Kernel Learning for Extrinsic Classification of Manifold Features

Raviteja Vemulapalli, Jaishanker K. Pillai and Rama Chellappa

Department of Electrical and Computer Engineering, University of Maryland, College Park.

Motivation

- Manifold features
  - Linear subspaces and symmetric positive definite matrices.
  - Used in many computer vision applications like image set-based recognition, texture classification, pedestrian detection, etc.
- Popular approaches like LDA, PLS, SVM, etc. can be extended to manifold features by using kernels.
- How can we find good kernels for the classification of manifold features?

Overview

- **Approach:** Use kernel learning approach to learn the kernel and the classifier jointly.
- **Criteria:**
  - For good classification performance, the risk function associated with the classifier in the mapped space should be minimized.
  - Since the features lie on a manifold with well defined structure, the mapping should be structure preserving.

The second criterion acts as a regularizer in learning the kernel.

$$\min_{W,K} \Gamma_c(W,K) + \lambda \Gamma_s(K)$$

Classifier risk function — Manifold structure cost

W: Classifier parameters, K: Kernel matrix

- **Manifold structure cost:**

$$\Gamma_s(K) = \sum_{1 \le i<j \le N} \left(K_{ii} + K_{jj} - K_{ij} - K_{ji} - d_{ij}^2 \right)^2.$$

Distance induced by the kernel K — Manifold geodesic distance

Extrinsic SVM - Multiple Kernel Learning

- Kernel K parameterized as a linear combination of known base kernels.

$$K = \sum_{m=1}^{M} \mu_m K^m; \ \ \mu_m \ge 0.$$

- SVM cost:

$$J_1(\vec{\mu}) = \max_{\vec{\alpha} \in \Omega} \left(\vec{\alpha}^T \vec{1} - \frac{1}{2} \vec{\alpha}^T \left(\vec{y}\vec{y}^T \circ \sum_{m=1}^{M} \mu_m K^m_{tr,tr} \right) \vec{\alpha} \right),$$

where $\Omega = \{\vec{\alpha} \ | \vec{0} \le \vec{\alpha} \le C\vec{1}, \vec{\alpha}^T\vec{y} = 0\}$, $\vec{y}$ are the training labels and $\vec{\alpha}$ are the SVM dual variables.

- Manifold structure cost:

$$J_2(\vec{\mu}) = \| \zeta \|_F^2 = \sum_{1 \le i<j \le N_{tr}} \left(\sum_{m=1}^{M} \mu_m \left(K^m_{ii} + K^m_{jj} - K^m_{ij} - K^m_{ji} \right) - d_{ij}^2 \right)^2$$

Distance induced by the base kernel K^m

- Kernel learning problem:

$$\min_{\vec{\mu} \ge 0} J(\vec{\mu}), \text{ where } J(\vec{\mu}) = J_1(\vec{\mu}) + \lambda J_2(\vec{\mu}).$$

- $J(\vec{\mu})$ is a differentiable convex function of $\vec{\mu}$.
- Kernel learning problem can be solved using gradient-based methods.

Base Kernels and Geodesic Distances

- **Linear subspaces - Grassmann manifold:**

$$K_P^{rbf}(S_1,S_2) = \exp(-\gamma \, \| Y_1Y_1^T - Y_2Y_2^T \|_F^2)$$

$$K_P^{poly}(S_1,S_2) = (\gamma \, \| Y_1^T Y_2 \|_F^2)^d$$

Y_1 and Y_2 are orthonormal matrices whose columns span the subspaces S_1 and S_2 respectively.

Distance: $d(S_1,S_2) = \| \vec{\theta} \|_2$, where $\vec{\theta} = [\theta_1, \ldots, \theta_n]$ are the principal angles between the subspaces S_1 and S_2.

- **Symmetric positive definite matrices:**

$$K_{log}^{rbf}(C_1,C_2) = \exp(-\gamma \, \| \log(C_1) - \log(C_2) \|_F^2)$$

$$K_{log}^{poly}(C_1,C_2) = (\gamma \, \text{trace}\,[\log(C_1)^T \log(C_2)])^d$$

Distance: $d(C_1,C_2) = \left(\sum_{i=1}^{d} ln^2(\lambda_i) \right)^{1/2}$, where λ_i are the generalized Eigen values of C_1 and C_2.

Experimental Results

- Several RBF and polynomial kernels used.
- YouTube: 3 image sets for training and 6 for testing.
- ETH80: 5 image sets for training and 5 for testing.

Table 1: Recognition rates for image set-based face and object recognition tasks using linear subspaces

Dataset	NN	S-MKL[1]	GDA[2]	Proj+PLS[3]	Proposed approach
YouTube	62.8	64.3	65.7	67.7	70.8
ETH80	93.2	93.7	92.8	95.3	96.0

Table 2: Recognition rates for image set-based face and object recognition tasks using covariance features

Dataset	NN	S-MKL[1]	CDL-LDA[3]	CDL-PLS[3]	Proposed approach
YouTube	40.7	69.7	67.5	70.1	73.2
ETH80	92.7	93.7	94.5	96.5	98.2

References

1. A. Rakotomamonjy, F. R. Bach, S. Canu, and Y. Grandvalet. SimpleMKL. JMLR, 9:2491–2521, 2008.
2. J. Hamm and D. D. Lee. Grassmann Discriminant Analysis: a Unifying View on Subspace-Based Learning. In ICML, 2008.
3. R. Wang, H. Guo, L. S. Davis, and Q. Dai. Covariance Discriminative Learning: A Natural and Efficient Approach to Image Set Classification. In CVPR, 2012.