Information Sciences 358–359 (2016) 29–43

Contents lists available at ScienceDirect

Information Sciences

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

A matrix-free smoothing algorithm for large-scale support vector machines[☆]

Tie Ni[a,*], Jun Zhai[b]

[a] College of Business Administration, Liaoning Technical University, Huludao 125105, PR China
[b] Department of Mathematics, The University of Chicago, Chicago 60615, USA

ARTICLE INFO

Article history:
Received 13 June 2015
Revised 25 March 2016
Accepted 3 April 2016
Available online 8 April 2016

MSC:
90C33
65K10
65F10

Keywords:
Support vector machines
KKT systems
Smoothing Newton algorithm
Preconditioner
Non-monotone line search

ABSTRACT

This paper is concerned with the iterative solution of sequences of Karush-Kuhn-Tucker (KKT) systems arising from smoothing Newton-type algorithms applied to support vector machines (SVMs). Every perturbed Newton step is crucial for the efficiency of the overall algorithm. However, the KKT systems based on smoothing reformulation become increasingly ill-conditioned as the smoothing parameter approaches to zero. By exploiting preconditioning techniques, we propose a matrix-free smoothing algorithm for solving large-scale support vector machines. Numerical results and comparisons are given to demonstrate the effectiveness and speed of the algorithm.

1. Introduction

Support Vector Machines (SVMs) are based on the statistical learning theory developed by [36,37] and have demonstrated outstanding performance in many applications. SVM [5,23] have proven successful in the classification and regression fields.

For classification, we consider the problem of classifying m points in the n-dimensional real space $\Re^n$, represented by $m \times n$ matrix A, according to the membership of each point A_i in the class $+1$ or -1 as specified by a given $m \times m$ diagonal matrix D with ones or minus ones along its diagonal. The linear support vector machine attempts to separate these finite points with a hyperplane such that the separation margin is maximized. We aim at determining a hyperplane $\mathcal{H}(\mathbf{w}, b) = \{x | \mathbf{w}^T x = b\}$, where $\mathbf{w} \in \Re^n \backslash \{\mathbf{0}\}$ and $b \in \Re$, which separates points into two classes.

For non-separable data, the classical soft-margin support vector machine has the following form:

$$\begin{aligned} &\min_{\mathbf{w}, b, \xi} \ \frac{1}{2}\|\mathbf{w}\|^2 + \frac{\nu}{2}\|\xi\|^2 \\ &\text{s.t. } D(A\mathbf{w} - be) + \xi \geq e. \end{aligned} \tag{1.1}$$

☆ This work is partially supported by the National Natural Science Foundation of China (Grant No. 71371091) and the Pre-Research Foundation of the College of Business Administration of Liaoning Technical University.

* Corresponding author. Tel.: +8615114295701; fax: +86 429 5310150.
E-mail addresses: lngdnt@163.com (T. Ni), junzhai0516@uchicago.edu (J. Zhai).

http://dx.doi.org/10.1016/j.ins.2016.04.010
0020-0255/

where

$$\xi := \max\{0, e - D(A\mathbf{w} - be)\}, \tag{1.2}$$

$e := (1, \ldots, 1)^T \in \Re^m$ and parameter $\nu > 0$ that trades off two competing goals of maximizing the separation margin and minimizing the misclassification error. From [24,40], the dual lagrangian formulation of the (1.1) leads to the following strongly convex quadratic programming:

$$\begin{aligned} \min_x \quad & \frac{1}{2}x^T DAA^T Dx + \frac{1}{2\nu}x^T x - e^T x \\ \text{s.t.} \quad & e^T Dx = 0 \\ & x \geq 0. \end{aligned} \tag{1.3}$$

The solution $(\mathbf{w}, b, \xi)$ of (1.1) can be recovered from a solution x of (1.3) by setting $\mathbf{w} = A^T Dx$, b to the optimal multiplier of the equality constraint $e^T Dx = 0$ and $\xi = (1.2)$. By [22,25], we can obtain the generalized support vector machine by applying a general kernel $K(A, A^T)$ to (1.1):

$$\begin{aligned} \min_{\mathbf{u}, b, \xi} \quad & f(\mathbf{u}) + \frac{\nu}{2}\|\xi\|^2 \\ \text{s.t.} \quad & D(K(A, A^T)D\mathbf{u} - be) + \xi \geq e, \end{aligned} \tag{1.4}$$

where f is a convex function in $\Re^m$. The corresponding nonlinear separating surface is:

$$K(x^T, A^T)D\mathbf{u} = b, \tag{1.5}$$

and the corresponding dual is expressed as:

$$\begin{aligned} \min_x \quad & \frac{1}{2}x^T DK(A, A^T)Dx + \frac{1}{2\nu}x^T x - e^T x \\ \text{s.t.} \quad & e^T Dx = 0 \\ & x \geq 0. \end{aligned} \tag{1.6}$$

Actually, if let $K(x^T, A^T) = x^T A^T$, $\mathbf{w} = A^T D\mathbf{u}$, then (1.5) reduces to $\mathbf{w}^T x = b$. Furthermore, if set $f(\mathbf{u}) = \frac{1}{2}\mathbf{u}^T DAA^T D\mathbf{u}$ and $K(A, A^T) = AA^T$, then (1.4) becomes (1.1) and (1.6) reduces (1.3).

Newton-based iterative methods for solving SVMs have been proposed in the literature, such as interior point method [9,38,39], semismooth method [10], smoothing method [22] and smoothing Newton method [30]. These methods generally work well on small and medium-scale problems, especially the smoothing Newton method. However, when the size of the problem becomes large, they inherently suffer from either numerical instability or memory caching inefficiency. In recent years, preconditioning technique has been widely used in various fields and progress has been reported. For instance, in [2], the authors studied updating constraint preconditioners for KKT systems of Quadratic Programming through low-rank corrections of the Schur complement. In [6], the authors proposed a primal-dual Newton conjugate gradients method based on preconditioning technique when applied to the compressed sensing problems. In [8], a matrix-free interior point method is applied to the compressed sensing problems. In [13], the authors gave estimates of bounds on eigenvalues of matrices arising from interior point methods. In [15], the authors presented a matrix-free trust-region SQP method to solve the equality constrained optimization problem. In [27,28] the authors analyzed the spectral estimates of the unreduced symmetric KKT systems arising from interior point methods when proper preconditioners are introduced. It is evident that preconditioning technique plays an important role in interior point methods.

The smoothing Newton algorithm has been a powerful tool for solving many optimization problems since it possesses not only good computational performance but also global and local superlinear convergence under some assumptions (see, for example, [14,16,19,33,34]). It should be noted that the smoothing Newton algorithm is a second-order algorithm, which requires more computations in each iteration. Furthermore, the KKT systems based on smoothing reformulation in smoothing Newton-type algorithms become increasingly ill-conditioned as the smoothing parameter approaches to zero(see Section 4 for theoretical details). Therefore, for the large-scale support vector machines, the efficiency of smoothing Newton algorithm is difficult to be guaranteed.

In this paper, we develop a new smoothing algorithm named matrix-free smoothing algorithm for solving the large-scale support vector machines. It is a hybrid algorithm combining the advantages of the first-order iterative methods for solving preconditioned linear system of equations and the second-order information based on smoothing technique. The two operations the preconditioner allowed to use are matrix-vector multiplications performed with the Hessian and Jacobian and its transposition. Those are necessary to achieve a matrix-free implementation of smoothing Newton algorithm.

The rest of the paper is organized as follows. In Section 2, the idea of smoothing reformulation and the notations are summarized. In Section 3, a popular framework of smoothing Newton algorithm is recalled. In Section 4, we analyze the estimate of condition number of the reduced linear system. In Section 5, we show the spectral analysis of preconditioned matrix associated with preconditioned reduced linear system. In Section 6, a matrix-free smoothing algorithm is proposed in detail for solving large-scale support vector machines. Numerical results are reported in Section 7. Some final remarks are given in Section 8.