

Improving the Classification Accuracy of RBF and MLP Neural Networks Trained with Imbalanced Samples

R. Alejo^{1,2}, V. Garcia^{1,2}, J.M. Sotoca¹, R.A. Mollineda¹, and J.S. Sánchez¹

¹ Dept. Llenguatges i Sistemes Informàtics, Universitat Jaume I
Av. Sos Baynat s/n, 12071 Castelló de la Plana (Spain)

² Lab. de Reconocimiento de Patronas, Instituto Tecnológico de Toluca
Av. Tecnológico S/N, 52140, Metepec, (México)

Abstract. In practice, numerous applications exist where the data are imbalanced. It supposes a damage in the performance of the classifier. In this paper, an appropriate metric for imbalanced data is applied as a filtering technique in the context of Nearest Neighbor rule, to improve the classification accuracy in RBF and MLP neural networks. We diminish atypical or noisy patterns of the majority-class keeping all samples of the minority-class. Several experiments with these preprocessing techniques are performed in the context of RBF and MLP neural networks.

1 Introduction

An imbalanced training sample (TS), can be defined as a sample in which the number of patterns of a (minority) class is much smaller than those in the other classes. This scenario strongly affects many types of classifiers, in particular, the artificial neural networks trained with procedures of iterative adjustment [7]. Following a common practice [10], we consider a simplified version with only two classes (majority-class and minority-class). Several proposals reduce the influence of class imbalance in training. In general, three categories [4] can be identified: *over-sampling* replicates examples in the minority-class, *under-sampling* eliminates examples from the majority-class and *biasing* the discrimination process to compensate the class imbalance.

A very established strategy to reduce the majority-class and compensate the imbalance has been exhaustively studied in the Nearest Neighbor (NN) rule [3]. The process removes redundant, atypical, and noisy patterns producing less confusing in the TS and improving the NN classification accuracy. These techniques do not produce a considerable reduction of the majority-class size and in general, do not solve the imbalanced distribution between classes. In this sense, a further preprocessing of the TS addressed to remove redundant patterns has been proposed [2], conducting to a significant reduction of the majority-class size. In the context of NN classifiers, these majority-class reduction techniques have led to better classifier performances than the plain use over all classes.

A number of papers have studied the imbalance problem in neural network frameworks. Three main general approaches have been proposed. One of them

is aimed to reduce the imbalance effects [11]. So, a first method multiplies the number of samples of the minority class, while a second one makes a two-step dynamic training of the neural networks considering first samples of the minority class and then, gradually adds samples of the majority class. A second approach focuses in adapting the backpropagation algorithm for imbalance situations [1], speeding up the convergence of the learning process for two-class problems. Finally, a third approach is directed to find appropriate parameters of the neural network models to improve their performance [5,7].

This work analyzes the behavior of two neural network models at classifying imbalance problems: Radial Basis Function (RBF) and Multilayer Perceptron (MLP). Through a resampling strategy, we remove examples of the majority-class from the overlap region, producing a local balance of the two classes. For this, the only requirement is that all samples of the minority class must be kept in the TS. As downsizing of the majority class can throw away significant information, an editing scheme is applied. Note that a global balance in the class sizes is not achieved.

The rest of the paper is organized as follow. In Sect. 2, the main characteristics of RBF and MLP neural networks are briefly described. The methodology of resampling used in the TS is presented in Sect. 3. Section 4 discusses the classification performance with a synthetic data set in different situations. as well as when applied to real databases. Finally, the main conclusions and possible future research are outlined in Sect. 5.

2 RBF and MLP Neural Networks

In the last years, the MLP (Backpropagation) has become popular in many tasks of machine learning, pattern recognition and data mining. In particular, it has been applied in image interpretation of remote sensing, whereas RBF neuronal networks have been used widely in applications with function approximation, interpolation with noise and tasks of classification. However, the knowledge about these models seems insufficient, what is translated into poor capacity of generalization in different applications.

By the simplicity of their architecture and training method, RBF networks are an attractive alternative respect to MLP. At the moment, RBF and MLP are two models of neural networks with great popularity in pattern recognition tasks. They are a clear example of feedforward neural networks with nonlinear layers. These techniques can be used as universal approximation [9], and are trained in a similar way with descendent gradient method [6]. In this kind of problems there always exists a RBF capable to make equal the MLP accuracy or viceversa. However, both networks have important differences [8].

1. The RBF has a single hidden layer, and the MLP can have one or more.
2. Generally, in the MLP all hidden and output nodes have the same neural model. On the other hand, in the RBF the hidden and output nodes have different neural models.

3. The parameters of the activation function for each hidden node in RBF are calculated with the Euclidean distance between the input vector and the prototype vector. Moreover, the parameters of the activation function of each hidden unit in a MLP is calculated from the sum of the product between the input vector and the synaptic weights of each unit.
4. The MLP generates a global approximation for the nonlinear association of input-output. Furthermore, the RBF networks generate a local approximation for the nonlinear association of input-output.

3 Methodology

The aim is increasing the classification accuracy of RBF and MLP neural networks in class imbalance problems, by improving the quality of the TS. The MLP neural network here used is a simplified version of the one proposed in [12], with a hidden layer and three hidden neurons. In our experiments, the learning rate and momentum are set to 0.9 and 0.7, respectively. The shutdown criterion settled down with an error smaller than 0.01 or maximum of 5000 training epochs. The RBF neural network was trained with the Backpropagation algorithm [13], four hidden neurons, and a learning rate equal to 0.9. The shutdown criterion settled down with an error smaller than 0.01 or maximum of 5000 training iterations.

For internally biasing the balance between classes in the overlap region, we have used the weighted distance used in [3] in resampling tasks. This weighted distance is defined as

$$d_w(y, x_0) = (n_i/n)^{1/m} d_E(y, x_0) \quad (1)$$

where $d_E(\cdot)$ is the Euclidean metric between the new sample to classify y , x_0 is a sample of the TS that belongs to the class i , n_i is the number of patterns of the class i , n is the total number of patterns in the TS, and m represents the dimensionality of the feature space.

In the preprocessing of the TS, we have utilized an editing technique based upon distances. These methods are an easy and simple strategy to eliminate noisy or atypical patterns from the TS. In this work, the classical Wilson's proposal (WE) is utilized for this purpose, finding the k nearest neighbors (with $k = 3$) of each instance from the TS. Three practical scenarios are proposed in the use of WE: editing with Euclidean distance in both classes, editing of the majority-class with Euclidean distance, and editing of the majority-class with the weighted distance shown above (Eq. 1).

With respect to the performance of the classifier, the average geometric mean is here used as the evaluation criterion. This measure is more appropriate in environments with imbalanced class distributions. The geometric mean is defined as follows:

$$g = \sqrt{a^+ \cdot a^-} \quad (2)$$

where (a^+) is the accuracy on the minority-class and (a^-) denotes the accuracy on the majority-class. This measure tries to maximize the accuracy on each of

the two classes while keeping these accuracies balanced. For instance, higher (a^+) with lower (a^-) results in a poor value of g .

4 Experiments and Discussion

To evaluate the effect of class overlapping on neural network classifiers with imbalanced classes, we have generated six synthetic databases with different levels of overlapping. Each domain is described by two classes with two dimension and uniform distributions: A0 = 0% (that is, non-overlapped), A20 = 20%, A40 = 40%, A60 = 60%, A80 = 80%, and A100 = 100%(that is, absolutely overlapped). Each artificial database consists of 500 patterns for training and 500 patterns for test (400 patterns for the majority-class and 100 patterns for the minority-class). The nature of the data is illustrated in Fig. 1.

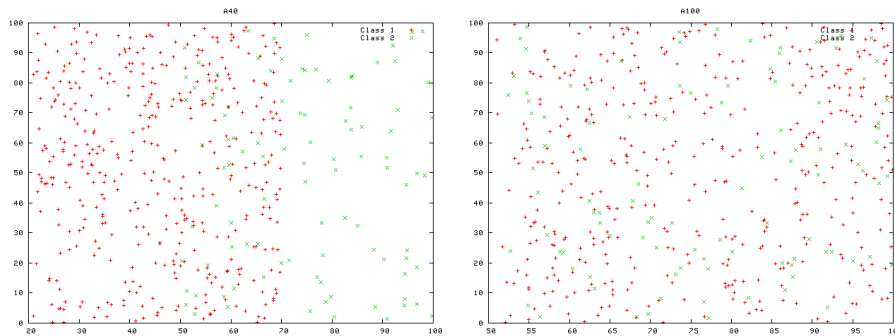


Fig. 1. Several degrees of imbalance: 40% and 100%

Furthermore, we also include three real databases (see Table 1) from the UCI Machine Learning Database Repository (<http://www.ics.uci.edu/~mllearn>). All data sets were transformed into two-class problems to provide a comparison with other published results [3]. A five-fold cross-validation error estimate method is employed in the classification tasks.

Table 1. A brief summary of the real databases

Data set	Features	Minority class	Majority class
Glass	9	29	185
Phoneme	5	1586	3818
Vehicle	18	212	634

In Table 2 and Table 3, the results with RBF and MLP neural networks are shown. Rows represent the results with the different preprocessing techniques applied to the TS respect to the original set. Moreover, the accuracy of

neural networks using the average geometric mean are represented for WE on the majority-class (Euclidean distance and weighted distance), and WE in both classes. This strategy is applied three times until the number of atypical or noise patterns is sufficiently small.

Table 2. RBF neural networks: average values of the geometric mean

		A0	A20	A40	A60	A80	A100
Original TS		98.99	89.32	74.83	61.64	37.42	0
WE (majority class)	1st application	98.99	90.61	78.47	61.42	35.60	0
	2nd application	98.99	89.15	77.22	61.82	35.60	0
	3rd application	98.99	89.15	78.04	61.88	38.10	0
weighted WE (majority class)	1st application	99.49	89.86	78.13	64.8	51.99	0
	2nd application	100	89.62	77.81	72.57	57.97	30.78
	3rd application	100	89.62	79.42	72.74	54.89	32.11
WE(both class)		98.49	89.44	72.11	58.3	37.42	0

The application of WE in the majority-class (see Table 2) equalizes or outperforms the classification accuracy for RBF classifier in all data sets. This improvement is remarkable in the case of A40, A60 and A80 databases for WE with Euclidean distance, and for all data sets when WE with the weighted distance is applied.

In general, the use of the weighted distance obtains better results than the Euclidean distance with WE in the majority class for the RBF neural networks. On the other hand, WE applied in both classes presents worse results, and in some cases does not improve the accuracy of the original TS.

Table 3. MLP neural networks: average values of the geometric mean

		A0	A20	A40	A60	A80	A100
Original TS		99.50	90.19	76.81	62.45	44.55	34.21
WE (majority class)	1st application	99.50	90.61	76.81	62.45	0	0
	2nd application	99.50	90.50	76.81	62.45	0	0
	3rd application	99.50	90.50	76.81	62.45	59.03	9.91
weighted WE (majority class)	1st application	100	87.98	79.61	64.23	60.77	0
	2nd application	100	87.62	74.65	65.91	54.67	0
	3rd application	100	87.62	76.68	70.25	55.10	0
WE(both class)		100	89.22	74.83	61.51	43.53	9.87

In the MLP neural network, we can observe that a data reduction in the majority-class does not enhance the classification accuracy meaningfully for this classifier (see Table 3). So, it is not clear the advantages of the weighted distance in WE for MLP neural network as editing technique to clean the TS.

In Table 4, a study of the classification accuracy for each individual class is presented in A60 database. Whenever the accuracy of the minority-class (a^+) increases, the accuracy of the majority-class (a^-) diminishes. The same characteristics was found for NN rule in problems with imbalanced classes [4].

Table 4. Partial accuracy for the A60 database. (a^-) the minority-class and (a^+) the majority-class.

		RBF		MLP	
		a^-	a^+	a^-	a^+
	Original TS	100.0	38.00	100.0	39.00
WE (majority class)	1st application	96.75	39.00	100.0	39.00
	2nd application	98.00	39.00	100.0	39.00
	3rd application	95.75	40.00	100.0	39.00
Weighted WE (majority class)	1st application	87.50	48.00	93.75	44.00
	2nd application	75.25	70.00	90.50	48.00
	3rd application	73.50	72.00	69.50	71.00
WE(both class)		100.00	34.00	97.00	39.00

Table 5. Size in the original TS and applying editing techniques

	A0	A20	A40	A60	A80	A100
Original TS	400	400	400	400	400	400
WE (majority class)	400	381	372	353	351	336
Weighted WE (majority class)	392	351	303	239	215	203
WE(both class)	400/94	385/82	379/53	369/41	364/21	355/5

When the weighted distance is used in WE, this reduction can be up to 50% of the size of the majority-class (a^-) (see Table 5). However, this does not mean a balance in the size of both classes. For example, WE with weighted distance in A100 database significantly reduces the size of the majority-class (a^-), but it does not reach a good balance in the size of both classes. Also, it is observed that the WE applied to both classes drastically reduces the size of the minority-class (a^+) to 5 patterns in A100 database.

Table 6 shows the classification accuracy before and after WE is applied in real data sets. In this case, we do not know the level of overlap between classes, and there is not information of atypical or noisy patterns. In RBF neural networks, the performance of the classifier improves when the majority-class of the TS is edited. Furthermore, the weighted distance is better than the Euclidean distance for the *Vehicle* and *Phoneme* data sets and worse for the *Glass* database.

In the case of MLP neural networks, one can see different behaviors and it only improves the results for the *Vehicle* database. In this classifier, it is not clear the benefits obtained when a WE in the majority-class is applied. The WE in both classes (see Table 6) obtains worse results than the original TS, except

Table 6. Average values of the geometric mean with real data sets

		Glass		Vehicle		Phoneme	
		RBF	MLP	RBF	MLP	RBF	MLP
Original TS		85.97	82.05	46.63	70.62	69.84	56.77
WE (majority class)	1st application	87.26	78.38	60.77	71.56	69.80	44.43
	2nd application	87.26	78.38	63.59	76.06	70.01	49.13
	3rd application	87.26	78.38	64.81	76.06	70.01	50.57
Weighted WE (majority class)	1st application	86.97	76.20	62.80	70.52	70.53	50.50
	2nd application	86.97	76.20	66.63	68.63	70.26	50.44
	3rd application	86.97	76.20	66.45	73.93	71.32	46.83
WE(both class)		81.52	86.43	39.86	67.23	67.89	47.80

with *Glass* database. It can be due to a smooth of the decision boundaries in the editing process. So, other issues such as the data complexity or the nature of the classifier must be analyzed in the editing process for imbalanced classes.

5 Conclusions

A preprocessing technique to filter the TS by removal of noise or atypical patterns is applied to clean the data, and enhance the classification accuracy in neural networks. In the case of RBF neural networks, the application of editing techniques with an adequate metric to imbalanced classes increases the classification accuracy. Despite the successful results in RBF neural networks, a common problem to all these downsizing techniques is that they do not allow a control on the number of patterns to be removed.

Nevertheless, this strategy to clean the decision boundaries has a worse behavior in the case of MLP neural networks. It can be due to a smoothing of the decision boundaries in the editing process. A study of the data complexity and the nature of the classifier is required to deepen the influence of the editing process with imbalanced classes.

Acknowledgments

This work has been supported in part by grants TIC2003-08496 from the Spanish CICYT, P1-1B2004-08 from Fundació Caixa Castelló-Bancaixa, and SEP-2003-C02-44225 from the Mexican CONACyT.

References

1. R. Anand, K. G. Mehrotra, C. K. Mohan, and S. Ranka. An Improved Algorithm for Neural Network Classification of Imbalanced Training Sets. *IEEE Transactions on Neural Networks*, vol. 4, no. 6, (1993) 962–969.
2. R. Barandela, N. Cortés, A. Palacios. The nearest neighbour rule and the reduction of the training sample size. 9th. Spanish Symposium on Pattern Recognition and Image Analysis, vol. 1, Benicassim, Spain, 2001, 103–108.

3. R. Barandela, J.S. Sánchez, V. García, E. Rangel. Strategies for learning in class imbalance problems. *Pattern Recognition*, vol. 36, no. 3 (2003) 849–851
4. R. Barandela, R.M. Valdovinos, J.S. Sánchez, F.J. Ferri. The imbalanced training sample problem: under or over sampling?. *Joint IAPR International Workshops on Structural, Syntactic, and Statistical Pattern Recognition (SSPR/SPR'04)*. Lecture Notes in Computer Science 3138, Springer-Verlag 2004, Lisbon (Portugal), 2004, 806–814.
5. V. L. Berardi, G. P. Zhang. The Effect of Misclassification Costs on Neural Network Classifiers. *Decision Sciences*, vol. 30, no. 3, 1999, 659–682
6. S.Q. Ding, C. Xiang. From multilayer perceptrons to radial basis function networks: a comparative study. *IEEE. Conference on Cybernetics and Intelligent Systems*, vol. 1, Singapore, 1-3 December, 2004, 69–74
7. X. Fu, L. Wang, K.S. Chua, F. Chu. Training RBF neural networks on unbalanced data. *IX International Conference on Neural Information Processing (ICONIP'02)*, Singapore, 2002, 1016–1020.
8. S. Haykin. *Neuronal Networks - a comprehensive foundation*. Ed. Prentice Hall, second edition, New Jersey, 1999, 278–282
9. J.M. Hutchinson, A. Lo, T. Poggio. A Nonparametric Approach to Pricing and Hedging Derivates Securities Via Learning Networks. Technical Report, Artificial Intelligence Laboratory and Center for Biological and Computational Learning, MIT, memo 1471, no. 92, 1994.
10. M. Kubat, S. Matwin. Addressing the curse of imbalanced training set: one-sided selection. *14th International Conference on Machine Learning*, Nashville, USA, 1997, 179–186
11. Y. Lu, H. Guo, and L. Feldkamp. Robust neural learning from unbalanced data examples. *IEEE International Joint Conference on Neural Networks*, 1998, 1816–1821
12. Y. H. Pao, *Adaptive Patter Recognition and Neuronal Networks*. Addison-Wesley, Reading. MA., 1989
13. F. Schwenker, H. A. Kestler, and G. Palm. Three learning phases for radial-basis-function networks. *Neural Networks*, vol.14 no. 4-5, May 2001, 439–458