Pattern Recognition Letters 89 (2017) 39–45

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Enhancing classification performance using attribute-oriented functionally expanded data

João Roberto Bertini Junior[a,*], Maria do Carmo Nicoletti[b,c]

[a] School of Technology, University of Campinas, Rua Paschoal Marmo, 1888, Jd. Nova Itália, Limeira, SP 13484-332, Brazil
[b] Computer Science Department, Federal University of São Carlos, Rod. Washington Luís Km 235, São Carlos, SP, 13565-905, PO Box 676, Brazil
[c] FACCAMP, R. Guatemala 167, Campo Limpo Paulista, SP, 13231-230, Brazil

ARTICLE INFO

Article history:
Received 3 August 2016
Available online 8 February 2017

MSC:
41A05
41A10
65D05
65D17

Keywords:
Improving classification performance
Functional expansion
Genetic algorithm

ABSTRACT

There are many data pre-processing techniques that aim at enhancing the quality of classifiers induced by machine learning algorithms. Functional expansions (FE) are one of such techniques, which has been originally proposed to aid neural network based classification. Despite of being successfully employed, works reported in the literature use the same functional expansion, with the same expansion size (ES), applied to each attribute that describes the training data. In this paper it is argued that FE and ES can be attribute-oriented and, by choosing the most suitable FE–SE pair for each attribute, the input data representation improves and, as a consequence, learning algorithms can induce better classifiers. This paper proposes, as a pre-processing step to learning algorithms, a method that uses a genetic algorithm for searching for a suitable FE–SE pair for each data attribute, aiming at producing functionally extended training data. Experimental results using functionally expanded training sets, considering four classification algorithms, KNN, CART, SVM and RBNN, have confirmed the hypothesis; the proposed method for searching for FE–SE pairs through an attribute-oriented fashion has yielded statistically significant better results than learning from the original data or by considering the result from the best FE–SE pair for all attributes.

1. Introduction

Functional Link Artificial Neural Networks (*FLANNs*) have been proposed by Pao [1], with the intent to overcome the complexities associated with conventional training processes of multi-layer Neural Networks (NN). The involved complexities are mostly related to the definition, before the training starts, of the network's topology i.e., the number of layers and the number of neurons per layer the network should have; this is, usually, approached as a trial-and-error task. FLANNs are feed-forward neural networks with a single layer which applies a transformation on the input data, by extending the set of attributes that describe the data. The transformation enlarges the attribute space by generating an improved data representation. Thus, the network hidden layer may implement a functional expansion, a tensor representation of the input data or a combination of both.

FLANNs, in spite of being a single-layer NN, have proved to be effective for dealing with non-linearly separable classification tasks [2]. They have been successfully used in a diverse range of applications, such as for solving the problem of channel equalization in digital communication systems as in [3], the prediction of machinery noise in opencast mines, as discussed in [4], the forecasting of stock markets, as proposed in [5] and image filtering as in [6], among many others.

In the literature also, the use of FLANNs combined with another technique is very common, such as in [7], where the functional neural fuzzy network (FNFN), for classification applications, is proposed. FNFN uses a FLANN for representing the consequent of the induced fuzzy rules i.e., a nonlinear combination of the FNFNs input variables. As the authors comment, their model can be defined as a single-layer NN able to define complex decision regions, by using FLANNs for inducing nonlinear decision boundaries. Hu [8] proposes to use a GA-based algorithm to train two FLNNs, for representing upper and lower bounds, respectively, for solving interval regression problems. A GA-based algorithm was used for searching for a convenient set of weights and other parameter values of both FLNNs. In [9] results are reported showing an improvement in the generalization ability of a FLANN, by applying an entropy-based method to perform feature selection, followed by the training of the FLANN, using differential evolution. The work described

* Corresponding author.
E-mail address: bertini@ft.unicamp.br (J.R. Bertini Junior).

http://dx.doi.org/10.1016/j.patrec.2017.02.003

in [10] proposes an artificial bee colony optimization process for training FLANNs, aiming at solving classification problems.

Authors in [11–13] adopted the idea of functionally expanded inputs, by employing several functional expansions to grow constructive neural networks (CoNNs), in an attempt to enhance their performance. Results from experiments are evidence that, most of the time, by enlarging the input data using functional expansions, the performance of the induced CoNNs has been enhanced. In their work, however, the same functional expansion, with the same expansion size (i.e., the same number of new attributes, per original attribute, included in the training set) was applied to all attributes describing the data.

Taking into account the good results obtained when inducing NNs and CoNNs using functionally expanded inputs, this paper reports a research on the impact on performance, of classifiers induced using four different learning algorithms, KNN, CART, SVM and RBNN, having as input functional expanded data. Specifically, the work tests and verifies the following hypothesis: is there a set of FE–ES pairs, each pair applied to a particular attribute, that enhances the learning performance, when compared to the use of a single pair FE–ES, applied to all attributes? Aiming at addressing such hypothesis, a GA-based algorithm has been proposed for searching for such attribute-oriented pairs.

The remainder of this paper is organized as follows: Section 2 is organized in two subsections. The first introduces the main aspects related to the attribute oriented FE–ES pair proposal and, the second, focuses on the GA-based algorithm used for searching for such pairs. The proposal and the computational system that implements it are referred to as GAFE for *Genetic Algorithm for Functional Expansion*. Initially, Section 3 briefly describes the experimental setup. Next, it presents and discusses the results of the conducted learning experiments using 15 data sets from the UCI Repository [14], aiming at establishing the effectiveness and usability of GAFE for enhancing the performance of classifiers, via functional expansion of the training set. Finally, Section 4 draws the conclusions of the work. The seven functional expansions (FEs) considered in the experiments are presented in Appendix.

2. Attribute-oriented functional expansions and searching for them via genetic algorithms – the GAFE system

This section is organized in two subsections, the first describes the attribute-oriented FE–ES pair proposal, for expanding training sets in supervised learning tasks, aiming at inducing better classifiers and, the second, presents the main details of the GA-based approach for searching for suitable FE–ES attribute-oriented pairs, implemented as the computational system GAFE.

2.1. Attribute-oriented functional expansions

The attribute-oriented functional expansion proposed in this paper is introduced next, using a simple example, where the training set is described by $p = 3$ numerical attributes, represented by variables x_1, x_2 and x_3. Fig. 1 shows a possible configuration of how FEs are commonly used (referred to as conventional), considering the trivial case of one particular FE (the Power Series (PW)) having a fixed expansion size (ES) of three. Notice that every attribute is subjected to the same transformation which, in this example, is the Power Series function with expansion size of $N = 3$, leading to a $p \times N$ input data representation.

In spite of the effectiveness of the scheme shown in Fig. 1, it is argued in this paper that, by applying a particular attribute-oriented transformation to each attribute, represented by a pair FE-ES, better classifiers could be induced due to the following consequences: (1) compared to the way FE are traditionally used, the search space is substantially enhanced, therefore it also enhances

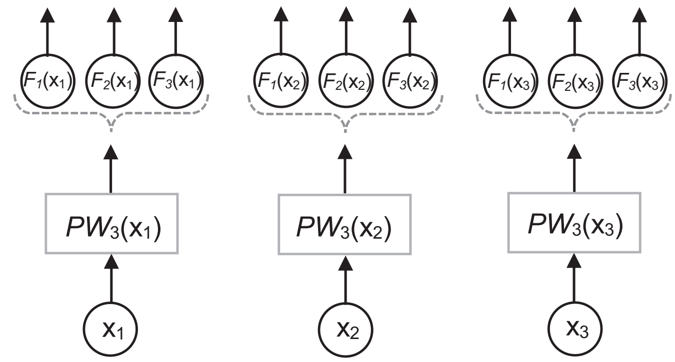

Fig. 1. An example of how FEs are commonly used. The Power Series (PW) function with expansion size of N = 3 is applied to the three attributes, x_1, x_2 and x_3 of each training instance in the original training set, expanding it to a nine dimensional instance.

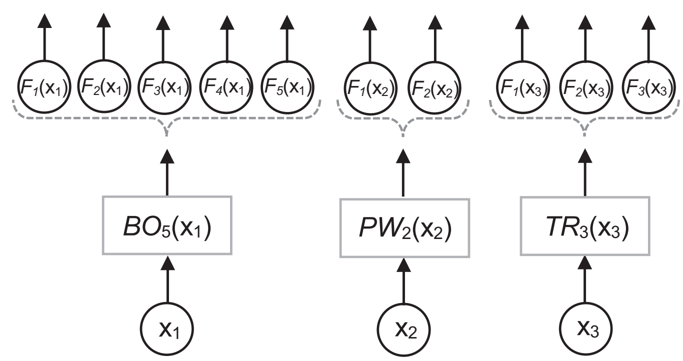

Fig. 2. Applying an attribute-oriented pair FE–ES to each attribute, x_1, x_2 and x_3.

the chance of finding a better representation of data; and (2) each attribute has its own intrinsic characteristic, which may be best represented by a particular pair of FE-ES. A representation of data that promotes classification may arise by assigning a particular set of pairs of FE-ES to the set of attributes, which not only provides a better representation of each attribute but also favors the relationship among them.

Fig. 2 sketches an example of this idea, which basically allows different pairs, FE–ES, be applied to each attribute. In the figure, x_1 was expanded by the Boubaker function (BO), with expansion size of $N = 5$, x_2 by the Power Series (PW), with $N = 2$ and x_3 by the Trigonometric function (TR), with $N = 3$.

Although any possible configuration can be implemented beforehand, as happens in the conventional approach, the problem lies in defining a suitable FE–ES pair that yields better results, when compared to those produced by applying the same FE–ES pair to all attributes. Note that, in the conventional case, a simple model selection can be carried out to define the suitable pair to be applied to all attributes.

Let N_F be the number of function expansions considered for model selection and let N be the maximum integer value considered for expansion size. A conventional model selection process involves testing $N_F \times N$ models. In the proposed approach, as attributes can go through different transformations, for a data set with p attributes there will be $p \times N_F \times N$ possible configurations. As p gets higher, however, it becomes prohibitive to perform a brute force model selection; instead, a GA-based searching method to find a good enough (maybe optimal) set of FE–ES pairs, each suitable to one of the p attributes involved, is proposed in this work and described next.