Hyper spectral Image classification using Dimensionality Reduction Techniques

Bilal Alhayani[1], Haci Ilhan[2]

Yildiz Technical University, Department of Electronics and Communication Engineering Isanbul, Turkey[1,2]

Abstract: Hyper spectral Imaging produces an image where each pixel is having narrow spectral bands with plentiful spectral information. Spectral bands refer to the large number of measured wavelengths bands of Electromagnetic Spectrum. The large number of spectral bands in hyper spectral data increases the computational burden. So, dimensionality reduction through spectral feature selection thoroughly affects the accuracy of the classification. The applications of hyper spectral images require to process given data and achieve two fundamental goals: 1) detect and classify the constituent materials for each pixel in the scene; 2) reduce the data volume (dimensionality), without loss of useful information, so that it can be processed efficiently by a human. We used the technique of DRR (Dimensionality Reduction via Regression) an unsupervised method for dimensionality reduction.

Keywords: Hyper spectral Imaging, Dimensionality Reduction, DRR.

I. INTRODUCTION

Hyper spectral Imaging is an important research area in Remote Sensing. It uses the potentiality of digital imaging and spectroscopy. Hyper spectral imaging collects information across the Electromagnetic Spectrum. This technology is used mostly in the Minerals, Coal, Oil, Gas Industry and remote sensing applications. Hyper spectral images provide a picture of the surface of an area so some useful conclusion can be drawn from the given hyper spectral picture. Humans can see only visible band of the light spectrum, but others bands of non-visible light spectrum with higher and lower frequencies exist to the left and right of visible spectrum which also provides the useful information. This spectrum is known as the Electromagnetic Spectrum and it is classified into the bands of ultra-violet, x-rays, radio waves, micro waves and infra-red radiations. These invisible bands are used in hyper spectral Imaging, and visible image is formed from these bands.

Earth's surface is covered with a variety of materials like soil, rocks, foliage and water, all composed of different minerals and chemical compounds. These materials all reflect only a small amount of the Electromagnetic Spectrum and using unique Electromagnetic Signatures these materials can be identified/ classified. Thus, every pixel in the hyper spectral image consists of a contiguous reflectance spectrum. Each pixel contains information (contiguous bands of electromagnetic spectrum from visible to infrared regions) can be used to identify/ classify the objects in the picture since it provides much more detailed information about the scene than the normal colour camera. Hyper spectral imaging can provide more detailed information from the picture of surveyed area compared to any other imaging technology. So the hyper spectral imaging provides an extremely enhanced ability to identify/ classify the objects in the picture based on their spectral properties. But, the classification of hyper spectral image is a very challenging task due to the huge number of band values for each pixel and hence more dimensionality. So dimensionality reduction becomes a crucial step for successfully implementing hyper spectral image classification.

II. DIMENSIONALITY REDUCTION IN HYPER SPECTRAL IMAGE

Hyper spectral image has very strong spectral correlation between two different bands and there is often a significant amount of redundant and misleading information present in the image and such information needs to be removed before any further processing can be carried out [1],[2],[3], and [4].so it is useful to reduce the data dimensionality.

A. Principal component analysis (PCA)
This is a popular tool for linear dimensionality reduction and feature extraction. PCA is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the

preceding components. The resulting vectors are an uncorrelated orthogonal basis set. The principal components are orthogonal because they are the eigenvectors of the covariance matrix, which is symmetric. PCA seeks to find a linear transformation which projects the data from a high-dimensional space to a lower dimensional subspace by maximizing the variance of the data in the projected subspace. The optimal number of projection is determined by the eigenvectors corresponding to the largest eigen values of the covariance matrix of the original data. PCA constitutes unsupervised dimensionality reduction and is commonly employed by researchers for classification. However, PCA provides, at best, suboptimal dimensionality reduction for classification tasks i.e., it is well understood that PCA can potentially discard information useful to the classification task at hand, particularly if such information is contained along the low-energy directions [5,6]. PCA is sensitive to the relative scaling of the original variables

Steps to calculate PCA
Given a set x_1, x2, x3 ...xn in R^n space
i) Take the whole dataset ignoring the class labels
ii) Computing the d-dimensional mean vector
iii) Computing the Scatter Matrix: The scatter matrix is computed by the following equation:

$$S = \sum_{k=1}^{n} (X_k - m)(X_k - m)^T$$

where m **is the mean vector**

$$m = \frac{1}{n} \sum_{k=1}^{n} X_k$$

iv) Computing the Covariance Matrix
v) Computing eigenvectors and corresponding eigenvalues
vi) Sorting the eigenvectors by decreasing eigenvalues
vii) Choosing k eigenvectors with the largest eigenvalues

B. Locality Preserving Projections(LPP)
It builds a graph incorporating neighbourhood information of the data set. Using the notion of the Laplacian of the graph, we then compute a transformation matrix which maps the data points to a subspace. This linear transformation optimally preserves local neighbourhood information in a certain sense. The representation map generated by the algorithm may be viewed as a linear discrete approximation to a continuous map that naturally arises from the geometry of the manifold [7]. The new algorithmic interesting from a number of perspectives

Given a set x1,x2,x3,......xm in Rn space, find a transformation matrix A that maps these m points to a set of points y1,y2,y3....ym in Rl space where (l<<n), such that yi represents xi where yi=ATxi

Steps to calculate LPP
i) Constructing the adjacency graph: Let G denote a graph with m nodes. We put an edge between nodes i and j if xi and xj are"close".

ii) Choosing the weights: Here, as well, we have two variations for weighting the edges. W is a sparse symmetric m×m matrix with Wij having the weight of the edge joining vertices i and j, and 0 if there is no such edge.

iii) Eigenmaps: Compute the eigenvectors and eigenvalues for the generalized eigen-vector problem
$$XLX^Ta = \lambda XDX^Ta$$

Where D is a diagonal matrix whose entries are column (or row, since W is symmetric) sums of W

Let the column vectors a0,..............,al-1 be the solution ordered according to their eigen values $\lambda_0 < ..\lambda_{l-1}$

Thus embedding is as follows

$$x_i \rightarrow y_i = A^T x_i A$$

C. Multidimensional scaling (MDS)
This is a means of visualizing the level of similarity of individual cases of a dataset. It refers to a set of related ordination techniques used in information visualization, in particular to display the information contained in a distance

matrix. An MDS algorithm aims to place each object in N-dimensional space such that the between-object distances are preserved as well as possible. Each object is then assigned coordinates in each of the N dimensions. The number of dimensions of an MDS plot N can exceed 2 and is specified a priori. Choosing N=2 optimizes the object locations for a two-dimensional scatter plot.

Steps in MDS

i) The classical MDS algorithm rests on the fact that the coordinate matrix X can be derived by eigenvalue decomposition from the scalar product matrix B-XX'. The problem of constructing B from the proximity matrix P is solved by multiplying the squared proximities with the matrix $J-I-n^{-1}$This procedure is called double cantering. The following steps summarize the algorithm

i) Setup the matrix of squared proximities $P^{(2)}-[p^2]$

ii) Apply the double cantering B $-1/2JP^{(2)}$ J using the matrix $J-I-n^{-1}ll'$ where n is the number of objects

iii) Extract the m largest positive eigenvalues $\lambda_1 \dots \lambda_m$ of B and corresponding m eigenvectors $e_1 \dots e_m$.

iv) A m-dimensional spatial configuration of the n object is derived from the coordinate matrix $X - E_m \Lambda_m^{1/2}$, where E_m is the matrix of m eigenvectors and Λ_m is the diagonal matrix of m eigenvalues of B respectively.

D. Kernel PCA

is the nonlinear form of PCA, which better exploits the complicated spatial structure of high-dimensional features.

I) suppose that instead of using the points we would first map them to some nonlinear feature space$\phi(x_i)$, Apply the Gaussian Kernel

ii) Extract principal component in that space (PCA)

iii) The result will be non-linear in the original data space

E. Dimensionality Reduction via Regression (DRR)

The DRR presented here tries to nonlinearly remove the information still shared by different PCA components. T is well known that, even though PCA leads to a domain with decor related dimensions, complete independence (or non redundant coefficients) is guaranteed only if the signal has a Gaussian probability density function (PDF). Here, a scheme to remove this redundancy (or uninformative data) is presented. The idea is simple: just predict the redundant information in each coefficient that can be extracted from the others. Only the non-predictable information (the residual prediction error) should be retained for data representation. Specifically, we start from the linear PCA representation outlined above: $\alpha = Vx$. Then, we propose to predict each coefficient, α_i , through a multivariate regression function, $f_i(.)$, that takes the higher variance components as inputs for prediction. The non-predictable information is:

$$y_i = \alpha_i - \widehat{\alpha_i} = \alpha_i - f_i(\alpha_1, \alpha_2, \dots \alpha_{i-1})$$

And this residual, yi is the i-th dimension of the DRR domain. This prediction + subtraction is applied d-1 times $\forall i=d$, d-1, 2 where d is the dimension of the input. As a result, the DRR, representation of each input vector x is

$$r = R(x) = (\alpha_1, y_2, y_3, . y_d)^T$$

III. EXPERIMENTATION

In this section, we give experimental evidence of the performance of the proposed algorithm along with four other algorithms. The challenging high dimensional data multispectral image Classification in which contextual information is stacked to the spectral information highly increases the dimensionality. The dataset used is Salinas Cube hypersectral data which is 512 X 217 pictures with 204 bands. The ground truth is given for the given dataset which is as shown below in figure1. Sixteen classes are identified in the image. A total of 54129 labelled samples are available. The classification accuracy is tested on Salinas's dataset, and the experimentation results are as shown below.

Dimensionality	Accuracy_PCA	Accuracy_LPP	Accuracy_MDS	Accuracy_KPCA	Accuracy_ DRR
100	86.1165	79.4602	85.5881	39.0382	78.5032
50	86.2495	84.1730	86.2495	50.0532	82.2115
30	86.3733	84.8787	86.3733	51.8248	87.1207
20	86.7816	84.1804	86.3733	51.0957	85.0912
10	90.8293	93.1275	90.8293	60.5165	87.1704
5	94.8697	94.3136	94.8697	89.4973	87.1289

Fig.1. Ground truth of Salinas's data

IV.CONCLUSION

This paper provides a various dimensionality reduction techniques in terms of feature extraction and classification accuracy for hyper spectral images. Commonly used feature extraction methods used such as principal component analysis(PCA), Locality Preserving Projections(LPP), Multidimensional scaling (MDS), Kernel PCA, Dimensionality Reduction via Regression (DRR) have been discussed for hyper spectral images. A new technique for dimensionality reduction has been implemented based on Dimensionality Reduction via Regression (DRR), with good accuracy.

REFERENCES

[1]   Jun Li, Prashanth Reddy Marpu, Antonio Plaza, Jose M. Bioucas-Dias, and Jon Atli Benediktsson," Generalized Composite Kernel Framework for Hyper spectral Image Classification," IEEE transactions. On geosciences and remote sensing, vol. 51, no. 9, September 2013.
[2]   B. Mojaradia, h.emamib, m.varshosazc and s. Jamali,"A Novel Band -Selection Method for Hyper spectral Data Analysis," The international archives of the photogrammetry, remote sensing and spatial information sciences. Vol. Xxxvii. Part b7. Beijing 2008.
[3]   Kun tan, Erzhu Li, Giandu, and   Peijun Du,"Hyperspectral Image Classification Using Band Selection And Morphological Profiles," IEEE journal of selected topics in applied earth observations and remote sensing, vol. 7, no. 1, January 2014.
[4]   Qian du and He Yang,"Similarity-based Unsupervised Band Selection for Hyper spectral Image Analysis," IEEE geoscience and remote sensing letters, vol. 5, no. 4, October 2008.
[5]   Craig Rodarmel and Jieshan,"Principal Component Analysis for Hyper spectral Image Classification," Surveying and Land Information Systems, vol. 62, no. 2, 2002.
[6]   S. Prasad and L. M. Bruce, "Limitations of principal component analysis for hyper spectral target recognition," IEEE Geosci. Remote Sens. Lett. vol. 5, no. 4, pp. 625–629, Oct. 2008.
[7]   M. Belkin and P. Niyogi, "Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering,"Advances in Neural   Information Processing Systems 14 Vancouver, British Columbia, Canada, 2002.
[8]   Valero Laparra, Jesús Malo, and Gustau Camps-Valls, "Dimensionality Reduction via Regression in Hyper spectral Imagery", IEEE journal of selected topics in signal processing, VOL. 9, NO. 6, SEPTEMBER 2015.