Expert Systems With Applications 54 (2016) 387–397

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Associative learning on imbalanced environments: An empirical study

L. Cleofas-Sánchez [a], J.S. Sánchez [a,*], V. García [b], R.M. Valdovinos [c]

[a] Institute of New Imaging Technologies, Department of Computer Languages and Systems, Universitat Jaume I, Castelló de la Plana, Spain
[b] División Multidisciplinaria de Ciudad Universitaria, Universidad Autónoma de Ciudad Juárez, Ciudad Juárez, Chihuahua, Mexico
[c] School of Engineering, Universidad Autónoma del Estado de México, Toluca, Mexico

ARTICLE INFO

Keywords:
Associative memory
Class imbalance
Resampling

ABSTRACT

Associative memories have emerged as a powerful computational neural network model for several pattern classification problems. Like most traditional classifiers, these models assume that the classes share similar prior probabilities. However, in many real-life applications the ratios of prior probabilities between classes are extremely skewed. Although the literature has provided numerous studies that examine the performance degradation of renowned classifiers on different imbalanced scenarios, so far this effect has not been supported by a thorough empirical study in the context of associative memories. In this paper, we fix our attention on the applicability of the associative neural networks to the classification of imbalanced data. The key questions here addressed are whether these models perform better, the same or worse than other popular classifiers, how the level of imbalance affects their performance, and whether distinct resampling strategies produce a different impact on the associative memories. In order to answer these questions and gain further insight into the feasibility and efficiency of the associative memories, a large-scale experimental evaluation with 31 databases, seven classification models and four resampling algorithms is carried out here, along with a non-parametric statistical test to discover any significant differences between each pair of classifiers.

1. Introduction

Associative memories have been an active research area in pattern recognition and neuroscience for over 50 years (Palm, 2013). Although typical applications of these connectionist models include image recognition and recovery, data analysis, control, inference and prediction (Danilo et al., 2015; Kareem & Jantan, 2011; Lou & Cui, 2007; Mu, Artiklar, Watta, & Hassoun, 2006; Nazari, Eftekhari-Moghadam, & Moin, 2014; Štanclová & Zavoral, 2005), the associative memories have lately emerged as useful classifiers for a large variety of problems in data mining and computational intelligence (Aldape-Pérez, Yañez Márquez, Camacho-Nieto, & Argüelles-Cruz, 2012; Aldape-Pérez, Yañez Márquez, Camacho-Nieto, López-Yáñez, & Argüelles-Cruz, 2015; Sharma et al., 2008; Uriarte-Arcia, López-Yáñez, & Yáñez Márquez, 2014).

From a practical point of view, classification refers to the assignment of a finite set of samples to predefined classes based on a number of observed variables or attributes. The effective design of classifiers is a complex task where the underlying quality of data becomes critical to further achieve accurate classifications. Besides other factors, the performance of classifiers strongly depends on the adequacy of the data in terms of several intrinsic characteristics such as number of examples, relevance of the attributes, class distribution and completeness of data. In particular, a very common situation in real-life classification problems is to find data sets where the amount of examples available for one class is quite different from that of the other classes. This significant difference in the size of the classes is usually known as class imbalance, and it has been observed that most traditional classifiers perform poorly on the minority class because they assume an even class distribution and equal misclassification costs (Japkowicz & Stephen, 2002).

Examples of applications with skewed class distributions can be found in many different areas ranging from bioinformatics and medicine to finance, network security and text mining. For instance, in credit risk and bankruptcy prediction, the number of observations in the class of defaulters is much smaller than the number of cases belonging to the class of non-defaulters (Kim, Kang, & Kim, 2015). Gene expression microarray analysis for cancer classification shows significant differences between the number of cancerous and healthy tissue samples (Blagus & Lusa, 2010). Web spam detection exhibits an asymmetric distribution between legitimate and spam pages (Fdez-Glez et al., 2015).

Within this context, the major research line has traditionally been directed to the development of techniques to tackle the class imbalance at both algorithmic and data levels (He & Garcia, 2009;

* Corresponding author. Tel.: +34964728350.
E-mail addresses: cleofas@uji.es (L. Cleofas-Sánchez), sanchez@uji.es, sanchez@ieee.org (J.S. Sánchez), vicente.jimenez@uacj.mx (V. García), li_rmvr@hotmail.com (R.M. Valdovinos).

http://dx.doi.org/10.1016/j.eswa.2015.10.001
0957-4174/

López, Fernández, García, Palade, & Herrera, 2013). The methods at the algorithmic level modify the existing learning models for biasing the discrimination process towards the minority class; the data level solutions consist of a preprocessing step to resample the original data set, either by over-sampling the minority class and/or under-sampling the majority class, until the classes are approximately equally represented. In general, the resampling strategies have been the most widely used because they are independent of the classifier and can be easily implemented for any problem (Cao, Zhao, & Zaiane, 2014; García, Sánchez, & Mollineda, 2012).

Over the last years, however, research on this topic has also put the emphasis on studying the effect of imbalance together with other data complexity characteristics such as overlapping, small disjuncts and noisy data (He et al., 2015; López et al., 2013; Napierala, Stefanowski, & Wilk, 2010; Prati, Batista, & Monard, 2004; Stefanowski, 2013). Another critical subject that has attracted increasing interest in the scientific community is how to assess the performance of a classification model in the presence of imbalanced data sets because most common metrics (e.g., accuracy and error rates) strongly depend on the class distribution and assume equal misclassification costs, which may lead to distorted conclusions (He & Garcia, 2009; Menardi & Torelli, 2014).

In addition to the aforementioned questions, research has also placed its attention on evaluating and comparing the performance of competing classifiers (Brown & Mues, 2012; Kwon & Sim, 2013; Liu, Yu, Huang, & An, 2011; López et al., 2013; Seiffert, Khoshgoftaar, Van Hulse, & Folleco, 2014), but many of these works do not take into consideration the impact of different levels of imbalance on the performance of each particular classification model, neither the implications of using this jointly with one resampling technique or another.

While research in class imbalance has mainly concentrated on well-known classifiers such as support vector machines (Akbani, Kwek, & Japkowicz, 2004; Hwang, Park, & Kim, 2011; Liu et al., 2011; Maldonado & López, 2014; Yu et al., 2015), kernel methods (Hong, Chen, & Harris, 2007; Maratea, Petrosino, & Manzo, 2014), k-nearest neighbors (Dubey & Pudi, 2013), decision trees (Kang & Ramamohanarao, 2014) and multiple classifier systems (Díez-Pastor, Rodríguez, García-Osorio, & Kuncheva, 2015; Galar, Fernández, Barrenechea, Bustince, & Herrera, 2012; Krawczyk, Woniak, & Schaefer, 2014; Park & Ghosh, 2014), very few theoretical or empirical analyses have been done so far to thoroughly establish the performance of associative memories when learning from class imbalanced data. Therefore, the present study intends to extend the very preliminary existing works (Cleofas-Sánchez, Camacho-Nieto, Sánchez, & Valdovinos-Rosas, 2014; Cleofas-Sánchez et al., 2013) by increasing the scope and detail at which a type of associative memory networks (the hybrid associative classifier with translation) performs in the framework of class imbalance. We believe that the extensive experimental analysis here carried out will allow to gain a deeper insight into the feasibility and efficiency of these associative memories and help researchers and practitioners to build effective associative learning models and develop techniques based on associative learning to handle the class imbalance problem. To sum up, the purpose of this paper is three-fold:

- To explore the performance of the hybrid associative memory with translation and compare this against other popular classification methods of different nature;
- to investigate how the imbalance ratio affects the performance of these associative memories; and
- to analyze the impact of several resampling strategies on the performance of the associative neural network.

The rest of this paper is organized as follows. Section 2 introduces the bases of the associative memory neural networks and in particular, of the hybrid associative classifier with translation. Section 3 briefly describes four resampling methods to deal with the imbalance problem, which will be further used in the experiments. The experimental set-up and databases are presented in Section 4, while the results and statistical tests are discussed in Section 5. Finally, Section 6 remarks the main conclusions and outlines some avenues for future research.

2. Associative memories

The associative memory is an early type of artificial neural network that takes the form of a matrix $\mathbf{M}$ generated from a finite set of p previously known associations, which is called fundamental set $\{(\mathbf{x}^\mu, \mathbf{y}^\mu) \mid \mu = 1, 2, \ldots, p\}$, where $\mathbf{x}^\mu$ are the fundamental input patterns of dimension n and $\mathbf{y}^\mu$ are the fundamental output patterns of dimension m. Then, x_j^μ and y_j^μ denote the jth component of an input pattern $\mathbf{x}^\mu$ and of an output pattern $\mathbf{y}^\mu$, respectively.

Functionality of the associative memories is accomplished in two phases: learning and recall. The learning process consists of building a matrix $\mathbf{M}$ with a value for each association $(\mathbf{x}^k, \mathbf{y}^k)$. In the recall phase, an output vector $\mathbf{y}$ is obtained from the associative memory; this vector is the most similar to the input vector $\mathbf{x}$. These memories have the capability of storing huge amounts of data that allow to recover the input examples with low computational efforts.

The associative memories can be of two types (Aldape-Pérez et al., 2012): heteroassociative (e.g., linear associator) and autoassociative (e.g., Hopfield network). The heteroassociative memories relates input patterns with output patterns of distinct nature and formats $(\mathbf{x}^\mu \neq \mathbf{y}^\mu)$, while the autoassociative memories are a particular case where $\mathbf{x}^\mu = \mathbf{y}^\mu$ and $n = m$.

Some of the most widely-studied models of associative memories are lernmatrix (Steinbuch, 1961), the linear associator (Anderson, 1972; Kohonen, 1972), the Hopfield network (Hopfield, 1982), the bidirectional associative memory (Kosko, 1987), and the morphological associative memory (Ritter, Sussner, & Diaz-de Leon, 1998). The popularity of these models comes from their capability of storing huge amounts of data that allow to properly recover the most similar patterns given an input vector with low computational efforts, but they have some difficulties to discriminate between classes.

2.1. The hybrid associative classifier with translation

With the aim of extending the applicability of classical associative memories to classification tasks, Santiago (2003) introduced the hybrid associative classifier with translation (HACT), which is based on the learning phase of the linear associator and the recall phase of the lernmatrix. The learning phase of the linear associator comprises two steps:

1. For each association $(\mathbf{x}^\mu, \mathbf{y}^\mu)$, compute the matrix $(\mathbf{y}^\mu) \cdot (\mathbf{x}^\mu)^T$, where $(\mathbf{x}^\mu)^T$ is the transpose input vector.
2. Sum the p matrices to obtain the memory $\mathbf{M} = \sum_{\mu=1}^{p} (\mathbf{y}^\mu) \cdot (\mathbf{x}^\mu)^T$, being $m_{i,j} = \sum_{\mu=1}^{p} (y_i^\mu)(x_j^\mu)^T$ its (i, j)th component.

On the other hand, the recall phase of the lernmatrix consists of finding the class of an input pattern $\mathbf{x}^\mu$, that is, to find the components of the vector $\mathbf{y}^\mu$ associated to $\mathbf{x}^\mu$, whose ith component is calculated according to the following expression:

$$y_i^\mu = \begin{cases} 1 & \text{if } \sum_{j=i}^{n} m_{i,j} x_j^\mu = \max_{h=1}^{p} \left[\sum_{j=i}^{n} m_{h,j} x_j^\mu \right] \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Note that if $\mathbf{x}^\mu$ belongs to class c, this expression leads to an m-dimensional vector with all components equal to zero except for the cth component whose value is equal to 1.

In addition, the HACT model incorporates the translation of the axes to a new origin located at the centroid of the fundamental input patterns. Let $A = \{\mathbf{x}^\mu \mid \mu = 1, 2, \ldots, p\}$ be a set of n-dimensional