Neurocomputing 174 (2016) 154–167

Contents lists available at ScienceDirect

Neurocomputing

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

Inductive bias for semi-supervised extreme learning machine

Federica Bisio [a,*], Sergio Decherchi [b], Paolo Gastaldo [a], Rodolfo Zunino [a]

[a] Dept. of Electric, Electronic and Telecommunications Engineering and Naval Architecture (DITEN), University of Genoa, Genova, Italy
[b] Istituto Italiano di Tecnologia, IIT, Morego, Genova, Italy

ARTICLE INFO

Article history:
Received 22 September 2014
Received in revised form
2 February 2015
Accepted 7 April 2015
Available online 13 August 2015

Keywords:
Extreme Learning Machine
Semi-supervised learning
Inductive bias

ABSTRACT

This research shows that inductive bias provides a valuable method to effectively tackle semi-supervised classification problems. In the learning theory framework, inductive bias provides a powerful tool, and allows one to shape the generalization properties of a learning machine. The paper formalizes semi-supervised learning as a supervised learning problem biased by an unsupervised reference solution. The resulting semi-supervised classification framework can apply any clustering algorithm to derive the reference function, thus ensuring maximum flexibility. In this context, the paper derives the biased version of Extreme Learning Machine (br-ELM). The experimental session involves several real world problems and proves the reliability of the semi-supervised classification scheme.
© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Inductive bias is of fundamental importance in learning theory, as it influences heavily the generalization ability of a learning system [1]. From a mathematical point of view, the inductive bias can be formalized as the set of assumptions that determine the choice of a particular class of functions to support the learning process. Therefore, it represents a powerful tool to embed the prior knowledge on the applicative problem at hand.

In literature, modifications to the original Extreme Learning Machine (ELM) [2] scheme have been proposed [3,4]. This paper addresses the advantages and the issues of introducing an inductive bias in the ELM when semi-supervised classifications problems are being tackled. In semi-supervised classification, one exploits both unlabeled and labeled data to learn a classification rule/function empirically [5]; the semi supervised approach should improve over the classification rule that is learnt by only using labeled data. The interest in semi-supervised learning has increased recently, especially because application domains exist (e.g., text mining, natural language processing, image and video retrieval, and bioinformatics) [6,7], in which large datasets are available but labeling is difficult, expensive, or time consuming.

Biased regularization provides a viable approach to implement an inductive bias in a kernel machine, as confirmed by the generalized 'Representer Theorem' [8]. Biased regularization of Support Vector Machines (SVMs) has been adopted in [9] for a personalized handwritten system and in [10] for a malware detection system. Ivanov-like biased regularization was adopted in [11] to shrink the generalization error bounds for SVMs under the so called cluster hypothesis (see [12] for a very in depth analysis). A similar result was obtained in the PAC Bayesian framework [13].

The research presented here shows that semi-supervised learning can benefit from biased regularization, too. First, a novel, general biased-regularization scheme is introduced that encompasses the biased version of ELM. Then, the paper proposes a semi-supervised learning model, which is based on that biased-regularization scheme and follows a two-step procedure. In the first step, an unsupervised clustering of the whole dataset (including both labeled and unlabeled data) obtains a reference solution; in the second step, the clustering outcomes drive the learning process in a biased ELM (br-ELM) to acquire the class information provided by labels. The ultimate result is that the overall learned function exploits both labeled and unlabeled data. The integrated framework applies to both linear and nonlinear data distributions: in the former case, one works under a cluster assumption on data; in the latter case, one works under a manifold hypothesis [14]. As a consequence, for a successful semi-supervised learning, unlabeled data are assumed to carry some intrinsic geometric structure, e.g., in the ideal case, a low-dimensional, non- linear manifold.

The proposed biased semi-supervised approach exhibits several features, such as modularity in the procedure that generates a biased solution, convexity of the cost function, predictable complexity, and out-of-sample extension. Moreover, the paper shows that the proposed framework may allow one to benefit from tight generalization bounds. Such result is noteworthy in that a learning machine featuring tight generalization bounds can provide powerful options to attain reliable model selection [11,15–17]; that property may prove especially useful in those cases when

* Corresponding author.
E-mail addresses: federica.bisio@edu.unige.it (F. Bisio), sergio.decherchi@iit.it (S. Decherchi), paolo.gastaldo@unige.it (P. Gastaldo), rodolfo.zunino@unige.it (R. Zunino).

http://dx.doi.org/10.1016/j.neucom.2015.04.104
0925-2312/© 2015 Elsevier B.V. All rights reserved.

strategies such as cross validation are difficult to be applied, e.g. in the small sample case (i.e., less than 50 labeled patterns). Indeed, this seems to represent an interesting novelty point in the scientific landscape of semi-supervised learning ELM models [18–20].

The experimental verification of the method involves three different datasets: USPS [21], Isolet [22], Statlog Landsat Satellite [23] and COIL-20 [24]. Experimental results confirm the effectiveness of br-ELM and prove that the proposed semi-supervised learning scheme compares positively with state-of-the-art algorithms, such as LapRLS [14], LapSVM [14], Transductive SVM (TSVM) [25], Correlated Nystrom Views (XNV) [26] and SS-ELM [20].

The paper is organized as follows. Section 2 gives a brief theoretical background on regularization based learning. Section 3 formalizes the biased regularization based learning scheme, and introduces biased ELM. Section 4 presents the semi-supervised classification framework based on biased regularization. Section 5 discusses experimental results and proposes a comparison with LapRLS, LapSVM, TSVM, and SS-ELM. Finally, Section 6 gives some concluding remarks.

2. Theoretical background

2.1. Regularization-based learning

Modern classification methods often rely on regularization theory, which was initially introduced in [27] and generalized to the nonlinear case by kernel methods [28]. In a regularized functional, a positive parameter λ rules the tradeoff between the empirical risk, $R_{emp}[f]$, (loss function) of the decision functions f (i.e., regression or classification) and a regularizing term. The cost to be minimized can be expressed as:

$$R_{reg} = R_{emp}[f] + \lambda\Omega[f] \tag{1}$$

where the regularization operator $\Omega[f]$ quantifies the complexity of the class of functions from which f is drawn. When dealing with maximum-margin algorithms, $\Omega[f]$ is implemented by the term $\|f\|$, which supports a square norm in the feature space. The Representer Theorem [8] proves that, when $\Omega[f] = \|f\|$, the solution of the regularized cost (1) can be expressed as a finite summation over a set of labeled training patterns $\mathrm{T}=\{(\boldsymbol{x},y)_i;\ i=1,\dots,P\}$, with $y \in \{-1,1\}$.

SVM [25] and Regularized Least Squares (RLS) [25] are popular methods belonging to this family of regularizing algorithms. In both cases, f belongs to a Reproducing Kernel Hilbert Space (RKHS) $\mathbf{H}$ [28] and a kernel function $K(\boldsymbol{x}_i,\boldsymbol{x}_j)$ allows treating only inner products of pattern pairs, disregarding the specific mapping of each single pattern. Accordingly, the solution of the regularized cost (1) can be expressed as

$$f(\mathbf{x}) = \sum_j \beta \cdot K(\mathbf{x}, \mathbf{x}_j). \tag{2}$$

The two learning algorithms differ in the choice of the loss function. The SVM model uses the ‘hinge’ loss function, whereas RLS operates on a square loss function.

The ELM framework indeed belongs to the class of regularized learning methods. In principle, the ELM model implements a single-hidden layer feedforward neural networks (SLFN) with N_h mapping neurons. The response of a neuron to an input stimulus $\boldsymbol{x}$ is implemented by any nonlinear piecewise continuous functions $a(\boldsymbol{x},\zeta)$, where ζ denotes the set of parameters of the mapping function. The overall output function is then expressed as

$$f(\mathbf{x}) = \sum_{j=1}^{N_h} w_j h_j(\mathbf{x}) \tag{3}$$

where w_j denotes the weight that connects the jth neuron with the output, and $h_j(\mathbf{x}) = a(\mathbf{x},\zeta_j)$. In ELM the parameters ζ_j are set randomly. As the training process reduces to the adjustment of the output layer, training ELMs is equivalent to solving a regularized least squares problem. Hence, the minimization problem can be expressed as

$$\min_f \left\{ \sum_{i=1}^{P} (y_i - f(\mathbf{x}_i))^2 + \lambda \|f\|^2 \right\}. \tag{4}$$

The vector of weights $\mathbf{w}$ is then obtained as follows:

$$\mathbf{w} = (\mathbf{H}^t\mathbf{H} + \lambda\mathbf{I})^{-1}\mathbf{H}^t\mathbf{y}. \tag{5}$$

Here, $\boldsymbol{H}$ is a $P \times N_h$ matrix with $h_{ij} = h_j(\boldsymbol{x}_i)$.

2.2. Generalization error in regularized classifiers

A crucial issue in classification problems is the tuning of the classifier regularization parameter(s). Using strategies based on cross validation procedures represent a popular and powerful solution to tackle model selection [29]. At the same time, the literature has showed [15–17] that a reliable option to attain effective model selection is represented by theoretical approaches that derive analytical expressions of the generalization error bounds. Such approaches may indeed prove useful when tackling limited-sample problems, as they do not require any data partitioning and are always based on the complexity on the hypothesis space, F. The bound value to the ‘true’ generalization error, $R[f]$, is asserted with confidence at least $1-\delta$, and is commonly written as the sum of several terms

$$R[f] = R_{emp}[f] + \chi + \psi \tag{6}$$

where $R_{emp}[f]$ is the error on the training set, χ measures the complexity of the space of classifying functions, and ψ penalizes the finiteness of the training sample. The Maximal Discrepancy (MD) [15] and the Rademacher Complexity (RC) [30] represent two well-known theoretical approaches that prove effective in estimating the upper bound of $R[f]$.

The following sections show that the proposed framework for semi-supervised learning based on biased regularization inherently impact on the termχ in (6). This in turn means that the learning model supporting the br-ELM can contribute to limit the complexity of the hypothesis space F, thus enhancing the performance in terms of generalization bounds.

3. A unifying framework for biased learning

The general biased regularization model consists in biasing the solution of a regularization-based learning machine by a reference function (e.g., a hyperplane). The nature of this reference function is a crucial aspect that concerns the learning theory in general. This section discusses two main aspects, that is, the formal definition of a general biased-regularization scheme, and the formalization of the ‘biased regularization ELM’ (br-ELM) within this scheme.

3.1. Biased regularization

In the linear domain one can define a generic convex loss function $L(X, Y, \mathbf{w})$, and a biased regularizing term; the resulting cost function is

$$L(X, Y, \mathbf{w}) + \frac{\lambda_1}{2}\|\mathbf{w} - \lambda_2\mathbf{w}_0\|^2 \tag{7}$$

where $\mathbf{w}_0$ is a “reference” hyper-plane, λ_1 is the classical regularization parameter that controls smoothness (i.e., λ in (1)), and λ_2

Download English Version:

https://daneshyari.com/en/article/406130

Download Persian Version:

https://daneshyari.com/article/406130

Daneshyari.com