Neurocomputing 265 (2017) 105–115

Contents lists available at ScienceDirect

Neurocomputing

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

The Naïve Associative Classifier (NAC): A novel, simple, transparent, and accurate classification model evaluated on financial data

Yenny Villuendas-Rey[a,b], Carmen F. Rey-Benguría[b], Ángel Ferreira-Santiago[c], Oscar Camacho-Nieto[a], Cornelio Yáñez-Márquez[c,*]

[a] Centro de Innovación y Desarrollo Tecnológico en Cómputo del Instituto Politécnico Nacional, Av. Juan de Dios Bátiz s/n, Nueva Industrial Vallejo, Gustavo A. Madero, 07700 Ciudad de México, D.F., México

[b] Center for Pedagogical Studies and Department of Computer Sciences, University of Ciego de Ávila, Carretera a Morón km 9 ½, CP 65100 Ciego de Ávila, Cuba

[c] Centro de Investigación en Computación del Instituto Politécnico Nacional, Avenida Juan de Dios Bátiz esq. Miguel Othón de Mendizábal, Nueva Industrial Vallejo, Gustavo A. Madero, CP 07738, Ciudad de México, D.F., México

ARTICLE INFO

Article history:
Received 29 February 2016
Revised 2 February 2017
Accepted 5 March 2017
Available online 5 June 2017

Keywords:
Supervised classification
Mixed data
Similarity
Bank deposits
Bank Telemarketing
Savings

ABSTRACT

In this paper the Naïve Associative Classifier (NAC), a novel supervised learning model, is presented. Its strengths lie in its simplicity, transparency, transportability and accuracy. The creation, design, implementation and application of the NAC are sustained by an original similarity operator of our own design, the Mixed and Incomplete Data Similarity Operator (MIDSO). One of the key features of MIDSO is its ability to handle missing values as well as mixed numerical and categorical data types. The proposed model was tested by performing numerical experiments using finance-related datasets including credit assignment, bank telemarketing, bankruptcy, and banknote authentication. The experimental results show the adequacy of the model for decision support in those environments, outperforming several state-of-the-art pattern classifiers. Additionally, the advantages and limitations of the NAC, as well as possible improvements, are discussed.

1. Introduction

A current topic among researchers in computer science is the creation, design, and implementation of simple and robust pattern recognition models. Such models have been shown to have several practical applications spanning diverse areas of human activity [1]. There exist four main tasks associated to intelligent algorithms for pattern recognition: recall, regression, classification and clustering. Of these, the former three belong to the supervised learning paradigm, while the latter is an unsupervised learning task [2].

Research in pattern classification has yielded several algorithms with widely varying conceptual bases. Many of them make use of the concept of metric and the properties of metric spaces in order to classify patterns [12]; likewise, several classifiers based on the statistical-probabilistic approach have emerged. These classifiers rely on Bayes' Theorem to make their predictions [3]. In addition, there exist classifiers pertaining to the neural approach, which consists of mathematical models of the brain's neurons [4]. Another of the most well-known and widely-used pattern classifiers today are Support Vector Machines (SVM), which heavily incorporate the concept of optimization in their functioning [5]. Regardless of this great diversity of algorithms, the continuing need for new, simple, and robust pattern classifiers is evident. The goal of this is to arrive at practical applications on various human activities in order to achieve improvements in the search for a sustainable world.

In this paper we introduce one such robust and simple supervised classifier in the form of a novel algorithm, the Naïve Associative Classifier (NAC), which boasts transparency, transportability, and accuracy. The NAC belongs to the associative approach to pattern classification [6], to which several other efficient and powerful classifiers belong, such as the Gamma Classifier [7], the Heaviside Classifier [8], and the Alpha–Beta Bidirectional Memories [9]. Along with the NAC, we introduce a novel similarity operator, MIDSO (Mixed and Incomplete Data Similarity Operator), which is able to handle missing values as well as mixed numerical and categorical data types.

* Corresponding author.
E-mail addresses: yenny.villuendas@gmail.com (Y. Villuendas-Rey), carmenrb@sma.unica.cu (C.F. Rey-Benguría), angel.fgsfds@gmail.com (Á. Ferreira-Santiago), ocamacho@ipn.mx (O. Camacho-Nieto), coryanez@gmail.com, cyanez@cic.ipn.mx (C. Yáñez-Márquez).

http://dx.doi.org/10.1016/j.neucom.2017.03.085
0925-2312/

The proposed classifier was validated by means of numerical experiments using finance-related datasets including credit assignment, bank telemarketing, bankruptcy, and banknote authentication data [10,11]. These experiments demonstrated the ability of the model for decision support in those environments by outperforming several of the state-of-the-art classifiers in this area. The advantages, limitations, and possible directions for the improvement of the NAC are discussed as well.

The rest of the paper is organized as follows. First, Section 2 presents an overview of competing pattern classification approaches against which the proposed model is compared. Next, Section 3 introduces the NAC, the proposed model, along with its similarity operator, MIDSO. The inner workings and computational properties of the model are discussed there. Section 4 follows with a description of the experimental design, including the enumeration of the algorithms that the NAC is compared against, and the datasets used for this purpose. A discussion of the experimental results is offered in Section 5 along with the conclusions of the work and ideas for future research.

2. Related works

In the current literature, the scientific community has a great number of pattern classification techniques at its disposal. However, there exists no such technique that outperforms all other ones over all possible datasets, as stated in the No Free Lunch theorems [12]. In the following subsections some of the most well-known approaches to supervised classification are described.

2.1. Artificial neural networks

Artificial Neural Networks (ANN) represents a learning paradigm based on mathematical models of the human brain's neurons. The seeds of this paradigm were planted in 1943 in the form of a mathematical model of an artificial neuron by McCulloch and Pitts [13]; however, this pioneering artificial neuron model did not possess a learning mechanism. Based on the work of McCulloch and Pitts, in 1957 Rosenblatt proposed the Perceptron [14], which is considered as the first artificial learning machine with the ability to recognize and classify objects. The Perceptron consisted of a set of input neurons which receive the patterns to recognize or classify and an output neuron to return the classification verdict. Multilayer Perceptron (MLP) is a model of neural network in which the nonlinear computing elements are arranged as a feedforward layered structure [15], and whose supervised training stage is carried out, typically, by using the backpropagation algorithm. The bases of this algorithm stem from the error-correcting learning rule [16].

Artificial Neural Networks have been successfully used to solve a wide variety of problems, including medical diagnosis [17,18], financial prediction [19] and image processing [20]. The main disadvantage of ANNs is their high computational complexity [21].

2.2. Naïve Bayes

The Naïve Bayes (NB) algorithm is based on Bayes' theorem. It uses conditional probability distributions for instance classification. Given an instance x, NB assumes that the attributes are independent from each other given the class of the instance. The NB algorithm modifies the way the probabilities associated to each feature are computed depending on the kind of data at hand: for categorical data, this probability is calculated using a discrete probability distribution function, while for numerical features a continuous probability distribution function is used. Similarly, the NB classifier excludes the objects with missing values from the frequency computation and probability estimation processes during training. For classification, the features with missing values are ignored. Although simple, this classifier exhibits a performance comparable to other much more complex classifiers [22].

2.3. K Nearest Neighbors

The k Nearest Neighbors (k-NN) model is part of the family of classification techniques called *lazy learners*. The training of this kind of algorithms is limited to storing in memory the patterns of the training set [23]. The k-NN classifier is perhaps the most popular lazy learning algorithm, due to its simplicity as well as for often achieving highly accurate results. The k-NN classifier is based on the idea that individuals from a population often share some similar properties and certain characteristics with the individuals around them, when placing them in a feature space. Thus, the classification of a pattern is carried out using the k closest instances of the training set based on a dissimilarity (or similarity) measure. Dissimilarity-based classifiers (such as the k-NN classifier) treat hybrid data separately from the classifier itself, by delegating this responsibility to the dissimilarity function used. Similarly, missing values may or may not be accounted for by the dissimilarity function [23]. Despite its simplicity, the k-NN classifier remains as a staple in the state of the art, and is currently used for a variety of classification problems [24–26].

2.4. Associative models

The associative approach for pattern recall and classification has an older than five-decade history [27], and associative models remain as a current research topic [28]. Recently, the Gamma associative classifier [29] was proposed to address supervised learning tasks, including regression. This classifier only handles numerical data without missing values. The Gamma classifier has been applied to a wide range of problems, among which medical applications stand out [30]; in addition, this classifier has also been extended to deal with Data Streams [31].

2.5. Decision trees

Decision Trees (DT) are supervised classification algorithms which create a tree structure during the training phase. The C4.5 algorithm [32] is one of the classical DT methods, along with the CART algorithm [33]. The C4.5 tree has attracted a huge amount of attention due to its ability to directly handle mixed and incomplete data types, and it is also considered among the top ten algorithms for solving data mining tasks [34]. It builds DTs from a dataset using the concept of entropy, and includes during the tree induction phase separates mechanisms for handling numerical and categorical features: in the case of the former, a binary threshold is chosen, giving birth to two branches of the tree; in the case of the latter, one branch is generated for each possible value of the categorical attribute. C4.5 chooses, at each node, the attribute of the pattern that most effectively splits its set of samples into subsets. The criterion of splitting is the difference in entropy (normalized information gain). The attribute with the highest normalized information gain value is chosen. In the case of the C4.5 tree, incomplete or missing values are not used for the information gain computation during the training phase. For classification, the missing values are ignored and the tree is traversed according to the values of the rest of the attributes.

DTs have been widely used to perform classification tasks in several areas, such as medicine [35,36] and geospatial processing [37,38].