A Novel Approach for Efficient SVM Classification with Histogram Intersection Kernel

Gaurav Sharma
gaurav.sharma@unicaen.fr

Frederic Jurie
frederic.jurie@unicaen.fr

GREYC CNRS UMR6072
University of Caen
Boulevard Marechal Juin
14032 Caen, France

We propose a novel approach for learning non-linear SVM corresponding to the histogram intersection kernel without using the kernel trick. We formulate the exact non-linear problem in the original space and show how to perform classification directly in this space. The learnt classifier incorporates non-linearity while maintaining O(d) testing complexity (for d-dimensional input space), compared to O($d \times N_{sv}$) when using the kernel trick. We show that the SVM problem with histogram intersection kernel is quasi-convex in input space and outline an iterative algorithm to solve it. The proposed approach has been validated in experiments where it is compared with other linear SVM-based methods, showing that the proposed method achieves similar or better performance at lower computational and memory costs.

The standard formulation for learning linear classifiers is the SVM primal formulation,

$$\min_{w} \left\{ \frac{\lambda}{2}||\mathbf{w}||_2^2 + \frac{1}{|\mathcal{I}_t|}\sum_i \xi_i \right\}, \quad \text{s.t. } y_i \mathbf{w}^T \mathbf{x}_i \geq 1 - \xi_i \text{ and } \xi_i \geq 0, \quad (1)$$

where $\mathbf{w}$ is the normal to the linear decision hyperplane and $\{(\mathbf{x}_i, y_i) | \mathbf{x}_i \in \mathcal{I}_t, y_i \in \{-1,+1\}\}$ are the training vector and label pairs. However, general visual tasks *e.g.* scene or object based classification of unconstrained images, are very challenging due to the presence of high variability due to viewpoint, lighting, pose *etc.* and require non-linear decision boundaries for competitive performance. Such non linear classifier are obtained by using the *kernel trick*, $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$, with the dual formulation,

$$\max_{\alpha} \left\{ \sum_i \alpha_i + \left(\frac{1}{2} - \frac{1}{\lambda}\right) \sum_{i,j} \alpha_i \alpha_j y_i y_j \, \mathbf{x}_i^T \mathbf{x}_j \right\}, \quad \text{s.t. } 0 \leq \alpha_i \leq \frac{1}{|\mathcal{I}_t|}. \quad (2)$$

While the dual formulation allows learning of non linear decision boundaries, the computation of classifier decision for a test vector $\mathbf{x}$, $f(\mathbf{x}) \propto \sum_i c_i K(\mathbf{x}, \mathbf{x}_i)$ (c_i being the model parameters), depends on kernel computation with *all support vectors* $\{\mathbf{x}_i \in \mathbb{R}^d | i = 1 \ldots N_{sv}\}$. Hence, the test time and space complexities becomes $O(d \times N_{sv})$ vs $O(d)$ for the linear case *i.e.* $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x}$. In practice, N_{sv} is of the order of number of training examples, and this leads to significant cost in terms of time and space. As a result of very successful histrogram based image representations *e.g.* Histogram of Oriented Gradients (HOG) and Bag-of-features, the related histogram intersection kernel $\sum_d \frac{x_d y_d}{|x_d||y_d|} \min(|x_d|, |y_d|) \;\; \forall \;\; \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ has been a natural choice, in visual tasks. This has led to many recent works proposing faster ways to compute the classifier and/or the decision function corresponding to the histogram intersection kernel *e.g.* [1, 2, 3].

In this paper, we take an (as far we know) unexplored route and show that it is possible to learn a nonlinear classifier implementing the histogram intersection kernel directly in the input space without using the dual formulation and the kernel trick, while achieving similar classification performance. We work with the primal optimization problem and view the kernel as a parametrized scoring function inducing a nonlinear boundary in the input space. The SVM primal optimization in feature mapped space reduces to the following, in input space,

$$L(\mathbf{w}) = \frac{\lambda}{2}||\mathbf{w}||_1 + \frac{1}{|\mathcal{I}_t|}\sum_i \max(0, m - y_i f(\mathbf{w}, \mathbf{x})) \quad (3a)$$

$$f(\mathbf{w}, \mathbf{x}) = \sum_d \frac{w_d}{|w_d|} min(x_d, |w_d|) \;\; (\text{as } \mathbf{x} \in \mathbb{R}^d_+). \quad (3b)$$

We show that the optimization problem becomes non convex, but remains *quasi* convex. To solve the problem, we outline an approximate optimization algorithm which starts by solving a, highly smoothed, convex approximation of the objective and then successively solves less smoothed objectives, initialized with the solution of the previous one, converging to

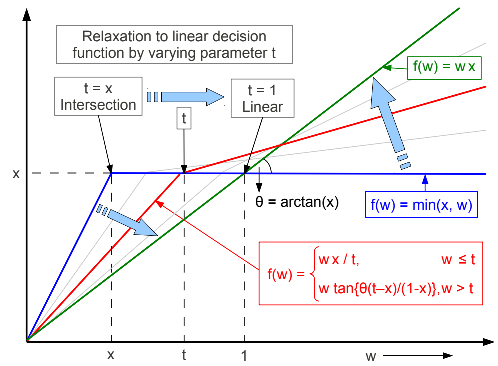

Figure 1: The linear decision function (green) can be seen as a convex relaxation of the histogram intersection like decision function (blue).

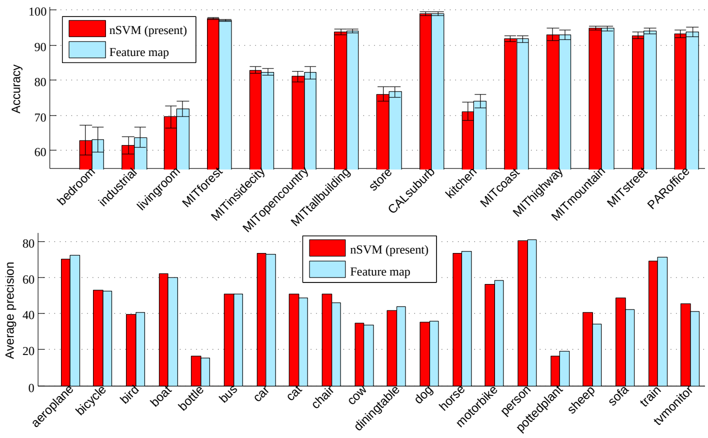

Figure 2: Performance on Scene 15 (top) and Pascal VOC 2007 dataset (bottom) for the proposed method (nSVM) and explicit feature maps [3].

	Time		Memory	
	Secs	Speedup	Kbytes	Reduction
Vedaldi and Zisserman [3]	3.8	1 (ref)	448	1 (ref)
Nonlinear SVM (present)	0.2	19×	64	7×

Table 1: Testing time and memory usage, on test set of aeroplane class of Pascal VOC 2007 dataset.

the current objective (Fig. 1). In practice, however, we find that a simple stochastic subgradient descent solver, working with the subgradient of the non convex objective, with very small steps and multiple pass over the data achieves competitive performance.

We validate the method by experiments on two challenging and widely used benchmarks datasets in image classification. While with the current approaches using explicit feature maps [2, 3] two costs are incurred at test time, (i) mapping the test vector and (ii) computing $O(d')$ dot product, in higher d'-dimensional feature space, with the proposed method we only need to perform $O(d)$, with $d < d'$, operations in original space. Our method achieves comparable performance (Fig. 2) to state-of-the-art methods for efficient SVM with histogram kernels while being faster and memory efficient (Tab. 1). Along with being conceptually interesting, our method thus leads to reduced space and time complexities.

[1] Subhransu Maji and Alexander C Berg. Max-margin additive classifiers for detection. In *ICCV*, 2009.

[2] Subhransu Maji, Alexander C Berg, and Jitendra Malik. Classification using intersection kernel support vector machines is efficient. In *CVPR*, 2008.

[3] A. Vedaldi and A. Zisserman. Efficient additive kernels using explicit feature maps. In *CVPR*, 2010.