Neural Networks 23 (2010) 887–891

Contents lists available at ScienceDirect

Neural Networks

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

A new local–global approach for classification

R.T. Peres, C.E. Pedreira *

COPPE-PEE – Engineering Graduate Program and School of Medicine, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

ARTICLE INFO

Article history:
Received 22 May 2009
Received in revised form 9 January 2010
Accepted 26 April 2010

Keywords:
Bayes classifier
Classification
Local–global
LVQ
SVM
Vector quantization
Pattern recognition
Prototype

ABSTRACT

In this paper, we propose a new local–global pattern classification scheme that combines supervised and unsupervised approaches, taking advantage of both, local and global environments. We understand as global methods the ones concerned with the aim of constructing a model for the whole problem space using the totality of the available observations. Local methods focus into sub regions of the space, possibly using an appropriately selected subset of the sample. In the proposed method, the sample is first divided in local cells by using a Vector Quantization unsupervised algorithm, the LBG (Linde–Buzo–Gray). In a second stage, the generated assemblage of much easier problems is locally solved with a scheme inspired by Bayes' rule. Four classification methods were implemented for comparison purposes with the proposed scheme: Learning Vector Quantization (LVQ); Feedforward Neural Networks; Support Vector Machine (SVM) and *k*-Nearest Neighbors. These four methods and the proposed scheme were implemented in eleven datasets, two controlled experiments, plus nine public available datasets from the UCI repository. The proposed method has shown a quite competitive performance when compared to these classical and largely used classifiers. Our method is simple concerning understanding and implementation and is based on very intuitive concepts.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

A global approach to classification may be pursued by trying to fit a probability density function, or a mixture of distributions, on the observed data. If this density can be estimated, one may end up with a multimodal distribution, each mode possibly corresponding to a class, and after estimating the *a priori* class probabilities, one could apply Bayes' rule (Duda, Hart, & Stork, 2001) to solve the problem in an optimal manner. Unfortunately, as it is well known, for most of real applications it is in general quite hard, or even impossible, to estimate a global probability density function, especially for high dimensionality spaces (Duda et al., 2001; Silverman, 1986). Furthermore, a pure global approach assumes that data is engendered by a phenomenon governed by a global fundamental law and does not take advantage of possible local generative structures. Accordingly, global models are constructed from all the available observations aiming to represent the entire problem space.

In a local classification approach, the aim is instead of constructing a global generative model from all observations, to build up local classification schemes, possibly using just a subset of the sample. The focus moves into partitions of the original problem.

Successful examples of local–global classification approaches are the kernel methods (Shawe-Taylor & Cristianini, 2004), for instance the Support Vector Machine (SVM) procedure (Vapnik, 1998). In SVM, a subset of the observations, the Support Vectors, is employed to determine a somehow optimal separating hyperplane. The local structure arises intrinsically through the kernels; the global framework comes by means of the generated hyperplanes. In Huang, Yang, King, and Lyu (2008), a local–global large margin classifier is proposed. A time series local–global approach can be found in Fariñas, Pedreira, and Medeiros (2004).

A different view comes from Vector Quantization (VQ) (Linde, Buzo, & Gray, 1980; Gray & Neuhoff, 1998) where the key idea is to use the whole sample (global) in an unsupervised environment to generate a partition that may enhance local structures. This results in a quantized approximation of the distribution, using a finite number of prototype vectors. In a supervised context, VQ can be naturally extended to Learning Vector Quantization (LVQ) (Kohonen, 2001), where prototypes location results from an update procedure based on the training dataset. Once the prototypes are set, one may associate one or more of those with each class, and classify an observation by using the nearest-neighbor rule (Duda et al., 2001). Some procedures were proposed to execute LVQ with an appropriately chosen subset of the sample (Pedreira, 2006; Peres & Pedreira, 2009).

Applications of local–global models may be found in diverse areas such as bioinformatics (Kasabov, 2007) or remote sensing imaging (Blanzieri & Melgani, 2008). Here, we propose a new local–global classification scheme that combines supervised and

* Corresponding address: UFRJ, COPPE-PEE, Av. Horácio Macedo, Prédio do CT, Bloco H, 3o andar, Ilha do Fundão, CEP 21941-914, Rio de Janeiro, Brazil. Tel.: +55 21 2562 8627.
E-mail addresses: rperes@lps.ufrj.br (R.T. Peres), pedreira@ufrj.br (C.E. Pedreira).

0893-6080/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2010.04.010

unsupervised approaches. We use a VQ unsupervised algorithm to divide the sample in local cells. In this first stage, the whole sample is used and some information on the data distribution is captured in a global mode. In this way, an assemblage of much easier problems is generated and locally solved, in a second stage, with a scheme inspired by Bayes' rule.

2. Methodology and datasets

The proposed approach belongs to the divide-and-conquer family. We create a set of sub-classifiers, applied in sub regions of the space, intending to make the classification task easier in a local level.

Let X be a sample of size m comprised by a set of observations $\{x_1, x_2, \ldots, x_m; x_i \in \Re^n, \forall i = 1, \ldots, m\}$. Let us consider a dichotomous classification environment where each observation x_i is associated to one, out of two, possible classes, C_1 or C_2. We denote $y(x_i)$ the label of observation x_i, being $y(x_i) = 1$ if x_i belongs to C_1, and $y(x_i) = 2$ if x_i belongs to C_2.

The proposed methodology is implemented in two stages. We first partition the observation space into cells using an unsupervised procedure. In a second stage, a supervised classification scheme is applied (locally) in each of the previously generated cells.

The first stage is done by the way of the LBG (Linde-Buzo-Gray) algorithm (Linde et al., 1980). Other algorithms, like for instance the k-means, may produce similar results. The LBG algorithm is an iterative procedure that alternatively conducts two actions, update and partition. It is started by setting the centroid of the sample as the code-vector in the initial iteration. It follows (partition) by adding (and subtracting) a perturbation and consequently splitting this code-vector into two. At each iteration, after completing the partition process, the update operation is done by redefining the cells, allocating each observation to its correspondent closest code-vector (in the sense of minimum squared Euclidean distortion), and updating the current code-vectors to the centroids of its associated observations. This process continues until a pre-established number of cells, let's say r, is reached. In this way, one ends up with r code-vectors $\{p_k \in \Re^n, k = 1, \ldots, r\}$ defining r cells $S_1, \ldots, S_r$.

The second stage consists of applying a classification procedure for each of the generated cells. A trivial situation occurs when a cell is homogeneous, i.e. all (training) observations in this cell belong to the same class. In this case, any testing observations attracted to this cell will be associated to the cell label. Otherwise, if the cell is heterogeneous, we introduce a local scheme inspired by the Bayes classifier (Duda et al., 2001).

Without loss of generality, we focus on an arbitrary testing observation x in cell S_k. Let us consider the two subsets ς_k^1 and ς_k^2 of S_k containing observations of classes C_1 and C_2 respectively:

$\varsigma_k^1 \equiv \{x_i \in S_k | y(x_i) = 1\}$ and $\varsigma_k^2 \equiv \{x_i \in S_k | y(x_i) = 2\}$. The relative frequencies f_k^1 and f_k^2 of classes C_1 and C_2 may now be calculated as

$$f_k^1 = \frac{\#\varsigma_k^1}{\#S_k} \quad \text{and} \quad f_k^2 = \frac{\#\varsigma_k^2}{\#S_k},$$

where # represents cardinality. Note that frequencies f_k^1 and f_k^2 are estimators of the *a priori* probability of classes C_1 and C_2 in cell S_k. The *a priori* probability ratio in S_k may now be estimated as:

$$\pi_k \equiv \frac{f_k^1}{f_k^2}.$$

We may next view the inverse of the distances of an observation x to the class means, as estimatives of the likelihood at x. So, we define

$$\hat{L}_x \equiv \left(d(x, m_k^1)\right)^{-1} / \left(d(x, m_k^2)\right)^{-1} \quad (1)$$

as an estimative for the likelihood ratio $L \equiv p(x|C_1)/p(x|C_2)$, where m_k^1 and m_k^2 are respectively the means of the observations with labels 1 and 2 in cell S_k.

We conclude by proposing the following decision rule for a testing observation x in cell k:

$$\begin{cases} x \to C_1 & \text{if } \hat{L}_x \geq (\pi_k)^{-1} \\ x \to C_2 & \text{otherwise.} \end{cases} \quad (2)$$

The proposed algorithm may be summarized as follows:

1. Segment the (training) sample space into r cells $S_1, \ldots, S_r$ by using the LBG algorithm.
2. Calculate the frequency ratio π and the class means m^1 and m^2 for all heterogeneous cells.
3. If an observation (in the testing set) lies in a homogeneous cell, attribute to this observation the label of this cell.
4. Otherwise, if observation x (in the testing set) lies in a heterogeneous cell, calculate $\hat{L}_x$ (as defined in (1)) and attribute a class in accordance with rule (2).

Four classification methods were implemented for comparison purposes with the proposed scheme: (i) LVQ; (ii) Feedforward Neural Networks (NN), (iii) Support Vector Machine (SVM) and (iv) k-Nearest Neighbors. The NNs were trained with Bayesian Regularization, with 10 initial neurons in the hidden layer and logistic activation function in both, the hidden and the output layers. For SVM we used radial basis function kernel.

2.1. On the datasets

In this sub-section, we briefly describe the datasets used to benchmark the proposed method performance. Besides the proposed algorithm, the four classification procedures, described in the previous sub-section, were implemented in eleven datasets, two controlled experiments, plus nine public available datasets from the UCI[1] repository, namely: Waveform, Letter-B, Statlog, Heart Diseases Diagnosis, Breast Cancer, Ionosphere, Pima Indians, Glass and Lung Cancer.

Experiments 1 and 2 are synthetic data. Experiment 1 consists of two classes divided by a cosine function. We generated 1030 observations of class C_1 and 1027 observations of class C_2 for in-sample, and 1060 observations labeled C_1 and 1041 observations labeled C_2 for out-of-sample. For experiment 2 two classes were generated through a circle and a roll with coincident centers (without superposition). In-sample: 123 observations of class C_1 and 2611 observations of class C_2. Out-of-sample: 127 observations of class C_1 and 2646 observations of class C_2.

The original dataset for experiment 3 had 3 classes of waveforms; here, we tested class 1 against the other two. This dataset is composed of 5000 observations, with 40 input features (3000 used for in-sample and 2000 for the out-of-sample phase).

Experiment 4 concerns letter recognition. The objective is to identify the 26 capital letters in the English alphabet. We set letter B as one class against the other 25 letters. The dataset consists of 20 000 observations with 16 input features, 10 600 of those used for in-sample and 9400 for out-of-sample testing.

The data for experiment 5 comes from the Statlog - landsat satellite. This dataset consists of the multi-spectral values of pixels in 3×3 neighbourhoods in a satellite image. The aim is to classify images, associated with the central pixel in each neighbourhood, given the multi-spectral values. We tested class 1 against the others. There are 6435 observations with 36 input features, 4435 were used for in-sample phase leaving 2000 for the out-of-sample testing.

The dataset for experiment 6, related to the diagnosis of coronary artery disease, was formed by an assemblage of four data sets as in Pedreira, Macrini, and Costa (2005). Each of these four data

[1] http://www.ics.uci.edu/~mlearn/MLSummary.html.

Download English Version:

https://daneshyari.com/en/article/404399

Download Persian Version:

https://daneshyari.com/article/404399

Daneshyari.com