Information Sciences 364–365 (2016) 156–166

Contents lists available at ScienceDirect

Information Sciences

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

A semi-supervised random vector functional-link network based on the transductive framework

Simone Scardapane*, Danilo Comminiello, Michele Scarpiniti, Aurelio Uncini

Department of Information Engineering, Electronics and Telecommunications (DIET), "Sapienza" University of Rome, Via Eudossiana 18, Rome 00184, Italy

ARTICLE INFO

Article history:
Received 15 May 2015
Revised 14 July 2015
Accepted 29 July 2015
Available online 22 August 2015

Keywords:
Random vector functional-link
Semi-supervised learning
Transductive learning
Manifold regularization

ABSTRACT

Semi-supervised learning (SSL) is the problem of learning a function with only a partially labeled training set. It has considerable practical interest in applications where labeled data is costly to obtain, while unlabeled data is abundant. One approach to SSL in the case of binary classification is inspired by work on transductive learning (TL) by Vapnik. It has been applied prevalently using support vector machines (SVM) as the base learning algorithm, giving rise to the so-called transductive SVM (TR-SVM). The resulting optimization problem, however, is highly non-convex and complex to solve. In this paper, we propose an alternative semi-supervised training algorithm based on the TL theory, namely semi-supervised random vector functional-link (RVFL) network, which is able to obtain state-of-the-art performance, while resulting in a standard convex optimization problem. In particular we show that, thanks to the characteristics of RVFLs networks, the resulting optimization problem can be safely approximated with a standard quadratic programming problem solvable in polynomial time. A wide range of experiments validate our proposal. As a comparison, we also propose a semi-supervised algorithm for RVFLs based on the theory of manifold regularization.

1. Introduction

Supervised learning (SL) is the task of inferring a function starting from a set of labeled examples of it [19]. Semi-supervised learning (SSL) is an extension of SL, in which the user is provided with an additional set of *unlabeled* data [12]. The central problem of SSL is to incorporate this additional set into the learning problem, so as to maximize accuracy over unseen patterns. Under this respect, SSL algorithms differentiate themselves mostly on the assumptions that are made to this end [19]. SSL is useful whenever labeled data is scarce (or costly to obtain), while unlabeled data is common and easy to collect. Examples abound in real world applications, including bioinformatics [23], image processing [32], computer vision [18] and text classification [22].

One common approach to SSL is manifold regularization (MR) [5,17,28,38]. In MR, it is assumed that the input data, while being observed in an Euclidean space $\mathbb{R}^d$ (called the ambient space), practically lies on an embedded manifold $\mathcal{M}$, called the intrinsic space, whose dimension might be much lower than d. Since the geometry of $\mathcal{M}$ only depends on the input data, it can be estimated using both labeled and unlabeled data together. Originally, Belkin and Niyogi [4] proposed this idea for dimensionality reduction in the ambient space. Later, it was shown how information on $\mathcal{M}$ can be used to incorporate an efficient regularization term in the learning problem [5], forcing the output of the resulting function to be close whenever two points are similar in their

* Corresponding author. Tel.: +39 06 44585495; fax: +39 06 4873300.
E-mail addresses: simone.scardapane@uniroma1.it, simonescardapane@gmail.com (S. Scardapane), danilo.comminiello@uniroma1.it (D. Comminiello), michele.scarpiniti@uniroma1.it (M. Scarpiniti), aurelio.uncini@uniroma1.it (A. Uncini).

http://dx.doi.org/10.1016/j.ins.2015.07.060
0020-0255/

intrinsic space. From this general framework, it is possible to derive a wide range of MR kernel-based algorithms, including the Laplacian Support Vector Machine (LAP-SVM) [28], and the Laplacian Regularized Least-Square (LAP-RLS) [5]. The main drawback of the MR approach is that, in order to derive the regularization term, it is necessary to estimate a matrix connected to $\mathcal{M}$, called the Laplacian, which in turn depends on the pairwise similarities between each pair of points. This is not straightforward in general, since its computation requires setting a number of free parameters, each of which heavily influences the resulting classification accuracy (see e.g. [36]). In fact, model selection is a crucial problem in SSL, where labeled data may not be sufficient to obtain a large validation set [31].

In the case of binary classification, an alternative approach to SSL is constituted by the transductive learning (TL) theory introduced by V. Vapnik [39,41]. TL is similar to SSL, in that we are provided with an additional set of unlabeled data. The main difference is that in TL we do not wish to infer a classifier over the full input space, but we only require the unknown labels of the additional dataset. Based on the principle of structural risk minimization, Vapnik proposed a transductive version of the support vector machine (denoted as TR-SVM) [22,39]. In TR-SVM, the unknown labels are introduced as additional variables in the optimization problem, and we minimize the training error over both labeled and unlabeled data. Despite being formulated in the TL framework, TR-SVM can actually be used without modifications for SSL, where it is known as semi-supervised SVM (S^3VM) [6,13].[1] Compared to MR approaches, the TR-SVM only requires an additional free parameter to be specified, but its optimization problem involves both real and binary variables, making it highly non-convex. Over the last years, multiple algorithms based on different relaxations have been proposed for its solution [13,14,27]. It is known, however, that the performance of each of these algorithms is highly domain-dependent, and each of them presents complex implementation details [13].

The basic idea underlying the TR-SVM, i.e. considering the unknown labels as variables in the learning problem, has rarely been applied to models laying outside the realm of the standard SVM formulation (two exceptions being applications to least-square SVM [2], and kernel ridge regression [11]). In this paper, we propose to apply it to a specific class of neural networks, known as random vector functional-link (RVFL) networks [29], or random-weights neural networks (RWNN) [3,35], to derive a novel semi-supervised algorithm for binary classification. An RVFL network is composed of a fixed layer of non-linearities, followed by an adaptable linear layer [20,29,34]. The main interest of RVFL in this context is that, under a least-square training criterion, the solution to its optimization problem can be expressed in closed form. Due to this, we show that the resulting semi-supervised algorithm can be formulated as a standard unconstrained 0-1 optimization problem [25]. While this is an NP-hard problem in general, we show that it can safely be approximated with a box-constrained quadratic (BCQ) problem, solvable in polynomial time [8]. The result is a semi-supervised training algorithm for RVFL networks, that we denote as transductive RVFL (TR-RVFL).[2] Through an extensive set of experimental evaluations, we show that it performs similarly, or better, than a conventional state-of-the-art algorithm based on the MR theory. However, it is easier to tune, and in many situations it can be trained faster than its counterpart.

The main novelty of this paper is the formulation of the optimization problem of TR-RVFL as a BCQ problem, which is solvable in polynomial time. Additionally, we also present an MR-based training algorithm for RVFL networks, and we use it for comparison purposes. In fact, to the best of the authors' knowledge, no work exists related to SSL with the use of RVFL networks. The only related work is [10], where the authors consider the use of functional link networks for semi-supervised clustering.

The rest of the paper is organized as follows. In Section 2, we present the basic concepts used in the rest of the paper, including the formulation of the SSL problem, and the theory of RVFL networks. For completeness, in Section 3, we derive an MR-based training algorithm for RVFL networks, which is used as a comparison in the experimental section. Section 4, which is the main innovative part of the paper, presents the TR-RVFL algorithm. We perform an extensive range of experiments in Section 5, before presenting our conclusive remarks in Section 6.

2. Preliminaries

In this Section, we present the basic theoretical tools used in the rest of the paper. In particular, we provide an overview of SSL in Section 2.1, followed by a derivation of a least-square training algorithm for RVFL networks in the standard supervised case in Section 2.2.

Before this, we provide a brief note on the notation which is adopted. In the following, vectors are denoted with lowercase bold letters, such as $\mathbf{a}$, while matrices are denoted by uppercase bold letters, such as $\mathbf{A}$. All vectors are considered column vectors. The notation A_{ij} denotes the (i, j)th entry of matrix $\mathbf{A}$. $\|\cdot\|_2$ is the standard Euclidean norm. We will also make use of the weighted Euclidean norm, defined for a generic vector $\mathbf{a}$ and matrix $\mathbf{B}$ as:

$$\|\mathbf{a}\|_{\mathbf{B}}^2 = \mathbf{a}^T\mathbf{B}\mathbf{a}\,. \tag{1}$$

Finally, we use $\mathbf{A} \succeq 0$ to denote a positive semi-definite (PSD) matrix, i.e. a matrix for which $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ for any vector $\mathbf{x}$ of suitable dimensionality.

2.1. Approaches to semi-supervised learning

We consider learning a binary classifier in the supervised and semi-supervised case. Thus, we are interested in finding a classification function $f : \mathbb{R}^d \rightarrow \{-1, +1\}$, given a set of examples of this relationship. In the SL setting [19], we have only access

[1] See also [12, Chapters 24 and 25] for additional comments on the link between transduction and SSL.

[2] We use the term 'transductive' to denote its relation to the TL framework. However, as for TR-SVM, the resulting algorithm can be used without modifications in an SSL context.