Neurocomputing 72 (2009) 973–984

Contents lists available at ScienceDirect

Neurocomputing

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

Decomposition techniques for training linear programming support vector machines

Yusuke Torii, Shigeo Abe *

Graduate School of Engineering, Kobe University, Kobe, Japan

ARTICLE INFO

Article history:
Received 8 January 2008
Received in revised form 14 April 2008
Accepted 18 April 2008
Communicated by S. Hochreiter
Available online 10 May 2008

Keywords:
Decomposition techniques
Linear programming
Primal–dual interior-point method
Simplex method
Support vector machines

ABSTRACT

In this paper, we propose three decomposition techniques for linear programming (LP) problems: (1) Method 1, in which we decompose the variables into the working set and the fixed set, but we do not decompose the constraints, (2) Method 2, in which we decompose only the constraints and (3) Method 3, in which we decompose both the variables and the constraints into two. By Method 1, the value of the objective function is proved to be non-decreasing (non-increasing) for the maximization (minimization) problem and by Method 2, the value is non-increasing (non-decreasing) for the maximization (minimization) problem. Thus, by Method 3, which is a combination of Methods 1 and 2, the value of the objective function is not guaranteed to be monotonic and there is a possibility of infinite loops. We prove that infinite loops are resolved if the variables in an infinite loop are not released from the working set and Method 3 converges in finite steps. We apply Methods 1 and 3 to LP support vector machines (SVMs) and discuss a more efficient method of accelerating training by detecting the increase in the number of violations and restoring variables in the working set that are released at the previous iteration step.

By computer experiments for microarray data with huge input variables and a small number of constraints, we demonstrate the effectiveness of Method 1 for training the primal LP SVM with linear kernels. We also demonstrate the effectiveness of Method 3 over Method 1 for the nonlinear LP SVMs.

1. Introduction

Support vector machines (SVMs) [36,35] are widely used for pattern classification. But in training an SVM we need to solve a quadratic programming problem with the number of variables equal to the number of training data. Thus, to speed up training for a large problem, we usually use a decomposition technique, in which the original variables are divided into working variables and fixed variables and a small problem with the working variables is iteratively solved [23,24]. A special case of the decomposition technique is the sequential minimal optimization (SMO) with the working set size of two [25]. The convergence of the decomposition technique for SVMs is theoretically proved [23,24,21,18,30] and there are many discussions on working set selection to speed up convergence of SMO [18,8,6,9] and general decomposition techniques with working set sizes larger than two [16,13,20,14]. By the decomposition techniques training of SVMs for large-scale problems is considerably speeded up.

As a variant of SVMs, linear programming SVMs (LP SVMs), in which the quadratic objective functions are replaced with linear objective functions, have been proposed [27,17,38]. In training LP SVMs, we need to solve LP problems with the number of variables more than three times the number of training data. But until now, there are not so many discussions on the decomposition techniques for LP SVMs. In [5], a decomposition technique is proposed, in which only a part of linear constraints are used for linear SVMs. This method confirms monotonic convergence of the objective function and is useful for the problems with a large number of constraints but a small number of variables. In [32], decomposition techniques for SVMs are extended to LP SVMs. Because direct implementation of the decomposition techniques leads to infinite loops, training speedup is done by modifying working set selection when the number of violations of complementarity conditions increases.

In this paper we propose three decomposition techniques for LP programs: Method 1, in which variables are divided into working variables and fixed variables but constraints are all used; Method 2, in which constraints are divided into working constraints and fixed constraints but variables are all used; Method 3, in which variables and constraints are divided into working and fixed variables and constraints, respectively. We prove that in Method 1, the values of the objective function are non-increasing for a minimization problem during training. While in Method 2 the values of the objective function are

* Corresponding author.
E-mail address: abe@eedept.kobe-u.ac.jp (S. Abe).

0925-2312/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2008.04.008

non-decreasing for a minimization problem. Therefore, for the combined method, Method 3, the values of the objective function are not monotonic and there is a possibility of infinite loops. We prove that if the variables in an infinite loop are kept in the working set during training, Method 3 converges in finite steps. We apply Methods 1 and 3 to LP SVMs and for Method 3, we discuss more efficient method for training. In computer experiments, we show that Method 1 can accelerate training of linear LP SVMs for microarray data, and Method 3 for training LP SVMs with a large number of training data.

The structure of the paper is as follows. In Section 2, we propose three decomposition techniques and clarify relations of the proposed decomposition techniques with that for SVMs. Then in Section 3, we apply these methods to LP SVMs and in Section 4, we demonstrate the effectiveness of the proposed methods using some benchmark data sets. Finally in Section 5 we conclude our work.

2. Decomposition techniques

If the size of a problem is very large, it is natural to consider dividing the problem into small sub-problems and solving the sub-problems iteratively. For an optimization problem, one way is to divide the problem into a working sub-problem and a fixed sub-problem, solve the working sub-problem, re-divide the problem into a working sub-problem and a fixed sub-problem, and iterate the procedure until the solution is obtained. However, to obtain the solution by this method, the objective function needs to be monotonic during the training process. If not, convergence in finite steps is required.

In the following, we discuss three decomposition techniques for an LP problem.

2.1. Formulation

We consider the following problem, which is a generalized version of an LP SVM:

$$\text{minimize} \quad \mathbf{c}^{\mathrm{T}}\mathbf{x} + \mathbf{d}^{\mathrm{T}}\xi \tag{1}$$

$$\text{subject to} \quad A\mathbf{x} \geqslant \mathbf{b} - \xi, \quad \mathbf{x} \geqslant \mathbf{0}, \quad \xi \geqslant \mathbf{0}, \tag{2}$$

where $\mathbf{c}$ is an m-dimensional constant vector, $\mathbf{d}$ is an M-dimensional vector and $\mathbf{d} > 0$, A is an $M \times m$ constant matrix, $\mathbf{b}$ is an M-dimensional positive constant vector, and ξ is a slack variable vector to make $\mathbf{x} = \mathbf{0}$ and $\xi = \mathbf{b}$ be a feasible solution. Therefore, the optimal solution always exists.

Introducing an M-dimensional slack variable vector $\mathbf{u}$ (2) becomes

$$A\mathbf{x} = \mathbf{b} + \mathbf{u} - \xi, \quad \mathbf{x} \geqslant \mathbf{0}, \ \mathbf{u} \geqslant \mathbf{0}, \ \xi \geqslant \mathbf{0}. \tag{3}$$

The dual problem of (8) and (9) is as follows:

$$\text{maximize} \quad \mathbf{b}^{\mathrm{T}}\mathbf{z} \tag{4}$$

$$\text{subject to} \quad A^{\mathrm{T}}\mathbf{z} + \mathbf{v} = \mathbf{c}, \quad \mathbf{z} + \mathbf{w} = \mathbf{d}, \quad \mathbf{v} \geqslant \mathbf{0}, \quad \mathbf{z} \geqslant \mathbf{0}, \quad \mathbf{w} \geqslant \mathbf{0}, \tag{5}$$

where $\mathbf{z}$ is an M-dimensional vector, $\mathbf{v}$ is an m-dimensional slack variable vector and $\mathbf{w}$ is an M-dimensional slack variable vector.

The optimal solution $(\mathbf{x}^*, \xi^*, \mathbf{u}^*, \mathbf{z}^*, \mathbf{v}^*, \mathbf{w}^*)$ must satisfy the following complementarity conditions:

$$x_i^* v_i^* = 0 \quad \text{for } i = 1, \ldots, m, \tag{6}$$

$$\xi_i^* w_i^* = 0, \quad z_i^* u_i^* = 0 \quad \text{for } i = 1, \ldots, M. \tag{7}$$

Now solving the primal or dual problem is equivalent to solving

$$\begin{aligned} & A\mathbf{x} = \mathbf{b} + \mathbf{u} - \xi, \quad \mathbf{x} \geqslant \mathbf{0}, \ \mathbf{u} \geqslant \mathbf{0}, \ \xi \geqslant \mathbf{0}, \\ & A^{\mathrm{T}}\mathbf{z} + \mathbf{v} = \mathbf{c}, \quad \mathbf{z} + \mathbf{w} = \mathbf{d}, \\ & \mathbf{z} \geqslant \mathbf{0}, \quad \mathbf{w} \geqslant \mathbf{0}, \ \mathbf{v} \geqslant \mathbf{0}, \\ & x_i v_i = 0 \quad \text{for } i = 1, \ldots, m, \\ & \xi_i w_i = 0, \quad z_i u_i = 0 \quad \text{for } i = 1, \ldots, M. \end{aligned}$$

Here, we call x_i active if $x_i > 0$ and inactive if $x_i = 0$. Likewise, the ith constraint is active if $u_i = 0$ and inactive if $u_i > 0$. Notice that even if we delete inactive variables and constraints, we can obtain the same solution as that of the original problem.

By the primal–dual interior-point method, the above set of equations is solved. By the simplex method, if we solve the primal or dual problem, the primal and dual solutions are obtained simultaneously [7]. Therefore, either by the primal–dual interior-point method or the simplex method, we obtain the primal and dual solutions.

2.2. Three decomposition techniques

Now we consider the following three decomposition methods to solve (1) and (3).

Method 1, in which, a subset of the variables in $\mathbf{x}$ is optimized using all the constraints, while fixing the remaining variables. Let the set of indices of the subset be W_{v} and the remaining subset be F_{v}, where $W_{\mathrm{v}} \cap F_{\mathrm{v}} = \emptyset$ and $W_{\mathrm{v}} \cup F_{\mathrm{v}} = \{1, \ldots, m\}$. Assuming $x_i = 0\,(i \in F_{\mathrm{v}})$, the original problem given by (1) and (3) reduces as follows:

$$\text{minimize} \quad \sum_{i \in W_{\mathrm{v}}} c_i x_i + \mathbf{d}^{\mathrm{T}}\xi \tag{8}$$

$$\text{subject to} \quad \sum_{j \in W_{\mathrm{v}}} A_{ij} x_j = b_i + u_i - \xi_i \quad \text{for} \quad i = 1, \ldots, M, \quad x_i \geqslant 0 \quad \text{for } i \in W_{\mathrm{v}}, \ \mathbf{u} \geqslant \mathbf{0}, \ \xi \geqslant \mathbf{0}. \tag{9}$$

The dual problem of (8) and (9) is as follows:

$$\text{maximize} \quad \mathbf{b}^{\mathrm{T}}\mathbf{z} \tag{10}$$

$$\text{subject to} \quad \sum_{j=1}^{M} A_{ji} z_j + v_i = c_i, \quad v_i \geqslant 0 \quad \text{for } i \in W_{\mathrm{v}}, \quad \mathbf{z} + \mathbf{w} = \mathbf{d}, \quad \mathbf{z} \geqslant \mathbf{0}, \ \mathbf{w} \geqslant \mathbf{0}. \tag{11}$$

Therefore from (10) and (11), if we solve (8) and (9), in addition to the solution of the primal problem, we obtain the solution of the dual problem except for $v_i\,(i \in F_{\mathrm{v}})$. Namely, except for $x_i v_i = 0\,(i \in F_{\mathrm{v}})$, the complementarity conditions given by (6) and (7) are satisfied. Using the first equation in (5) for $i \in F_{\mathrm{v}}$, we can calculate $v_i\,(i \in F_{\mathrm{v}})$. Because we assume that $x_i = 0\,(i \in F_{\mathrm{v}})$, if $v_i \geqslant 0\,(i \in F_{\mathrm{v}})$, v_i satisfy the constraint and the obtained primal solution is optimal. But if some of v_i are negative, the obtained solution is not optimal.

If the obtained solution is not optimal, we move the indices associated with inactive variables from W_{v} to F_{v}, move, from F_{v} to W_{v}, the indices associated with the violating variables, and iterate the previous procedure.

By this method, the optimal solution at each iteration step is obtained by restricting the original space

$$\{\mathbf{x} \mid A\mathbf{x} \geqslant \mathbf{b} - \xi, \mathbf{x} \geqslant \mathbf{0}, \xi \geqslant \mathbf{0}\} \tag{12}$$

to

$$\{\mathbf{x} \mid A\mathbf{x} \geqslant \mathbf{b} - \xi, \ \xi \geqslant \mathbf{0}, \ x_i \geqslant 0 \quad \text{for } i \in W_{\mathrm{v}}, \ x_i = 0 \quad \text{for } i \in F_{\mathrm{v}}\}. \tag{13}$$

If the solution is not optimal, we repeat solving the sub-problem with the non-zero $x_i\,(i \in W_{\mathrm{v}})$ and with the violating variables