INTEGRATING SPATIAL PROXIMITY WITH MANIFOLD LEARNING

Wonkook Kim, Student Member, IEEE, and Melba M. Crawford, Fellow, IEEE

School of Civil Engineering
and
Laboratory for Applications of Remote Sensing,
Purdue University
West Lafayette, IN 47907

Dimension reduction is a useful preprocessing step for many types hyperspectral image analysis, including visualization, regression, clustering and classification. By dimension reduction, high dimensional data are mapped into a lower dimensional space while the important features of the original data are preserved according to a given criterion. Although linear dimension reduction methods such as principal component analysis (PCA), and multidimensional scaling (MDS) have been satisfactory for the analysis of multispectral images, increase in spectral resolution enables the characterization of nonlinearity in hyperspectral data, which calls for the use of nonlinear dimension reduction methods such as manifold learning. Most manifold learning algorithms [1-5] only deal with the attributes of a data point and assume that samples are statistically independent, ignoring spatial correlation between samples that are within close proximity. In this paper, we develop a manifold learning scheme where spatial proximity of data samples is incorporated into original spectral similarity through a composite kernel.

The use of composite kernels has yielded successful results in applications such as hypertext categorization [6] and hyperspectral image classification [7]. In these investigations, a composite kernel is used to integrate disparate information (link structure information for hypertext, and spatial information for hyperspectral data) to generate an improved kernel for support vector machines. Although the use of a composite kernel has been investigated for the classification, it has not been exploited for the manifold learning of spatial data. By accommodating spatial correlation, the representation of data is enhanced, which is potentially useful for the subsequent analysis, including classification.

In the original setting of kernel based manifold learning methods, similarity between samples is defined in the entries of a kernel matrix based on spectral values of data samples. The key idea of the proposed composite kernel design is based on the observation that spatial proximity of the data samples affects the similarity through an interaction with the spectral proximity. Since not all spatially close samples are necessarily similar objects, spatial proximity alone cannot provide adequate information to characterize similarity. Instead, it can be said that spatial proximity increases existing similarity, and the magnitude of the increase is given as a function of both the spectral similarity and the spatial similarity. This motivates the use of an interaction kernel. When a spectral kernel, $K_{Spectral}$, and a spatial kernel, $K_{Spatial}$, are derived from the spectral proximity and the spatial proximity of a pixel respectively, the interaction kernel, K_{Inter}, is obtained by combining the two kernels through a composite kernel operation, f.

$$K_{Inter} = f(K_{Spectral}, K_{Spatial})$$

The interaction kernel is then multiplied by the original spectral kernel to obtain a final composite kernel, K'.

$$K' = K_{Spectral} \cdot K_{Inter}$$

A set of manifold coordinates is derived from the composite kernel for a given target dimension. Although non-kernel based methods such as isometric mapping (Isomap) [1], Laplacian eigenmap (LE) [2], locally linear embedding (LLE) [3], and diffusion maps [4] do not involve the use of kernels, it is known that they can be rearranged with kernel matrices [8] and described as kernel PCA (KPCA) [5]. Therefore, in this paper, we apply the proposed kernel to kernel PCA, which can be expanded to the other methods.

For the base kernels, $K_{Spectral}$ and $K_{Spatial}$, heat kernels are used to define similarity between samples. A heat kernel is a natural choice because it has favorable behavior relative to the geometric properties of Riemannian manifolds [9]. To find the optimal parameters involved in the proposed composite kernel, we consider a parameter tuning method which exploits existing class information. The parameters are tuned in such way that the kernel maximizes the separation between different classes, while suppressing the variation of the samples in the same class.

Evaluation of resulting manifold coordinates is performed with the nearest-neighbor (NN) classification method. Two scenarios are considered in the experiment with two sets of labeled data for an image: (a) spatially collocated labeled data and (b) spatially disjoint labeled data. In a usual sampling scheme, the training and the test data are sampled randomly from a pool of labeled samples. However, this inflates classification accuracies when the labeled samples are spatially collocated because spectral variation is often not significant within the region within which they are collocated. In our second experiment, spatially disjoint groups of the labeled samples are first identified, and the samples from different areas are used for training and testing. Therefore, it can be said that the spatially disjoint experiment measures the generalized performance of the coordinates for other areas that the existing labeled samples do not cover. For the two scenarios, the classification performance of the proposed method is compared to the results of PCA and KPCA.

Preliminary experiments were performed with a wide range of parameters. Results show that among the parameter values, an optimal parameter set could be identified that performs well both for the collocated and the disjoint case. In both experiments, SKPCA showed consistent improvement over the other non-spatial methods, whereas KPCA only produced comparable accuracies to the PCA results in the optimal case. The classification results of the disjoint experiment indicate that a small number of features in SKPCA can recover the classification accuracy that is obtained with the set of full bands, while the other methods show a significant decrease in accuracy when a small number of features was used. Experimental results indicate that better data representation can be obtained for spatially correlated data when spatial proximity is considered in the similarity between samples.

REFERENCES

[1] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[2] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003.

[3] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by local linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[4] R. R. Coifman and S. Lafon, "Diffusion maps," *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 5–30, Jul. 2006.

[5] B. Scholkopf, A. J. Smola, and K. R. Muller, "Kernel principal component analysis," *Lecture Notes in Computer Science*, vol. 1327, pp. 583–588, Oct. 1997.

[6] T. Joachims and N. Cristianini, "Composite Kernels for Hypertext Categorisation," in *Proc. Eighteenth International Conference on Machine Learning (ICML 2001)*, Williamstown, MA, Jun. 28–Jul. 1, 2001.

[7] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 93–97, Jan. 2006.

[8] J. Ham, D. D. Lee, S. Mika, and B. Schölkopf, "A kernel view of the dimensionality reduction of manifolds," in *Proc. twenty-first international conference on Machine learning (ICML 2004)*, Banff, Alberta, Canada, Jul 4–8, 2004.

[9] P. Auscher, *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces*. American Mathematical Society, 2004, pp. 65–66.