Generalized RBF feature maps for Efficient Detection

Sreekanth Vempati[1]
http://web.iiit.ac.in/~v_sreekanth

Andrea Vedaldi[2]
http://www.robots.ox.ac.uk/~vedaldi

Andrew Zisserman[2]
http://www.robots.ox.ac.uk/~az

C. V. Jawahar[1]
http://www.iiit.ac.in/~jawahar

[1] Center for Visual Information Technology,
IIIT Hyderabad, India
http://cvit.iiit.ac.in

[2] Visual Geometry Group,
Oxford University, UK
http://www.robots.ox.ac.uk/∼vgg

Kernel methods yield state-of-the-art performance in certain applications such as image classification and object detection. For these applications, the gold-standard kernels are the so called *generalized radial-basis function (RBF) kernels.* A typical example of one such kernel is the exponential-χ^2 kernel

$$K(\mathbf{x},\mathbf{y}) = e^{-\frac{1}{2\sigma^2}\chi^2(\mathbf{x},\mathbf{y})}, \quad \chi^2(\mathbf{x},\mathbf{y}) = \frac{1}{2}\sum_{l=1}^{d}\frac{(x_l-y_l)^2}{x_l+y_l}.$$

These kernels combine the benefits of two other important classes of kernels: the homogeneous additive kernels (e.g. the χ^2 kernel) and the RBF kernels (e.g. the exponential kernel).

However, large scale problems require machine learning techniques of at most linear complexity and these are usually limited to linear kernels. Recently, Maji and Berg [2] and Vedaldi and Zisserman [4] proposed explicit feature maps to approximate the additive kernels (intersection, χ^2, etc.) by linear ones, thus enabling the use of fast machine learning technique in a non-linear context. An analogous technique was proposed by Rahimi and Recht [3] for the translation invariant RBF kernels. In this paper, we complete the construction and combine the two techniques to obtain explicit feature maps for the *generalized RBF kernels*.

The generalized RBF kernels extend the RBF kernels to use a metric not necessarily Euclidean. Recall that, for any Positive Definite kernel $K(\mathbf{x},\mathbf{y})$, the equation

$$D^2(\mathbf{x},\mathbf{y}) = K(\mathbf{x},\mathbf{x}) + K(\mathbf{y},\mathbf{y}) - 2K(\mathbf{x},\mathbf{y}) \tag{1}$$

defines a corresponding squared metric. Given an RBF kernel $K_{\mathrm{RB}}(\mathbf{x},\mathbf{y}) = k(\|\mathbf{x}-\mathbf{y}\|_2^2)$, one can then obtain a corresponding generalized variant

$$K_{\mathrm{RB}D^2}(\mathbf{x},\mathbf{y}) = k(D^2(\mathbf{x},\mathbf{y})). \tag{2}$$

Constructing an approximate feature map for (2) involves two steps. The first step involves the construction of the feature map $\hat{\Psi}(\mathbf{x})$ for the kernel (as in 1) by using the method of [4] to approximate distance measure D^2. The second step involves the construction of the feature map for $K_{\mathrm{RB}D^2}$ using random Fourier features [3]. The complete procedure of computing the approximate feature vector $\hat{\Psi}_{\mathrm{RB}\chi^2}(\mathbf{x})$ from the given feature vector $\mathbf{x}$ for the exp-χ^2 kernel is given in Figure 1.

A limitation of the random Fourier features is the relatively large number of projections required to obtain good accuracy. The simplest way to select useful random Fourier features is to use an appropriate regularizer for SVM training. In this paper, we considered two formulations based on l^1 regularization to select useful random Fourier features. Also, there exists efficient implementations as a part of LIBLINEAR [1] for these formulations. We refer to these l^1 regularization based formulations as $\mathrm{SVM}^{\mathrm{sparse}}$ and $\mathrm{LR}^{\mathrm{sparse}}$, and to the standard SVM as $\mathrm{SVM}^{\mathrm{dense}}$.

We evaluate the proposed feature maps as part of the construction of an object detector on the PASCAL VOC 2007 data. We work on top of the state-of-the-art multiple-stage detector proposed in [5] using only PHOG features. The multiple stages are a cascade of a linear, χ^2 and exp-χ^2 detector. We use our feature map for exp-χ^2 to speed-up the third stage of the cascade (exp-χ^2), which is also noted to be the bottleneck in [5]. We show that feature maps can improve on fast additive kernels, and investigate the trade-offs in complexity and accuracy. Figure 2 shows the comparison of performance and testing time using exponential and additive kernels for both exact and approximate versions. It is found that both the dense and sparse approximations perform nearly as well as the exact exp-χ^2 kernel. The testing time of the approximate sparse SVM is two to three times faster than the dense SVM.

Compute a $2m$ dimensional approximate finite feature map for the exponential-χ^2 kernel $K(\mathbf{x},\mathbf{y}) = \exp(-\frac{1}{2\sigma^2}\chi^2(\mathbf{x},\mathbf{y}))$.

Preprocessing: Draw m random vectors $\boldsymbol{\omega}$ sampled from a $(2n+1)d$ isotropic Gaussian of variance $1/\sigma^2$.

Given: A vector $\mathbf{x} \in \mathbb{R}^d$.
Compute: The approximate feature map $\hat{\Psi}_{\mathrm{RB}\chi^2}(\mathbf{x})$

1: Construct the $2n+1$ dimensional vector $\hat{\Psi}(\mathbf{x})$ by setting for $j = 0,\ldots,2n$

$$[\hat{\Psi}(\mathbf{x})]_j = \begin{cases} \sqrt{xL\,\mathrm{sech}(0)}, & j = 0, \\ \sqrt{2xL\,\mathrm{sech}(\pi\frac{j+1}{2}L)}\cos\left(\frac{j+1}{2}L\log x\right) & j > 0 \text{ odd}, \\ \sqrt{2xL\,\mathrm{sech}(\pi\frac{j}{2}L)}\sin\left(\frac{j}{2}L\log x\right) & j > 0 \text{ even}, \end{cases} \tag{3}$$

2: Construct the $2m$ dimensional vector $\hat{\Psi}_{\mathrm{RB}\chi^2}(\mathbf{x})$ by setting for $j = 1,\ldots,2m$

$$[\hat{\Psi}_{\mathrm{RB}\chi^2}(\mathbf{x})]_j = \begin{cases} \frac{1}{\sqrt{m}}\cos\left(\boldsymbol{\omega}_{\frac{j+1}{2}}^\top \hat{\Psi}(\mathbf{x})\right), & j \text{ odd}, \\ \frac{1}{\sqrt{m}}\sin\left(\boldsymbol{\omega}_{\frac{j}{2}}^\top \hat{\Psi}(\mathbf{x})\right), & j \text{ even}. \end{cases} \tag{4}$$

Figure 1: **Feature map for the exponential-χ^2 kernel.** The resulting vector is $2m$ dimensional. Here n controls the χ^2 approximation, and is typically chosen as a small number, e.g. $n = 1$ (and in this case $L \approx 0.8$, see [4] for details on how to choose this parameter). The algorithm requires only two modifications for any other RBF-D^2 kernel. First, (3) should be adjusted so as to match the metric D^2 (closed forms are given in [4]). Second, the projections $\boldsymbol{\omega}_j$ should be sampled from the density $\kappa_{\mathrm{RB}}(\boldsymbol{\omega})$ corresponding to the desired RBF profile.

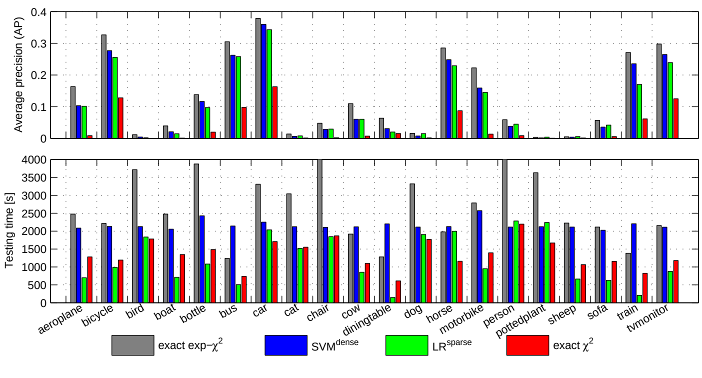

Figure 2: Comparison of performance for exponential and additive kernels along with their approximations for twenty classes of the VOC 2007 challenge.

[1] R.-E. Fan, K.-W. Chang, C.-J Hsieh, X.-R. Wang, and C.-J. Lin. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9, 2008.

[2] S. Maji and A. C. Berg. Max-margin additive classifiers for detection. In *Proc. ICCV*, 2009.

[3] A. Rahimi and B. Recht. Random features for large-scale kernel machines. In *Proc. NIPS*, 2007.

[4] A. Vedaldi and A. Zisserman. Efficient additive kernels via explicit feature maps. In *Proc. CVPR*, 2010.

[5] A. Vedaldi, V. Gulshan, M. Varma, and A. Zisserman. Multiple kernels for object detection. In *Proc. ICCV*, 2009.