Basis Mapping Based Boosting for Object Detection

Haoyu Ren, Ze-Nian Li
School of Computing Science, Simon Fraser University.

Boosting family algorithms achieve considerable performance for object detection tasks. However, since the boosting procedure focuses on the hard samples gradually, it gets more and more difficult to find the weak classifiers that can efficiently improve the classification power of the strong classifier. As a result, the training may converge very slowly or can not converge at all. In this paper, we propose a novel basis mapping approach in the boosting framework to solve the above problem. The basis mapping maps the original samples into a constrained region referring to the current hard-to-classify samples (namely *basis sample*) in each boosting round, which makes positive patterns with less inner-class variation and easier to be discriminated from negative patterns. As a result, boosting on such mapped region is much more effective than that on the original sample space. In addition, we show that the weak classifier based on basis mapping is an approximation of using kernel methods, while keeping the computation cost same as the linear methods, so that both the detection accuracy and the training efficiency of the boosted classifier will be improved.

We define the basis mapping as a process that condenses the sample space into a region by referring to a hard sample, namely *basis sample*. Formally, we formulate the basis mapping which maps the original space $\mathbf{R}^m$ to a new space

$$\Phi(\mathbf{x}) = \varphi(\mathbf{x}, \mathbf{x}_{basis}) \quad \varphi : \mathbf{R}^m \times \mathbf{R}^m \to \mathbf{R}^n. \tag{1}$$

We present a kind of mapping that restricts the mapped samples to a "hypersphere" around the $\Phi(\mathbf{x}_{basis})$ with radius $2||\Phi(\mathbf{x}_{basis})||$ as

$$\forall \mathbf{x} \in \mathbf{R}^m : ||\Phi(\mathbf{x}) - \Phi(\mathbf{x}_{basis})|| \leq 2||\Phi(\mathbf{x}_{basis})||, \tag{2}$$

where $||\bullet||$ represents the sum of absolute $\mathbf{x}^{(i)}$, and $\mathbf{x}^{(i)}$ is the ith dimension of $\mathbf{x}$. We further a sufficient condition of (2) to constrain the mapping

$$||\Phi(\mathbf{x}) - \Phi(\mathbf{x}_{basis})|| \leq ||\Phi(\mathbf{x})|| + ||\Phi(\mathbf{x}_{basis})|| \leq ||2\Phi(\mathbf{x}_{basis})||. \tag{3}$$

Therefore, we use (4) as a constraint of the mapping function

$$\forall \mathbf{x} \in \mathbf{R}^m : ||\Phi(\mathbf{x})|| \leq ||\Phi(\mathbf{x}_{basis})||. \tag{4}$$

Substituting φ into (4), it can be seen that $||\varphi(\bullet, \bullet)||$ is a kind of similarity measure for vectors in $\mathbf{R}^m$. In our case, the histogram features are used. Therefore, we adopt the similarity metrics of histograms, to conduct the function $||\varphi(\bullet, \bullet)||$. Histogram intersection [1] is usually used as a similarity metric for histogram-based representations of images. We define the Histogram Intersection Mapping (HIM) as equation (5). Each dimension of the mapped vector can be calculated as

$$\Phi^{(i)}(\mathbf{x}) = \varphi^{(i)}_{HIM}(\mathbf{x}, \mathbf{x}_{basis}) = min(\mathbf{x}^{(i)}, \mathbf{x}^{(i)}_{basis}). \tag{5}$$

To evaluate the effectiveness of the HIM, we train a classifier using HOG descriptor and LogitBoost algorithm on INRIA pedestrian dataset. The sample distributions on the first selected HOG descriptor are plotted in Fig. 1, where the referred basis sample is (9, 10). It could be seen that the HIM maps the original samples into a condensed space, where the pattern distributions become much more separable. So that it is easier to learn a classification hyperplane in the mapped space.

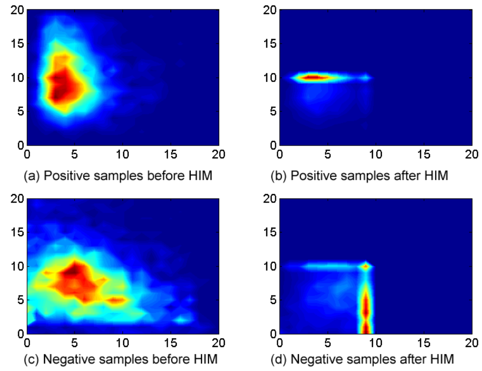

Figure 1: Sample distribution before and after HIM

Next we show that the weak classifier based on basis mapping is an approximation of applying additive kernels methods as weak classifiers in the boosting algorithm. In boosting training, we consider the learning of weak classifier f on $\mathbf{x} \in R^m$. Generally, linear classification in the implicit space can be implemented in the original space through the *kernel trick*. Given a kernel function K and the corresponding function ψ in the Mercer Condition, denote the optimal classification hyper-plane in the implicit space by $\mathbf{w}^*$, and the optimal classification function $f^*(\mathbf{x}) = \mathbf{w}^* \bullet \psi(\mathbf{x})$. In the extreme case, if there is a vector $\mathbf{x}^* \in \mathbf{R}^m$ satisfies $\psi(\mathbf{x}^*) = \mathbf{w}^*$,

$$f^*(\mathbf{x}) = \mathbf{w}^* \bullet \psi(\mathbf{x}) = \psi(\mathbf{x}^*) \bullet \psi(\mathbf{x}) = K(\mathbf{x}, \mathbf{x}^*). \tag{6}$$

So the only problem is to find out such an $\mathbf{x}^*$. Unfortunately, in most of the cases, ψ is not invertible or even ψ itself could not be explicitly described. But in boosting framework, we could approximate $\mathbf{x}^*$ by selecting one of the current training samples $\mathbf{x}'$. The optimal f^* is then approximated using

$$f^*(\mathbf{x}) \approx f(\mathbf{x}) = \mathbf{w}' \bullet \psi(\mathbf{x}) = K(\mathbf{x}, \mathbf{x}'), \tag{7}$$

where $\mathbf{w}' = \psi(\mathbf{x}')$. This implies that by referring to an appropriate sample $\mathbf{x}'$, the linear classification in the implicit space could be approximated.

Then we turn back to the basis mapping. HIM is constructed based on an additive kernel, and each dimension is independent with each other, so (7) could be written as

$$f(\mathbf{x}) = K(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{m} \varphi(\mathbf{x}^{(i)}, \mathbf{x}'^{(i)}). \tag{8}$$

As mentioned above, in the boosting framework, we could use $\mathbf{x}'$ to approximate $\mathbf{x}^*$. This is achieved by evaluating different hard samples in current training stage to get the best one $\mathbf{x}_{basis}$. Then (8) is achieved by (9)

$$f(\mathbf{x}) = \sum_{i=1}^{m} \varphi(\mathbf{x}^{(i)}, \mathbf{x}'^{(i)}) = \sum_{i=1}^{m} \varphi(\mathbf{x}^{(i)}, \mathbf{x}^{(i)}_{basis}). \tag{9}$$

We further fit a linear classifier based on (10) as the final weak classifier

$$f(\mathbf{x}) = \sum_{i=1}^{m} a^{(i)} \varphi(\mathbf{x}^{(i)}, \mathbf{x}^{(i)}_{basis}) + b. \tag{10}$$

(10) is the linear classification on the mapped space $\Phi(\mathbf{x})$ around the basis sample. So we get the conclusion that the basis mapping is an approximation of additive kernel classification in the original space, which significantly has better discrimination power than simple decision stump or linear weak classifiers. In general, the performance of a boosted classifier mainly depends on the weak classifiers [2]. So the proposed basis mapping will contribute to the overall accuracy of the boosted classifier.

[1] S. Maji, A. C. Berg, and J. Malik. Classification using intersection kernel support vector machines is efficient. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[2] R. E. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. *Machine learning*, 37:297–336, 1999.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.