

title

Tinghui Zhou

## 1 Introduction

The proposed approach consists of two major components: 1) providing an efficient and effective data description scheme based on the idea of Minimum Enclosing Balls (MEBs); 2) developing a new SVM variant (referred as Ball SVM for now) that is suitable for learning from data described by the MEBs.

One application of this approach is a much faster version of the image matching framework proposed in [1]. As pointed out in the paper, a big limitation of their approach is the speed, since “it requires training an SVM (with hard-negative mining) at query time ... too slow for many practical applications...”. However, noting that the negative set in their case remains the same for all query images, an interesting question would be: is it possible to describe the negative set using an alternative representation such that it’s efficient to be trained on? To do that, we propose a data description scheme using a set of MEBs, which tries to characterize the data with an efficient representation that can be utilized by our proposed SVM variant at query time.

Several benefits of this approach:

- Less memory storage. The negative set can be discarded once its corresponding MEBs are extracted.
- Faster SVM training at query time. Originally the hard-negative mining needs to be performed over millions of image patches. With our approach, the size of negative set reduces to the number of MEBs, which would be several orders of magnitude smaller.

## 2 Data description with Minimum Enclosing Balls

In this section, we describe a method for data description using a set of Minimum Enclosing Balls (MEBs).

## 2.1 Problem Formulation

Suppose we are given a set of data points  $X = \{x_1, \dots, x_N\}$ . The goal is to find a set of balls  $B = \{(c_1, r_1), \dots, (c_K, r_K)\}$ , with  $c_k$  and  $r_k$  denote the center and radius of the  $k$ -th ball respectively, such that they enclose as many data points with as little volume as possible.

To fit a MEB to each cluster, we use the method proposed by Tax and Duin [2], where they solve this problem via a formulation similar to Support Vector Classifier:

$$\begin{aligned} \min_{c,r} \quad & r^2 + C \sum_i \xi_i \\ \text{s.t.} \quad & \|x_i - c\|^2 \leq r^2 + \xi_i, \quad \xi_i \geq 0 \quad \forall i \end{aligned} \quad (1)$$

## 2.2 Solving SVDD with gradient descent

The SVDD objective given in Eq. 1 can also be formulated as follows:

$$\min_{c,r} \quad r^2 + C \sum_i \max(0, \|x_i - c\|^2 - r^2). \quad (2)$$

Let  $F$  denote the corresponding objective function. We minimize  $F$  by iteratively updating  $r$  and  $c$ . Specifically, the gradient over  $c$  given fixed  $r$  is given by

$$\frac{\partial F}{\partial c} = C \sum_i g_c(x_i),$$

where

$$g_c(x_i) = \begin{cases} 0 & \text{if } \|x_i - c\|^2 - r^2 \leq 0 \\ 2(c - x_i) & \text{otherwise.} \end{cases}$$

Similarly, the gradient over  $r$  given fixed  $c$  is

$$\frac{\partial F}{\partial r} = 2r + C \sum_i g_r(x_i),$$

where

$$g_r(x_i) = \begin{cases} 0 & \text{if } \|x_i - c\|^2 - r^2 \leq 0 \\ -2r & \text{otherwise.} \end{cases}$$

Note that in Page 5 of [2], it's claimed that  $C \geq 1$  indicates the hard-margin solution. This is because if there's  $\geq 1$  points that are not enclosed, the gradient over  $r$  as in Eq. 2.2 will be negative. So  $r$  will become larger and larger until it can enclose all the points. In other words, the valid range for  $C$  is  $(0, 1)$ .

### 3 Ball SVM

Let  $B = \{(c_1, r_1), \dots, (c_K, r_K)\}$  denote the set of MEBs describing the background (negative) data  $X_- = \{x_1, \dots, x_N\}$ , and  $n_k, k = 1, \dots, K$  denote the number of points enclosed by the  $k$ -th ball. When data from the positive set  $X_+ = \{\tilde{x}_1, \dots, \tilde{x}_P\}$ <sup>1</sup> is given, we would like to find a hyperplane that separates  $X_+$  from  $X_-$ , with  $X_-$  now approximated by  $B$ . This goal can be formally formulated as follows:

$$\min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 + C_B \sum_{k=1}^K \ell_B(c_k, r_k, \mathbf{w}) + C_P \sum_{p=1}^P \ell_P(\tilde{x}_p) \quad (3)$$

where  $\ell_B(c_k, r_k, n_k, \mathbf{w}) = \max(0, n_k(1 + r_k + \mathbf{w}^T c_k))$  is the loss induced by ball  $(c_k, r_k)$ , and  $\ell_P(\tilde{x}_p) = \max(0, 1 - \mathbf{w}^T \tilde{x}_p)$  is the loss induced by  $\tilde{x}_p$ .  $C_B$  and  $C_P$  are regularization parameters to be determined via cross-validation.

A geometric interpretation of  $\ell_B(c_k, r_k, \mathbf{w})$  is that it penalizes the learned hyperplane for assigning the ball to the positive side and/or intersecting the ball, and is weighted by the number of points enclosed by the ball. In other words,  $\ell_B(c_k, r_k, \mathbf{w}) = 0$  if and only if the hyperplane correctly assigns the center of the ball  $c_k$  to the negative side, and is at least  $1 + r_k$  distant from it.

#### 3.1 Gradient Descent for Ball SVM

Let  $F$  denote the objective function given in Eq. 3. The gradient over  $\mathbf{w}$  is given by

$$\frac{\partial F}{\partial \mathbf{w}} = \mathbf{w} + C_B \sum_{k=1}^K g_B(c_k, n_k) + C_P \sum_{p=1}^P g_P(\tilde{x}_p)$$

where

$$g_B(c_k, n_k) = \begin{cases} 0 & \text{if } r_k + \mathbf{w}^T c_k \leq -1 \\ n_k c_k & \text{otherwise,} \end{cases}$$

and

$$g_P(\tilde{x}_p) = \begin{cases} 0 & \text{if } \mathbf{w}^T \tilde{x}_p \geq 1 \\ -\tilde{x}_p & \text{otherwise,} \end{cases}$$

---

<sup>1</sup> $P = 1$  in the case of exemplar image matching.

### 3.2 K-means SVM (Baseline)

As a baseline, we use the K-means cluster centers to represent negatives. The objective function then becomes

$$\min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 + C_M \sum_{k=1}^K \ell_M(c_k, n_k, \mathbf{w}) + C_P \sum_{p=1}^P \ell_P(\tilde{x}_p), \quad (4)$$

where  $\ell_B(c_k, n_k, \mathbf{w}) = \max(0, n_k(1 + \mathbf{w}^T c_k))$ . Therefore, the gradient over  $\mathbf{w}$  becomes

$$\frac{\partial F}{\partial \mathbf{w}} = \mathbf{w} + C_M \sum_{k=1}^K g_M(c_k, n_k) + C_P \sum_{p=1}^P g_P(\tilde{x}_p)$$

where

$$g_M(c_k, n_k) = \begin{cases} 0 & \text{if } \mathbf{w}^T c_k \leq -1 \\ n_k c_k & \text{otherwise,} \end{cases}$$

and  $g_P(\tilde{x}_p)$  remains the same as previously.

## References

- [1] A. Shrivastava, T. Malisiewicz, A. Gupta, and A. A. Efros, “Data-driven visual similarity for cross-domain image matching,” in *ACM SIGGRAPH ASIA*, 2011.
- [2] D. M. Tax and R. P. Duin, “Support vector data description,” *Machine Learning*, 2004.