ARTICLE IN PRESS

ISA Transactions xxx (2018) 1–8

Contents lists available at ScienceDirect

ISA Transactions

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

Practice article

A multiple kernel classification approach based on a Quadratic Successive Geometric Segmentation methodology with a fault diagnosis case

Leonardo M. Honório [a,*], Daniele A. Barbosa [a], Edimar J. Oliveira [a], Paulo A. Nepomuceno Garcia [a], Murillo F. Santos [b]

[a] *Department of Energy, Federal University of Juiz de Fora, Brazil*
[b] *Federal Institute of Technology, Cataguases, Brazil*

ARTICLE INFO

Article history:
Received 19 June 2017
Revised 25 November 2017
Accepted 1 January 2018
Available online XXX

Keywords:
Successive Geometric Segmentation
Support Vector Machine
Quadratic classification
Multiple kernel classifier
Transformer fault diagnosis

ABSTRACT

This work presents a new approach for solving classification and learning problems. The Successive Geometric Segmentation technique is applied to encapsulate large datasets by using a series of Oriented Bounding Hyper Box (OBHBs). Each OBHB is obtained through linear separation analysis and each one represents a specific region in a pattern's solution space. Also, each OBHB can be seen as a data abstraction layer and be considered as an individual Kernel. Thus, it is possible by applying a quadratic discriminant function, to assemble a set of nonlinear surfaces separating each desirable pattern. This approach allows working with large datasets using high speed linear analysis tools and yet providing a very accurate non-linear classifier as final result. The methodology was tested using the UCI Machine Learning repository and a Power Transformer Fault Diagnosis real scenario problem. The results were compared with different approaches provided by literature and, finally, the potential and further applications of the methodology were also discussed.

1. Introduction

Machine learning techniques, such as Artificial Neural Networks (ANN) [1–3], Deep Learning (DL) [4,5], Clustering [6,7], Support Vector Machine (SVM) [8] and others have been successfully used in a wide range of applications [9–12]. Regarding ANN, although the literature shows several different proposals, Multilayer Perceptron (MLP) is still one of the most popular approaches [1,13,14]. The MLP presents a well-known and simple procedure; once the appropriate architecture is defined, an optimization process finds a set of weights that separates desirable patterns. In spite of all ANN researches, the definition of the best architecture is not always trivial and, in practice, it is often obtained through heuristic tests, thus remaining an important research object [15,16].

Considering SVM and other kernel-based techniques, in general, these approaches tend to achieve better generalization than traditional ANN techniques for most classification problems. However, two concerns draw attention to these methods; A) the kernel definition [17,18]; B) the scalability due to the fact that the most difficult classification scenarios require the resolution of a quadratic programming problem in order to find the separation hyperplane, leading to a $\mathcal{O}(n^2)$ space complexity, where n is the data size [19] and; C) the unbalanced data between classes [20].

In addition to the features related to each approach, both RNA-based and SVM-based techniques are still subject to traditional nonlinear optimization problems, such as; sensitivity to the initialization point, local minimums, numerical instability when the Hessian is near singularity, and the solution close to the security barrier. These problems are well known in any nonlinear optimization scenario [21] and are still the focus of several studies in the machine learning field [22].

Trying to find a new approach to overcome some of the problems mentioned above, the Successive Geometric Segmentation (SGS) technique was first presented in Ref. [23]. The SGS begins representing each pattern by an envelope defined by an OBHB [24] which, ideally, is the minimum box capable of encapsulating a given dataset. If one envelope is linearly separable from other, it is possible to define at least one hyperplane associated with this classification. However, when an envelope is not linearly separable, its data can be divided into at least two subsets, with smaller and more accurate

* Corresponding author.
E-mail address: leonardo.honorio@ufjf.edu.br (L.M. Honório).

https://doi.org/10.1016/j.isatra.2018.01.013
0019-0578/

new envelopes being generated. This process repeats until the data of a given class is linearly separated from the other classes or until a stop criterion is reached. Moreover, if all resulting envelopes from one class are linearly separable from all others from a different class it is possible to automatically assemble an ANN that maps a classifier. In addition, due to its well-defined geometric approach, it identifies specific regions on the solution space providing several advantages, such as; it is not necessary to solve traditional optimization processes, and it is possible for a specialist to analyze the generated network, since it has a well-defined geometric meaning. As a result, the original SGS performed well [23] and improved versions were successfully applied in real scenarios as presented in Refs. [25,26].

Despite the good results, SGS is not satisfactory in situations where there is an overlapping boundary region between two classes. In this case the increased number of separation planes reduces the algorithm performance.

To overcome these problems, the present work introduces the Quadratic Successive Geometric Segmentation (QSGS) which is a new approach based on the SGS plus a Quadratic discriminant analysis. The first set of significant modifications is related to the new envelope breaking and definition. When a collision is detected between two OBHBs, the breaking process uses a Kernel Density Estimation (KDE) function to choose optimum breaking points to generate fewer and more specialized OBHBs. Thus, one modification is that, since one of the most important characteristics in any learning and clustering algorithm is the correct removal of points with low relevance to the original sample [27], the encapsulated box is designed to represent just the most significant data. In this way, it is possible to ignore entries such as noise, sensor errors or just non-standard samples, improving the estimation of the correct data distribution over the solution space. This strategy creates a very effective density-based data clustering [28] specially in cases with few data entries, where any instance can significantly change the spacial distribution. This feature is accomplished by using a Kernel Density Estimation (KDE) [29,30] technique that excludes individuals with pertinences below a given threshold.

The second modification deals with the collision test methodology, which was originally performed through the Separating Axis Theorem (SAT) by testing all axes of each involved box [24]. This technique usually identifies more than one separation axis per test, leading to an optimization problem where the objective is to identify the set of axes that best represents the desired classifier. As any combinatorial problem, this step presents a high computational cost. To avoid this situation, a technique based on quadratic discriminant analysis will be applied to find the best separation surface between two or more classes. To accomplish that, each envelope is now approximated by a normal multivariate distribution. It uses the OBHBs center and covariance matrix to evaluate a Quadratic Discriminant Function (QDF) for each envelope [31]. The processing of all those QDFs generates a Gaussian separation surface which is able to identify each desirable class.

There are several advantages in the proposed modification; a) while approaches such as ANN, DL and SVM are dependent, besides the parameters, of a given topology, the QSGS finds its own topology by organizing and clustering the provided data. b) although the final solution is a nonlinear separation surface, the initial strategy is still based on linear segmentation, meaning a fast and precise classification approach. c) by using the QDF alternative in order to generate the separation surface it is possible to define likelihood thresholds, indicating that entries with low pertinence will not be classified, ensuring lower false-positive cases. d) one major drawback of the original SGS was the computational performance to find the ANN assembling, which was $\mathcal{O}(n^2)$. By considering each OBHB as a specialized kernel that generates a separation surface, the complexity is now proportional to the number of final OBHBs. As the new breaking process generates fewer boxes, the computational time is severed decreased.

To demonstrate the ideas above, this work is organized as follows: Section 2 briefly shows the modified breaking algorithm responsible for identify clusters in the solution space; Section 3 presents the proposal of the new QSGS method, which creates a set of non-linear discriminant surfaces. To better understand this new approach, a tutorial case using the well-known banana dataset is adopted to illustrate the theoretical sections. To evaluate the QSGS performance two groups of tests are used in section 4. First, at subsection 4.1 the well know UCI dataset for testing the performance of machine learning algorithms is used. The second group is shown in subsection 4.2 and focus on the real and important problem of power transformers fault diagnosis, where the proposed algorithm is also tested by using the IEC TC10 open access database, which is based on real acquisitions. The conclusions and future possibilities are discussed at section 5.

2. Modified Successive Geometric Segmentation methodology

The main idea behind the original SGS methodology is to find a set of clusters in the solution space capable of representing the original data. These clusters work as a data abstraction layer and can be used as classifiers. Moreover, each cluster is an OBHB and represents a subset from a given class and is separable from all other clusters of different classes by a series of independent linear surfaces through a MLP as shown in Ref. [32].

The methodology presents three drawbacks that compromises its performance: a) the geometrical segmentation, if not properly done, could lead to an overspecialization, b) noisy data close to boundaries can change the OBHBs geometrical displacement and, c) the approach of using a series of linear separation for classification tends to lose abstraction in regions that lays far away from the separation surfaces or with low data density.

To handle those problems, this work proposes, respectively: a) to change the way of performing the segmentation by using a statistical KDE analysis, which is more effective in breaking and identifying the stop criterion, b) to identify and discard outliers, also by a statistical KDE analysis and c) to use the final OBHBs to generate a nonlinear separation surface by using quadratic discriminants.

To better illustrate the methodology, the high level procedures used in the algorithm is shown in Fig. 1, and a tutorial case will be demonstrated using a binary classification problem based on the banana dataset [33].

Considering a generic supervised classification problem, each pattern $\mathcal{P}^i$ labels a set $X^i = \{(x^i_{1,1}, \dots, x^i_{1,m}); \quad \dots \quad ; (x^i_{n,1}, \dots, x^i_{n,m})\}^T \in \mathbb{R}^{n \times m}$, where n and m are the number of elements and attributes/features respectively. The objective of SGS is to break X^i in smaller subsets $X^i = \{X^{i,1} \cup \cdots \cup X^{i,k}\}$ where each $X^{i,j} \in \mathbb{R}^{n_j \times m}$ will

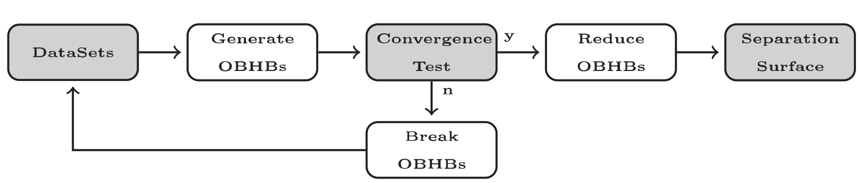

Fig. 1. General algorithm.