Combining One-Class Classifiers for Robust Novelty Detection in Gene Expression Data

Eduardo J. Spinosa and André C.P.L.F. de Carvalho

Universidade de São Paulo (USP), Instituto de Ciências Matemáticas e de Computação (ICMC),
Av. do Trabalhador São-Carlense, 400, São Carlos, 13560-970, Brasil
ejspin@icmc.usp.br**, andre@icmc.usp.br
http://www.icmc.usp.br/

Abstract. One-class classification techniques are able to, based only on examples of a normal profile, induce a classifier that is capable of identifying novel classes or profile changes. However, the performance of different novelty detection approaches may depend on the domain considered. This paper applies combined one-class classifiers to detect novelty in gene expression data. Results indicate that the robustness of the classification is increased with this combined approach.

1 Introduction

Supervised learning algorithms learn from labeled examples in a training set and later, on a test phase, attempt to classify new unseen examples based on the knowledge acquired in the training phase. In a traditional approach, the absence of good representative examples of a certain class in the training set leads to a poor performance of the classifier on that particular class. In an extreme situation, if a class does not have any examples at all, a traditional classifier will assign objects of that class to one of the known classes, even though it might not be an appropriate choice.

Therefore, the ability to detect a new class or sub-class is an important aspect for a machine learning system. Slight modifications in the data distribution might indicate, for instance, the appearance of a new class, or a profile modification in a class that has already been modeled. The capability to identify these changes is known as *Novelty Detection* (ND) [9], *Outlier Detection* or *One-Class Classification* [11] [12].

The term *One-Class* refers to the key characteristic of ND techniques, which is the fact that the training is carried out based only on examples from a single class that represents the normal profile. In other words, the algorithm learns to identify a novelty profile without having seen any examples of such a class. The power of novelty detection lies exactly on this aspect: in the training phase, no examples of any novel profile are presented. As a consequence, the performance

** Alternate e-mail: ejspin@yahoo.com

J.C. Setubal and S. Verjovski-Almeida (Eds.): BSB 2005, LNBI 3594, pp. 54–64, 2005.

of one-class classifiers cannot be directly compared to that of two-class classifiers, since the latter uses examples of both classes on the training phase.

Different approaches to ND have been proposed [9] and applied to a variety of tasks. In this paper, some of these approaches are combined to produce a single decision, as explained in Section 2. Section 3 presents and analyzes experiments involving gene expression data. Section 4 reviews the most important conclusions.

2 A Combined Approach to Novelty Detection

The problem of ND consists in the discovery of new profiles that were not present in the training samples. Thus, the classifier is induced based only on positive examples of a target class. All other examples are removed from the training set as these examples are considered outliers.

Of the various approaches to ND described in the literature, five of them have been chosen for this work: *Parzen Window* [10], *K-NN (K-Nearest Neighbor)* [6], *K-Means* [3], *SOM (Self-Organizing Map)* [8] and *PCA (Principal Components Analysis)* [3]. Each of these one-class classifiers uses one of three different strategies, according to the classification proposed by Tax [12]. Other classifications of ND techniques are available in the literature [9].

Parzen Window is a *density estimation* technique that, based on a data distribution scheme, defines a threshold to distinguish between normal and novel profiles. K-NN constructs hypersphere *boundaries* to involve data of the target class, therefore considering outliers any elements that fall outside these boundaries. K-Means, SOM and PCA are classified as *reconstruction* techniques. K-Means is a clustering algorithm that builds a boundary around prototype objects. SOM is based on a Neural Network architecture called Self-Organizing Map, in which prototypes are constrained to a lower-dimensional space in order to be later visualized. PCA performs a transformation of the original input attributes to a smaller number of uncorrelated, thus more meaningful, attributes.

Each of these techniques alone may perform better in a specific domain, and may also depend on a good parameter setting. Therefore, from a user's point of view, it might be hard to discover which approach is more likely to work best when experimenting with a variety of datasets.

The combined approach proposed in this work aims to increase classification robustness by taking into account the opinion of a set of one-class classifiers, instead of relying on a single approach, that might favor one class over the other.

Initially, all classifiers in the set are trained with a set containing only examples of the target class. For the same dataset, each class is considered the target class at one time, and examples of all other classes are labeled as outliers and used for testing purposes only. In the test phase, when target and outlier examples are present, the opinion of each classifier is taken and recorded. The final decision for each example (normal or novelty) is taken by the set of the classifiers. If the majority considers that the example belongs to the target class, then it is labeled *normal*, otherwise it is marked *novelty*.

Many statistic measures are taken throughout this process to ensure a good analysis of the results. A desired situation is one where the classifier is able to detect new profiles with high accuracy, but continues to classify normal examples with a good level of confidence. In other words, the aim is to minimize the false negative and false positive rates. However, this optimum point is not easily achieved, once some classifiers might be more restrictive than others in the definition of the normal profile.

Therefore, the major motivation for the combined approach is the belief that, when the opinions of more than one classifier are considered, the undesirable individual tendencies toward a specific class will be less important in the whole picture, since the final decision is taken by the majority. By doing so, it is expected that the optimum point described previously will be more easily achieved.

Previously, initial good results, not reported here, have been obtained with various standard datasets from the UCI Machine Learning Repository [4]. These results inspired a series of experiments carried out with gene expression data, presented in the following section.

3 Experiments

The main goal of the experiments described in this section is to compare the individual ND performance of each one-class classifier against the performance of the combined approach described previously. All classifiers used are available in DDtools, the Data Description Toolbox for Matlab [13], and this technique has been previously tested on standard datasets from the UCI Machine Learning Repository [4].

The experiments presented in this section have been conducted with the following gene expression datasets:

- breast - Classification of breast tumor samples based on the positive or negative status of the estrogen receptor (ER) [14]. The database is composed of 44 examples with 7129 attributes each.
- colon - Distinction between tumor and normal colon tissue samples based on gene expression [2]. The original database is composed of 62 examples and 2000 attributes.
- leukemia - Identification of two types of Leukemia (ALL and AML) from values of gene expression [7]. The original database contains 72 examples and 7129 attributes.
- lymphoma - Distinction between germinal center and activated diffuse large B-cell lymphoma based on gene expression profiling [1]. The original database is composed of 47 examples and 4026 attributes.

Throughout the analysis, classes are referred with numbers instead of labels, according to the association shown in Table 1.

Table 1. Classes numbers

Base	Class 1	Class 2
breast	ER-	ER+
colon	Tumor	Normal
leukemia	ALL	AML
lymphoma	Germinal Center	Activated

3.1 Methodology

Stratified 10-fold cross-validation has been used in all experiments to ensure that results represent the average behavior, not a specially successful or unsuccessful case. The same folds were used in all experiments to allow replicability.

According to the number of incorrect predictions, two error rates were calculated: the normal error rate, that considers examples of the normal profile incorrectly classified as outliers, and the novelty error rate, which indicates the percentage of outliers that have been incorrectly considered members of the normal profile. The results obtained are presented and discussed as follows.

3.2 Analysis of the Results

Initially, experiments were performed with 2 original datasets and a set of 5 classifiers: Parzen Window, K-NN, K-Means, SOM and PCA. Table 2 presents these results. In each cell, the mean error rate of the 10 folds tested is followed by the standard deviation. These statistics are available for each classifier alone, and for the combined approach. As previously explained, for all datasets, each class has been considered the normal profile at a time. For example, when class 1 is the normal profile, examples of class 2 are not present in the training phase. In fact, class 2 represents the novelty that the classifier is supposed to identify in the testing phase. Then, the same procedure is carried out considering class 2 as the normal profile.

The first aspect to notice in the results is the poor performance of all classifiers. In general, they consider almost all test examples as being either normal (very high novelty error rate) or novelty (very high normal error rate). For instance, when the Parzen Window technique obtains a novelty error rate equal to 1.00 and a normal error of 0.00, it means that it is classifying all test samples as normal, which is completely inadequate. The opposite is seen with the SOM technique in the *lymphoma* dataset, with normal error rates as high as 0.87. Neither one nor the other behavior is useful, and each shows that the classifier has not been able to estimate the distribution of the data. This situation, i.e. where all data are either considered normal or novelty, can be caused, among other things, by a classifier that is either inadequate for that particular data domain or badly configured. However, in this specific situation, a very high number of attributes (2000 in the *colon* dataset and 4026 in *lymphoma*) could also be the complicating factor.

Table 2. Results with 2 original datasets and a set of 5 classifiers. In each cell, the mean error rate is followed by the standard deviation

Base: colon		Normal Error		Novelty Error	
Normal Class: 1	parzen	0.00	0.00	1.00	0.00
	knn	0.10	0.17	0.95	0.16
	kmeans	0.15	0.17	1.00	0.00
	som	0.15	0.17	1.00	0.00
	pca	0.18	0.26	0.95	0.16
	Combined	**0.13**	**0.18**	**1.00**	**0.00**
Normal Class: 2	parzen	0.00	0.00	1.00	0.00
	knn	0.10	0.21	0.63	0.27
	kmeans	0.13	0.22	0.65	0.27
	som	0.13	0.22	0.63	0.27
	pca	0.20	0.26	0.50	0.26
	Combined	**0.10**	**0.21**	**0.63**	**0.27**
Base: lymphoma		Normal Error		Novelty Error	
Normal Class: 1	parzen	0.00	0.00	1.00	0.00
	knn	0.10	0.21	0.67	0.29
	kmeans	0.18	0.24	0.75	0.27
	som	0.87	0.22	0.00	0.00
	pca	0.27	0.24	0.50	0.34
	Combined	**0.13**	**0.22**	**0.67**	**0.29**
Normal Class: 2	parzen	0.00	0.00	1.00	0.00
	knn	0.08	0.18	0.80	0.26
	kmeans	0.17	0.22	0.80	0.26
	som	0.85	0.24	0.03	0.11
	pca	0.25	0.36	0.50	0.42
	Combined	**0.17**	**0.22**	**0.80**	**0.26**

To investigate that, a preprocessing phase has been added. In that phase, the number of attributes has been reduced to a calculated optimum amount, different for each dataset, based on the same technique used in [7]. This procedure aimed to minimize the error rates of ND. As a positive side effect, it also largely reduced the computational cost.

Table 3 shows the results after attribute reduction, with the same set of classifiers seen previously in Table 2. The *colon* dataset has been reduced to *colon16*, with 16 attributes, and the *lymphoma* dataset has been reduced to *lymphoma32*, with 32 attributes. With a few exceptions, the majority of the error rates decreased, which confirms that the high dimensionality of the original dataset did not allow the induction of reliable ND classifiers. This table also includes results obtained from 2 other reduced datasets, *breast128* and *leukemia64*, with 128 and 64 input attributes respectively.

In this second round of experiments, K-NN and K-Means achieved low error rates, except for the novelty class of *colon16*, which is known to be a difficult dataset. The PCA based classifier obtained good results on all datasets, even for the *colon16* dataset, when the normal examples belong to class number 1. Unfortunately, in that case, most of the classifiers were not as successful. Parzen Window was the worse of all classifiers, displaying the same behavior seen previously in Table 2. However, this negative influence did not have a strong impact on the overall performance of the combined approach.

Table 3. Results with the 4 reduced datasets and 5 classifiers

Base: breast128		Normal Error		Novelty Error	
Normal Class: 1	parzen	0.00	0.00	1.00	0.00
	knn	0.15	0.34	0.03	0.11
	kmeans	0.18	0.34	0.00	0.00
	som	0.18	0.34	0.00	0.00
	pca	0.18	0.34	0.00	0.00
	Combined	**0.15**	**0.34**	**0.00**	**0.00**
Normal Class: 2	parzen	0.00	0.00	1.00	0.00
	knn	0.15	0.24	0.13	0.22
	kmeans	0.15	0.24	0.00	0.00
	som	0.15	0.24	0.00	0.00
	pca	0.18	0.24	0.00	0.00
	Combined	**0.15**	**0.24**	**0.00**	**0.00**
Base: colon16		Normal Error		Novelty Error	
Normal Class: 1	parzen	1.00	0.00	0.00	0.00
	knn	0.10	0.17	0.50	0.34
	kmeans	0.13	0.13	0.53	0.34
	som	0.15	0.13	0.48	0.30
	pca	0.18	0.26	0.27	0.24
	Combined	**0.18**	**0.17**	**0.40**	**0.33**
Normal Class: 2	parzen	1.00	0.00	0.00	0.00
	knn	0.10	0.21	0.85	0.17
	kmeans	0.15	0.24	0.68	0.24
	som	0.15	0.24	0.63	0.27
	pca	0.18	0.24	0.63	0.18
	Combined	**0.15**	**0.24**	**0.65**	**0.21**
Base: leukemia64		Normal Error		Novelty Error	
Normal Class: 1	parzen	1.00	0.00	0.00	0.00
	knn	0.12	0.17	0.07	0.14
	kmeans	0.09	0.11	0.03	0.11
	som	0.11	0.11	0.03	0.11
	pca	0.14	0.13	0.03	0.11
	Combined	**0.13**	**0.11**	**0.03**	**0.11**
Normal Class: 2	parzen	1.00	0.00	0.00	0.00
	knn	0.07	0.21	0.47	0.24
	kmeans	0.08	0.18	0.14	0.19
	som	0.13	0.22	0.17	0.22
	pca	0.28	0.35	0.09	0.15
	Combined	**0.17**	**0.22**	**0.14**	**0.19**
Base: lymphoma32		Normal Error		Novelty Error	
Normal Class: 1	parzen	1.00	0.00	0.00	0.00
	knn	0.08	0.18	0.12	0.19
	kmeans	0.13	0.32	0.05	0.16
	som	1.00	0.00	0.00	0.00
	pca	0.22	0.24	0.28	0.26
	Combined	**0.30**	**0.32**	**0.05**	**0.16**
Normal Class: 2	parzen	1.00	0.00	0.00	0.00
	knn	0.15	0.34	0.23	0.34
	kmeans	0.15	0.24	0.00	0.00
	som	0.97	0.11	0.00	0.00
	pca	0.18	0.24	0.00	0.00
	Combined	**0.23**	**0.34**	**0.00**	**0.00**

The SOM technique only showed difficulty in the *lymphoma32* dataset. However, even with two classifiers providing totally misleading results, the effect on the performance of the combined approach in the *lymphoma32* dataset was little. This shows superior robustness of the combined approach against the choice of a single classification strategy.

Table 4. Results with a set of 3 classifiers, one from each strategy

Base: breast128		Normal Error		Novelty Error	
Normal Class: 1	parzen	0.00	0.00	1.00	0.00
	knn	0.15	0.34	0.03	0.11
	kmeans	0.18	0.34	0.00	0.00
	Combined	**0.15**	**0.34**	**0.03**	**0.11**
Normal Class: 2	parzen	0.00	0.00	1.00	0.00
	knn	0.15	0.24	0.13	0.22
	kmeans	0.15	0.24	0.00	0.00
	Combined	**0.15**	**0.24**	**0.13**	**0.22**
Base: colon16		Normal Error		Novelty Error	
Normal Class: 1	parzen	1.00	0.00	0.00	0.00
	knn	0.10	0.17	0.50	0.34
	kmeans	0.13	0.13	0.53	0.34
	Combined	**0.20**	**0.20**	**0.40**	**0.33**
Normal Class: 2	parzen	1.00	0.00	0.00	0.00
	knn	0.10	0.21	0.85	0.17
	kmeans	0.15	0.24	0.68	0.24
	Combined	**0.15**	**0.24**	**0.68**	**0.24**
Base: leukemia64		Normal Error		Novelty Error	
Normal Class: 1	parzen	1.00	0.00	0.00	0.00
	knn	0.12	0.17	0.07	0.14
	kmeans	0.09	0.11	0.03	0.11
	Combined	**0.19**	**0.15**	**0.03**	**0.11**
Normal Class: 2	parzen	1.00	0.00	0.00	0.00
	knn	0.07	0.21	0.47	0.24
	kmeans	0.08	0.18	0.14	0.19
	Combined	**0.15**	**0.25**	**0.14**	**0.19**
Base: lymphoma32		Normal Error		Novelty Error	
Normal Class: 1	parzen	1.00	0.00	0.00	0.00
	knn	0.08	0.18	0.12	0.19
	kmeans	0.13	0.32	0.05	0.16
	Combined	**0.22**	**0.33**	**0.05**	**0.16**
Normal Class: 2	parzen	1.00	0.00	0.00	0.00
	knn	0.15	0.34	0.23	0.34
	kmeans	0.15	0.24	0.00	0.00
	Combined	**0.20**	**0.35**	**0.00**	**0.00**

To assess the impact of the number of classifiers in the set on the final results, experiments were also performed with a set of 3 classifiers, one representing each of the ND strategies (density estimation, boundary and reconstruction).

The results, displayed in Table 4, show a small increase in the error rates in the *breast128* and *colon16* datasets. On the other hand, for the *lymphoma32* dataset there was a small reduction in the error rates. For the *leukemia64* dataset the results were similar. Considering that the number of classifiers was reduced from 5 to 3, and that the relative influence of the Parzen Window on the overall result was increased, the performance of the combined approach has not been seriously affected.

A different set, in which Parzen Window has been replaced by PCA, has also been tested and the results are presented in Table 5. In most cases, the performance has been improved. However, although Parzen Window, which apparently has not shown any contribution to the combined result, has been replaced in the combination by PCA, a technique which has shown superior performance, the impact on the combined result was not as high as could be expected. In fact, this stability indicates the potential of the combined approach. With the combination, extremes

Table 5. Results with another set of 3 classifiers

Base: breast128		Normal Error		Novelty Error	
Normal Class: 1	knn	0.15	0.34	0.03	0.11
	kmeans	0.18	0.34	0.00	0.00
	pca	0.18	0.34	0.00	0.00
	Combined	**0.15**	**0.34**	**0.00**	**0.00**
Normal Class: 2	knn	0.15	0.24	0.13	0.22
	kmeans	0.15	0.24	0.00	0.00
	pca	0.18	0.24	0.00	0.00
	Combined	**0.15**	**0.24**	**0.00**	**0.00**
Base: colon16		Normal Error		Novelty Error	
Normal Class: 1	knn	0.10	0.17	0.50	0.34
	kmeans	0.13	0.13	0.53	0.34
	pca	0.18	0.26	0.27	0.24
	Combined	**0.15**	**0.17**	**0.40**	**0.33**
Normal Class: 2	knn	0.10	0.21	0.85	0.17
	kmeans	0.15	0.24	0.68	0.24
	pca	0.18	0.24	0.63	0.18
	Combined	**0.10**	**0.21**	**0.73**	**0.18**
Base: leukemia64		Normal Error		Novelty Error	
Normal Class: 1	knn	0.12	0.17	0.07	0.14
	kmeans	0.09	0.11	0.03	0.11
	pca	0.14	0.13	0.03	0.11
	Combined	**0.06**	**0.10**	**0.03**	**0.11**
Normal Class: 2	knn	0.07	0.21	0.47	0.24
	kmeans	0.08	0.18	0.14	0.19
	pca	0.28	0.35	0.09	0.15
	Combined	**0.07**	**0.14**	**0.14**	**0.19**
Base: lymphoma32		Normal Error		Novelty Error	
Normal Class: 1	knn	0.08	0.18	0.12	0.19
	kmeans	0.13	0.32	0.05	0.16
	pca	0.22	0.24	0.28	0.26
	Combined	**0.13**	**0.22**	**0.12**	**0.19**
Normal Class: 2	knn	0.15	0.34	0.23	0.34
	kmeans	0.15	0.24	0.00	0.00
	pca	0.18	0.24	0.00	0.00
	Combined	**0.15**	**0.24**	**0.00**	**0.00**

can be avoided and, consequently, the robustness of the system as a whole can be improved. Although the best possible results may not be achieved, unstable situations in which a classification technique favors one specific profile over the other can be avoided, i.e. normal over novelty or novelty over normal. As mentioned previously, this is an important issue when dealing with one-class classification, since the challenge is to identify new profiles with a high level of confidence while maintaining a good performance on the normal profile.

Finally, to provide a better visualization of the decisions taken throughout the process, individual errors made by each classifier on each example of the test set have been recorded for each fold and later reassembled. Figure 1 displays those errors in a graphical format, where white squares represent examples correctly classified and black squares mark errors. Examples are placed along the horizontal axis and classifiers vertically.

It is easily noticed that the larger number of errors is concentrated in the dataset *colon16* when the second class represents the normal profile. In this situation, all classifiers except Parzen Window make similar mistakes, which can also be confirmed by the error rates displayed in Table 3.

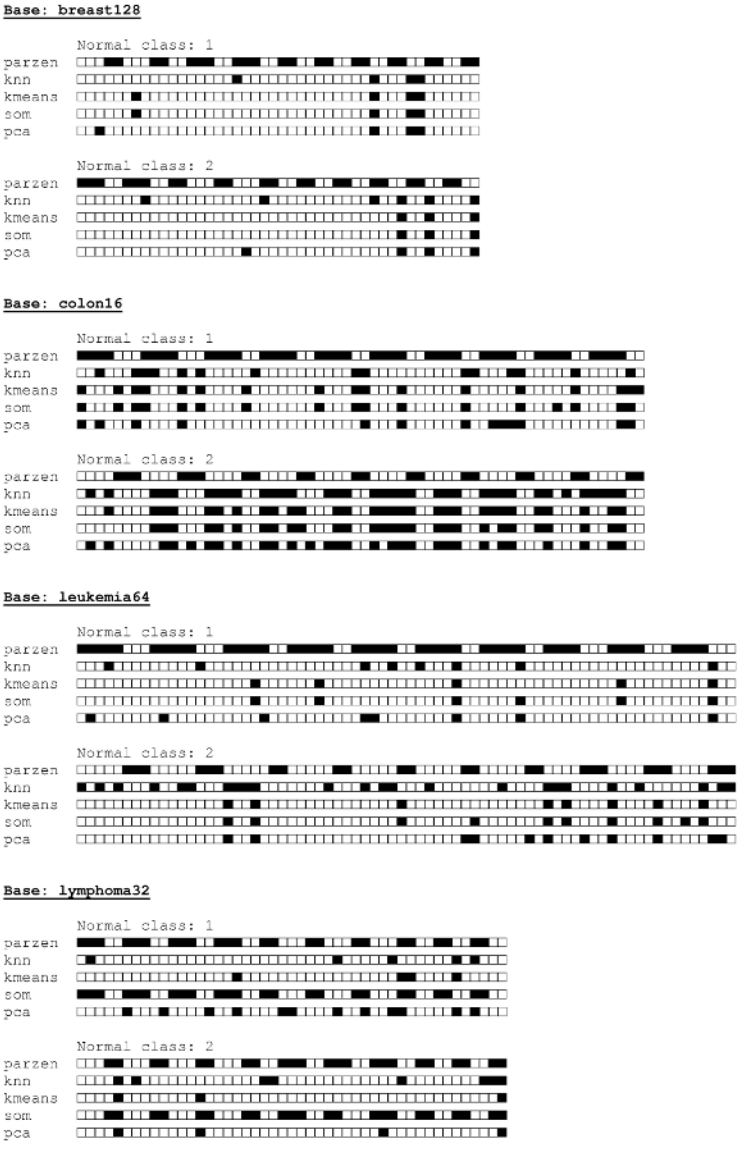

Fig. 1. Individual errors (black squares) of each classifier (vertical axis) on each example (horizontal axis)

Through these graphs it is also clearer to see that a classification strategy that shows good results in one dataset might not be successful in another, even considering datasets of the same domain (gene expression). For example, the horizontal lines which represent the performance of the SOM technique display

a very different amount of classification mistakes, depending on the dataset and on the normal class considered. This picture reinforces that the combination of classifiers leads to more robust results, since the decision is always taken by the majority of them.

An example of a desired situation is shown in the first plot of *leukemia64* and in both plots of *breast128*, where low error rates have been achieved. The vertical alignment of the errors indicate that all classifiers are having similar difficulties. These are problematic examples, which can be further investigated with a different technique, or isolated to be analyzed by a specialist.

However, if a larger number of scattered errors is present, the final performance of the combined approach might still be good. This is due to the fact that each classifier is filling-in other classifiers faults, which exemplifies the importance to combine classifiers built with various techniques, since the diversity of classifiers in the set may determine the robustness of the system as a whole.

4 Conclusion

One-class classification techniques are able to, based only on examples of a normal profile, induce a classifier that is capable of detecting novelty.

This paper has shown the use of a simple strategy which combines the opinions of a set of one-class classifiers for the task of ND in gene expression data. The results obtained suggest that the use of such a combined approach improves the robustness of the overall decision. By considering the opinion of the majority of a set of classifiers instead of just one, this technique avoids individual tendencies that certain approaches might present in some datasets or domains.

The improvement achieved so far inspire further investigations. As analyzed, the diversity of classifiers in the decision set seems to be an important aspect in the final performance of the combined approach. Another possible way of improving the results might be the addition of a selection phase, after which only the ND approaches that better fit the problem at hand would be considered. An assessment of the impact of both technical and biological noise on the differential performance of the classifiers has been suggested, and also inspires further experimentation.

Still, other combinations of one-class classifiers are yet to be explored in bioinformatics, following previous initiatives [12], as the authors continue to explore ND techniques for the identification of novel classes and profile changes.

Acknowledgement

The authors wish to thank CNPq for the financial support and Bruno Feres de Souza for the preprocessing of datasets [5].

References

1. A. A. Alizadeh, M. B. Eisen, R. E. Davisintegral, C. Maintegral, I. S. Lossos, A. Rosenwaldintegral, J. C. Boldrick, H. Sabetintegral, T. Tranintegral, X. Yuintegral, J. I. Powell, L. Yang, G. E. Marti, T. Moore, J. H. Jr, L. Lu, D. B. Lewis, R. Tibshirani, G. Sherlock, W. C. Chan, T. C. Greiner, D. D. Weisenburger, J. O. Armitage, R. Warnke, R. Levy, W. Wilson, M. R. Grever, J. C. Byrd, D. Botstein, P. O. Brown, and L. M. Staudt. Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling. *Nature*, 403:503–511, 2000.
2. U. Alon, N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack, and A. J. Levine. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of National Academy of Sciences USA*, 96:6745–6750, 1999.
3. C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
4. C. L. Blake and C. J. Merz. UCI repository of machine learning databases, 1998.
5. B. F. de Souza. Seleção de características para svms aplicadas a dados de expressão gênica. Master thesis, Universidade de São Paulo (USP), Instituto de Ciências Matemáticas e de Computação (ICMC), 2005.
6. R. O. Duda and P. E. Hart. *Pattern Classification*. Wiley Interscience, 2nd edition, 2001.
7. T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander. Molecular classification of cancer: Class discovery and class prediction by gene expression. *Science*, 286:531–537, 1999.
8. T. Kohonen. *Self-Organizing Maps*. Springer, 3rd edition, 2001.
9. S. Marsland. Novelty detection in learning systems. *Neural Computing Surveys*, 3:157–195, 2003.
10. E. Parzen. On the estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.
11. B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, 2001.
12. D. M. J. Tax. *One-class classifiers*. PhD thesis, Delf University of Technology, Faculty of Information Technology and Systems, 2001.
13. D. M. J. Tax. DDtools, the data description toolbox for matlab. `http://www-ict.ewi.tudelft.nl/~davidt/dd_tools.html`, March 2005. version 1.1.2.
14. M. West, C. Blanchette, H. Dressman, E. Huang, S. Ishida, R. Spang, H. Zuzan, J. A. O. Jr., J. R. Marks, and J. R. Nevins. Predicting the clinical status of human breast cancer by using gene expression profiles. *Proceedings of National Academy of Sciences USA*, 98(20).