Hybrid Associative Memories for Imbalanced Data Classification: An Experimental Study

L. Cleofas-Sánchez[1], V. García[2], R. Martín-Félez[2], R.M. Valdovinos[3], J.S. Sánchez[2], and O. Camacho-Nieto[4]

[1] Centro de Investigación en Computación, Instituto Politécnico Nacional, Av. Juan de Dios Bátiz s/n, Col. Nueva Industrial Vallejo, 07738 México D.F., México
[2] Institute of New Imaging Technologies, Department of Computer Languages and Systems, Universitat Jaume I, Av. Vicent Sos Baynat s/n, 12071 Castellón de la Plana, Spain
[3] Centro Universitario Valle de Chalco, Universidad Autónoma del Estado de México, Hermenegildo Galena 3, 56615 Valle de Chalco, México
[4] Centro de Innovación y Desarrollo Tecnológico en Cómputo, Instituto Politécnico Nacional, Av. Juan de Dios Bátiz s/n, Col. Nueva Industrial Vallejo, 07700 México D.F., México

Abstract. Hybrid associative memories are based on the combination of two well-known associative networks, the lernmatrix and the linear associator, with the aim of taking advantage of their merits and overcoming their limitations. While these models have extensively been applied to information retrieval problems, they have not been properly studied in the framework of classification and even less with imbalanced data. Accordingly, this work intends to give a comprehensive response to some issues regarding imbalanced data classification: (i) Are the hybrid associative models suitable for dealing with this sort of data? and, (ii) Does the degree of imbalance affect the performance of these neural classifiers? Experiments on real-world data sets demonstrate that independently of the imbalance ratio, the hybrid associative memories perform poorly in terms of area under the ROC curve, but the hybrid associative classifier with translation appears to be the best solution when assessing the true positive rate.

Keywords: Class Imbalance, Associative Memory, Neural Network.

1 Introduction

An associative memory [1] is a type of neural network that allows to recall the previously stored training example $\mathbf{x}^i$ that most closely resembles the one presented to the network. This connectionist model has demonstrated to be very effective for information storage and retrieval [2–4], but it has not been much studied in the framework of classification. Among the simplest and first studied associative memory models are the lernmatrix [5] and the linear associator [6,7], which are considered as hetero-associative memories capable of producing exact recall. Both these models can also work as classifiers, but they present some drawbacks that make difficult their application to many real-life problems: the lernmatrix needs to be provided with binary input vectors $\mathbf{x}^i \in \{0, 1\}$, whereas the linear associator requires the input vectors to be orthonormal and linearly independent.

J.A. Carrasco-Ochoa et al. (Eds.): MCPR 2013, LNCS 7914, pp. 325–334, 2013.

In order to benefit from the advantages of these associative memories and overcome their shortcomings, several extensions have been developed. These include the hybrid associative classifier (HAC) and the hybrid associative classifier with translation (HACT) [8], which combine the procedure used by the linear associator in the learning phase with the recall stage of the lernmatrix. While these two classification models have been used with some success in a number of applications, there still exist open questions regarding their limitations that deserve a more thorough investigation. For example, the present paper addresses the issue of imbalanced data classification [9], which appears as a much more challenging task for this type of associative memories.

Many complex pattern recognition and data mining problems are characterized by imbalanced data, where at least one class is heavily under-represented as compared to others. Following the common practice in the area [10, 11], we will here consider only binary classification problems where the examples from the majority class are often referred to as the negative examples and those from the minority class as the positive examples, since these usually represent the concept of most interest.

The importance of the class imbalance problem comes from the fact that in general, it hinders the performance of most standard learning algorithms because they are often biased towards the majority class and have a poor performance on the minority class. Besides the classifiers are commonly built with the aim of reducing the overall error, what may lead to erroneous conclusions; for example, an algorithm that achieves an accuracy of 99% will be worthless if it fails on classifying all positive examples.

Many classifiers have been investigated in the context of class imbalance, ranging from the nearest neighbor rule and decision trees to support vector machines and various topologies of neural networks [11–15]. However, to the best of our knowledge, the use of associative memory models has not received adequate attention from researchers on this topic. In fact, we have found only a recent work [16] that analyzes the performance of the HACT approach after under-sampling the imbalanced data set, but it presents several limitations such as the reduced number of databases used in the experiments, the lack of comparisons with other state-of-the-art classifiers and especially, the fact that it does not take care of the imbalance ratio (i.e. the ratio of the majority to the minority instances) and its effect on the HACT performance.

The purpose of this paper is to gain insight into the behavior of the HAC and HACT associative models when these are used for the classification of imbalanced data, pursuing to fully understand how the class imbalance affects the performance of these classifiers. To this end, we provide a large pool of experiments on 58 real-world benchmarking data sets that have different degrees of imbalance, comparing those hybrid associative memories with other well-known artificial neural networks: a Bayesian network (BNet), a multilayer perceptron (MLP) and a radial basis function (RBF). We conducted our experiments by evaluating three performance metrics: the area under the ROC curve, the true positive rate and the true negative rate.

2 Two Hybrid Associative Memories

In this section we provide a brief introduction to the associative memory models that will be further experimented with, covering only the general concepts and notation needed to understand their foundations. For a complete description of associative memories, the reader may review any of the many books on this subject (e.g. [17, 18]).

In general, an associative memory can be defined as a mapping matrix $\mathbf{M}$ so that an input vector $\mathbf{x}^i \in \mathbb{R}^n$ (with n components) will be transformed into an output vector $\mathbf{y}^i \in \mathbb{R}^m$ (with m components), that is

$$\mathbf{y}^i = \mathbf{M}\mathbf{x}^i \qquad i = 1, \ldots, p \tag{1}$$

where p denotes the number of input vectors.

The stored samples will be represented in the form of pairs of associations $(\mathbf{x}^i, \mathbf{y}^i)$ between the input and output vectors, $\mathbf{x}^i$ and $\mathbf{y}^i$, and are often called fundamental pattern. The set of p pairs (fundamental patterns) constitutes the fundamental set of associations.

The matrix $\mathbf{M}$ has to be determined through an iterative procedure in the learning phase. Afterwards, during the recall or recovery phase, an unknown pattern $\mathbf{x}^0$ will be applied to the input of the matrix in order to produce the vector $\mathbf{y}^0$, which is expected to be a good approximation of the true output $\mathbf{y}$.

Hybrid Associative Classifier (HAC). As previously pointed out, the HAC model [8] arises from the combination of the lernmatrix and the linear associator with the aim of overcoming the practical drawbacks of these associative neural networks. Apart from these obvious advantages, it is worth remarking that the HAC model presents some other interesting properties such as simplicity, requirements of low computational cost and the ability to support real-valued input vectors [8].

During the learning phase, the HAC memory imitates the process of the linear associator: each sample that belongs to class k is represented by a vector with zeros in all components except the k'th element that equals 1. In this way, the outer product of vectors $\mathbf{x}^i$ and $\mathbf{y}^i$ gives the corresponding associations between them. Then the matrix $\mathbf{M}$ of size $n \times m$ will be obtained as the sum of all p outer products as

$$\mathbf{M} = \sum_{i=1}^{p} (\mathbf{y}^i)(\mathbf{x}^i)^{\mathbf{T}} \tag{2}$$

After computing the mapping matrix $\mathbf{M}$, the recovery of a given input sample will be performed following the process of the lernmatrix model in order to estimate its class label.

It has to be pointed out, however, that a practical drawback of the HAC model comes from the possible large differences in the magnitude of the input vectors because in such a case, the vectors with a lower magnitude will be assigned to the class of the vectors with a larger magnitude.

Hybrid Associative Classifier with Translation (HACT). This is a modification of the HAC model that tries to face several of its limitations. More specifically, if the input samples are clustered in the same quadrant, the performance of the HAC memory will be affected negatively. Thus the HACT approach [8] starts with a translation of the coordinate axes whose origin is taken to lie in the mean vector of all the input vectors as computed by

$$\overline{\mathbf{x}} = \frac{1}{p}\sum_{i=1}^{p}\mathbf{x}^i \tag{3}$$

In this way, the new coordinate axes are parallel to the original coordinate axes, but eliminates the clustering of samples in a unique quadrant. Then the input and test vectors in the new coordinate system will be obtained as $\mathbf{x}^i = \mathbf{x}^i - \overline{\mathbf{x}}$. After the corresponding translation of axes, the learning and recovery phases will be the same as those described for the HAC model.

3 Experimental Set-Up

As already discussed, the aim of this work and the experiments conducted here is to investigate whether two models of hybrid associative memories, which are based on the lernmatrix and the linear associator, are suitable or not for imbalanced data classification, and to what extent the degree of imbalance may affect their performance.

Table 1. Description of the data sets used in the experiments

Data sets	Features	Samples	IR	Data sets	Features	Samples	IR
Glass1	9	214	1.82	Ecoli-0-3-4-6_vs_5	7	205	9.25
Pima	8	768	1.87	Ecoli-0-3-4-7_vs_5-6	7	257	9.28
Iris0	4	150	2.00	Yeast-0-5-6-7-9_vs_4	8	528	9.35
Glass0	9	214	2.06	Vowel0	13	988	9.98
Yeast1	8	1484	2.46	Ecoli-0-6-7_vs_5	6	220	10.00
Haberman	3	306	2.78	Glass-0-1-6_vs_2	9	192	10.29
Vehicle3	18	846	3.00	Ecoli-0-1-4-7_vs_2-3-5-6	7	336	10.59
Glass-0-1-2-3_vs_4-5-6	9	214	3.20	Led-0-2-4-5-6-7-8-9_vs_1	7	443	10.97
Vehicle0	18	846	3.25	Ecoli-0-1_vs_5	6	240	11.00
Ecoli1	7	336	3.36	Glass-0-6_vs_5	9	108	11.00
New-thyroid2	5	215	5.14	Glass-0-1-4-6_vs_2	9	205	11.06
Ecoli2	7	336	5.46	Glass2	9	214	11.59
Segment0	19	2308	6.02	Ecoli-0-1-4-7_vs_5-6	6	332	12.28
Glass6	9	214	6.38	Cleveland-0_vs_4	13	177	12.62
Yeast3	8	1484	8.10	Ecoli-0-1-4-6_vs_5	6	280	13.00
Ecoli3	7	336	8.60	Shuttle-0_vs_4	9	1829	13.87
Page-blocks0	10	5472	8.79	Yeast-1_vs_7	7	459	14.30
Ecoli-0-3-4_vs_5	7	200	9.00	Glass4	9	214	15.47
Yeast-2_vs_4	8	514	9.08	Ecoli4	7	336	15.80
Ecoli-0-6-7_vs_3-5	7	222	9.09	Page-blocks-1-3_vs_4	10	472	15.86
Ecoli-0-2-3-4_vs_5	7	202	9.10	Glass-0-1-6_vs_5	9	184	19.44
Glass-0-1-5_vs_2	9	172	9.12	Yeast-1-4-5-8_vs_7	8	693	22.10
Yeast-0-3-5-9_vs_7-8	8	506	9.12	Glass5	9	214	22.78
Yeast-0-2-5-6_vs_3-7-8-9	8	1004	9.14	Yeast-2_vs_8	8	482	23.10
Yeast-0-2-5-7-9_vs_3-6-8	8	1004	9.14	Yeast4	8	1484	28.10
Ecoli-0-4-6_vs_5	6	203	9.15	Yeast-1-2-8-9_vs_7	8	947	30.57
Ecoli-0-1_vs_2-3-5	7	244	9.17	Yeast5	8	1484	32.73
Ecoli-0-2-6-7_vs_3-5	7	224	9.18	Ecoli-0-1-3-7_vs_2-6	7	281	39.14
Glass-0-4_vs_5	9	92	9.22	Yeast6	8	1484	41.40

The empirical analysis has been performed over a total of 58 benchmarking data sets taken from the KEEL Data Set Repository (`http://www.keel.es/dataset.php`) [19]. Note that all the original multi-class databases have firstly been transformed into two-class problems. Table 1 summarizes the main characteristics of the data sets, including the imbalance ratio (IR), i.e. the number of negative examples divided by the number of positive examples. As can be seen, the databases chosen for the experiments go from a low imbalance of 1.82 in Glass1 to a high/moderate imbalance of 41.40 in the case of Yeast6.

In order to gain sufficient insight into the behavior of the associative memory models, three other neural networks (BNet, MLP, RBF) have been used as baselines for comparison purposes. These were taken from the Weka toolkit [20] with their default parameter values. For the experiments here carried out, we have adopted a 5-fold cross-validation method to estimate three classification performance measures commonly used in skewed domains: the area under the ROC curve (AUC), the true positive rate (TPrate) and the true negative rate (TNrate). Each data set has been divided into five stratified blocks of size $N/5$ (where N denotes the total number of samples in the database), using four folds for training the connectionist classifiers and the remaining block as an independent test set. Therefore the results reported in tables of Section 4 correspond to those three measures averaged over the five runs.

Table 2. Confusion matrix for a two-class problem

	Predicted positive	Predicted negative
Positive class	True Positive (TP)	False Negative (FN)
Negative class	False Positive (FP)	True Negative (TN)

Given a 2×2 confusion matrix as that illustrated in Table 2, the performance measures used in the experiments can be calculated as follows: $TPrate = \frac{TP}{TP+FN}$, $TNrate = \frac{TN}{TN+FP}$, and $AUC = \frac{TPrate+TNrate}{2}$. Note that the latter corresponds to the AUC defined by a single point on the ROC curve.

4 Experimental Results

Table 3 reports the AUC values obtained by the neural network models on each database, along with the average across the whole collection of data sets. From these results, the first observation is that the HAC memory yields a 50% of AUC, which indicates that all samples of one class have been misclassified while all of the other have been correctly classified. This effect has not been found in the case of the HACT model, but its performance in terms of AUC is lower than that achieved by the three other neural networks on most databases. When paying attention of the average values, the MLP model clearly performs the best (80.70% of AUC), but the results of the Bayesian network and the RBF are not too far from that of the HACT approach.

Table 3. Experimental results using the AUC

Data set	HAC	HACT	BNet	MLP	RBF	Data set	HAC	HACT	BNet	MLP	RBF
Glass1	50.00	56.02	67.51	68.6	62.24	Ecoli-0-3-4-6_vs_5	50.00	79.12	83.11	88.65	91.96
Pima	50.00	57.58	69.01	74.69	70.30	Ecoli-0-3-4-7_vs_5-6	50.00	79.05	73.78	88.92	84.06
Iris0	50.00	95.50	100	100	100	Yeast-0-5-6-7-9_vs_4	50.00	74.94	56.91	72.79	53.36
Glass0	50.00	71.53	79.93	77.01	67.63	Vowel0	50.00	77.39	88.43	99.44	86.78
Yeast1	50.00	66.92	67.59	66.94	60.74	Ecoli-0-6-7_vs_5	50.00	79.75	82.25	86.50	87.25
Haberman	50.00	62.74	55.42	58.10	55.11	Glass-0-1-6_vs_2	50.00	63.14	50.00	47.71	48.00
Vehicle3	50.00	65.10	67.63	74.26	63.63	Ecoli-0-1-4-7_vs_2-3-5-6	50.00	76.81	80.51	87.03	79.01
Glass-0-1-2-3_vs_4-5-6	50.00	92.69	88.26	92.03	89.41	Led-0-2-4-5-6-7-8-9_vs_1	51.25	81.66	88.24	89.30	83.06
Vehicle0	50.00	74.64	81.74	94.95	84.51	Ecoli-0-1_vs_5	50.00	77.72	87.04	89.54	89.54
Ecoli1	50.00	87.36	85.01	85.83	88.35	Glass-0-6_vs_5	50.00	86.34	78.39	100	94.50
New-thyroid2	50.00	75.71	92.85	95.15	98.01	Glass-0-1-4-6_vs_2	50.00	64.62	50.00	48.67	49.74
Ecoli2	50.00	82.34	86.08	89.24	90.72	Glass2	50.00	65.49	50.00	51.03	48.97
Segment0	50.00	75.82	98.78	99.39	97.71	Ecoli-0-1-4-7_vs_5-6	50.00	79.30	51.84	84.87	83.19
Glass6	50.00	89.41	91.17	84.92	87.44	Cleveland-0_vs_4	50.00	47.92	62.63	87.22	84.90
Yeast3	50.00	78.92	85.42	85.85	87.06	Ecoli-0-1-4-6_vs_5	50.00	77.31	86.93	79.05	89.23
Ecoli3	50.00	81.96	84.01	78.34	66.82	Shuttle-0_vs_4	50.00	91.19	100	99.60	99.11
Page-blocks0	50.00	48.70	89.73	87.59	74.52	Yeast-1_vs_7	50.00	65.25	46.43	62.61	54.53
Ecoli-0-3-4_vs_5	50.00	80.00	84.44	88.60	91.66	Glass4	50.00	82.57	64.92	87.34	86.59
Yeast-2_vs_4	50.00	74.67	87.40	82.50	87.89	Ecoli4	50.00	81.51	82.34	89.21	89.05
Ecoli-0-6-7_vs_3-5	50.00	77.00	89.00	82.50	68.50	Page-blocks-1-3_vs_4	50.00	80.17	96.56	97.89	91.99
Ecoli-0-2-3-4_vs_5	50.00	80.22	86.40	89.17	89.20	Glass-0-1-6_vs_5	50.00	88.29	90.43	79.14	89.71
Glass-0-1-5_vs_2	50.00	63.63	50.00	52.48	50.24	Yeast-1-4-5-8_vs_7	50.00	59.65	50.00	51.37	50.00
Yeast-0-3-5-9_vs_7-8	50.00	69.43	59.78	64.69	61.45	Glass5	50.00	88.05	91.34	89.51	84.02
Yeast-0-2-5-6_vs_3-7-8-9	50.00	69.89	75.08	73.38	67.66	Yeast-2_vs_8	50.00	77.32	77.39	77.06	79.78
Yeast-0-2-5-7-9_vs_3-6-8	50.00	75.75	83.89	86.22	88.86	Yeast4	50.00	73.32	62.84	64.39	50.00
Ecoli-0-4-6_vs_5	50.00	78.97	89.18	88.92	86.69	Yeast-1-2-8-9_vs_7	50.00	65.03	57.96	56.46	51.67
Ecoli-0-1_vs_2-3-5	50.00	77.54	50.56	80.67	79.21	Yeast5	50.00	78.65	91.77	83.60	63.30
Ecoli-0-2-6-7_vs_3-5	50.00	77.95	80.01	81.01	81.01	Ecoli-0-1-3-7_vs_2-6	50.00	80.85	84.63	84.81	84.63
Glass-0-4_vs_5	50.00	90.81	99.41	100	94.41	Yeast6	50.00	74.89	83.30	73.85	50.00
Average							50.02	75.45	77.16	80.70	77.05

By the analysis of the behavior of these classifiers as a function of the imbalance ratio, one can guess that there is not necessarily a direct relationship between the classification performance and the degree of imbalance. For instance, the balanced accuracies for the Ecoli-0-1-3-7_vs_2-6 database, which has an imbalance ratio of 39.14, are significantly higher than those for Glass1, even though this presents a very low imbalance ratio of 1.82. In some sense, it appears that databases may also suffer from other intrinsic problems such as class overlapping, small disjuncts, feature noise and lack of representative data, which in turn may affect classification performance much more strongly than the presence of class imbalance.

In order to accomplish a better understanding of the performance of these neural network models, Tables 4 and 5 report the true positive and true negative rates, respectively. These measures allow to analyze the behavior of a classifier on each individual class, thus drawing out whether it is biased towards one class or another. This is especially important in the context of imbalanced data because the examples from the minority class, which usually correspond to the most interesting cases, are more likely to be misclassified. In addition, it is often preferable to achieve a higher true positive rate rather than a higher true negative rate and consequently, the AUC by itself is not sufficiently informative when evaluating the performance of a set of classifiers.

Table 4. Experimental results using the true positive rate

Data set	HAC	HACT	BNet	MLP	RBF	Data set	HAC	HACT	BNet	MLP	RBF
Glass1	0	80.17	47.34	59.60	50.00	Ecoli-0-3-4-6_vs_5	0	95.00	70	80	85.00
Pima	0	44.36	58.22	67.18	55.20	Ecoli-0-3-4-7_vs_5-6	0	96.00	48.00	80.00	72.00
Iris0	0	100	100	100	100	Yeast-0-5-6-7-9_vs_4	0	86.36	18.00	48.72	8.00
Glass0	0	100	80.00	70.00	42.84	Vowel0	0	97.78	78.86	98.88	75.56
Yeast1	0	76.68	46.14	43.84	27.28	Ecoli-0-6-7_vs_5	0	95.00	65.00	75.00	75.00
Haberman	0	59.26	17.52	28.20	15.98	Glass-0-1-6_vs_2	0	100	0	0	0
Vehicle3	0	60.33	63.64	58.94	41.92	Ecoli-0-1-4-7_vs_2-3-5-6	0	93.33	62.66	76.00	58.66
Glass-0-1-2-3_vs_4-5-6	0	94.00	80.18	87.74	84.36	Led-0-2-4-5-6-7-8-9_vs_1	2.50	100	78.20	81.06	67.84
Vehicle0	0	100	95.94	90.98	80.92	Ecoli-0-1_vs_5	0	100	75.00	80.00	80.00
Ecoli1	0	94.83	83.16	76.68	91.02	Glass-0-6_vs_5	0	100	70.00	100	90.00
New-thyroid2	0	91.43	85.70	91.42	97.14	Glass-0-1-4-6_vs_2	0	100	0	0	0
Ecoli2	0	96.36	77.44	82.72	87.08	Glass2	0	100	0	6.66	0
Segment0	0	100	98.20	99.10	97.90	Ecoli-0-1-4-7_vs_5-6	0	100	44.00	72.00	68.00
Glass6	0	96.67	86.66	72.00	78.66	Cleveland-0_vs_4	0	33.50	26.04	78.18	71.52
Yeast3	0	98.79	72.94	74.28	77.32	Ecoli-0-1-4-6_vs_5	0	100	75.00	60.00	80.00
Ecoli3	0	97.14	79.98	59.98	34.30	Shuttle-0_vs_4	0	99.20	100	99.20	98.40
Page-blocks0	0	19.15	85.32	76.92	50.84	Yeast-1_vs_7	0	76.67	13.34	26.64	10.00
Ecoli-0-3-4_vs_5	0	100	70.00	80.00	85.00	Glass4	0	90.00	33.32	76.68	76.68
Yeast-2_vs_4	0	90.18	76.54	66.54	78.36	Ecoli4	0	100	65.00	80.00	80.00
Ecoli-0-6-7_vs_3-5	0	88.00	80.00	67.00	41.00	Page-blocks-1-3_vs_4	0	68.67	100	96.00	86.00
Ecoli-0-2-3-4_vs_5	0	100	75.00	80.00	80.00	Glass-0-1-6_vs_5	0	100	90.00	60.00	80.00
Glass-0-1-5_vs_2	0	95.00	0	13.34	5.00	Yeast-1-4-5-8_vs_7	0	66.67	0	3.34	0
Yeast-0-3-5-9_vs_7-8	0	86.00	20.00	34.00	24.00	Glass5	0	100	90.00	80.00	70.00
Yeast-0-2-5-6_vs_3-7-8-9	0	77.68	54.36	49.42	37.32	Yeast-2_vs_8	0	70.00	55.00	55.00	60.00
Yeast-0-2-5-7-9_vs_3-6-8	0	89.95	70.68	73.78	79.94	Yeast4	0	90.18	29.28	29.46	0
Ecoli-0-4-6_vs_5	0	95.00	80.00	80.00	75.00	Yeast-1-2-8-9_vs_7	0	80.00	16.68	13.34	3.34
Ecoli-0-1_vs_2-3-5	0	96.00	10.00	65.00	65.00	Yeast5	0	100	86.4	68.08	26.92
Ecoli-0-2-6-7_vs_3-5	0	90.00	63.00	64.00	64.00	Ecoli-0-1-3-7_vs_2-6	0	100	70.00	70.00	70.00
Glass-0-4_vs_5	0	100	100	100	90.00	Yeast6	0	94.29	71.42	48.58	0
Average							0.04	88.96	60.16	64.75	57.42

For instance, the results of HAC in Table 4 demonstrate that this hybrid associative model is of no value at all because it fails on the classification of all positive examples. This makes clear that the AUC of 50% reported in Table 3 is due to the awful true positive rate of this classifier and the very high rate achieved on the negative class (see Table 5). On the contrary, the true positive rate of HACT suggests that this can be a good tool for the classification of data with class imbalance because it yields a true positive rate of close to 89% in average, that is, even higher than that of the best performing algorithm (MLP) in terms of AUC.

It is also interesting to note that in general, the highest differences between HACT and MLP are found in the most strongly imbalanced data sets. Unfortunately, in these cases, the true negative rate of the HACT model is lower than that of the MLP, but we should recall that there exist numerous real-world applications in which the minority class represents the concept of most interest and therefore, it will be crucial to correctly classify the positive examples even if this might entail a certain degradation of the true negative rate.

Table 5. Experimental results using the true negative rate

Data set	HAC	HACT	BNet	MLP	RBF	Data set	HAC	HACT	BNet	MLP	RBF
Glass1	100	31.88	87.68	77.6	74.48	Ecoli-0-3-4-6_vs_5	100	63.24	96.22	97.30	98.92
Pima	100	70.80	79.80	82.20	85.40	Ecoli-0-3-4-7_vs_5-6	100	62.11	99.56	97.84	96.12
Iris0	100	91.00	100	100	100	Yeast-0-5-6-7-9_vs_4	100	63.53	95.82	96.86	98.72
Glass0	100	43.05	79.86	84.02	92.42	Vowel0	100	57.01	98.00	100	98.00
Yeast1	100	57.16	89.04	90.04	94.20	Ecoli-0-6-7_vs_5	100	64.50	99.50	98.00	99.50
Haberman	100	66.22	93.32	88.00	94.24	Glass-0-1-6_vs_2	100	26.29	100	95.42	96.00
Vehicle3	100	69.87	71.62	89.58	85.34	Ecoli-0-1-4-7_vs_2-3-5-6	100	60.28	98.36	98.06	99.36
Glass-0-1-2-3_vs_4-5-6	100	91.38	96.34	96.32	94.46	Led-0-2-4-5-6-7-8-9_vs_1	100	63.33	98.28	97.54	98.28
Vehicle0	100	49.29	67.54	98.92	88.10	Ecoli-0-1_vs_5	100	55.45	99.08	99.08	99.08
Ecoli1	100	79.88	86.86	94.98	85.68	Glass-0-6_vs_5	100	72.68	86.78	100	99.00
New-thyroid2	100	60.00	100	98.88	98.88	Glass-0-1-4-6_vs_2	100	29.25	100	97.34	99.48
Ecoli2	100	68.31	94.72	95.76	94.36	Glass2	100	30.99	100	95.40	97.94
Segment0	100	51.64	99.36	99.68	97.52	Ecoli-0-1-4-7_vs_5-6	100	58.59	59.68	97.74	98.38
Glass6	100	82.16	95.68	97.84	96.22	Cleveland-0_vs_4	100	52.50	99.22	96.26	98.28
Yeast3	100	59.05	97.90	97.42	96.80	Ecoli-0-1-4-6_vs_5	100	54.62	98.86	98.10	98.46
Ecoli3	100	66.78	88.04	96.70	99.34	Shuttle-0_vs_4	100	83.18	100	100	99.82
Page-blocks0	100	78.24	94.14	98.26	98.20	Yeast-1_vs_7	100	53.84	79.52	98.58	99.06
Ecoli-0-3-4_vs_5	100	60.00	98.88	97.20	98.32	Glass4	100	75.13	96.52	98.00	96.50
Yeast-2_vs_4	100	59.17	98.26	98.46	97.42	Ecoli4	100	63.03	99.68	98.42	98.10
Ecoli-0-6-7_vs_3-5	100	66.00	98.00	98.00	96.00	Page-blocks-1-3_vs_4	100	91.67	93.12	99.78	97.98
Ecoli-0-2-3-4_vs_5	100	60.44	97.80	98.34	98.40	Glass-0-1-6_vs_5	100	76.57	90.86	98.28	99.42
Glass-0-1-5_vs_2	100	32.26	100	91.62	95.48	Yeast-1-4-5-8_vs_7	100	52.63	100	99.40	100
Yeast-0-3-5-9_vs_7-8	100	52.85	99.56	95.38	98.90	Glass5	100	76.10	92.68	99.02	98.04
Yeast-0-2-5-6_vs_3-7-8-9	100	62.10	95.80	97.34	98.00	Yeast-2_vs_8	100	84.65	99.78	99.12	99.56
Yeast-0-2-5-7-9_vs_3-6-8	100	61.55	97.10	98.66	97.78	Yeast4	100	56.46	96.40	99.32	100
Ecoli-0-4-6_vs_5	100	62.94	98.36	97.84	98.38	Yeast-1-2-8-9_vs_7	100	50.06	99.24	99.58	100
Ecoli-0-1_vs_2-3-5	100	59.09	91.12	96.34	93.42	Yeast5	100	57.29	97.14	99.12	99.68
Ecoli-0-2-6-7_vs_3-5	100	65.90	97.02	98.02	98.02	Ecoli-0-1-3-7_vs_2-6	100	61.69	99.26	99.62	99.26
Glass-0-4_vs_5	100	81.62	98.82	100	98.82	Yeast6	100	55.49	95.18	99.12	100
Average							100	61.94	94.16	96.65	96.68

5 Conclusions and Further Research

This paper pursues to investigate the suitability of associative memories for the classification of data in imbalanced domains. In particular, the present work has concentrated on two hybrid models, the hybrid associative classifier (HAC) and the hybrid associative classifier with translation (HACT), which come from the combination of the learning phase of the linear associator with the recall phase of the lernmatrix.

In contrast to the lernmatrix and the linear associator, two of the main characteristics of HAC and HACT refer to the potential of using real-valued input vectors that do not require to be orthonormal and linearly independent. These appealing properties allow the application of the hybrid associative models to a huge number of real-life problems. However, they have not been thoroughly studied in the context of imbalanced data classification and therefore, it is not possible to fully assert their suitability in domains where it is common to find such a complexity in data (i.e., credit risk evaluation, fraud detection in mobile telephone communications and prediction of rare diseases).

The experiments carried out over a collection of 58 real-world databases with the two hybrid associative models and three classical neural networks (Bayesian network, MLP and RBF) have demonstrated both the non-suitability of the HAC approach and the good performance of the HACT memory. In fact, this model has achieved the highest true positive rate, which is often the most important measure when working on imbalanced data because it denotes the percentage of examples from the minority class that have been correctly classified.

This work has to be viewed just as a preliminary analysis of the hybrid associative memories in classification problems with skewed class distributions. Other avenues for further research still remain open and therefore, it will be necessary a more exhaustive experimentation that will allow to give response to a number of important issues: (i) How do other complexities in data affect the classification performance on imbalanced data sets?, (ii) Are the gains in the true positive rate of the HACT model statistically significant?, (iii) Does the HACT memory outperform other non-neural classifiers such as support vector machines and decision trees?, and (iv) How do other associative memory models perform in the presence of class imbalance?

Acknowledgment. This work has partially been supported by the Spanish Ministry of Education and Science and the Generalitat Valenciana under grants TIN2009–14205 and PROMETEO/2010/028 and Secretaría de Investigación y Posgrado del Instituto Politécnico Nacional , respectively.

References

1. Anderson, J.A.: A simple neural network generating an interactive memory. Mathematical Biosciences 14(3-4), 197–220 (1972)
2. Deng, T., Zhao, L., Feng, L., Xue, W.: Information re-finding by context: A brain memory inspired approach. In: Proc. 20th ACM International Conference on Information and Knowledge Management, Glasgow, Scotland, UK, pp. 1553–1558 (2011)
3. Pantazi, S.V.: Methodological review: Unsupervised grammar induction and similarity retrieval in medical language processing using the Deterministic Dynamic Associative Memory (DDAM) model. Journal of Biomedical Informatics 43(5), 844–857 (2010)
4. Yaakobi, E., Bruck, J.: On the uncertainty of information retrieval in associative memories. In: Proc. IEEE International Symposium on Information Theory, Cambridge, MA, pp. 106–110 (2012)
5. Steinbuch, K.: Die Lernmatrix. Kybernetik 1, 36–45 (1961)
6. Anderson, J.A., Rosenfeld, E. (eds.): Neurocomputing: Foundations of Research. MIT Press, Cambridge (1988)
7. Kohonen, T.: Self-Organization and Associative Memory. Springer, New York (1989)
8. Santiago-Montero, R.: Clasificador Híbrido de Patrones basado en la Lernmatrix de Steinbuch y el Linear Associator de Anderson-Kohonen. Master's thesis, Centro de Investigación en Computación, Instituto Politécnico Nacional, México D.F. (2003)
9. Japkowicz, N., Shaju, S.: The class imbalance problem: A systematic study. Intelligent Data Analysis 6(5), 429–449 (2002)
10. Galar, M., Fernández, A., Barrenechea, E., Bustince, H., Herrera, F.: A review on ensembles for the class imbalance problems: Bagging, boosting, and hybrid-based approaches. IEEE Trans. on Systems, Man, and Cybernetics – Part C: Applications and Reviews 42(4), 463–484 (2012)
11. García, V., Sánchez, J.S., Mollineda, R.A.: On the effectiveness of preprocessing methods when dealing with different levels of class imbalance. Knowledge-Based Systems 25(1), 13–21 (2012)

12. Akbani, R., Kwek, S., Japkowicz, N.: Applying support vector machines to imbalanced datasets. In: Proc. 15th European Conference on Machine Learning, Pisa, Italy, pp. 39–50 (2004)
13. Cieslak, D.A., Chawla, N.V.: Learning decision trees for unbalanced data. In: Proc. 19th European Conference on Machine Learning, Antwerp, Belgium, pp. 241–256 (2008)
14. Khoshgoftaar, T.M., Van Hulse, J., Napolitano, A.: Supervised neural network modeling: An empirical investigation into learning from imbalanced data with labeling errors. IEEE Trans. on Neural Networks 21(5), 813–830 (2010)
15. Hoens, T.R., Qian, Q., Chawla, N.V., Zhou, Z.-H.: Building decision trees for the multi-class imbalance problem. In: Tan, P.-N., Chawla, S., Ho, C.K., Bailey, J. (eds.) PAKDD 2012, Part I. LNCS, vol. 7301, pp. 122–134. Springer, Heidelberg (2012)
16. Cleofas-Sánchez, L., Guzmán Escobedo, M., Valdovinos, R.M., Yáñez Márquez, C., Camacho-Nieto, O.: Using hybrid associative classifier with traslation (HACT) for studying imbalance data sets. Engineering and Research 32(1), 53–57 (2012)
17. Hasson, M.H.: Associative Neural Memories: Theory and Implementation. Oxford University Press (1993)
18. Yáñez Márquez, C., Díaz de León, J.L.: Introducción a las Memorias Asociativas. Instituto Politécnico Nacional, México D.F. (2003)
19. Alcalá-Fernández, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sánchez, L., Herrera, F.: KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. Journal of Multiple-Value Logic and Soft Computing 17(2-3), 255–287 (2011)
20. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. SIGKDD Explorations Newsletter 11(1), 10–18 (2009)