A Class-Based Feature Selection Method for Ensemble Systems

Karliane M O Vale, Filipe G Dias, Anne M P Canuto and Marcílio C P Souto
Informatics and Applied Mathematics Department, Federal University of RN Natal, RN - BRAZIL, 59072-970
karlianev@gmail.com, pimpao@gmail.com, anne@dimap.ufrn.br, marcilio.souto@gmail.com

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.

1. Introduction

An intense investigation of many approaches and methodologies for pattern recognition systems has been come into sight as a result of the increasing complexity and widening applicability of such systems. However, 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 [1,4]. 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 [2,8]. 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. In the former, a filter is used to define the importance of the attributes, while the latter uses a machine learning method, along with a selection procedure, in order to define the quality of the attribute set selected. The main problem of the wrapper approach is that it is classifier-dependent, in which the use of a different classifier leads to the choice of a different attribute set.

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 two different synthetic datasets.

2. Ensembles

The main goal of using ensembles is to improve the accuracy of a pattern recognition system [2,19]. There are two main issues in the design of an ensemble: the ensemble components and the combination. With respect to the first issue, the correct choice of the set of base classifiers is fundamental to the overall performance of an ensemble. The ideal situation would be a set of base classifiers with uncorrelated errors - they would be combined in such a way as to minimize the effect of these failures. In other words, the base classifiers should be diverse among themselves. Diversity can be reached in different ways: variations of the parameters of the base classifiers (e.g., initial weights and topology of a neural network model), use of different samples of the dataset as training sets, use of the different types of base classifiers, among others.

Once a set of classifiers has been created, the next step is to choose an effective way of combining their outputs. There are a great number of combination methods reported in the literature [2,15,19,21]. According to their functioning, there are two main strategies of combination

978-0-7695-3326-1/08 $25.00 © 2008 IEEE
DOI 10.1109/HIS.2008.109

methods: fusion-based and selection-based methods. In this paper, a fusion-based will be used.

3. Feature Selection in Ensembles

Feature selection methods try to reduce the dimensionality of the attributes of a dataset, spotting the best ones. The attribute subset selection 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, for instance, feature selection methods aim to reduce redundancy among the attributes of a pattern and to increase the diversity in such systems. In this paper, feature selection methods will be used in ensemble system. Hence, hereafter, the term feature selection will be used in the context of ensembles.

Recently, several authors have investigated the use of feature selection methods in ensembles, such as in [3,6,7,10,11,12,14,15,17,19,20]. 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 [12,16,17,18], 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 [5,3,6,9,15,19,20], 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.

In the filter approach, usually, the attributes are ranked based on a certain criterion and the top N attributes are picked. In other words, a general 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), we are distributing the difficulties of one class 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.

In using class-based ranking, each classifier of an ensemble is associated with one class, increasing the processing time of the recognition process. However, it is believed that this will be compensated by the decrease in the number of features for each class and by the increase in the accuracy of the classification method.

There are some works in the literature which use class-based feature selection, such as the favorite class method [16]. 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.

3.1. The proposed Method

Aiming to smooth out the aforementioned problems, a feature selection method is proposed. 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 based its decision on attributes which are important only for this class.

The second procedure ranking can be illustrated on figure 2. In this method, a new ranking strategy is used for 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 proposed process, 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 ranked in the analyzed class, this will be positively counted for the attribute. In addition, if an attribute is lowly ranked on the other classes, this will be positively counted for the attribute. Otherwise, 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:

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

Where:

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

And

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

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

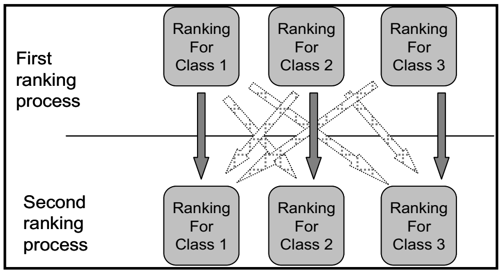

Figure 2: the process of a class-based feature selection method

Based on equation 1, the second ranking process is made and the first N attributes are picked.

Figure 2 shows the general functioning of the proposed two-step feature selection method, applying to a problem with three classes in which the solid arrows indicates the reward influence (eq .2) and dashed arrows indicates the punish influences (eq. 3).

4. Experimental Works

In this paper, an analysis of the proposed feature selection method is performed. In other to do this, the original filter method (only the first ranking) is also analyzed. 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, Pearson Correlation and Variance. The idea of using Pearson Correlation is similar to the used in the Favorite Class Method [16], while the use of variance is because it is a criterion that does not need the class label vector to calculate its ranking [18]

Two databases are used for the execution of the experiments: Gaussian3 and Simulated6. They are both synthetic databases that simulate microarray data and were created to test the ML algorithms in the gene expression analysis [13]. 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 [13].

Four types of 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. The choice of the aforementioned classifiers was due to the diversity in the classification criteria used by each method chosen.

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 ensembles of size six are used for Simulated6. For each system size, several different configurations are used, which varied from non-hybrid (homogeneous) to non-hybrid (heterogeneous) structures of ensembles. In order to make the choice of the structures of the ensembles more systematic, 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. For instance, the homogeneous ensembles represent the average values for NB, k-NN, MLP and DT ensembles.

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 ensemble systems using the original feature selection methods. To do this comparison, a statistical test was applied, which is called hypothesis test (t-test) [8]. It is a test which involves testing two learned hypotheses on identical test sets. In order to perform the test, a set of samples (classifier results) from both algorithms should be used to calculate error mean and standard deviation. Based on the information provided, along with the number of samples, the significance of the difference between the two sets of samples, based on a degree of freedom (α), is defined. In this paper, the confidence level is 95% (α = 0.05).

5. Results and Discussion

Table 1 shows the accuracy and standard deviation of the ensemble systems when using no feature. In other words, all components of the systems use all attributes (600). As it can be observed from Table 1, the HYB 4 structure was not applied to the Gaussian dataset. It is because the Gaussian dataset has 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.

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

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

	SIMULATED		**GAUSSIAN**	
	Acc(%)	SD(%)	Acc(%)	SD(%)
NH	91.11	7.63	81.25	13.57
HYB 2	91.67	7.34	81.34	9.95
HYB 3	92.62	7.37	95.41	6.63
HYB 4	93.89	7.38	---	---

5.1. The Gaussian3 Dataset

Table 2 shows the accuracy and standard deviation of the ensemble systems when using both feature selection methods, applied to the Gaussian3 dataset. In both methods, 100 attributes were allocated for each classifier. As it can be noticed from Table 2, the ensemble systems provided accuracies which are higher than the one obtained by the ensemble systems with no feature selection (Table 1). It is important to emphasize that, in this dataset, there are 200 exclusive attributes that determine each class. In this sense, in Table 2, only half of the attributes are used by the classifiers.

Still in Table 2, the ensemble systems using the proposed feature selection method have provided higher accuracies than the ensemble systems using the original method. In analyzing the improvement in accuracy reached by the proposed method, in relation to the used criterion, the Pearson Correlation has provided the highest improvement in accuracy (average of improvement of 6.3% for Pearson Correlation and of 4.9% for variance).

Table 2: Accuracy (Acc) and Standard Deviation (SD) of the ensemble systems using feature selection methods for the Gaussian3 dataset with 100 attributes each classifier

	Variance			
Ensemble structure	Original		Proposed	
	Acc(%)	SD	Acc(%)	SD
NH	88.75	9.06	92.08	4.64
HYB 2	85.19	6.00	92.36	2.26
HYB 3	87.64	2.46	91.80	2.00
	Pearson Correlation			
Ensemble structure	Original		Proposed	
	Acc(%)	SD	Acc(%)	SD
NH	78.33%	10.21	90.42	7.20
HYB 2	83.26%	5.79	88.40	4.49
HYB 3	86.32%	0.97	87.99	2.02

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 ensemble systems with the proposed method were statistically significant in all analyzed cases (p-values of 0.02; 2.86E-10; 1,24E-12;1,75E-8; 1,48E-5 and 0.002, for lines 4 to 6 and 10 to 12 in Table 2, respectively). These results show that the improvements reached by the proposed method are significant, based on a statistical test, for all cases of Table 2.

Table 3 shows the accuracy and standard deviation of the ensemble systems when using both feature selection methods and, in both methods, 200 attributes were allocated for each classifiers. In this Table, all exclusives attributes were used by the classifiers.

Table 3: Accuracy (Acc) and Standard Deviation (SD) of the ensemble systems using feature selection methods for the Gaussian3 dataset with 200 attributes each classifier

	Variance			
Ensemble structure	Original		Proposed	
	Acc(%)	SD(%)	Acc(%)	SD
NH	87.50	13.34	88.75	11.96
HYB 2	87.39	3.80	91.64	1.97
HYB 3	89.39	2.01	94.78	1.5
	Pearson Correlation			
Ensemble structure	Original		Proposed	
	Acc(%)	SD(%)	Acc(%)	SD
NH	89.58	11.68	89.17	10.77
HYB 2	92.01	3.91	94.31	1.43
HYB 3	97.33	2.60	97.50	1.29

As it can be seen from Table 3, the ensemble systems using the proposed feature selection method have provided higher accuracies than the systems with the original method. However, this improvement in the accuracy was lower than for systems using 100 attributes (average improvement of 2.16% for 200 attributes and 5.6% for 100 attributes). It is believed that the improvement was substantial for 100 attributes because of the proposed method indeed chooses the important attributes to a certain class. When using 100 attributes, the proposed method tends to choose the attributes that are important to determine a class. Conversely, when using 200 attributes, both methods tend to choose the same attributes, since that there are 200 exclusive attributes that determine each class.

In relation to the improvement reached by the proposed method for each used criterion, unlike Table 2, Variance has provided the highest improvement in accuracy of the proposed method (average of 3.63% for Variance and of 0.69% for Pearson Correlation). Therefore, it seems that when few attributes are to be chosen, Correlation behaves better than Variance. Conversely, when more attributes are to be chosen, Variance behaves better than Correlation.

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. 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 4 out of 6 analyzed cases (p-values of 0.33; 8.91E-9; 1.44E-22; ---; 0.0004 and 0.26, for lines 4 to 6 and 10 to 12 in Table 3, respectively).

5.2. The Simulated6 Dataset

Table 4 shows the accuracy and standard deviation of the ensemble systems when using both feature selection methods, applied for the Simulated6 dataset. For both selection methods, 50 attributes were allocated for each classifier.

Table 4: Accuracy (Acc) and Standard Deviation (SD) of the ensemble systems using feature selection methods for the Simulated6 dataset with 50 attributes each classifier

	Variance			
	Original		Proposed	
	Acc(%)	SD	Acc(%)	SD
NH	49.58	6.08	72.08	2.56
HYB 2	49.21	7.20	73.42	4.49
HYB 3	47.64	2.95	70.25	3.20
HYB 4	48.80	2.27	71.00	1.70
	Pearson Correlation			
	Original		Proposed	
	Acc(%)	SD	Acc(%)	SD
NH	66.92	8.48	67.50	9.01
HYB 2	63.56	5.23	65.42	5.41
HYB 3	62.92	8.14	63.26	8.09
HYB 4	64.47	8.08	64.86	6.40

As it can be noticed from Table 4, the ensemble systems provided accuracies which are lower than the ones obtained by the ensemble systems with no feature selection (Table 1). This fact is because of the number of attributes was reduced drastically, affecting the accuracies of the individual classifiers and, as a consequence, the accuracy of the ensemble systems. However, the accuracies of the proposed method were much higher than the original method, mainly for Variance. This shows that the use of the proposed method leads to the choice of important attributes, even when the number of attributes was not sufficient for an efficient recognition process.

In relation to the improvement reached by the proposed method for each used criterion, Variance has provided the highest improvement in accuracy of the proposed method (average of 22.9% for Variance and of 0.8% for Pearson Correlation). It is important to emphasize that the improvement reached by the proposed method for Variance was indeed high. As this dataset has noisy attributes, the original feature selection method picked many noisy attributes, while these were discarded for the proposed feature selection method. This was reflected in the accuracy of the ensemble systems using the proposed method.

In order to evaluate whether the significance of the improvement delivered by the proposed method, the hypothesis tests (t-test), using a confidence level of 95%, is performed. As a result of the hypothesis test, it was observed that the improvements reached by the proposed method were statistically significant in 5 out of 8 analyzed cases. Among these, all four cases for Variance and HYB 2 for Pearson Correlation (p-values of 6.47E-35; 7.2E-30; 9.61E-48; 5.45E-61; 0.38; 1.63E-6, 0.27 and 0.28, for lines 4 to 7 and 11 to 14 in Table 4, respectively).

It is important to emphasize that as the values shown in the tables of this paper represent average values of several system configurations, the number of samples used in the t-test is high. In HYB 4, for instance, the number of possible combinations is 24 and each combination used 10-fold cross-validation. This means that 240 groups were used in the t-test. For Pearson correlation, although the standard deviations were high, the difference in accuracies was statistically significant due to the large number of samples.

Table 5 shows the accuracy and standard deviation of the ensemble systems when using both feature selection methods and, in both methods, 100 attributes were allocated for each classifiers. From Table 5, it can be seen that, once more, the ensemble systems using the proposed feature selection method have obtained higher accuracies than the ensemble systems with the original method. In addition, mainly for Variance, the accuracies were higher than the ensembles with no feature selection.

Table 5: Accuracy (Acc) and Standard Deviation (SD) of the ensemble systems using feature selection methods for the Simulated6 dataset with 100 attributes each classifier

	Variance			
	Original		Proposed	
	Acc(%)	SD	Acc(%)	SD
NH	94.17	1.05	96.25	1.49
HYB 2	93.96	1.91	96.60	1.88
HYB 3	93.91	1.65	96.83	1.41
HYB 4	94.19	1.22	96.22	1.05
	Pearson Correlation			
	Original		Proposed	
	Acc(%)	SD	Acc(%)	SD
NH	83.33	1.92	90.00	1.95
HYB 2	84.68	1.52	89.91	1.02
HYB 3	85.69	1.24	90.35	1.97
HYB 4	84.03	1.96	91.53	1.32

The overall improvement reached by the proposed method was, on average, 4.58%, which is lower than for ensembles using 50 attributes (Table 4). This is because the similar level of accuracy obtained by the ensemble systems when using Variance. Unlike Table 4, in Table 5, the highest improvement in accuracy was reached by Pearson Correlation (6.01%), and, then, by Variance (3.14%). When comparing the results from Tables 4 (50 attributes) and 5 (100 attributes) for both criteria, for Variance, the improvement in accuracy obtained by the proposed method in Table 5 was lower than in Table. On the other hand, for Pearson Correlation, the improvement in accuracy was much higher than in Table 4.

The result of the hypothesis test, comparing the proposed method with the original one, has shown that the improvements reached by the ensemble systems with the proposed method were statistically significant in all analyzed cases (p-values of 1.52E-10; 1.1E-8; 4.85E-13; 5.23E-12; 1.09E-25; 6.6E-30, 6.61E-21 and 7.6E-35, for lines 4 to 7 and 11 to 14 in Table 5, respectively).

6. Final Remarks

This paper proposed a class-based feature selection method to be used in ensemble systems. 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. In order to analyze the feasibility of the proposed method, an empirical evaluation was performed. In this analysis, ensemble systems using the proposed method were compared with systems using a standard feature selection method, as well as systems with no feature selection methods. All three types of systems were applied to two synthetic datasets.

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. These improvements were statistically significant (t-test) in most of the analyzed cases (22 out of 28 cases). 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] A. Canuto.” Combining neural networks and fuzzy logic for applications in character recognition”. PhD thesis, Univ of Kent, 2001.

[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] X Fu and L. Wang: Data dimensionality reduction with application to simplifying RBF network structure and improving classification performance. IEEE Trans. System, Man, Cybern, Part B: Cybernetics 33, 399-409, 2003.

[6] Gerra-Salcedo, C.,Whitley, D.: Genetic approach to feature selection for ensemble creation. In: Proceedings of the Genetic and Evolutionary Computation Conference, pp. 236–243, (1999).

[7] T K Ho. The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. 20(8), 832–844 (1998).

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

[9] L I Kuncheva and L C Jain. Designing classifier fusion systems by genetic algorithms. IEEE Trans. Evol. Comput. 4(4), 327–336 (2000).

[10] B Liu, C Wan, and L.P. Wang, An efficient semi-unsupervised gene selection method via spectral biclustering, IEEE Trans. Nano-Bioscience, vol.5, pp.110-114, 2006.

[11] H Liu.; E Dougherty.; J Dy.; K Torkkola.; E Tuv; H Peng; C Ding; F Long; M Berens; L Parsons; Z Zhao; L Yu.; G Forman; Evolving feature selection, IEEE Intelligent Systems, vol.20, pp.64-76, 2005.

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

[13] 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. Machine Learning, vol.52:91–118, 2003.

[14] I-S Oh; J-S Lee and B-R Moon: Hybrid genetic algorithms for feature selection. IEEE Trans. Pattern Analysis and Machine Intelligence 26, 1424 – 1437, 2004.

[15] D.Opitz, Feature selection for ensembles, in: Proc. 16th Nat. Conf. on A. Intelligence, AAAI Press, pp. 379–384, 1999.

[16] N. Oza. and K Tumer. Input Decimation Ensembles: Decorrlation through Dimensionality Reduction. In Proc. Of 12nd Int. Workshop on Multiple Classifier Systems (MCS), 2001, LNCS 2096, pp.238-247.

[17] 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.

[18] L Santana; D Oliveira; A Canuto and M Souto. A Comparative Analysis of Feature Selection Methods for Ensembles with Different Combination Methods. In: International Joint Conference on Neural Networks (IJCNN), pp 643-648, 2007.

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

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

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