GCS Kernel For SVM-Based Image Recognition

Sabri Boughorbel[1], Jean-Philippe Tarel[2]
François Fleuret[3], and Nozha Boujemaa[1]

[1] IMEDIA Group, INRIA Rocquencourt, 78153 Le Chesnay, France
[2] DESE, LCPC, 58 Bd Lefebvre, 75015 Paris, France
[3] CVLAB, EPFL, 1015 Lausanne, Switzerland

Abstract. In this paper, we present a new compactly supported kernel for SVM based image recognition. This kernel which we called Geometric Compactly Supported (GCS) can be viewed as a generalization of spherical kernels to higher dimensions. The construction of the GCS kernel is based on a geometric approach using the intersection volume of two n-dimensional balls. The compactness property of the GCS kernel leads to a sparse Gram matrix which enhances computation efficiency by using sparse linear algebra algorithms. Comparisons of the GCS kernel performance, for image recognition task, with other known kernels prove the interest of this new kernel.

1 Introduction

Support Vector Machine is one of the successful kernel methods that has been derived from statistical learning theory. Incremental version of SVM has been introduced in [1] allowing faster on-line learning. We focus in this paper on how to improve the computational efficiency of SVM training using compactly supported kernels. We propose a new compactly supported kernel. In §2, we introduce and derive the new kernel, we named Geometric Compactly Supported (GCS) kernel. We provide experimental results proving that GCS kernel leads to good accuracy for image recognition task while being computationally efficient.

2 CS Kernel

A kernel $\varphi(x, y)$ is said to be compactly supported (CS) whenever it vanishes from a certain cut-off distance $2r$ between x and y.

$$\varphi(x,y) = \begin{cases} \varphi(x,y) \text{ if } \|x-y\| < 2r \\ 0 \text{ if } \|x-y\| \geq 2r \end{cases}$$

The main advantage of CS kernels is that their Gram matrices $[\varphi(x_i, x_j)]_{i,j}$ are sparse. If such matrices are associated with a linear system they can be solved efficiently using sparse linear algebra methods. Genton was the first to point out the possible gain in efficiency provided by CS kernels with machine learning techniques [2].

Triangular $K_T(x,y)$	$1-\frac{\lvert x-y\rvert}{2r},\ \lvert x-y\rvert<2r$ positive definite in $\mathbb{R}$
Circular	$\arccos(\frac{\lVert x-y\rVert}{2r})-\frac{\lVert x-y\rVert}{2r}\sqrt{1-\left(\frac{\lVert x-y\rVert}{2r}\right)^2},\ \lVert x-y\rVert<2r$ positive definite in $\mathbb{R}^2$
Spherical	$1-\frac{3}{2}\frac{\lVert x-y\rVert}{2r}-\frac{1}{2}\left(\frac{\lVert x-y\rVert}{2r}\right)^3,\ \lVert x-y\rVert<2r$ positive definite in $\mathbb{R}^3$

Table 1. Examples of compactly supported kernels in $\mathbb{R}$, $\mathbb{R}^2$ and $\mathbb{R}^3$.

Triangular, circular and spherical kernels, see Tab. 1 for definitions, which are used in geostatistic applications, have been also studied in the context of machine learning [2]. However, the use of these kernels is limited to dimensions from one to three, since they are not positive definite for higher dimensions. Indeed, we provide next a counterexample given in [3] proving that the triangular kernel $K_T(x,y)$ is not positive definite for features living in $\mathbb{R}^2$. Let's take $x,\ y \in \mathbb{R}^2$, thus $|x-y|$ is replaced by $\|x-y\|$ (the L-2 norm of $\mathbb{R}^2$) in definition of $K_T(x,y)$, see Tab. 1. By choosing $x_{i,j} \in \mathbb{R}^2$ from a 8×8 square grid of spacing $\sqrt{2}r$, and $c_{i,j}$ alternatively $+1$ and -1, we have:

$$\sum_{j_1,j_2=1}^{8}\sum_{k_1,k_2=1}^{8} c_{j_1,j_2}c_{k_1,k_2}K_T(x_{j_1,j_2},x_{k_1,k_2}) = -1.6081 < 0$$

Therefore K_T is not positive definite on $\mathbb{R}^2$. A few attempts have been carried out to derive compactly support (CS) kernels for high dimensions [2]. In [4], experimentations using a CS kernel are described proving that such kernel does not give usually good performances in the context of non-linear regression (SVR) of functions. Notice that just truncating positive definite kernels does not generally lead to positive definite kernels. We now introduce a new compactly supported kernel that can be viewed as an extension of triangular, circular and spherical kernels to higher dimensions.

3 The GCS kernel

The derivation of the new Geometric Compactly Supported (GCS) kernel is based on the intersection of two n-dimensional balls. Basically, we use the fact that the intersection volume of two n-dimensional balls leads to a compactly supported and positive definite kernel.

The properties of positiveness and compactness of the GCS kernel are presented in the following proposition.

Proposition 1. *Let x and $y \in \mathbb{R}^n$, $\Psi_n(x,y)$ denotes the intersection volume of two balls having the same radius r, centered in x and y. Thus $\Psi_n(x,y)$ is compactly supported and positive definite kernel.*

Proof. Intersection volume $\Psi_n(x, y)$ can be written as the following integral:

$$\begin{aligned}\Psi_n(x,y) &= \int_{a\in\mathbb{R}^n} \mathbb{1}_{\{\|x-a\|\leq r\}}\mathbb{1}_{\{\|y-a\|\leq r\}}da \\ &= \int_{a\in\mathbb{R}^n} \underbrace{f_{a,r}(x)f_{a,r}(y)}_{\text{positive definite}}\,da\end{aligned} \tag{1}$$

where $f_{a,r}(z) = \mathbb{1}_{\{\|z-a\|\leq r\}}(z)$. Thus, $\Psi_n(x, y)$ is a positive definite kernel as a mixture of positive definite kernels. Each time the feature point $a \in \mathbb{R}^n$ is in the intersection of the two balls of centers x and y, the function under the integral equals to one, otherwise it equals to zero. As a consequence, the summation over $a \in \mathbb{R}^n$ gives the intersection volume of the two balls. Whenever $\|x-y\| > 2r$, the balls intersection is empty, so $\Psi_n(x, y) = 0$. Therefore, Ψ_n is a CS and positive definite kernel.

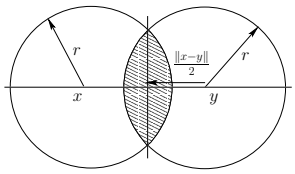


Fig. 1. Intersection volume calculus (case $n = 2$).

To simplify notations, we omitted the radius hyper-parameter r in $\Psi_n(x, y)$. Next, we derive a more explicit formula for the GCS kernel which allows to compute the kernel in a fast recursive way. The volume $\mathcal{V}_n(r)$ of a n-dimensional ball with radius r can be calculated recursively as follows:

$$\mathcal{V}_n(r) = \int_0^r \mathcal{V}_{n-1}(\sqrt{r^2 - t^2})dt, \text{ for } n \geq 2 \tag{2}$$

Thus, a general formula of the n-dimensional ball volume can be written as follows:

$$\mathcal{V}_n(r) = \begin{cases} \frac{1}{(\frac{n}{2})!}\pi^{\frac{n}{2}}r^n & \text{if } n \text{ is even} \\ \frac{2^{\frac{n+1}{2}}}{n!!}\pi^{\frac{n-1}{2}}r^n & \text{if } n \text{ is odd} \end{cases}$$

where $n!! = n(n-2)(n-4)\ldots 1$ is the double factorial when n is odd. The same recursive approach as in (2) can be used for the derivation of the intersection volume $\Psi_n(x, y)$. We can see in Fig. 1 that the median plane of x and y is a symmetric plane for the intersection volume, so $\Psi_n(x, y)$ can be written, similarly to (2), but now integrating from $\frac{\|x-y\|}{2}$ to r rather than from 0 to r:

$$\begin{aligned}\Psi_n(x,y) &= 2\int_{\frac{\|x-y\|}{2}}^{r} \mathcal{V}_{n-1}(\sqrt{r^2-t^2})dt \\ &= A_n(r)\int_{\frac{\|x-y\|}{2}}^{r} \frac{1}{r}\left(1-\left(\frac{t}{r}\right)^2\right)^{\frac{n-1}{2}} dt \\ &= A_n(r)\Phi_n(x,y)\end{aligned}$$

where $A_n(r)$ is only a normalization factor, which is independent of x and y. SVM decision function is invariant to the product of any positive constant with the kernel. Thus, in the following, we get rid of A_n, and we denote by $\Phi_n(x,y)$ the remaining term which defines the GCS kernel $K_{GCS}(x,y)$.

By variable changing $t = r\sin\theta$, $\Phi_n(x,y)$ can be written as follows:

$$\Phi_n(x,y) = \int_{\arcsin(\frac{\|x-y\|}{2r})}^{\frac{\pi}{2}} (\cos\theta)^n d\theta \tag{3}$$

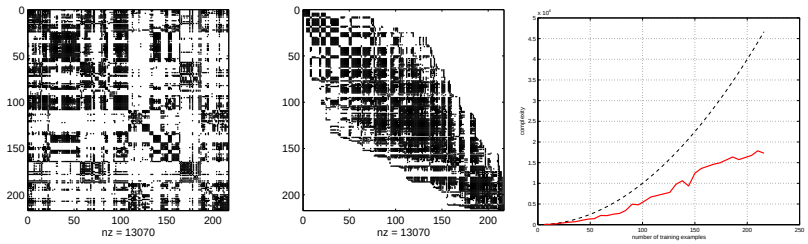
By integrating by part, we are able to derive a recursive computation of $\Phi_n(x,y)$. For a dimension n, we define $\varphi_{n,k}(x,y)$ as the value of $\Phi_n(x,y)$ at kth iteration. As a consequence, we have $\Phi_n(x,y) = \varphi_{n,n}(x,y)$. We prove easily that for $\|x-y\| < 2r$:

$$\begin{cases} \varphi_{n,k}(x,y) = \frac{k-1}{k}\varphi_{n,k-2}(x,y) - \frac{1}{k}\frac{\|x-y\|}{2r}\left(1-\left(\frac{\|x-y\|}{2r}\right)^2\right)^{\frac{k-1}{2}} \\ \varphi_{n,2}(x,y) = \arccos(\frac{\|x-y\|}{2r}) - \frac{\|x-y\|}{2r}\sqrt{1-\left(\frac{\|x-y\|}{2r}\right)^2} \\ \varphi_{n,1}(x,y) = 1-\frac{\|x-y\|}{2r} \end{cases} \tag{4}$$

For $\|x-y\| \geq 2r$, $\varphi_{n,k}(x,y) = 0$. Notice that Φ_1 on $\mathbb{R}$ is the triangular kernel, Φ_2 on $\mathbb{R}^2$ is the circular kernel and Φ_3 on $\mathbb{R}^3$ is the spherical kernel. Thus, it can be viewed as the generalization of the spherical kernel to higher dimensions. The GCS kernel Φ_n is positive definite on $\mathbb{R}^n$ as proved before. Nevertheless $\varphi_{n,k}$, $1 \leq k \leq n-1$ are not positive definite kernels on $\mathbb{R}^n$. Indeed, the counterexample given in beginning of §. 2, tells us that $\varphi_{2,1}$, which is the triangular kernel for data living in $\mathbb{R}^2$, is not positive definite. Functions $\varphi_{n,k}$ are only intermediaries functions useful to compute the GCS kernel Φ_n recursively.

4 Complexity Reduction

One of the interest of GCS kernel is the reduction of algorithmic complexity. Actually, we can take advantages of the Gram matrix sparsity to enhance training computation stage. Fig. 2-a shows the sparsity of the GCS Gram matrix. Fig .2-b presents the rearrangement of the Gram matrix using a Cuthill-McKee permutation which leads to a banded matrix [5]. Fig. 2-c presents computational savings for the SVM training stage: it plots the complexity of the quadratic term of the SVM dual problem with respect to the size of training sample n. The usual



(a) (b) (c)

Fig. 2. (a)-(b) Symmetric reverse Cuthill-McKee permutation of sparse GCS Gram Matrix for radius hyperparameter $r = 0.4$. (c) Complexity of quadratic term of the SVM dual problem using banded matrix representation with respect to the size of training sample.

computation leads to a quadratic complexity of $O(n^2)$, for each iteration. The use of a sparse Gram matrix leads to a better complexity of $O(nB(n))$, where $B(n) < n$ is the bandwidth of the rearranged Gram matrix. Moreover, after rearranging the Gram matrix, only the banded Gram matrix is kept in memory.

5 Experiments

Fig. 3. Image examples from the 6 classes used for experiments.

Figure 3 shows some images from Corel database that we used for experiments. This database gathers 3200 images in 6 different classes. Images are represented by 64-bin RGB color histogram. We compare 4 kernels namely: Laplace kernel $K_{Lapl}(x, y) = exp(-\frac{\|x^a - y^a\|}{\sigma})$, Polynomial kernel $K_{\text{Poly}}(x, y) = (1 + \langle x^a \cdot y^a \rangle)^d$, K_{GCS} defined by (4) and $K_{CS} = K_{GCS} K_{Lapl}$. The parameter a applies a non-linear remapping of feature space which is shown to improve drastically performances for image recognition task. We set $a = 0.25$ as in [6]. For K_{GCS}, we tune the radius r. For K_{CS}, we set the radius to $r = 4$ such that we obtain Gram matrix sparsity of 90%, then we tune the σ of inside K_{Lapl}. Table 2 shows that K_{GCS} and K_{CS} yield to similar results to K_{Lapl} which known as the best kernel in the state of the art. Optimal radius of the GCS kernel does not give sparse enough Gram matrices, however combined with K_{Lapl}, K_{CS} has a sparsity of 90%. Table 3 shows the class-confusion matrix obtained with K_{CS}, values on the diagonal gives the number of correctly classified images.

	valid. err.	test err.
K_{Lapl}	25.32±0.19	25.34±0.42
K_{GCS}	25.19±0.30	**25.07±0.64**
K_{Poly}	27.81±0.17	27.92±0.43
K_{CS}	26.12± 0.37	25.30±0.44

Table 2. Validation and test errors comparisons for the different kernels on Corel database.

	Animals	Birds	Buildings	Night scenes	Roses	Water scenes
Animals	536	104	80	69	128	99
Birds	25	62	2	17	35	24
Buildings	37	12	291	95	25	27
Night scenes	13	8	28	82	12	25
Roses	140	189	34	53	542	56
Water scenes	34	36	29	45	27	172

Table 3. Class-confusion matrix obtained for kernel K_{CS} on Corel database.

6 Conclusion

In this paper, we have presented a new compactly supported kernel namely the GCS kernel. The construction of this kernel is based on a geometric approach using the intersection volume of two n-dimensional balls. Hence, the GCS kernel can be viewed as the generalization of spherical kernels to higher dimensions. It yields to good recognition performance when the radius is tuned similar to that of Laplace kernel, however optimal radius does not lead to sparse Gram matrix. To recover sparsity, we combine the GCS kernel and the Laplace kernel to obtain an efficient kernel with a highly sparse Gram matrix of 90% of zeros.

References

1. C. Gentile, "A new approximate maximal margin classification algorithm," *Journal of Machine Learning Research*, vol. 2, pp. 213–242, 2001.
2. M. Genton, "Classes of kernels for machine learning: a statistics perspective," *Journal of Machine Learning Research*, vol. 2, pp. 299–312, 2002.
3. N. Cressie, *Statistics for Spatial Data*, New York, Wiley, 1993.
4. B. Hamers, J. A.K. Suykens, and B. De Moor, "Compactly supported rbf kernels for sparsifying the gram matrix in ls-svm regression models," in *International Conference on Artificial Neural Networks (ICANN'02)*, Madrid Spain, 2002, pp. 720–726.
5. J. Gilbert, C. Moler, and R. Schreiber, "Sparse matrices in matlab: design and implementation," *SIAM Journal on Matrix Analysis*, vol. 13, no. 1, pp. 333–356, 1992.
6. O. Chapelle, P. Haffner, and V. Vapnik, "Svms for histogram-based image classification," *IEEE Transactions on Neural Networks*, vol. 9, 1999.