13.60	82.22±15.00	86.31±10.60	90.59±8.30	90.43±7.90	95.00±6.10
H3	82.89±13.50	82.53±13.90	87.68±12.00	93.13±8.00	91.62±7.80	94.85±7.50
		SP			VP	
NH	89.58±4.80	88.33±7.90	90.00±6.10	90.00±6.80	90.00±6.10	89.58±6.20
H2	90.20±9.70	92.43±8.70	94.58±6.90	90.97±8.10	91.11±8.30	94.86±5.60
H3	96.23±7.90	96.30±7.60	97.10±6.40	93.33±8.20	92.46±8.70	94.13±7.70
		Simulated6				
		SO			VO	
	MV	SUM	W-SUM	MV	SUM	W-SUM
NH	69.16±15.00	84.58±13.20	89.16±12.40	80.00±11.80	92.50±8.50	90.00±11.60
H2	69.95±16.60	85.18±13.10	86.52±12.30	83.98±10.30	93.24±8.40	89.76±10.50
H3	68.68±16.70	84.16±13.30	86.45±12.60	87.63±11.20	92.84±9.60	90.69±10.50
H4	69.58±15.50	84.58±12.90	83.33±12.90	84.09±11.80	91.38±10.80	88.33±11.70
		SP			VP	
NH	81.25±9.10	86.66±9.10	83.75±7.40	90.41±10.40	93.33±7.90	87.91±9.90
H2	81.15±9.40	86.71±9.60	83.88±9.00	91.20±10.10	93.42±8.40	90.32±10.00
H3	81.59±10.40	87.15±9.90	84.44±9.00	91.66±10.10	93.75±8.10	91.38±9.30
H4	80.27±8.70	85.90±8.40	82.50±7.40	92.70±9.60	94.16±7.70	90.83±9.60

As it can be seen from Table 2, all the ensemble systems using feature selection have a similar accuracy than the ensembles without feature selections (Table 1) for Gaussian3 dataset. For Simulated6 dataset, the ensemble systems provided accuracies which are lower than the ones obtained by the ensemble systems with no feature selection (Table 1), when using Spearman Correlation. This fact is because this synthetic dataset has noisy attributes. When these noisy attributes are related with the class label, these attributes will have high correlation with the specific class, being selected to be part of one classifier.

Still in Table 2, as it happened with systems with no feature selections, the hybridization of the ensemble systems has caused an increase in the accuracy of these systems. In addition, the use of variance has caused an improvement in the accuracy of the ensemble systems, when compared with Spearman correlation, for both feature selection methods. The ensemble systems using the proposed feature selection method have provided higher accuracies than the systems with the original method, for both datasets, for most of the cases. Finally, the use of weights was positive for the ensemble systems, since it caused an increase in the accuracy of these systems, when using the original and proposed feature selection method.

In order to evaluate whether the improvement in accuracy delivered by the proposed feature selection method is significant, the hypothesis tests (t-test) is performed. In this test, the accuracy of the proposed feature selection method is compared with the original method, using a confidence level of 95%. As a result of the hypothesis test, it was observed that the improvements reached by the proposed method were statistically significant in 6 out of 18 analyzed cases for Gaussian3 and 9 out 24 analyzed cases for Simulated6. In most of the cases, the statistically significant improvements were reached when using Spearman correlation (all 6 cases of Gaussian3 and 6 cases of Simulated6). It is important to emphasize that the small number of statistically significant improvements is due to the high scale of the standard deviation, which is caused by the small number of samples in these datasets.

Table 3 shows accuracy and standard deviation of the ensemble systems when using a threshold to define the number of attributes for the subsets of attributes. For Gaussian3, the thresholds were defined for each criterion and it was 0.75 for Spearman and 0.5 for Variance. The average number of attributes for Spearman was 72 (original) and 45 (proposed) and it was 83 (original) and 52 (proposed) for Variance. This means that the number of attributes in the subsets was lower than the ones used in Table 2, mainly for Spearman. For simulated6, the threshold used was 0.8 for Spearman and 0.6 for variance, reaching an average size of 51 (original) and 48.16 (proposed) for Spearman and 56.33 (original) and 39.5 (proposed) for Variance.

Table 3: Accuracy and Standard Deviation of the ensemble systems using feature selection methods for the synthetic datasets with a variable size of the subsets of attributes

			Gaussian3			
		SO			VO	
	MV	SUM	W-SUM	MV	SUM	W-SUM
NH	87.00±10.80	73.70±15.70	87.91±8.90	75.80±12.40	67.90±12.70	74.50±14.70
H2	83.10±13.40	71.80±16.50	88.70±9.60	73.40±13.20	68.10±15.10	73.00±13.70
H3	82.00±15.20	70.50±15.80	88.10±11.30	74.80±14.30	69.20±15.10	73.90±14.20
		SP			VP	
NH	87.00±10.00	77.50±14.70	84.10±11.20	79.10±14.20	78.70±13.40	74.10±14.70
H2	80.20±13.70	76.30±16.00	83.30±12.50	76.70±15.50	79.50±15.50	74.60±14.70
H3	80.20±15.60	76.20±16.30	85.70±12.90	75.60±15.30	79.20±15.00	73.30±14.20
			Simulated6			
		SO			VO	
	MV	SUM	W-SUM	MV	SUM	W-SUM
NH	32.00±13.90	56.60±16.50	72.90±12.60	28.70±15.60	46.20±17.90	50.00±14.80
H2	35.40±14.70	57.60±16.40	71.80±13.70	31.20±15.90	42.20±18.00	46.50±17.00
H3	37.20±14.40	57.60±16.30	70.60±13.80	32.80±16.70	42.40±17.80	47.20±17.10
H4	36.00±14.30	56.10±16.30	65.50±15.20	32.70±16.10	44.00±16.20	47.30±15.80
		SP			VP	
NH	51.60±11.60	73.70±12.10	72.90±10.70	36.20±18.10	47.50±19.30	49.50±16.00
H2	52.60±13.00	73.40±13.60	68.50±11.60	36.70±16.40	47.00±18.90	50.90±15.20
H3	53.60±13.10	73.20±13.00	69.70±12.20	37.00±16.60	46.80±18.00	50.60±15.80
H4	56.80±13.40	74.30±13.50	62.60±11.20	36.80±16.50	44.20±18.30	53.30±15.50

From Table 3, it can be seen that, in a general perspective, there is a decrease in the accuracy of the ensemble systems, when compared with the fixed numbers (Table 2). This is because these datasets are synthetic, in which all classes have the same level of difficulty. The use of subsets of variable size did not affect positively the accuracy of the ensemble systems. It is important to emphasize that the proposed method used fewer attributes than the original one and it provided higher accuracy. This shows that the proposed method is more efficient than the original one, for both datasets.

As a result of the hypothesis test, it was observed that the improvements reached by the ensemble systems using the proposed method were statistically significant 5 out of 18 analyzed cases for Gaussian3 and 10 out 24 cases for Simulated6.

4.2 Real Datasets

In this part of the experiments, two real datasets will be analyzed, which are: St Jude Leukemia and Splice. Table 4 shows accuracy and standard deviation of the ensemble systems when using both feature selection methods with a fixed number of attributes (160 attributes for St Jude and 20 attributes for Splice). Like the synthetic datasets, all the ensemble systems using feature selection have a similar accuracy than the ensembles without feature selections (Table 1). In addition, the use of weights was positive for the ensemble systems, since it caused an increase in the accuracy of these systems, when using the original and proposed feature selection method (apart from variance for St Jude dataset). Finally, the more heterogeneous an ensemble is, the higher accuracy it has.

In relation to the chosen criterion, for St Jude, both criteria (Spearmann and Variance) have a similar performance. However, an interesting point is the poor performance of the original method when using variance for Splice dataset. This might be because this dataset is composed of nominal attribute and a simple ranking of the attributes was not enough to detect important attributes. However, the reward/punishment process of the proposed smoothed out this problem.

The result of the statistical test shows that the ensemble systems with the proposed method had statistically improvements in most of the cases (17 out of 24), when compared with the ensemble systems with the original feature selection method and in all analyzed cases for Splice dataset.

Table 5 illustrates accuracy and standard deviation of the ensemble systems when using a threshold to define the number of attributes for the subsets of attributes. For St Jude dataset, the threshold used was 0.7 for Spearman and 0.65 for variance, reaching an average size of 207 (original) and 74.83 (proposed) for Spearman and 38.16 (original) and 37.87 (proposed) for Variance. For Splice dataset, the threshold used was 0.4 for Spearman and 0.6 for variance, reaching an average size of 43.33 (original) and 10.66 (proposed) for Spearman and 43 (original) and 18 (proposed) for Variance.

Table 4: Accuracy (Acc) and Standard Deviation (SD) of the ensemble systems using feature selection methods for real datasets with a fixed number of attributes for each classifier

	St Jude Leukemia					
	SO			VO		
	MV	SUM	W-SUM	MV	SUM	W-SUM
NH	95.30±4.10	95.55±3.30	94.25±3.10	93.53±3.80	93.63±3.90	93.23±3.80
H2	96.09±3.50	96.36±3.10	94.96±3.10	94.81±3.80	95.32±3.70	94.32±3.80
H3	96.25±3.30	96.36±3.20	95.17±3.00	95.58±3.90	96.00±3.70	94.70±3.70
H4	96.35±3.10	96.51±3.00	95.22±3.20	95.88±4.00	96.46±3.70	95.35±3.80
	SP			VP		
NH	96.25±3.50	96.35±3.50	96.55±3.50	95.15±3.70	95.66±3.50	95.36±3.80
H2	97.24±3.30	97.40±3.20	97.54±3.00	95.64±3.90	96.57±3.50	96.54±3.30
H3	97.66±2.90	97.68±2.90	97.80±2.80	97.52±3.40	97.68±3.20	97.46±3.10
H4	97.83±2.80	97.83±2.80	97.78±2.80	97.87±3.30	97.95±3.10	97.61±3.10
	Splice					
	SO			VO		
	MV	SUM	W-SUM	MV	SUM	W-SUM
NH	84.41±2.10	85.72±2.20	85.76±2.00	53.20±1.10	60.47±2.00	53.08±0.70
H2	84.98±2.10	85.38±2.20	86.47±2.00	53.51±1.10	60.63±2.00	53.43±0.60
H3	85.97±2.00	86.16±2.00	87.64±1.90	53.57±1.60	60.81±1.90	53.38±0.50
	SP			VP		

NH	89.38±1.80	89.51±1.70	89.26±1.50	79.80±2.30	83.35±1.90	81.66±2.40
H2	89.68±1.90	91.14±1.80	91.50±1.70	80.26±2.20	85.39±1.90	83.87±1.90
H3	90.48±1.90	91.89±1.60	92.16±1.60	81.63±2.00	88.40±1.70	87.69±1.60

Table 5: Accuracy and Standard Deviation of the ensemble systems using feature selection methods for St_jude dataset with with variable number of attributes for each classifier

St Jude Leukemia						
		SO			VO	
	MV	SUM	W-SUM	MV	SUM	W-SUM
NH	94.44±3.99	94.45±3.67	93.73±3.55	91.91±4.88	91.41±4.70	90.70±4.47
H2	95.15±3.62	95.31±3.45	93.85±3.42	93.07±5.00	93.33±4.40	92.25±3.78
H3	95.45±3.59	95.97±3.46	94.38±3.62	94.13±4.43	94.59±4.19	93.09±3.51
H4	95.95±3.57	96.37±3.46	94.51±3.76	94.51±4.20	95.12±3.71	93.46±3.22
		SP			VP	
NH	95.34±4.09	96.55±3.71	96.46±6.60	93.82±4.70	94.44±4.69	94.04±4.62
H2	96.25±3.89	96.72±3.54	96.72±3.47	95.23±4.27	95.56±3.80	95.27±3.66
H3	96.56±3.68	97.11±3.33	97.03±3.18	95.94±3.88	96.33±3.48	95.90±3.37
H4	96.46±3.85	97.16±3.21	96.93±3.07	96.53±3.37	96.72±3.26	96.43±3.10
Splice						
		SO			VO	
	MV	SUM	W-SUM	MV	SUM	W-SUM
NH	84.87±2.13	84.09±1.87	84.27±1.97	81.67±1.82	82.52±2.15	83.22±1.90
H 2	87.44±2.16	87.76±1.97	88.84±1.88	82.33±1.70	83.88±1.84	85.02±1.75
H 3	89.94±2.30	91.16±2.00	91.96±1.88	83.75±1.61	85.76±1.64	86.47±1.57
		SP			VP	
NH	88.64±2.76	89.78±2.67	89.99±2.49	80.46±2.79	84.11±2.52	82.91±2.57
H2	89.06±2.71	90.00±2.65	90.36±2.54	81.17±2.59	85.54±2.11	84.40±2.23
H3	89.41±2.73	90.37±2.64	90.73±2.60	82.24±2.45	87.39±1.74	86.87±1.90

Once again, the ensemble systems have similar accuracies than the ensemble systems with fixed number of attributes. Both criteria (Variance and Spearmann correlation) have a similar performance, with improvements when using the proposed method and when using weights. When applying the statistical test to compare the proposed method with the original one, the proposed method had statistically improvements in most of the cases for St Jude (17 out of 24) and Splice (13 out of 18).

5 Final Remarks

This paper presents an analysis on filter-based feature selection methods for ensembles. One of them uses a class-based method (original), while the other one uses a two-step ranking procedure (proposed). This proposed method selects important attributes to a corresponding class and the ensemble systems need to have, at least, one classifier to correctly recognize each class. Through this analysis, it could be observed that the use of the proposed method resulted in an improvement in the accuracy of the ensemble systems, when compared with the original method, for homogeneous and heterogeneous structures. These improvements were statistically significant (t-test) in most of the analyzed cases, mainly for the real datasets. The small number of statistically significant improvements in the synthetic datasets is due to the small number of samples (it causes instability of the systems). In addition, it provided accuracies which were higher than the ones of the ensemble systems with no feature selection in some cases.

This shows that the choice of important attributes are of fundamental importance for the performance of ensemble systems, and this was reached by the proposed feature selection method.

References

[1] C.L Blake and C.J Merz. “UCI Repository of machine learning databases”. Univ of California, Dept of Information and Computer Science. [http://www.ics.uci.edu/~mlearn/MLRepository.html]

[2] A Canuto, M Abreu, L Oliveira, J Xavier Junior and A Santos. Investigating the Influence of the Choice of the Ensemble Members in Accuracy and Diversity of Selection-based and Fusion-based Methods for Ensembles. **Pattern Recognition Letters**, 28(4), pp. 472-486, 2007.

[3] D.Caragea, A.Silvescu, and V.Honavar. “Decision tree induction from distributed, heterogeneous, autonomous data sources”. In **Conf on Int Systems Design and App (ISDA)**, 2003.

[4] J Czyz and M Sadeghi and J Kittler and L Vandendorpe. Decision fusion for face authentication, **Proc First Int Conf on Biometric Authentication**, 686-693, 2004.

[5] L I Kuncheva. Combining Pattern Classifiers. Methods and Algorithms, Wiley, 2004.

[6] P.J Modi and P.W Tae Kim. “Classification of Examples by multiple Agents with Private Features”. **Proc of IEEE/ACM Int Conf on Intelligent Agent Technology**, 223-229, 2005.

[7] J J Rodriguez, L I Kuncheva and C.J. Alonso, Rotation Forest: A new classifier ensemble method, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 28 (10), 2006, 1619-1630.

[8] L Santana; D Oliveira; A Canuto and M Souto. A Comparative Analysis of Feature Selection Methods for Ensembles with Different Combination Methods. **IJCNN**, pp 643-648, 2007.

[9] Tsymbal; M. Pechenizkiy and P.Cunningham. Diversity in search strategies for ensemble feature selection **•Information Fusion**, Vol 6(1), pp. 83-98, 2005.

[10] A Tsymbal, S Puuronen and D W Patterson. Ensemble feature selection with the simple Bayesian classification. **Inf. Fusion 4**, 87–100 (2003).

[11] K Tumer and N C Oza. Input decimated ensembles. **Pattern Anal. Appl.** 6, 65–77 (2003)

[12] S. Monti, P. Tamayo, J. Mesirov, and T. Golub. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. **Mach Learning**, vol.52:91–118, 2003.

[13] Yeoh, E.-J., et.all.. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. **Cancer Cell**, 1:2, 2002.