Contents lists available at ScienceDirect

Neurocomputing

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

New classifier based on compressed dictionary and LS-SVM

Chen Sun *, Licheng Jiao, Hongying Liu, Shuyuan Yang

Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, Xidian University, Xi'an, Shaanxi Province 710071, China

ARTICLE INFO

Article history:
Received 20 May 2015
Received in revised form
24 March 2016
Accepted 4 August 2016
Communicated by Dr. H. Yu

Keywords:
Sparsity
Compressed dictionary
LS-SVM
Restricted isometry property
Compressive sensing

ABSTRACT

Inspired by the compressive sensing (CS) theory, a new classifier based on compressed dictionary and Least Squares Support Vector Machine (LS-SVM) is proposed to deal with large scale problems. The coefficients of support vectors can be recovered from a few measurements if LS-SVM is approximated to sparse structure. Using the known Cholesky decomposition, we approximate the given kernel matrix to represent the coefficients of support vectors sparsely by a low-rank matrix that we have used as a dictionary. The proposed measurement matrix being coupled with the dictionary forms a compressed dictionary that proves to satisfy the restricted isometry property (RIP). Our classifier has the quality of low storage and computational complexity, high degree of sparsity and information preservation. Experiments on benchmark data sets show that our classifier has positive performance.

1. Introduction

With the exponential growth and availability of data, it becomes difficult to classify using traditional tools. Big data [1] reveals data with much sparsity, thus, classifiers with high efficiency and sparsity are increasingly necessary. Support Vector Machine (SVM), a classical and useful tool in machine learning and pattern recognition, has attracted great attention since its introduction by Vapnik [2–5]. The training procedure amounts to solving a constrained quadratic programming problem. For a small scale problem, it is fast to find the optimal solutions. But for a large scale problem, it would take a long time to train. Least squares support vector machine (LS-SVM) [6,7] is presented as a least squares version for SVM. LS-SVM considers equality type constraints instead of inequalities from the SVM approach. As a result the solution follows directly from solving a set of linear equations instead of quadratic programming. Another advantage of LS-SVM is that it involves fewer tuning parameters.

Although the gain of LS-SVM in efficiency is significant, the sparseness of support values is lost [8,9]. Many researchers tried to solve the problem in LS-SVM and gave their solutions. Several pruning algorithms have been developed to impose the sparseness in LS-SVM. Suykens et al. [10] proposed to prune the training samples that have the smallest absolute Lagrange multipliers. Kruif and Vries [11] presented to omit the sample bearing the smallest error. Zeng and Chen [12] proposed a SMO-based pruning method. However, these algorithms have large retraining costs. Yang et al. [13] proposed a training algorithm via compressive pruning to avoid retraining. But all these algorithms are not feasible for large-scale problems. For handing large-scale problems, Jiao et al. [14] presented a fast sparse approximation scheme (FSALS-SVM), which built the decision function by adding one basis function from a kernel-based dictionary per time, but the scale of optimization problem was large. The primal formulation of LS-SVM was proposed by using Nystrom approximation with a set of prototype vectors to introduce sparsity described in Suykens et al. [15]. Brabanter et al. [16] developed an optimized fixed-size LS-SVM version. Mall et al. [17] further optimized the sparsity in a second stage by means of L_0-norm. A sparse conjugate direction pursuit approach was proposed in [18] where they iteratively constructed a model and stopped early to obtain sparse solution. They focused on the primal problem and over-determined linear system, while we proposed the model focusing on an under-determined linear system.

Relevance vector machines (RVMs) are sparse Bayesian methods whose sparsity can be attained by imposing hierarchical prior on parameters. RVM, by Agarwal and Triggs [19], discerned basis functions which were 'relevant' for making good predictions. The sparse Bayesian inference method has high computational complexity and requires large memory. When the prior distribution of parameters deviated from the actual, the results would not be good.

Inspired by the recently developed Compressed Sensing (CS)

* Corresponding author.
E-mail address: smilesunc@163.com (C. Sun).

http://dx.doi.org/10.1016/j.neucom.2016.08.024
0925-2312/

Please cite this article as: C. Sun, et al., New classifier based on compressed dictionary and LS-SVM, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.08.024

Table 1
Comparisons of **Gaussian**, **Binary** and the proposed measurement matrices in terms of MSE.

	Ringnorm	Satimage	Waveform	USPS	Adult	FC
Gaussian	0.7217	0.0722	0.9454	0.9862	0.9991	0.9800
Binary	0.7225	0.0724	0.9452	0.9860	0.9990	0.9826
The proposed	0.7160	0.0715	0.9419	0.9857	0.9986	0.9784

Table 2
Comparisons of **CDLS-SVM**, **FSALS-SVM**, **CG** and **RVM** in terms of memory requirement, training cost, and prediction cost.

	CDLS-SVM	FSALS-SVM	LS-SVM (CG)	RVM
Memory	$O(rn)$	$O(ln)$	$O(n)$	$O(n^2)$
Training	$O(r^2n)$	$O(l^2n)$	$O(\#it^*n^2)$	$O(n^3)$
Prediction	$O(l)$	$O(l)$	$O(n)$	$O(l)$

l: number of support vectors; n: number of samples; r: rank of the kernel matrix; $\#it$: number of iterations.

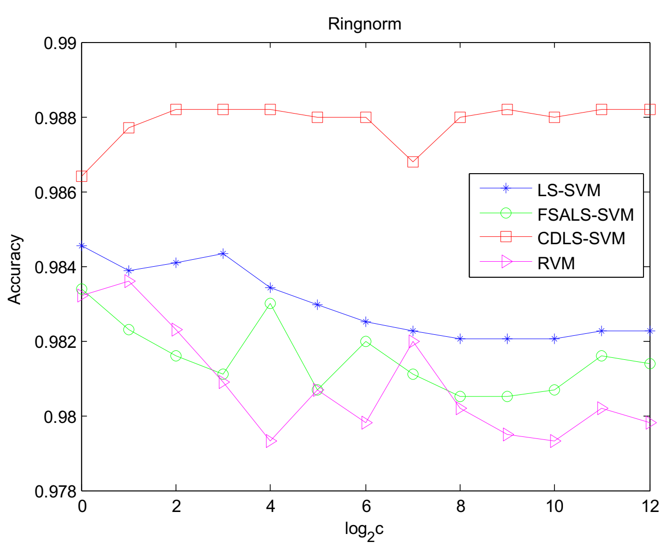

Fig. 1. Comparison of FSALS-SVM, LS-SVM, CDLS-SVM and RVM on accuracy.

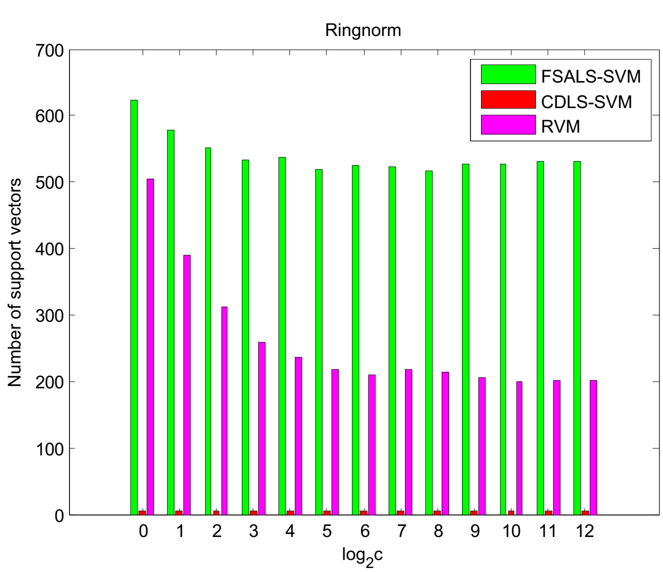

Fig. 2. Comparison of FSALS-SVM, CDLS-SVM and RVM on number of support vectors.

theory [20–25], a new classifier based on compressed dictionary is proposed. CS provides an effective way to recovering signals of general interest from a small number of measurements lower than the traditional sampling number of Shannon sampling theory. But it relies on two principals: sparsity and incoherence. The topology of LS-SVM can be described by the coefficients of support vectors. If it can be approximated to sparse structure, we can use the compressed measurement of LS-SVM to produce compact structure. As long as the compressed dictionary satisfies the RIP, coefficients of support vectors can be recovered with little degradation in performance. Consequently, the scale of the optimization problem and the requirements of storage are remarkably reduced.

With respect to other available approaches, the characteristics of the proposed CDLS-SVM are as follows.

(1) The proposed method seeks the sparse solutions and compressed version of signals directly, which saves much training time, in contrast to the pruning methods [10–12] which have spent large efforts to obtain the full information on the support vectors and afterward thrown away most of information to achieve test speed.
(2) Taking a low-rank matrix as the dictionary, we design a novel measurement matrix and a compressed dictionary to deal with large scale problems. The measurement matrix is learned from data and the constructed compressed dictionary proves to satisfy the RIP. It largely reduces the requirements of storage.
(3) The proposed method can be operated easily and conveniently without casting any prior hypothesis on data.

The rest of the paper is organized as follows. In Section 2, the background of CS is discussed. CDLS-SVM is proposed in Section 3. Experimental results are reported in Section 4. Finally, in Section 5, we give some concluding remarks.

2. Compressive sensing

CS builds upon the empirical observation that many signals can be well approximated using only a few non-zero coefficients in a suitable dictionary. Nonlinear optimization enables recovery of such signal from few measurements. To make this possible, CS relies on two principals: sparsity and incoherence [22,23].

A vector $\mathbf{x} \in \mathbb{R}^n$ is called k-sparse if $\mathbf{x}$ has at most k non-zero components $\|\mathbf{x}\|_0 \le k$or can be represented sparsely under a dictionary $\mathbf{D}$, $\mathbf{x} = \mathbf{D}c$, $\|c\|_0 \le k$.

Taking $m \ge 2k$ measurements of a signal $\mathbf{x}$ corresponds to applying a measurement matrix $\mathbf{\Phi}$, $\mathbf{y} = \mathbf{\Phi x} = \mathbf{\Phi D}c \in \mathbb{R}^m$ where $m \ll n$. If the matrix $\mathbf{\Phi D}$ satisfies the restricted isometry property (RIP) [20–23], accuracy recovery of the signal is possible. The greater the incoherence between the measurement matrix $\mathbf{\Phi}$ and the dictionary $\mathbf{D}$, the smaller the number of measurements needs.

3. CDLS-SVM

The dictionary in LS-SVM is not appropriate to represent the signal sparsely, thus the CDLS-SVM model is proposed in this section. The key idea is as follows.

Firstly, we approximate the given kernel matrix by a low-rank matrix. Then, take the low-rank matrix as a new dictionary. Since the rank of the dictionary is lower than the number of equations and the number of variables, it is essentially an underdetermined problem. The linear programming problem has multiple solutions, including some sparse solutions. Thus, the first principal satisfies.

Finally, we designed a measurement dictionary which proves to

Please cite this article as: C. Sun, et al., New classifier based on compressed dictionary and LS-SVM, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.08.024